Evaluering av ledelsesutviklingsprogram

Hvordan AFF bør evaluere sine ledelsesutviklingsprogram, med bakgrunn i oppdatert forskning og ekspertkompetanse

AFF

UNIVERSITY OF OSLO

Rapport utformet av Solveig Vik Hofseth, Gina Sutterud Warland, Line Stig Alexandersen, Mona Roos Hovden-Undseth, Karen Mathea Odland og Mohammed Haris Akhtar.

Prosjektforum ved Universitet i Oslo, Høsten 2021.

Antall ord: 19.378

Forord

Denne rapporten er utarbeidet av en gruppe studenter ved masterprogrammet Organisasjon, ledelse og arbeid som en del av emnet SVPRO4000: Prosjektforum på Universitetet i Oslo. Rapporten er skrevet på oppdrag fra AFF, som ønsket en kartlegging av forskning på evaluering av ledelsesutviklingstiltak, i tillegg til en modell for hvordan de kan evaluere sine ledelsesutviklingsprogram. Vi vil gjerne takke AFF for tilliten og muligheten til å gjennomføre dette lærerike prosjektet.

Vi har hatt flere viktige bidragsytere på dette prosjektet. Først vil vi takke vår kontaktperson i AFF, Richard Havdahl, for godt samarbeid og kommunikasjon gjennom hele prosessen. Vi vil også takke vår veileder, Thomas Hoholm, for gode og konstruktive tilbakemeldinger, samt faglige innspill. En stor takk går også til informantene våre, som har bidratt med viktig teoretisk og praktisk kompetanse innen evaluering av ledelsesutvikling. Til slutt vil vi takke vår seminarleder Beate Karlsen, for lærerike seminarer om samarbeid og gruppedynamikk.

Denne rapporten er skrevet av Solveig Vik Hofseth, Gina Sutterud Warland, Line Stig Alexandersen, Mona Roos Hovden-Undseth, Karen Mathea Odland og Mohammed Haris Akhtar.

Oslo, november 2021.

Sammendrag

Denne rapporten er skrevet i forbindelse med emnet Prosjektforum ved UiO, høstsemesteret 2021. Formålet var å kartlegge forskning på evalueringsfeltet spesielt tilknyttet ledelsesutvikling, i tillegg til å gi AFF en modell for hvordan de kan evaluere sine ledelsesutviklingsprogram, forankret i oppdatert forskning og «beste praksis». Modellen er utviklet spesielt for AFFs bedriftsspesifikke program, men vil også kunne være relevant for andre ledelsesutviklingsprogrammer.

En viktig del av oppdraget vårt var å kartlegge forskning på evalueringsfeltet. Det har vi gjort gjennom strukturerte litteratursøk. Forskningen er presentert i vårt teorikapittel. Her redegjør vi for sentral evalueringsteori og læringsteori, i tillegg til at vi presenterer viktige teoretiske modeller og metoder for evaluering. For å undersøke «beste praksis» på evaluering, har vi gjennomført dybdeintervjuer med 8 eksperter på ledelsesutvikling og evaluering. Ekspertene består av fire teoretikere og fire praktikere, som har kunnskap om og erfaring med gjennomføringen av evalueringer av ledelsesutviklingsprogram.

Funnene våre viser at evaluering av ledelsesutvikling er vanskelig å generalisere, og at det alltid vil være behov for tilpasning i henhold til konteksten ledelsesutviklingen gjennomføres i. Det er essensielt for evalueringen at det settes konkrete og målbare mål i forkant av programmet, og at disse målene gjenspeiles i programmets aktiviteter. Slik kan man i etterkant vurdere om programmet har gitt ønsket effekt. Det er hensiktsmessig å benytte en kombinasjon av kvalitative og kvantitative metoder, slik at man får oversikt over store trender på organisasjonsnivå, i tillegg til dybdeforståelse av menneskenes individuelle opplevelser av programmet. I tillegg er det essensielt å gjennomføre målinger i forkant, under og i ettertid av programmet, for å undersøke årsakssammenheng mellom program og utvikling. Midtveisevaluering er også hensiktsmessig for å kunne justere programmets innhold, for å sikre at det gir ønsket effekt. Funnene våre har resultert i en femstegsmodell som AFF kan evaluere sine ledelsesutviklingsprogrammer med. Steg 1 er en avklaringssamtale med kunden, med formål om å avklare mål og innhold i programmet. Steg 2 er en førmåling, som skal vise hvordan bedriften ligger an i forkant av programmet. Steg 3 er underveismåling, som brukes til å justere programmet og evaluere progresjon. Steg 4 er ettermåling, som sammenlignes med

førmålingen for å måle utbyttet av programmet. Steg 5 er en avslutningssamtale med kunden i etterkant av programmet. Hensikten med dette er å evaluere målsetningen for fremtidige programmet, i tillegg til at AFF kan presentere resultatene for kunden.

Innholdsfortegnelse

Forord 1

Sammendrag 2

1.0 Innledning 6
1.1 Aktualitet og bakgrunn for oppgaven 6
1.2 Problemstilling 6
1.3 Oppgavens struktur 7
1.4 Avgrensninger 7
1.4.1 Skillet mellom lederutvikling og ledelsesutvikling 7
1.4.2 Bedriftsspesifikke program 8

2.0 Teoretisk rammeverk 8
2.1 Evalueringslogikk 8
2.2 Hva er evaluering? 9
2.2.1 Formativ og summativ evaluering 10
2.3 Formål med evaluering 10
2.4 Læringsteori 11
2.5 Tradisjonell evalueringsteori 12
2.5.1 Kirkpatricks Four Levels of Learning Evaluation Model 13
2.5.1.1 Utfordringer ved Kirkpatricks modell 14
2.5.1.2 Videreutvikling av Kirkpatricks modell 14
2.5.2 Brinkerhoffs Success Case Method 15
2.6 Kontekst, endring og prosess 15
2.6.1 Et paradigmeskifte: fra kompetanse til kontekst 16
2.6.1.1 En balanse mellom ytelse og læring 16
2.6.2 CIRO 17
2.6.3 360-graders lederevalueringer 18
2.6.4 Evalueringsmodell basert på Theory of Change 19
2.6.5 Prosessevaluering og four-phase approach 19
2.6.6 Value of Learning Model 21
2.6.7 SMART-mål 21
2.6.8 Deloppsummering 22

3.0 Metode 22
3.1 Forskningsdesign 22
3.1.1 Litteratursøk 22
3.1.2 Dybdeintervju 23
3.2 Utvalg 24
3.3 Gjennomføring av intervju 27
3.4 Forskerrollen og etiske hensyn 28
3.5 Koding 29

3.6 Dataenes kvalitet 30
3.7 Utforming av modell 30

4.0 Analyse 31
4.1 Kontekst 31
4.2 Sette målbare mål 33
4.3 Hvordan måle utbyttet av programmet? 37
4.4 Kvantitative og kvalitative metoder i evalueringen 41
4.5 Før-, underveis- og ettermåling 44
4.5.1 Viktigheten av underveismåling grunnet læring 46
4.5.2 Når bør etter-målingen være? 47
4.6 Evaluering i teorien vs. praktisk gjennomførbar modell 49

5.0 Bakgrunnen for utformingen av modellen 51
5.1 Modell 52
5.2 Modellen beskriver et rammeverk som behøver tilpasninger 54
5.3 Indikatorer 55
5.4 Bruk av 360-graders evaluering 55
5.5 Bruk av midtveismålinger 56

6.0 Avslutning 57
6.1 Konklusjon 58
6.2 Videre forskning 58

7.0 Litteraturliste 59

8.0 Vedlegg 62
Vedlegg 1: Generisk modell for evaluering 62
Vedlegg 2: Modell for evaluering med case-eksempler 65
Vedlegg 3: Mandat fra AFF 73
Vedlegg 4: Informasjonsskriv 74
Vedlegg 5: Hoved-intervjuguide 79
Vedlegg 6: Godkjenning fra NSD 82

1.0 Innledning

1.1 Aktualitet og bakgrunn for oppgaven

Evaluering og oppfølging av kompetansetiltak er en forutsetning for at satsingen på kompetanse skal kunne regnes som reelt strategisk. Systematisk evaluering kan dermed være et kontrollverktøy som kan gi grunnlag for å vurdere effekten av utviklingstiltak, behovet for forbedringer, og om investeringene i kompetansekapitalen har resultert i verdifull avkastning for kunden. Det er derfor paradoksalt at evaluering ofte utelates. Dette kan ofte være i sammenheng med mangler av konkrete mål, svak metodekompetanse eller ressursmessige begrensninger, som gjør det utfordrende å danne planmessig evaluering og oppfølging (Lai, 2013:183).

AFF er en av de største bidragsytere i det norske ledelses- og organisasjonsutviklingsmiljøet. Gjennom partnerskap med NHH vektlegger stiftelsen stor faglig tyngde, og søker kontinuerlig å utvikle sitt eget og sine kunders kunnskapsgrunnlag innen ledelses- og organisasjonsutvikling (AFF, u.å.). Med et viktig samfunnsoppdrag om å utvikle ledere og organisasjoner i Norge, er det dermed også essensielt for AFF å sikre seg at utviklingsprogrammene deres har best mulig effekt. Derfor ønsker AFF å sette lys på sine evalueringsrutiner og oppdatere disse. Den nåværende evalueringspraksisen i organisasjonen er i hovedsak basert på deltakernes opplevde utbytte og grad av tilfredshet. Det fremkommer at denne evalueringsformen ikke gir en helhetlig og reell måling av det som skal evalueres, og medfører en rekke metodiske problemer. Metodisk kvalitetssikrede verktøy vil være nødvendig for en vellykket evalueringsprosess og kan gi økt legitimitet.

1.2 Problemstilling

Formålet med prosjektet er å levere en anvendelig modell og fremgangsmåte til oppdragsgiver, AFF, som de kan benytte når de skal evaluere sine bedriftsspesifikke ledelsesutviklingsprogram (se Vedlegg 1). Modellen skal være basert på "beste praksis", forankret i oppdatert forskning og teori, samt ekspertvurderinger fra evalueringsfeltet. Oppgavens problemstilling er som følger:

> *Hvordan bør AFF, med bakgrunn i oppdatert forskning og ekspertkompetanse, evaluere sine ledelsesutviklingsprogram?*

1.3 Oppgavens struktur

Innledningsvis vil det i oppgavens første kapittel redegjøres for aktualitet og bakgrunn for oppgaven, problemstilling, og avgrensninger. I det andre kapittelet beskriver vi hva evaluering innebærer og hvordan det kan forstås. Vi vil presentere relevant teori på evalueringsfeltet, som tar utgangspunkt i tradisjonell evalueringsteori, og tre forhold som presenterer evalueringsfeltet i nyere tid: kontekst, endring og prosess. I det tredje kapittelet legger vi frem den metodiske tilnærmingen for oppgaven, og beskriver våre veivalg og utfordringer i prosessen.

Analysen utgjør oppgavens fjerde kapittel, der vi formidler og analyserer de empiriske funnene fra ekspertintervjuene. I kapittel fem vil vi gjøre rede for valg og avveininger, slik at vi kan presentere en generisk modell som AFF kan benytte for å evaluere sine bedriftsspesifikke ledelsesutviklingsprogram. Avslutningsvis vil en konklusjon trekkes i kapittel seks, der det også redegjøres for studiens generaliseringspotensial og forslag til videre forskning på området.

1.4 Avgrensninger

1.4.1 Skillet mellom lederutvikling og ledelsesutvikling

Vi anser det som viktig å definere sentrale begreper ved problemstillingen, med den hensikt å få en sammenfallende oppfatning av hva begrepet inneholder. Det er vanlig å skille mellom begrepene lederutvikling og ledelsesutvikling. Lederutvikling kan forstås på flere måter, men kan i en bredere forstand defineres som enhver form for utvikling eller utviklingsfase hos leder som fremmer, oppmuntrer og bidrar til ytterligere kunnskap og ekspertise innenfor lederpotensial og prestasjoner (Kuvaas & Dysvik, 2016:144). En kan si at lederutvikling betegnes som utvikling av den enkelte lederen, mens ledelsesutvikling derimot kan ses i sammenheng med en aktuell organisasjon (Day & Dragoni, 2015). Ledelsesbegrepet inneholder derfor en bredere forståelse av ledelse og inkluderer i større grad et organisatorisk perspektiv (Ulrich, 2014:6).

AFF er bevisst på skillet mellom lederutvikling og ledelsesutvikling. Samtidig kan ikke dette betegnes som generell praksis, da de ikke er konsekvente i begrepsbruken og benytter begrepene om hverandre. I lys av dette ønsker vi å gjøre et teoretisk skille mellom begrepene, der vi benytter oss av ledelsesutviklingsbegrepet. Dette gjør vi, da vi anser en organisasjons kollektive ledelse som viktigere i et langsiktig perspektiv. Ledelse kan dermed betegnes som mer enn en posisjon, og oppstår gjennom samhandling (Ulrich, 2014:7).

1.4.2 Bedriftsspesifikke program

Mandatet fra AFF åpnet opp for muligheten til at prosjektgruppen kunne velge å inkludere både ledelses- og organisasjonsutviklingsprogram, eller avgrense der vi kun fokuserer på evalueringen av AFFs sine bedriftsspesifikke ledelsesutviklingsprogram. Av hensyn til prosjektets kvalitet og med tanke på tidsbegrensningen på prosjektet, har prosjektgruppen valgt å avgrense til sistnevnte. Vi vil imidlertid ikke utelukke muligheten for at denne rapporten er nyttig også for AFFs andre program.

2.0 Teoretisk rammeverk

I denne delen av oppgaven skal vi presentere det teoretiske rammeverket for rapporten. I den første delen vil vi beskrive hvilke premisser vi legger til grunn for hva evaluering innebærer og hvordan det kan forstås. Dette gjør vi ved å presentere evalueringslogikk, skillet mellom formativ og summativ evaluering, samt ulike formål ved evaluering. Et sentralt mål ved AFFs ledelsesutviklingsprogram, er at lederne skal tilegne seg kunnskap som bidrar til en reell atferdsendring. Vi baserer oss derfor på en antakelse om at denne formen for atferdsendring og kunnskapstilegnelse skal være på bakgrunn av en læringsprosess. I lys av dette vil den andre delen av det teoretiske rammeverket belyse læringsteori, som viser hvordan deltakerne kan dra nytte av programmet og hvordan organisasjoner som helhet kan lære.

I den tredje delen blir tradisjonell evalueringsteori presentert for å få en forståelse av dominerende modeller på evalueringsfeltet. Samtidig vil vi belyse aktuelle bidrag som er videreutviklet på grunnlag av disse modellene. Avslutningsvis vil vi rette fokuset mot tre forhold som presenterer evalueringsfeltet i nyere tid: kontekst, endring og prosess. Dette vil gi et nyansert og helhetlig bilde av evalueringsteorier, som er faglig forankret i oppdatert forskning. Underveis trekker vi ut de mest sentrale momentene av den tradisjonelle forståelsen og evalueringsfeltet i nyere tid, som senere vil legge grunnlag for videre analyse og drøftelse. Dette vil danne utgangspunktet for å kunne skape en skreddersydd evalueringsmodell for AFF.

2.1 Evalueringslogikk

Evalueringslogikk er en logikk som beskriver evaluering som en prosess, der en vurderer fortjenesten, verdien eller betydning av et objekt (Owen, 2007:9). Fournier (1995:16) beskriver evalueringslogikken gjennom en firetrinnsprosess. Det første trinnet handler om å definere

verdifulle kriterier, som kan benyttes til å kartlegge hvilke dimensjoner deltakeren skal forbedre seg på. Trinn to er fastsettelse av standarder, der en bestemmer hvor godt noe skal utføres basert på bestemte kriterier. Videre vil en på det tredje trinnet måle en prestasjon og sammenligne dette med en standard. Her er det kritisk at en måler den faktiske ytelsen. Det fjerde trinnet handler om syntetisering og integrering av evidens i vurderingen av verdien, der en bedømmer hva en fikk ut av programmet. En styrke ved denne modellen er dens generalitet, altså dens evne til å møte vekslende krav uten å forandre egenskaper. Dette kan gi et grunnlag der en kan fordype seg i ulike tilnærminger til evalueringspraksis (Owen, 2007:10).

Det er også viktig å være oppmerksom på betydningen av å gjøre verdigrunnlaget eksplisitt. Dette er fordi evalueringsprosessen følges av den verdiposisjonen som er tatt, og en må være klar over verdisystemet som ligger til grunn for evalueringen. Det siste steget der en bedømmer programmets verdi, utledes dermed fra en eller flere av følgende punkter: programmets mål, deltakernes behov, interessene til ulike interessentgrupper, og måling av effektivitet.

Oppsummert kan en ikke overvurdere betydningen av verdier som grunnlag for evalueringspraksis. Det sentrale spørsmålet ved en evaluering vil være hvilke verdier eller verdirammer som kan brukes til å bedømme programmet. Dersom interessenter av evalueringen har ulike verdier, kan evalueringsprosessen bli farget av verdiperspektivet som velges. Dette kan få implikasjoner i praksis, og representerer en utfordring i å håndtere dette mangfoldet av synspunkter når evalueringen planlegges og besluttes (Owen, 2007:10-12).

2.2 Hva er evaluering?

Ifølge Sverdrup (2014:12) er det vanlig å ta utgangspunkt i evaluering som en forskningsbasert systematikk og metode, der en ønsker å fastslå en form, virkning eller effekt av et tiltak. Hensikten er å undersøke om tiltaket som ble iverksatt, førte til de fastsatte virkningene. Denne tilnærmingen har fokus på enkle, lineære studier av virkninger og effekter, som er utformet slik at evaluatoren har en nøytral og distansert rolle.

I lys av problemstillingen, ønsker vi imidlertid å benytte oss av en nyere tilnærming av begrepet evaluering. Vi ønsker å forstå evaluering som en systematisk aktivitet som gjennomføres for å tilføre nytte, gjerne i kombinasjon med brukerevaluering, der evaluatorens rolle karakteriseres av nærhet og involvering (Sverdrup, 2014:12). Premisset for vår evalueringsmodell, er at

evalueringen skal gjennomføres av AFF på sine egne ledelsesutviklingsprogram. Slik vil evaluatoren være tett på den en evaluerer, og nøytralitet og distansering er derfor ikke mulig.

2.2.1 Formativ og summativ evaluering

I 1967 introduserte Michael Scriven skillet mellom formativ og summativ evaluering (McDavid & Hawthorn, 2006:21). Formativ evaluering refererer til underveisevaluering, prosessanalyse eller iverksettingsanalyse. Denne formen for evaluering innebærer undersøkelse og vurdering av iverksatte tiltak, herunder foreløpige resultater og forslag til justeringer. En benytter underveisevalueringer blant annet når en ønsker tilbakeføring av erfaringer raskest mulig (Askim, Døving & Johnsen, 2013:76).

En summativ evaluering, også kalt etterevaluering, rapporterer et tiltak etter at det er avsluttet ved å dokumentere virkninger, effekter eller resultater av tiltak. Summative evalueringer kan undersøke både ressursbruk og effekter ved et tiltak, og kan derfor vurdere tiltakets kostnadseffektivitet (Askim et al., 2013:77). I forbindelse med denne type evaluering blir evalueringer i hovedsak rapportert til overordnet myndighet og allmennhet, og av den grunn forventes det en kritisk distanse til tiltakshaveren og evalueringsobjektet.

2.3 Formål med evaluering

Det kan være flere ulike formål med å iverksette evaluering. Det kan bl.a. være relevant i forbindelse med tre forhold: for å kontrollere, som et symbol og for å legitimere (Mehlbye, Rieper & Togeby, 1993:25). Kontroll refererer til formålet om å kontrollere hvorvidt det er en effektiv ressursutnyttelse, og om det er de riktige gruppene som mottar innsatsen, samt om medarbeiderne gjør jobben sin godt nok. Symbol, kalt symbolverdi, handler om at formålet med evalueringen er å dokumentere at en "gjør noe" og at beslutninger tas på et seriøst grunnlag. Videre kan evaluering ha som formål om å legitimere beslutninger som allerede er tatt (Mehlbye et al., 1993:25).

Det er vesentlig å påpeke at et sentralt hovedmål med evaluering er å fremskaffe informasjon som videre kan danne grunnlag for vurdering av iverksatte tiltak og programmer. Formålet med dette er at beslutningstakere anvender informasjonen fra evalueringer til å underbygge beslutninger og læring (Halvorsen & Madsen, 2013:22). Bastøe (2020:114) poengterer at evaluering vanligvis har to hovedhensikter: ansvarliggjøring og læring. Ansvarliggjøring

refererer til å etterprøve om målene er oppnådd, det vil si om en har gjort det som var lovet. I denne sammenhengen handler det også om å identifisere hvem som er ansvarlig for resultater, eller mangel derav. Læring er knyttet til å systematisere erfaringer, identifisere hvorfor noe lykkes eller ikke, for å så bruke kunnskapen i utformingen og gjennomføringen av fremtidige tiltak (Bastøe, 2020).

2.4 Læringsteori

For å belyse læringsteori vil vi redegjøre for skillet mellom individuell og organisatorisk læring, før vi retter fokuset mot et av de første og mest innflytelsesrike bidragene innen lærende organisasjoner, utviklet av Argyris og Schöns (1978; 1996). Teorien utfordrer grunnleggende antakelser, ved å peke på utfordringer knyttet til individers tanker og handlinger i organisasjonen (Argyris & Schön, 1996:30). Bakgrunnen for at vi har tatt utgangspunkt i denne læringsteorien, er at den belyser forskjellen på individuelle og organisatoriske læringsprosesser, noe som er et sentralt element i vår oppgave.

2.4.1 Individuell og organisatorisk læring

Individuell læring fokuserer på læring som individuell tilegnelse og utvikling av kunnskap, som videre påvirker eller endrer individets atferd (Filstad, 2016). Utfordringen ved dette perspektivet er at det tradisjonelt sett har hatt et utelukkende fokus på individet som lærende. Dermed har det vært en dominerende forestilling om at så lenge individene lærer i en organisasjon, så gjør organisasjonen det også. Det er imidlertid ingen automatikk i at individuell kompetanse resulterer i organisatorisk kompetanse, da organisatorisk læring er et resultat av kunnskapsdeling, felles praksis, og utøvelse av arbeidsoppgaver individuelt og i samarbeid med hverandre (Filstad, 2016:38).

2.4.2 Bruksteori og uttrykt teori

Argyris og Schön (1978; 1996) beskriver i sitt bidrag om lærende organisasjoner en handlingsteori som danner grunnlaget for to former for organisatorisk læring; enkeltkrets- og dobbeltkretslæring. Organisatorisk læring blir omtalt som det som skjer når individer i en organisasjon opplever en problematisk situasjon, og undersøker dette på vegne av organisasjonen (Argyris & Schön, 1996:16). En lærende organisasjon må videre forstås ut i fra organisatoriske handlingsteorier, da en organisasjon verken husker, tenker eller lærer slik individer gjør. For at organisasjonslæring skal genereres må det skje en endring i individenes

handlingsteori. Organisatoriske handlingsteorier deles inn i henholdsvis bruksteorier og uttrykte teorier (Argyris & Schön, 1996:17). Bruksteori er de styrende verdiene gjennom handlinger i organisasjonen. Den sier derfor noe om den faktiske atferden i organisasjonen, da den reflekterer handlingsmønsteret hos organisasjonsmedlemmene. Uttrykt teori refererer til det organisasjonen uttrykker av verdier, og gjenspeiler de ansattes forestillinger knyttet til egen og andres atferd i organisasjonen (Argyris & Schön, 1996:13-14). Det er nødvendigvis ikke slik at handlingsteoriene samsvarer, ettersom det kan være avvik mellom forklaringene av handlingene til organisasjonsmedlemmene og handlingene som faktisk foretas.

2.4.3 Enkelkrets- og dobbelkretslæring

Ifølge Argyris og Schön (1996) fører skillet mellom bruksteori og uttrykt teori videre til to former for organisatorisk læring, enkeltkrets- og dobbeltkretslæring. Enkelkretslæring forekommer ved at organisasjonsmedlemmene endrer og tilpasser sin atferd når de opplever lav måloppnåelse, som følge av en feil eller uønsket konsekvens av en handling (Argyris & Schön, 1996:20). Da det i enkelkretslæring hovedsakelig er fokus på å oppnå et mål, gjøres det uten at de styrende verdiene blir reflektert over eller vurdert (Filstad, 2016:45). Problemet løses uten at det blir brukt tid og ressurser på å reflektere over årsaken til at situasjonen oppstod. Dermed blir graden av refleksjon hos organisasjonsmedlemmene begrenset, og grunnleggende verdier og normer adresseres ikke. I dobbelkretslæring reflekterer organisasjonsmedlemmene derimot over hvorfor problemet oppstod. Ved å gjøre dette blir de grunnleggende verdiene og antagelsene i organisasjonen utfordret og endret som en del av læringsprosessene (Argyris & Schön, 1996:21-22). Organisasjonslæring blir en prosess hvor medlemmene oppdager misforhold og feil i organisasjonen, som deretter rettes ved rekonstruering av organisasjonens bruksteori (Filstad, 2016:46).

2.5 Tradisjonell evalueringsteori

I denne delen vil vi gjøre rede for tradisjonell evalueringsteori. Vi vil spesielt rette fokuset mot to tradisjonelle evalueringsmodeller: Kirkpatricks (1959) “Four Levels of Learning Evaluation Model” og Brinkerhoffs (2003) “The Success Case Method”. For å kunne gi et mer nyansert bilde vil vi også belyse teorier og modeller som har videreutviklet Kirkpatricks tilnærming til evaluering.

2.5.1 Kirkpatricks Four Levels of Learning Evaluation Model

Modellen ble utviklet av Kirkpatrick på 1950-tallet, og har siden vært en av de mest anerkjente evalueringsmetodene. Ifølge denne metoden bør en evaluere på 4 ulike nivåer: reaksjon, læring, atferdsendring og resultat (Kirkpatrick & Kirkpatrick, 2016:9-10).

Reaksjonsnivået innebærer hvordan deltakerne reagerer på det aktuelle programmet, og hvorvidt de er tilfredse (Singh & Ramdeo, 2020:409). Evaluering på dette nivået innebærer at deltakerne betraktes som kunder eller oppdragsgivere, og hensikten er å måle hvor tilfredse deltakerne er med tjenesten de har blitt tilbudt (Lai, 2013:192). Læringsnivået kan defineres som «graden av tiltenkt kunnskap, ferdigheter, innstilling, selvtillit og følelse av forpliktelse deltakeren har tilegnet seg gjennom det aktuelle programmet» (Kirkpatrick & Kirkpatrick, 2016:15). En utfordring ved egenvurdering av læring er at hukommelsen vår er rekonstruktiv, og mennesker har en sterk tendens til å undervurdere egen læring, spesielt når det gjelder ren kunnskapslæring (Lai, 2013:193). Hvis en klarer å finne fram til en metode som måler reaksjon og læring, kan det gi vesentlige legitimerings- og markedsføringseffekter, noe som igjen vil ha en verdi for å underbygge betydningen av det aktuelle tiltaket (Lai, 2013:194).

Nivået for atferdsendring handler om graden av anvendelse av det deltakerne har lært gjennom programmet (Kirkpatrick & Kirkpatrick, 2016:13). Som Lai (2013:194) påpeker i sitt litteraturbidrag, er det langt mellom liv og lære. Derfor er det helt vesentlig å evaluere effekter på atferdsnivået, det vil si om læringseffekten av et tiltak blir anvendt i praksis og gir endrede handlings- og atferdsmønstre (Lai, 2013:194). Det pekes på fem betingelser for at det skal oppstå en atferdsendring som følge av ny læring. Deltakeren må ha et ønske om forandring, de må vite hva som skal endres og hvordan de skal gjøre det, de må befinne seg i et godt arbeidsklima, de må få nødvendig hjelp til å anvende den nye læringen og hensiktsmessige belønningssystem for atferdsendringer må være på plass (Lai, 2013:194).

Det siste nivået i Kirkpatricks evalueringsmodell, resultat, måler i hvilken grad det aktuelle tiltaket har bidratt til å nå organisatoriske resultater (Kirkpatrick & Kirkpatrick, 2016:12). Eksempler på slike resultater kan være høyere kvalitet, reduserte kostnader, bedre måloppnåelse og høyere verdiskapning. Formålet med evaluering på det siste nivået er å anslå hvilken organisasjonsmessig nytte det aktuelle programmet harBates bidratt med, og hvorvidt tiltaket er lønnsomt (Lai, 2013:195).

2.5.1.1 Utfordringer ved Kirkpatricks modell

Kirkpatricks modell (1959) legger vekt på at en skal måle effekt på hvert enkelt nivå for å gjennomføre en optimal evaluering. Til tross for dette har evalueringer en tendens til å unngå måling på nivåene for atferd og resultater, da det er mer ressurskrevende på høyere nivåer. Giangreco, Carugati og Sebastiano (2010) argumenterer bl.a. for at Kirkpatricks modell krever en viss form for stabilitet, noe en rekke virksomheter i den post-industrielle kunnskapsøkonomien ikke har.

I praksis er modellen hovedsakelig brukt til å måle effekten av lederutvikling på individnivå, og vektlegger i mindre grad virkningene av ledelsesutvikling på organisasjonen som helhet. Ofte måles deltakernes tilfredshet umiddelbart etter programmets slutt, og det pekes på at modellen er lite anvendelig i komplekse miljøer (Bates, 2004).

En generell kritikk av modellen er at den er forenklet og ufullstendig, noe som kan forklare utvikling av en rekke andre modeller som har hentet inspirasjon fra fire-nivåmodellen, men som likevel foreslår modifiserte versjoner av denne. Med et ønske om å redusere gapet mellom evalueringsteori og praksis, har forskere forsøkt å finne en evalueringsmetode som både fanger opp komplekse effekter av utviklingsprogrammer, samtidig som evaluatoren får en forståelse av hva deltakerne har lært. En måte å gjøre dette på er å evaluere programmets effekt på deltakernes evne til å reflektere over egen atferd (Boyatzis, 2008). King og Nesbit (2015) argumenterer for at økt bruk av kvalitative analysemetoder kan gi en dypere forståelse av deltakernes erfaringer og endrede perspektiver over tid.

2.5.1.2 Videreutvikling av Kirkpatricks modell

I løpet av de siste tiårene har eksperter på lederutvikling blitt utfordret til å vise hvordan deres kurs bidrar økonomisk i virksomheten (Topno, 2012:21). Med bakgrunn i dette, har Jack Phillips (1996) modifisert og videreutviklet Kirkpatricks modell. Phillips supplerer med et femte nivå til Kirkpatricks modell, som han kaller ROI, "Return on Investment". ROI vil si avkastningen på investeringen, der en regner ut summen av de økonomiske kostnadene og virkningene av tiltaket.

Phillips er opptatt av fremgangsmåten for innsamling av data, der en samler inn data før programmet, som videre sammenlignes med resultatene i etterkant. Videre er modellen opptatt

av hvordan en isolerer betydningen av organisasjonsutviklingstiltak, slik at en sikrer korrelasjon mellom tiltaket og resultatet. Modellen ønsker også å redegjøre for immaterielle fordeler med organisatoriske utviklingstiltak, ved å omdanne disse virkningene til kvantitative variabler (Singh & Ramdeo, 2020:411).

Det er imidlertid rettet kritikk mot Phillips modell, da målingene skjer etter at tiltaket er implementert. Dermed har denne utregningen begrenset verdi, ettersom investeringen allerede er skjedd. Det argumenteres for at en slik utregning er mer hensiktsmessig når en skal vurdere om en skal gjennomføre et tiltak, heller enn som en del av evalueringen (Singh & Ramdeo, 2020:412).

2.5.2 Brinkerhoffs Success Case Method

Success Case Method, utviklet av Brinkerhoff (2003), fokuserer på evaluering av hvert enkelt tiltak som har til hensikt å skape endring. Det betyr at den ikke kun er begrenset til lederutviklingsprogrammer, og kan bli benyttet for enhver endring i virksomheter, ved eksempelvis kjøp av utstyr eller implementering av nye prosesser. Denne modellen kjennetegnes av at evalueringsmetoden baserer seg på innhenting og analyse av kvalitative data. Metoden innebærer å identifisere hvilke typer endringstiltak som var suksessfulle og hvilke som var mislykket. Ved å foreta en sammenligning mellom det som viser seg å være vellykket og det mislykkede, vil organisasjonen få en indikasjon på hva som bør gjøres eller endres ved neste tiltak for å oppnå suksess (Singh & Ramdeo, 2020:411).

Ifølge Brinkerhoff er metoden godt egnet for å få svar på hovedsakelig fire spørsmål: Hva er det som egentlig skjer, hvilke resultater bidrar programmet til å produsere, hvilken verdi har disse resultatene, og hvordan kan initiativet forbedres? (Passmore & Velez, 2012). Metoden blir imidlertid sjeldent benyttet ved systemomfattende organisatoriske endringer, men heller til å evaluere individuelle endringstiltak (Singh & Ramdeo, 2020:411).

2.6 Kontekst, endring og prosess

For å belyse organisasjoners kompleksitet ser vi det som hensiktsmessig å fokusere på tre forhold: kontekst, endring og prosess. Dette kan bidra til å redusere gapet mellom teori og praksis, der en ønsker å fange opp ulike dimensjoner i organisasjoners kontekst, endringer på individ- og organisasjonsnivå, og organisasjoners påfølgende tilpasninger for å forbedre

evalueringer. Disse elementene kan være med på å utfordre det tradisjonelle perspektivet på evaluering.

2.6.1 Et paradigmeskifte: fra kompetanse til kontekst

Siden årstusenskiftet har det vært et skifte i forståelsen av ledelse i en bredere kontekst, i motsetning til det mer tradisjonelle og individualistiske fokuset på ledelse som et dyadisk forhold mellom leder og medarbeider. Det nye skiftet peker på at organisasjoner har et økende behov for å utvikle ledere som kan operere på tvers av organisasjoner, samt i skiftende kulturer, kontekster og situasjoner i læringsprosessen. Dette forsterker dermed utfordringen til det tradisjonelle og individualistiske perspektivet på ledelse (Edwards & Turnbull 2013:46).

Ifølge Edwards og Turnbull (2013:46), er problemet med dagens tenkning rundt evaluering av lederutviklingsprogrammer, at en ikke tar tilstrekkelig hensyn til at læring påvirkes av den kulturelle konteksten ledere opererer i. I lys av dette, foreslås det en økt forståelse av kulturelle faktorer, både organisatorisk, regionalt og nasjonalt, når en evaluerer lederutviklingsprogrammer. Deres tilnærming inkluderer både et helhetlig syn på evaluering for å se utvikling i sin bredere kontekst, men borer også dypt inn i de ulike interaksjonene og nettverkene som eksisterer. Den søker også å forstå hvordan ledelsesdynamikken påvirkes av en organisasjons historie, og er derfor sensitiv for kulturelle dimensjoner som tid, språk, symboler, ritualer, mytologi og maktstrukturer. Det nye paradigmeskiftet beveger seg vekk fra ROI-diskursen, ettersom en ikke kan bedømme programmenes effektivitet uten en forståelse av "råstoffet" i bedriften og kulturen som disse individene opererer i. På sitt mest komplekse søker en å forstå ikke bare individuell innvirkning, men også innvirkning på organisasjonen og samfunnet, over tid, sted, relaterte nettverk og organisasjonsnivåer. En må vurdere flere interessenter med flere oppfatninger og erkjenne at ledelse aldri kan være en absolutt vitenskap, men heller en kunst (Edwards & Turnbull, 2013:57).

2.6.1.1 En balanse mellom ytelse og læring

Tradisjonell ledelsesevaluering har også i stor grad fokusert på individuelle kompetanseendringer og selvevaluering, og har tatt lite hensyn til kognitive endringer, endret motivasjon og selvbevissthet. Det er et betydelig behov for økt kunnskap om målingen av kognitive attributter og motivasjonsattributter som bidrar til styrke ledelsesprosesser. For å få en helhetlig evaluering av ledelsesutviklingstiltak er det behov for en balanse mellom ytelse og læring på både individ- og gruppenivå. Ytelse er en viktig pådriver for investering i

ledelsesutviklingstiltak, men faren ved å kun måle ytelse, er at en ikke fanger opp viktig læring, eller at en fanger opp endring som egentlig skyldes eksterne faktorer. Det betyr imidlertid ikke at en kun bør fokusere på læringsfaktorer, da ytelse og læring utfyller hverandre. Gode resultater i ytelse kan indikere at en "gjør de riktige tingene", mens læring kan indikere at en "gjør tingene på riktig måte" (Wallace, Torres & Zaccaro, 2021:1-3).

Som et svar på dette behovet har forfatterne utviklet et rammeverk som representerer utviklingskriterier på flere nivåer og flere dimensjoner. Formålet er å kunne måle både utvikling i ferdigheter og kognitiv utvikling. En oppfordrer lederutviklere og evaluatorer til å sette klare mål før programstart, som implementeres i evalueringen. Evaluering bør være en kombinasjon av kvantitative og kvalitative data. I tillegg vil læringen og modningen hos leder være en prosess som skjer over tid, og en anbefaler derfor å ha flere målinger over en lengre tidsperiode (Wallace et al., 2021:10).

Wallace et al. (2021:3) illustrerer sin tilnærming gjennom et klassifiseringsskjema for å vise læringsutbyttet fra ledelsesprogrammer. Der ser vi at koblingene mellom de individuelle og kollektive aktivitetene fanger interaksjoner på tvers av nivåer og sammenhenger mellom leder- og ledelsesutvikling. Ledere påvirker kollektive prosesser som eksempelvis sensemaking og tilrettelegging for motivasjon. Dette vil igjen påvirke endringer i lederes evner til å utføre disse aktivitetene og kan føre til utvikling av ledelse på et kollektivt nivå.

Modellen tar for seg direkte ytelsesresultater, førsteordens-læring, andreordens-læring og nullordens-læring. "Direkte ytelsesresultater" handler om individuell atferdsendring og organisatorisk ytelse. Likevel representerer ikke dette læring i seg selv, men er en form for forløpet for læring. "Førsteordens læring" refererer til endringer i kunnskap, ferdigheter og evner i individer og prosesser i organisasjonen, som fører til ledelse. Videre er "andreordens-læring" resultatet av modningsprosesser som resulterer i endrede ledelsesidentiteter og epistemologier. Dette styrker videre "nullordens-læring", som fungerer som et grunnlag for videre utvikling (Wallace et al., 2021:3).

2.6.2 CIRO

I 1970 ble CIRO-modellen for evaluering av lederutviklingsprogram foreslått (Warr, Bird & Rackham, 1970). Denne modellen var basert på evaluering av fire aspekter ved lederutvikling: kontekst, input, reaksjon og resultat. CIRO-modellen fokuserer på måling både før og etter at programmer er gjennomført. Hovedstyrken til modellen er at målene (kontekst) og verktøyet

(input) vurderes. Kontekstevaluering fokuserer på faktorer som identifisering av behov og fastsetting av mål tilknyttet organisasjonens kultur og klima. Input-evalueringen er opptatt av utforming og levering av aktivitetene i programmet. Videre fokuserer reaksjonsevaluering på å innhente og benytte seg av informasjon om erfaringene av programmene. Til slutt innebærer resultatevaluering hva en har oppnådd på bakgrunn av programmet, der en vurderes på tre nivåer: umiddelbar, underveis og endelig evaluering. Umiddelbar evaluering forsøker å måle endringer i kunnskaper, ferdigheter eller holdninger før en deltaker returnerer til arbeidsplassen. Underveisevaluering handler om effekten av kurset på deltakernes jobbytelse og hvordan læringen overføres tilbake på arbeidsplassen. Til slutt forsøker den endelige evalueringen å vurdere effekten av kurset når det gjelder resultatene på avdelings- og organisasjonsnivå (Topno, 2012:21).

2.6.3 360-graders lederevalueringer

360-graders lederevalueringer innebærer å innhente vurderinger av den enkelte leders jobbrelaterte atferd, der en vektlegger ønsket utvikling fra flere kilder i lederens omgivelser. Dette kan gjøres uformelt, der lederens overordnede stiller spørsmål til kolleger, underordnede eller kunder som har grunnlag til å uttale seg om lederen det gjelder. Deretter benyttes disse vurderingene i den overordnedes vurdering av personen. Ved en mer formell art, innhenter en vurdering fra lederens overordnede, og en stiller krav til minimum tre vurderere per kilde. Dette er for å ivareta kildens konfidensialitet, og av hensyn til validitet og reliabilitet i målingene (Kuvaas & Dysvik, 2016:146).

Hensikten med formelle 360-graders vurderinger, er å gi lederen en samlet og balansert tilbakemelding, som skal bidra til utvikling. Det er en rekke forhold som kan påvirke resultatene, der det bl.a. er vanlig at de som vurderer ikke har tilstrekkelige eller riktige kvalifikasjoner til å foreta nøyaktige evalueringer. Ofte blir vurderingene også preget av "politisk korrekte" vurderinger dersom medarbeiderne frykter at deres konfidensialitet blir truet, der resultatene igjen kan medføre alvorlige konsekvenser for en selv eller deres kolleger (Kuvaas & Dysvik, 2016:145-146).

2.6.4 Evalueringsmodell basert på Theory of Change

Watkins, Lysø og deMarrais (2011) introduserer en evalueringsmodell basert på "Theory of Change", som utforsker og fanger opp endringer på individ- og organisasjonsnivå, samt utvikling basert på mer åpne mål. Dette er på bakgrunn av at forskere har opplevd det som utfordrende å utvikle evalueringsmodeller for tradisjonell lederevaluering, som måler komplekse programresultater utover individuell læring (Watkins et al., 2011:209).

Theory of Change modellen innebærer at en først spør de ansvarlige for programmet om å identifisere hva som skal endres hos deltakerne etter programmet. Deretter skal en angi relevante aktiviteter, og vise hvordan de er knyttet til tiltenkte resultater av programmet som skal føre til endring (Watkins et al., 2011). Dette gjøres på følgende fire nivåer: Kortsiktig individnivå: endringer i kunnskaper og ferdigheter; intermediært individnivå: endringer i deltakeratferd; organisasjonsnivå: endringer i organisasjonen; og systemnivå: endringer som resulterer i bl.a. nye retningslinjer og prosedyrer. Av dette ser en at modellen vektlegger endringer på ulike nivåer, fremfor konkrete resultatmål. Modellen gjenspeiler kompleksiteten i åpne og lite målbare resultater, og fokuserer på organisatorisk kapasitetsbygging fremfor konkrete mål (Watkins et al., 2011:214-215)

Modellen til Watkins et al. (2011:216) baserer seg videre på kvalitative evalueringsmetoder med dybdeintervjuer. Dette er for å best mulig evaluere endringer og virkninger av programmene, samt imøtekomme det høye nivået av kompleksitet i organisasjoner. I tillegg har de benyttet seg av et verktøy som kalles "Critical Incident Technique", som er en teknikk for å samle inn data der spørsmålene blir formulert som "tenk på en gang du.." (Watkins et al., 2011:216). Hensikten med dette er at deltakeren skal fortelle om spesifikke situasjoner hvor de har brukt kunnskapen de har lært fra programmet innen hvert av kjernekompetanse-områdene. Evaluatorene kan ut ifra disse dataene identifisere i hvilken grad programmet faktisk fører til tiltenkte utfall.

2.6.5 Prosessevaluering og four-phase approach

I forskningsartikkelen til Lingham, Rischely og Rezania (2006:334) argumenteres det for at evalueringer bør tilpasses spesifikt det programmet det skal evaluere. Det anbefales videre at programevalueringer ikke bare omfatter opplæringsprosessen som helhet, men også tilbakemeldinger fra deltakerne når det gjelder innhold og anvendelighet av programmene

underveis. Et godt evalueringssystem anses for å være en samarbeidsprosess mellom organisatoriske ledere, kursholdere, deltakere og evaluatorer, slik at en har en felles forståelse av hva formålet med programmet er. For å kunne måle effekt, er det behov for en bedre forståelse av funksjonen til evalueringen en tar i bruk, dets formål og andre viktige relevante aspekter, som eksempelvis organisasjonens og deltakernes behov. Ved å gjøre dette tilpasses evalueringen så spesifikt som mulig til det programmet den skal evaluere (Lingham et al., 2006:335). Programmet som blir evaluert i artikkelen strekker seg over lenger tid, består av flere seksjoner, og har ulike grupper som gjennomfører seksjonene parallelt. Det vil si at når gruppe 1 er ferdig med seksjon 1, begynner de på seksjon 2, samtidig begynner gruppe 2 på seksjon 1 og så videre (Lingham et al., 2006:342). For å evaluere ledelsesutviklingsprogram fremlegges det en evalueringsprosess inndelt i fire faser, kalt four-phase approach.

Fase 1 *Design av evalueringsprosess før programstart*, innebærer at kursholder og organisatorisk leder samarbeider og utvikler en evaluering som er spesifikt tilpasset hvem som skal delta og hva en ønsker å oppnå med programmet. Evaluatoren deltar allerede fra start, og avklarer hvordan hen skal være delaktig underveis i ledelsesutviklingsprogrammet (Lingham et al., 2006:338). I fase 2 *Starten av programmet,* følger evaluatoren evalueringsopplegget de har blitt enige om før kursstart. Videre har de i dette tilfellet kommet frem til å intervjue 20 % av deltakerne om hvordan de har opplevd programmet så langt. Det ble også gjennomført deltakende observasjon, der evaluatoren var til stede ved de første leksjonene og noterte utfordringer (Lingham et al., 2006:340)

Videre er i fase 3 *Feedback tas til etterretning og programmet endres*, blir intervjuene og observasjonen fra gruppe 1, som har gjennomført de to første leksjonene, anvendt til å gi til tilbakemeldinger til kurslederne slik at de kan forbedre de første seksjonene til nye deltakergrupper. I tillegg tar de utgangspunkt i temaer som har kommet opp i intervjuene, og bruker disse til å utvikle hypoteser om hvordan programmet vil gi best læringsutbytte. Ut i fra hypotesene lager de så ulike spørsmål som kvantitativt skal måle deltakernes utbytte av programmet (Lingham et al., 2006:340-341). I fase 4 *Pågående trening og evaluering,* besvarer deltakerne spørsmålene etter hver leksjon de deltar på. Slik kan en følge utviklingen deres gjennom programmet, og se om holdningene endrer seg. Samtidig kan de ulike seksjonen kontinuerlig forbedres til nye grupper som skal gjennomføre programmet (Lingham et al., 2006:342).

Denne studien viser at ved å benytte seg av en slik prosessevaluering inndelt i fire faser, blir evalueringen løpende med på å forbedre ledelsesutviklingsprogrammet (Lingham et al., 2006) Da programmet styrkes mens det pågår, er det ikke nødvendig å vente med å se på resultatene til programmet er avsluttet. Samtidig er de kvantifiserte dataene fra atferdsspørsmålene som deltakerne blir stilt etter hver seksjon, med på å vise konkrete sammenhenger mellom innholdet i programmet og deltakernes atferds- og holdningsendringer. Et godt evalueringssystem er en samarbeidsprosess mellom organisasjonsledere, kursholdere, deltakere og evaluatorer (Lingham et al., 2006).

2.6.6 Value of Learning Model

Andersons value of learning modell (2007) peker på ubalansen mellom forventningene til utviklingsprogrammer og forretningsmålene. Modellen oppmuntrer til samsvar mellom målene ved utviklingstiltakene og de strategiske målene til organisasjonen i evalueringen. Anderson (2007) beskriver en tre-trinns evalueringssyklus for læring på organisasjonsnivå, heller enn for spesifikke individuelle tiltak. Første trinn handler om å bestemme endringstiltak opp mot strategiske prioriteringer. Videre er neste steg å benytte ulike metoder for å vurdere og evaluere læringsbidrag. På det tredje trinnet etablerer en de mest relevante tilnærmingene for organisasjonen og bestemmer hvilke tiltak som skal iverksettes. Tiltakene skal gjerne også ha effekt på større strategiske mål, slik som avkastning på investeringer (Singh & Ramdeo, 2020:410-411).

2.6.7 SMART-mål

SMART er en metode for å skape og formulere oppnåelig og målbare mål (Chen, 2015:231). Chen trekker frem bruken av SMART-mål som et fordelaktig verktøy nettopp i settinger der kunden bestiller program med et formål om generell forbedring (Chen, 2015:232). SMART ble opprinnelig utviklet for å gi en modell til å sette mål i prosjektledelse. SMART er et akronym, der S står for spesifikk (specific), M står for målbar (measurable), A for oppnåelig (achievable), R for relevant og T for tidsbundet (time-bound).

2.6.8 Deloppsummering

Med utgangspunkt i teorikapittelet, tar vi med oss noen betraktninger videre til oppgavens analysekapittel. Her vil de empiriske funnene fra intervjuene våre sees opp mot de teoretiske aspektene, og vi vil forsøke å sammenfatte de to perspektivene. Vi vil blant annet ta med oss betydningen av å vektlegge kontekstuelle forhold, og hensikten med å evaluere aktiviteter i ulike faser. Kirkpaticks modell danner bakgrunnen for en drøfting rundt hensikten med å kartlegge deltakernes tilfredshet, hvordan læring og atferdsendring kan måles for å vurdere utbytte, og forskjellen på å utforske endringer på individ- og organisasjonsnivå. Vi vil videre vurdere betydningen av å sette målbare mål før evalueringen skal starte, og vurdere i hvilke situasjoner det er hensiktsmessig å ta i bruk kvalitative metoder, og i hvilke det er mer hensiktsmessig med kvantitative. Til slutt vurderer vi forskjellen på teoretiske og praktiske betraktninger tilknyttet hvordan ledelsesutviklingsprogram bør evalueres, og diskuterer hvorvidt teoretiske modeller i praksis er anvendbare.

3.0 Metode

3.1 Forskningsdesign

Da formålet med vår studie er å undersøke og forstå hvordan AFF på best mulig vis kan evaluere sine ledelsesutviklingsprogram, fant vi det hensiktsmessig å benytte oss av et *eksplorativt design*. Dette er et design som er egnet for å utforske et tema en har lite kjennskap til fra før, og heller ingen klare formeninger eller antakelser om hvilke mulige sammenhenger som finnes (Tjora, 2017:257). Med dette som utgangspunkt, bestemte vi oss for å undersøke tematikken ved hjelp av en kvalitativ metode, bestående av én del med litteratursøk, og én del ekspertintervju. Slik opplevde vi at vi best ville få bredde og oversikt over evalueringsfeltet, som kunne sette oss i stand til å besvare oppgaven fra AFF på en kvalifisert måte.

3.1.1 Litteratursøk

Som en del av mandatet fra AFF etterspørres det: "kartlegging av forskning på evaluering av utviklingstiltak", derfor har vi benyttet oss av strukturerte søkerekker for å systematisk hente inn relevant og oppdatert forskning og teori på evalueringsfeltet. Vi benyttet søkerekkene i utvalgte søkeportaler etter veiledning med en bibliotekar knyttet til emnet SVPRO4000. Søkeprotalene var Oria (en fellesportal for fysisk og digitalt materiale for de fleste norske fag- og forskningsbibliotek), Emerald Insight (en fulltekstdatabase av tidsskrifter innenfor emner,

som bl.a. ledelse) og Web of Science (en tverrfaglig database, som deriblant dekker samfunnsvitenskapen). De strukturerte søkerekkene ble laget ut ifra relevante begreper innhentet fra mandatet, som; evaluering, ledelse, leder, organisasjonsutvikling, Kirkpatrick og Brinkerhoff. Disse begrepene ble brukt både på norsk og engelsk i diverse kombinasjoner for å lage forskjellige søkerekker. Ut ifra resultatet av søkerekkene kunne vi utvikle ekstra søkerekker til å inkludere nyfunnet begrep/teoretikere/modeller, eksempelvis: Kaufman, CIRO, Anderson og Phillips. Til slutt hadde vi totalt syv søkerekker. For å sikre kvaliteten på artiklene vi fant gjennom det strukturerte litteratursøket, benyttet vi oss av NSDs register over vitenskapelige publiseringskanaler for å sjekke nivået på de aktuelle tidsskriftene. Dette kan gi en viss indikator på kvaliteten av forskningen. Kvalitetssikringen ble forsterket da vi også sjekket antall siteringer på artiklene så langt det var mulig.

3.1.2 Dybdeintervju

Gjennom intervjuene hadde vi som mål å få tilgang til å forstå aktuelle fagfolk sine opplevelser og personlige erfaringer tilknyttet til evaluering, samtidig som flere av dem er forskere på relaterte felt. Vi vurderte at dybdeintervju og fokusgruppeintervju ville være de best egnede datainnsamlingsmetodene for å få innsikt i, samt studere, informantenes meninger, holdninger og erfaringer, og anså derfor disse intervjuformene for å være mest hensiktsmessige (Tjora, 2017:113). Av hensyn til logistikken tilknyttet å få gjennomført et fokusgruppeintervju med informanter som potensielt satt spredt rundt om i landet, og at vi ikke hadde så god tid på å gjennomføre intervjuene, falt valget på å legge fokusgruppeintervju til side, og kun gjennomføre semi-strukturerte dybdeintervjuer (Tjora, 2017:114). Vi vurderte den store fordelen med semi-strukturerte intervju til å være at vi kunne ha en relativt fri samtale rundt noen forhåndsbestemte temaer vi ønsket å få vite mer om, hvor informanten fikk mulighet til å reflektere rundt, og utdype sine svar, og dermed gi oss tilgang til mer detaljert data (Tjora, 2017:113).

For å holde en viss struktur på intervjuene, benyttet vi oss av en intervjuguide. Vi holdt intervjuene to og to, og intervjuguiden ble et nyttig verktøy som både sikret at vi dekket alle viktige temaer og ga en viss forutsigbarhet (Tjora, 2017:158). Noen spørsmål ble tilpasset hver enkelt informant med utgangspunkt i deres kompetanse og erfaringer, slik at vi kunne få et dypdykk i denne, mens andre spørsmål ble stilt felles til alle informantene uavhengig av hva slags bakgrunn de hadde.

Det var viktig for oss å ikke legge føringer for informantene ved å gå detaljert inn på evalueringsteoriene vi hadde lest om på forhånd. Grunnen til dette var at vi ønsket å unngå et scenario hvor de antok hvilke svar vi ønsket basert på teoriene vi eventuelt hadde tatt opp under intervjuet.

3.2 Utvalg

Rekrutteringen av informanter til studien ble gjort med utgangspunkt i at vi ønsket å samle et strategisk utvalg, som kunne bidra til å kaste lys på problemstillingen med bakgrunn i kvalifiserte og reflekterte erfaringer fra evalueringsfeltet (Tjora, 2017:130). For å få en bred innsikt i feltet, både fra et teoretisk og et praktisk perspektiv, valgte vi å søke etter to grupper av informanter. Den første gruppen kalte vi «teoretikere». Disse var akademikere som hadde forsket på enten evaluering eller ledelsesutvikling, og som hadde mesteparten av sin kunnskap på feltet fra teoretiske studier. Den andre gruppen var «praktikere». Dette var personer som enten jobber eller har jobbet med evaluering, og som har mesteparten av sin evalueringskunnskap fra praktiske gjennomføringer av evalueringer. Dybdeintervjuene våre var dermed en form for eliteintervjuer, der informantene ble valgt på bakgrunn av sin ekspertkompetanse.

Årsaken til at vi ønsket å snakke med teoretikere, til tross for at flere av disse hadde publisert skriftlige kilder om evaluerings- og ledelsesutviklingsfeltet som vi leste og satt oss inn i, var at vi var interesserte i deres personlige erfaringer og opplevelser tilknyttet tematikken. Vi ønsket ikke ta ta dem i bruk som en lærebok, men vi ville få til en reflekterende samtale der vi diskuterte problemstillinger med bakgrunn i deres forståelse og innsikt i tematikken vi diskuterte.

For å søke etter teoretikere, startet vi med å notere oss hvem som hadde skrevet bøkene og forskningsartiklene vi fant gjennom vårt strukturerte litteratursøk. Vi undersøkte deretter hva slags fagbakgrunn de hadde, hva slags forskningsartikler de ellers hadde skrevet, og hvem de hadde samarbeidet med. Slik dannet vi et oversiktsbilde over teoretikerens kjennskap til temaet, og kontaktet dem vi anså å ha størst kunnskap om evalueringsfeltet. Spesielt interessante var teoretikerne som hadde studert evaluering i forbindelse med ledelsesutviklingsprogram, da dette spesifikt var det vi søkte kunnskap om. Vi gjorde også egne søk på nettet, og prøvde å kartlegge hvilke teoretikere som kunne komme med gode refleksjoner til studien. I dette søket

ble universiteter og høyskoler sine nettsider spesielt nyttige, da de ofte inneholder omfattende presentasjoner av ansatte og deres kvalifikasjoner.

Rekrutteringen av praktikere ble gjort ved å ta i bruk Google og prøvde å få oversikt over personer, organisasjoner og firmaer som tilbyr å gjennomføre evalueringer. Også her var vi spesielt interessert i dem som hadde spesifikk kjennskap til evaluering av ledelsesutvikling, men vi var også åpne for å snakke med folk som hadde mer generell kjennskap til evaluering. Bakgrunnen for dette var at vi antok at fagpersoner som jobber, eller har jobbet med evaluering, kunne komme med gode betraktninger rundt hvordan en generelt bør gå frem for å utvikle en anvendelig evalueringsmodell. Dette anså vi å være nyttig for vårt prosjekt, der det å finne ut hvordan en evalueringsmodell kan utvikles til å bli funksjonell er avgjørende. Vi antok at praktikere ville ha flere perspektiver tilknyttet dette enn teoretikerne, da disse er nødt til å ta hensyn til slike forhold når de utvikler evalueringsmodeller.

Vi formulerte en tekst som inneholdt en kort beskrivelse av prosjektet, et lenger informasjonsbrev som forklarte studiens formål i detalj og hva slags informanter vi søkte etter. Da vi kontaktet potensielle informanter, oppfordret vi dem til å selv vurdere om de hadde kunnskap og innsikt de mente kunne bidra til studien. Ved å formulere oss på denne måten håpet vi å sikre at de som takket ja til å stille til intervju var personer som hadde faglig kjennskap til evalueringsfeltett. Spesielt viktig var dette når det gjaldt praktikerne, hvor det ikke var like lett å kartlegge deres formalkompetanse tilknyttet evaluering, sammenlignet med teoretikerne som stort sett hadde åpne CVer med opplistinger av publikasjoner og formalkompetanser.

Vi startet med å kontakte teoretikere med ulik fagbakgrunn, i et ønske om å få ulike perspektiver på studiens tematikk. Da responsen imidlertid ikke var like stor som vi håpet, forhørte vi oss etter hvert med flere teoretikere med samme fagbakgrunn. Ettersom intervjuene ble gjennomført tok vi også i bruk snøballmetoden for å rekruttere ytterligere informanter, og spurte informantene om de kjente til personer vi burde snakke med (Tjora, 2017:135). Denne metoden var med på å sette oss i kontakt med flere nye informanter.

Noen av de aktuelle informantene vi kontaktet, takket nei. En av disse personene oppga at årsaken til dette var at deres arbeidsplass var å anse som en konkurrent til AFF. En annen

person valgte å takke nei, fordi de var imot bruk og utvikling av en generisk modell for evaluering. Dette er personer som vi vurderte til å kunne bidra med gunstige data til prosjektet, både med tanke på deres bakgrunn og erfaringer.

Utvalget vårt besto til slutt av fire praktikere og fire teoretikere:

Jan Ketil Arnulf er dekan for etter- og videreutdanning og professor ved Institutt for ledelse og organisasjon på BI. Fokuset for Arnulfs forskning har ligget på hvordan ledelse utøves og skapes, samt organisasjonsmålinger. Arnulf har tatt både master- og doktorgrad i psykologi ved Universitetet i Oslo, og har jobbet på BI de siste 17 årene.

Hans Meisingset har jobbet med lederutvikling som konsulent, bedriftsinternt, og jobber i dag med lederutvikling i Forsvaret. Tidligere har han både jobbet med, og studert effekten av, lederutvikling i Posten i forbindelse med videreutdanning (masternivå) på BI. Meisingset har pedagogisk utdannelse både fra Norges Idrettshøgskole og krigsskoleutdannelsen i Forsvaret.

Lars Ivar Slemdal jobber som konsulent innenfor "Management Consulting" i Rambøll, hvor hans arbeidsoppgaver i stor grad er evalueringer og kartlegginger av ulike slag. Han har bl.a. tidligere evaluert et lederutviklingsprogram for Helsedirektoratet som har blitt utviklet og holdt av BI. Slemdal har en bachelorgrad i pedagogikk og en mastergrad i Organisasjon, ledelse og arbeid - begge fra Universitetet i Oslo.

Marit Horgen har jobbet med lederutvikling i 20-25 år hos bl.a. AFF og Politiet, og jobber for tiden i Forsvaret. Det er fra Forsvaret og Norges Idrettshøgskole hun har sin utdanning i fagkombinasjonen idrett, ledelse og organisasjonspsykologi. I tillegg har hun videreutdanning fra både Høgskolen i Lillehammer, Universitetet i Bergen og BI innen lederutvikling.

Caroline Howard Grøn er amanuensis ved "Kronprins Frederiks Center for Offentlig Ledelse" ved Aarhus Universitet og forsker på offentlig ledelse. Hun har tatt master- og doktorgrad i statsvitenskap fra Institut for Statskundskab ved Københavns Universitet, hvor hun også tidligere har vært ansatt som amanuensis. Grøn har vært med til å bidra til utviklingen av et nytt offentlig lederevalueringsverktøy i forbindelse med den danske regjeringens ledelses- og kompetansereform.

Tom Lindtein er spesialrådgiver på BI Corporate (avdelingen for bedriftsinterne program) og har erfaring som ledelseskonsulent i internasjonale bedrifter i både Asia, Midtøsten og Europa. Tidligere har Lindtein også ledet BI Corporate i åtte år og vektlegger et sterkt fokus på verdibidrag for kunder. Lindtein er utdannet sivilingeniør i industriell økonomi og teknologiledelse ved NTNU, i tillegg har han tatt diverse etterutdannelser i bl.a. strategi og internasjonal handel.

Alexander Madsen Sandvik er førsteamanuensis ved Institutt for Strategi og ledelse ved NHH. Hans forskningsinteresser er hovedsakelig ledelse og organisatorisk atferd. Han er utdannet ingeniør fra Universitetet i Bergen og har videreutdannet seg ved NHH, der han tok en doktorgrad i Ledelse av kunnskapsarbeidere. Sandvik var nevnt i mandatet fra AFF, som ønsket informant, da han tidligere har hatt ansvaret for en evaluering av AFFs ledelsesutviklingstiltak.

Einar Lier Madsen har vært forskningsleder ved Nordlandsforskning, og har tidligere jobbet med å evaluere diverse bedriftsrettede offentlige virkemidler hos Innovasjon Norge. Han var med på å etablere, og har ledet, Norsk Evalueringsforening i fire år. Madsen er utdannet siviløkonom og har tatt en doktorgrad i Entreprenørskap fra Handelshøgskolen i Bodø ved Nord Universitet.

3.3 Gjennomføring av intervju

Vi anså det som hensiktsmessig å gjennomføre intervjuene to og to, da dette ville føre til mer effektivitet og større grad av struktur enn om vi hadde stilt alle seks, men hvor man var flere enn én og dermed kunne samarbeide om å stille gode spørsmål og å få mest ut av samtalene. Intervjuene ble i hovedsak gjennomført via videokonferanseprogrammet «Zoom», mens to ble holdt via telefon og ett ble gjennomført fysisk. I forbindelse med koronapandemien har man blitt vant med digitale kommunikasjonsløsninger. Digitale videokonferanser er effektive både fordi det gjør at informanten kan snakke med oss fra hvor som helst i verden, og fordi informanten ikke trenger å beregne tid i forkant eller etterkant av intervjuet til å møte de som skal intervjue ham eller henne. Vi vurderte det derfor å være en effektiv og praktisk måte å gjennomføre intervjuene på både for oss og informantene.

Det første intervjuet som ble gjennomført via telefon opplevde vi utfordrende. Det var informanten selv som ønsket at intervjuet skulle foregå via telefon, og derfor falt valget på dette. Det var imidlertid vanskelig å kommunisere når man ikke så hverandre, og samtalen fløt ikke like godt som ved videointervju. Det var lett å avbryte hverandre når man ikke kunne lese hverandres kroppsspråk, og det var av og til utfordrende å komme til ordet av samme grunn (Tjora, 2017:169). Etter dette intervjuet ble vi derfor bevisste på at dette var en intervjuform vi ønsket å unngå i fremtiden (Tjora, 2017:171). Da vi allikevel ble nødt til å holde nok et telefonintervju som følge av dårlig internettforbindelse, gikk imidlertid samtalen bedre. I forkant av dette intervjuet hadde informanten forespurt å få tilsendt intervjuguiden, og samtalen forholdt seg derfor veldig strukturert ut ifra denne. Dette gjorde trolig at telefonintervju nummer to var lettere å gjennomføre enn det første.

For å sikre at det som ble sagt under intervjuene ble korrekt gjengitt ved transkribering og koding av dataene, tok vi lydopptak av intervjuene. Informantene ble informert om bruken av lydopptak og det ble gjort kun etter deres samtykke. De ble også forsikret om at lydopptakene ville bli slettet umiddelbart etter transkribering (Tjora, 2017:166-167).

3.4 Forskerrollen og etiske hensyn

I studien vår ble gjennomføring av intervju gjort for å innhente kvalifiserte betraktninger fra fagpersoner om studiens tematikk, som vi kunne bruke til å se opp mot teorien vi hadde innhentet på feltet. For å gi tyngde og troverdighet til informantenes betraktninger, ønsket vi å være åpne om deres navn, stillingstittel og faglige erfaringer på feltet (Tjora, 2017:178).

Før rekrutteringen av informanter begynte, formulerte vi et informasjonsbrev som uttrykte hva studien gikk ut på og hva vi ønsket å få ut av intervjuene. Ved å sende informasjonsbrev til alle mulige informanter, kunne vi sikre at alle informantene ville være informerte og komfortable rundt hva intervjuet skulle brukes til, og dermed være trygg på hva det ville innebære å takke ja til å delta. De fikk også tilbud om å være anonym dersom de ønsket, og de måtte samtykke til at intervjuene kunne tas opp med lydopptaker. Det var imidlertid ingen av informantene som ønsket å være anonyme.

For å gjøre det så komfortabelt som mulig å stå frem med navn i studien, var vi svært bevisste på å styre unna sensitive temaer i intervjuene. I utgangspunktet var tematikken for intervjuet faglig, der det ikke var nødvendig å komme inn på sensitiv informasjon om informantene. Flere

av informantene, i hovedsak praktikerne, måtte imidlertid inkludere en del fortellinger fra eget arbeidsliv for å kunne sette konteksten rundt refleksjonene de kom med. Dette er tematikk som *kan* være sårbar for noen, spesielt i de tilfeller der konkrete situasjoner kommer opp. Både av hensyn til dette, og av hensyn til at personer vi har gitt en «eksperttittel» naturlig nok ønsker å uttale seg korrekt, tilbudte vi informantene å gjennomføre sitatsjekk før vi publiserte rapporten. Slik kunne de ha kontroll over hva som ble publisert i deres navn, og få mulighet til å ordlegge seg på en annen måte dersom de ønsket (Tjora, 2017:177-178). Vi anså sitatsjekk for å være nyttig for oss også, fordi det kunne være med på å sikre at vi faktisk hadde forstått det informanten forsøkte å si (Tjora, 2017:179).

3.5 Koding

For at alle seks på gruppa skulle få en felles forståelse av intervjuene, til tross for at man ikke hadde vært med på å gjennomføre alle selv, var vi opptatt av å ha en grundig og transparent analyseprosess. Første steg i denne var å ta utgangspunkt i lydopptakene fra intervjuene, og å transkribere disse. Slik fikk vi en ordrett gjengivelse av hva intervjuene hadde dreid seg om, som var fri for tolkning og egne betraktninger. Å transkribere intervjuene nøye og ordrett var også nyttig med tanke på informantenes ønske om sitatsjekk. Gjennom gode transkripsjoner sikret vi at informantene ble sitert korrekt i henhold til hvordan de ordla seg (Tjora, 2017:173-175).

Etter intervjuene var transkribert, kodet vi disse empirinært. Empirinære koder lages ut i fra det informantene *faktisk* sier, og er dermed kun utviklet ut i fra data, og ikke fra teori, hypoteser, forskningsspørsmål eller planlagte temaer. Slik unngår man tolkninger og premature konklusjoner (Tjora, 2017:198). De empirinære kodene ble formulert som korte setninger, som konkret og kortfattet fortalte hva essensen i et utvalg setninger eller korte avsnitt var (Tjora, 2017:197). Slik kunne vi få et raskt inntrykk av hva informanten snakket om i et utdrag av intervjuet uten å måtte gå tilbake til transkripsjonene. Vi benyttet oss av Nvivo, et kvalitativt dataanalyseprogram, til å kode intervjuene (Tjora, 2017:229).

Etter at alle intervjuene var kodet empirinært, tematiserte vi kodene ut i fra hva slags fenomener de belyste. Ønsket var at en slik tematisering ville gjøre det lettere for oss å analysere datamaterialet fra de åtte ulike intervjuene, da vi lettere kunne finne frem til når informantene hadde snakket om felles tematikk. Slik ble det lettere å se de ulike perspektivene i sammenheng.

Kodeprosessen utgjorde starten av analyseprosessen, og analysen er styrt av empirien fra intervjuene.

3.6 Dataenes kvalitet

For å sikre at studien skulle ha høy pålitelighet og gyldighet, har vi vært opptatt av å ta nøye gjennomtenkte valg i alle ledd av studien og gjøre rede for disse gjennom et metodekapittel med høy transparens (Tjora, 2017:231-232). Det er viktig å nevne at vi la et premiss for alle intervjuene våre, der informantene våre ble oppfordret til å svare med utgangspunkt i dette. Gjennom å sette oss inn i evalueringsteorien har vi forstått at evalueringsteorien er svært kontekstuell, og at det kan være vanskelig å komme med en ideell og generisk modell som passer inn i alle evalueringssituasjoner. Vi forklarte derfor informantene hva mandatet vårt fra AFF var, og at vi - om mulig - ønsket at de skulle besvare våre spørsmål i lys av oppgavetematikken vår. I og med at vi ønsket å få tak på ekspertenes erfaringer tilknyttet tematikken, og vi ønsket å bruke denne til å utvikle vår egen evalueringsmodell, anså vi det som nyttig å legge dette premisset. Det gjorde at informantene forsto hvilken kontekst vår modell skulle utformes i, og at de i større grad reflekterte rundt hensiktsmessige metodiske grep for vår modellutvikling, istedenfor generelle betraktninger. Å legge et slikt premiss kan dog ha styrt informantene i en retning med tanke på hvordan de la frem sine svar, og kan dermed ha påvirket hva de trakk frem i intervjuene. Dessuten er det et sentralt premiss å lese analysen slik man forstår informantene og deres uttalelser i lys av deres kontekst og synspunkt. Det er naturlig at en praktiker vil vektlegge praktiske løsninger og en teoretiker vil begrunne seg med utgangspunkt i forskning og akademia.

3.7 Utforming av modell

Utformingen av modellen er basert på både teoretiske og analytiske avveininger. Sentrale valg blir diskutert og begrunnet i kapittel 5. Et førsteutkast av modellen ble presentert til oppdragsgiver for bl.a. å sikre modellens anvendelighet. Etter dialog med AFF fikk vi tilgang til en presentasjon av planene til et anonymisert ledelsesutviklingsprogram, for kunne komme med kvalifiserte eksempler til modellen. Dog inneholdt denne presentasjonen ikke detaljerte beskrivelser av hvilke konkrete aktiviteter som skulle foregå i løpet av programmet, og det var heller ikke mulig for oss å ta bedriftens strategi i betraktning i utformingen av eksemplene. Slike kontekstuelle hensyn burde tas i betraktning av AFF når de benytter seg av modellen. Behovet for dette diskuteres og argumenteres for i oppgavens analysekapittel.

4.0 Analyse

4.1 Kontekst

Gjennom intervjuprosessen fremkommer det at flere informanter anser det som viktig å tilpasse evalueringen etter kontekst. Med kontekstttilpasning menes her at evalueringen er tilpasset det spesifikke ledelsesutviklingsprogrammet som skal evalueres, eksempelvis ved å ta hensyn til programmets formål, metodikk og/eller deltakere. Informant Lars Ivar Slemdal trekker frem kontekst i sitt svar på spørsmål knyttet til hva han mener er viktigst å tenke på når man gjennomfører evalueringer:

> *Det er* [viktig] *å forstå konteksten godt, først og fremst, å forstå hvilket land* [man befinner seg i]*, eller omgivelsene rundt utviklingsprogrammene. Og så er det naturlig å sette seg godt inn i hva, hvordan og hvorfor programmet gjennomføres. Det er noe med å forstå tiltaket eller endringene som skal gjennomføres godt, både med mål for klientene og mål for programmet* [...] *Forstå innholdet i både aktiviteter og hvordan de gjennomføres og* [skaper] *god forståelse for programmet.*

Her eksemplifiserer Slemdal flere kontekstuelle faktorer, som f.eks. land, omgivelser, og programmets mål og innhold. En del av de ytterligere eksemplene fra informantene våre står i likhet med de Edwards og Turnbull kom med i sitt bidrag, som f.eks. kultur, organisasjonsstruktur, makt og omgivelser (2013:46). Dette viser at det er flere kontekstuelle faktorer som bør tas hensyn til.

Informant Marit Horgen fortalte oss at lederutviklingsprogrammene forandrer seg med jevne mellomrom avhengig av forskjellige faktorer. Derfor oppgir hun det som helt naturlig at innhold og spørsmål i en evaluering må tilpasses deretter. Våre funn antyder dermed at informantene våre ikke er tilhengere av en tilnærming som innebærer en generisk evalueringsmodell for ledelsesutviklingsprogrammer. Samtidig legges det vekt på at det er mulig å finne en balanse, der man tar utgangspunkt i en generell utforming, og tilpasser etter kontekst. Horgen sier følgende om dette:

> *Vi har et fast sett av spørsmål i forhold til utviklingsmoduler som vi har hatt ansvar for,* [...] *samtidig så endrer vi litt på innholdet og programmet og skrur litt på det, og da må vi jo tilpasse også evalueringsspørsmålene.* [...] *Det må skreddersys til en viss grad, men selvfølgelig så er det også noen standardformuleringer, hvis vi tenker på en sånn spørreundersøkelsestype som kan gå igjen.*

Her vektlegger Horgen viktigheten av standardisering i evaluering, f.eks. ved utforming av spørsmål i et spørreskjema. Naturligvis er det ikke hensiktsmessig å endre på alle spørsmålene hver gang man gjør en evaluering, men det påpekes likevel at hensynet til kontekst må være til stede. Formålet med å tilpasse etter kontekst kan bl.a. være å forbedre evalueringens kvalitet. Horgen nevner videre at det også er viktig for deltakerne å ha en god opplevelse av evalueringsprosessen. Hun antyder at standardiserte spørsmål fort blir gjennomskuet av de som skal bidra til evalueringen, og hvis det blir for mange slike spørsmål vil det hemme engasjement og motivasjon for å bidra. Slemdal beskriver på sin side at evaluering tilpasset til kontekst vil øke sannsynligheten for å kunne innhente viktig data, og legge til rette for at den blir mest mulig presis.

Som skrevet tidligere, ble tilgjengelige ressurser nevnt i forbindelse med kontekstilpasset evaluering. Det betyr at de som gjennomfører evalueringen ikke kun bør ta hensyn til deltakernes kontekst, men bør vurdere sin egen kontekstuelle situasjon ved å identifisere tilgjengelige ressurser. Einar Lier Madsen sier bl.a. at evalueringsmetoder i stor grad avhenger av hvor mye ressurser man har. De kanskje mest åpenbare ressursene man må vurdere, som ble nevnt under intervjuene, er tid og penger. Informantene er samstemte om at ideell evaluering er ressurskrevende å gjennomføre grunnet flere årsaker, men spesielt på grunn av mangel på tid og penger. Dermed må personen(e) som gjennomfører evalueringen medregne nok tid og penger hvis de ønsker å gjøre en omfattende evaluering, eksempelvis når de vil kombinere flere evalueringsmetoder eller benytte seg av en metode som er tilsynelatende omfattende.

Videre kan ressurser også være andre forhold, da det finnes utallige faktorer som kan betegnes som en ressurs. Andre ressurser som ble nevnt av informantene er kunnskap innen evalueringsfeltet, eller antall deltakere som skal bidra til evalueringen. Sistnevnte kan mangle i antall enten grunnet praktiske omstendigheter eller mangel på motivasjon for å delta. Informantene anser det som relevant at man involverer mennesker som best kjenner til lederne i programmet sin kontekstuelle situasjon. Det kan f.eks. være deres medarbeidere og nærmeste leder, som kan ha muligheten til å observere deres atferd i reelle situasjoner og innehar kjennskap til organisasjonen de jobber i. Informantenes syn samsvarer med både kontekstfokuset i CIRO-modellen og Edwards og Turnbull (2013) sine poeng. Begge understreker viktigheten av fokus på kontekst, og knyttet til dette nevner de at man må etterspørre innsyn i den aktuelle organisasjonens kultur og klima, og tilpasse evalueringen deretter (Warr et al., 1970). I likhet med informantene våre, anser disse teoretikerne det som

hensiktsmessig at man, under evalueringen, tar hensyn til innvirkningen kultur og andre organisatoriske faktorer kan ha på lederen. Dermed vil involvering av mennesker som befinner seg i det organisatoriske miljøet sammen med deltakerne være helt essensielt for å oppnå evalueringens formål.

4.2 Sette målbare mål

Gjennom intervjuene blir viktigheten av å sette mål trukket frem: Vi kan forstå målene som grunnlaget og objektet for en evaluering - kort oppsummert av Madsen: "Skal dere evaluere noe, så må dere vite hva målene er.". Mål er en essensiell rød tråd i både ledelsesutviklingsprogrammer og i evalueringsprosesser, og det er formuleringen av mål som er starten på en evaluering, slik som Jan Ketil Arnulf presiserer: "Når man drar i gang med sånne evalueringer, så må man tenke på selv på forhånd, altså *når* man planlegger en aktivitet, hva er det ved aktiviteten som man vil evaluere?". I evalueringsteorien argumenteres det for at den som evaluerer ideelt sett skal være en del av programmet og dets prosess gjennom hele programmets livssyklus (Boulmetis & Dutwin, 2011:49). Derfor er det også essensielt å begynne evalueringen samme tid som planleggingen av programmet.

I forbindelse med å sette og formulere mål, er det flere elementer som må tas i betraktning. Flere av informantene snakker om sammenhengen mellom en definisjon på ledelse og formulering av mål. Caroline Howard Grøn påpeker at det ikke er tilstrekkelig å gi en definisjon av hva god ledelse er og om hva som er den ideelle ledelsesformen:

> *Et argument om at sige, det der med at god ledelse er det samme alle steder, det holder ikke. God ledelse er kontingent og afhænger af hvilken organisation du er i, og ledelsesevaluering skal kunne understøtte den ledelsesudvikling som er relevant i en given organisation.*

Dette sitatet kan vi se i lys av Fourniers (1995) prosessforståelse av evalueringslogikken, der man starter å definere verdifulle kriterier og mål som skal være grunnlaget for de dimensjonene deltakerne skal forbedre seg på. Et videre trinn handler om å utforme standarder for hvor bra deltakerne skal gjøre det ut ifra definerte kriterier og mål. Elementet av å utforme standarder er viktig når definisjonen på ledelse ikke innehar en standard i sin natur.

Arnulf problematiserer også begrepet ledelse:

> *Det store problemet med ledelse er at ingen er enige om hva ledelse er.* [...] *Ledelse er i en særklasse av tåkete ord, som egentlig betyr veldig lite, og det største problemet med begrepet ledelse er at det er slått sammen med sin egen effekt.*

Videre argumenterer Arnulf for at fokuset for en evaluering ikke burde sentreres rundt begrepet ledelse, men heller de aktivitetene som brukes og trenes i til å forbedre og utvikle deltakerne i et ledelsesprogram:

> [...] *Mitt største problem med ordet ledelse, er at du kan ikke trene på ledelse. Ledelse er som ordet kunst; du kan ikke trene på kunst. Du kan trene på å spille gitar, eller du kan trene på å male akvareller, og alt sammen er kunst, men du kan ikke trene på kunst. På samme måte kan du ikke trene på ledelse. Du kan trene lederne på f.eks. å coache sine medarbeidere, på kommunikasjon og å få medarbeiderne til å føle seg sett og stimulert.*

Med utgangspunkt i disse problematiseringene av begrepet ledelse, er det derfor sentralt at man i utviklingen av mål utformer en standard for hva deltakerne skal oppnå, ettersom en definisjon av ledelsesbegrepet i seg selv ikke inneholder en slik standard. Dessuten kan vi med utgangspunkt i ovenstående sitat også argumentere for at man ikke skal sette opp for generelle mål i ledelsesprogrammet og som grunnlag for evalueringen. Arnulf trekker frem at generelle mål er en spesiell fare, når bedrifter kommer med et ønske om generell forbedring grunnet harde konkurranseforhold i bedriftens omgivelser:

> *Noen bedrifter har ikke spesifisert problemer, men ønsker at ledere skal prestere bedre generelt. De er omgitt av stor konkurranse og har inntjeningskrav. Kan sette opp finansielle kriterier, men da holder det ikke å bare måle før og etter, men må ha en idé om hvordan ledernes forbedrede kunnskaper skal kunne gripe inn i forretningsprosessen.*

Alexander Sandvik påpeker at det å sette opp for generelle mål også gir problemer med å sikre kausalitet mellom ledelsesutviklingtiltakene og en eventuell effekt disse tiltakene kan ha hatt:

Hvis det er et veldig generelt lederprogram der man utvikler ledere helt generelt, så må man nok lage mer generelle mål. Da tror jeg effektene kan være vanskelige å finne, fordi det er mange ting som tilskriver ledelse i en eller annen form. Men jo mer konkrete de er, så tror jeg det er mulig å måle og forstå.

For å sikre at målene ikke er for generelle og dermed forårsaker vansker med å måle effekten av et ledelsesprogram, kan man benytte seg av modellen SMART goals. I dette tilfellet er det spesielt relevant å se S'en i SMART, som vektlegger at mål skal være spesifikke. Sandvik legger også til at jo mer konkret innhold det er i ledelsesutviklingsprogrammet, jo enklere er det å sette opp mål og finne effekten av disse. Videre poengterer Grøn viktigheten av at målene skal være relevant i forhold til innholdet i ledelsesprogrammet:

Det andet der selvfølgelig også er vigtig det er parameter – hvad er det man er blevet trænet i? Og hvis jeg træner en leder i transformationsledelse, så er det ikke helt fair å evaluere om lederen er blevet bedre til økonomistyring, selv om det også er noget vi tænker er vigtig at lederen skal kunne.

At målene må samsvare med innholdet i kurset, som Grøn påpeker, kan vi også finne teoretisk medhold for ved å se på R'en i SMART som poengterer at et mål må relatere til formålet ved programmet. Det er av samme årsak at målene ikke må spørre generelt om lederen har blitt en bedre leder. Dette er både fordi vi ikke har en entydig definisjon på hva en "god leder" er, som vi tidligere har diskutert, og fordi lederen ikke har blitt trent i å være en god leder, men blitt trent i spesifikke ledelsesverktøy og ledelsesstiler.

Når en evaluator og ledelsesutvikler skal sette mål, er det viktig å stille seg spørsmålet om hvem som skal være med til å bestemme målene som skal settes. Arnulf argumenterer for at både deltakerne på kurset og ledelsen i bedriften som bestiller kurset, burde reflektere over hva som skal være innholdet i ledelsesprogrammet, slik at man sikrer en felles forståelse av hva formålet med programmet er: "Slike refleksjoner burde både de som deltar i programmet og de som betaler for det gjøre seg – for det er jo ikke alltid sammenfallende hva de tenker.". Dette er i tråd med Linghams et al. argument om at evaluering bør være en samarbeidsprosess mellom organisatoriske ledere, kursholdere, deltakere og evaluatorer (2006:335).

Arnulf påpeker at dersom kunden får ha innflytelse på målenes innhold, vil det være gunstig med tanke på at organisasjonen også skal kunne legge til rette for at den læringen som har skjedd på programmet kan implementeres. Samtidig som det også er viktig at den som gjennomgår kurset tas i betraktning, når man setter mål:

> *Hvis du sender noen på et opplæringsopphold, så er det forskning som tilsier at når de kommer hjem igjen, så er det halvparten av de som slutter etter et halvt år. Det er ubehagelig å lære ting som organisasjonen ikke setter pris på.* [...] *Løsningen er å konkretisere mål og mening i forkant av programmet. Hvorfor gjør bedriften dette? Hvorfor skal medarbeiderene være med, og hvilke ferdigheter skal de lære?* [...] *Med andre ord er det ikke åpenbart hva programmet skal være. Derfor bør man spørre hva ledernes ønsker er. Det handler om bestillingskompetanse.*

Lindtein, som representerer et strategisk og forretningsmessig perspektiv, argumenterer for at ledelsesutviklingsprogram har liten verdi om ikke resultatene er i tråd med bedriftens overordnede strategi. Det er ikke tilstrekkelig, ifølge Lindtein, å utvikle bedre ledere, dersom man ikke kan vise til organisatoriske effekter av tiltaket. Program som kun baserer seg på «individets egenutvikling» kan oppleves å ha verdi for individet, men har liten effekt i organisasjonen. Også Andersons value of learning-modell poengterer ulempene ved en ubalanse mellom forventningene til utviklingsprogrammer og forretningsmålene. Han oppfordrer derfor til å sikre samsvar mellom målene ved utviklingstiltakene og de strategiske målene til organisasjonen (Anderson, 2007).

En annen problemstilling som Arnulf trekker frem er vanskeligheten med å gjennomføre såkalte "helhetlige evalueringer", og fremhever at man heller burde trekke ut de viktigste og mest relevante målene man vil evaluere. På denne måten vil man unngå å evaluere ledelsesutviklingsprogrammet i sin helhet:

> *Jeg vil advare mot å drive med sånn man kan kalle for en «helhetlig evaluering», jeg ville heller gå inn og spesifisere hva man ønsker med de enkelte aktivitetene og elementene i programmet, og om de elementene har oppnådd skrittvis det man ønsket seg. Da kan man til slutt ta hele pakken – har dette hatt noen konsekvenser utover seg selv?*

Vi kan forstå dette som at målene for evalueringen skal konsentrere seg rundt de aktivitetene som skal foregå på kursingen. Når man har evaluert hvorvidt disse målene er oppnådd, kan man videre gå inn og evaluere hvorvidt denne måloppnåelsen har hatt en positiv effekt på organisasjonen.

4.3 Hvordan måle utbyttet av programmet?

I dette delkapitlet vil vi presentere og analysere våre funn på hvordan utbyttet av programmet kan måles. Flere av informantene våre argumenterer for at deltakerens tilfredshet som indikator for læring i forbindelse med et ledelsesutviklingsprogram er problematisk. Tom Lindtein forteller med referanse til Kirkpatricks modell, at akademia tenderer mot en grunnleggende tro på at verdi kan skapes i form av en lineær prosess fra klasserom til resultater og verdi. Han argumenterer imidlertid for at tilfredshet og kunnskap er to helt ulike ting, og at tilfredshet aldri må forveksles med et godt resultat av ledelsesutviklingsprogrammet. Dette er fordi en deltaker godt kan ha hatt en god opplevelse med programmet, uten at de har lært noe. En deltaker kan også ha lært mye i et lederutviklingsprogram uten at han/hun blir en bedre leder eller skaper bedre resultater av den grunn. Forskning viser også at kursdeltagere ofte opplever en form for eufori rett etter kursets slutt, som fører til at de vil gi en mer positiv feedback rett etter programmets slutt enn de ville gjort senere (Giangreco et al., 2010). Grøn argumenterer også for tilfredshetens begrensinger som indikator på positivt utbytte, og legger til at ledere og deres medarbeidere ofte har ulike syn på lederens innsats:

> *Jeg synes ikke, at* [tilfredshet] *er en indikation på at de er blevet en bedre leder. Særlig også fordi at vi har masser af studier som viser at lederne faktisk ikke rigtigt, eller rigtig mange af dem har ikke en særlig præcis forståelse af, hvordan deres ledelse virker for deres medarbejdere.*

Lindtein forteller også at han i noen tilfeller har sett at program som er særlig utfordrende, og ikke nødvendigvis skaper tilfredshet hos deltakerne, kan gi et større læringsutbytte. Når kursdeltakere blir utfordret og må tre ut av komfortsonen kan det skape større muligheter for læring, hevder han.

Det finnes imidlertid argumenter for at tilfredshet kan være hensiktsmessig å måle. Hans Meisingset argumenterer for at tilfredshet hos deltakeren indikerer at programmet hadde en god læringskultur, og at rammene for utvikling har vært gode. Tilfredshet sier også noe om deltakerens motivasjon til å utvikle seg som leder, ifølge Grøn:

> *Hvis man skal købe et kursusforløb som organisation, så ville det da være dumt at købe noget som lederen ikke kunne lide. Det er den første forudsætningen for at de får lyst til at arbejde med deres lederskab, det er at de faktisk engagerer sig i hvad det er for et kursusforløb de bliver sat i gennem.*

Å måle tilfredshet er altså ikke problemfritt, ifølge informantene våre. Dette er det også forskning som tyder på (Sahni, 2020). Det er imidlertid heller ikke uten verdi. På den ene siden er tilfredshet ikke en god indikasjon på at kurset har hatt en læringseffekt hos deltakerne, men på den andre siden er det viktig å lage et program som deltakerne vil ønske å delta på. Vi fortolker dette som at tilfredshet er hensiktsmessig å måle, men at man må være bevisste på manglene ved dette. Måling av tilfredshet kan gi en indikasjon på om kurset har gitt mersmak hos deltakerne, men for å undersøke om deltakerne faktisk har tatt med seg kunnskap videre og lært noe, blir man nødt til å gå dypere til verks.

Læring i forbindelse med ledelsesutvikling defineres av flere av våre informanter som endring i atferd. Sandvik argumenterer for at lederens atferdsendring kan undersøkes både ved selvevaluering og ved å spørre medarbeidere og overordnede. Ved å spørre både deltakeren og deltakerens medarbeidere, kan man fange opp ulike typer atferdsendring. Lederen kan til en viss grad rapportere selv om ting han eller hun gjør annerledes, men det er ikke alltid at handlingene er bevisste. Da kan det være mer hensiktsmessig å spørre medarbeidere. F.eks. mener han at det er logisk å spørre medarbeiderne om en leder har blitt mer karismatisk, mens dersom det er snakk om konkrete handlinger som å hjelpe sine medarbeidere, kan dette godt måles ved selvrapportering, f.eks. gjennom en dagbok.

Horgen argumenterer for at atferdsendringen må knyttes til hva som er hensiktsmessig atferd i lederrollen. For å undersøke effekten av et ledelsesutviklingsprogram kan det være nyttig å undersøke om lederen har endret atferden sin på ulike områder tilknyttet de spesifikke aktivitetene som fant sted under programmet. Jo mer konkrete utviklingsaktiviteter lederne

driver med, jo enklere er det å måle læringen. Dette var vi inne på i delkapittel 4.2. Videre argumenterer Horgen og flere andre informanter for at det er mest hensiktsmessig å undersøke atferdsendring ved å gjennomføre en 360-graders evaluering. Grunnen til det kan være at medarbeidere vil kunne se lederen utenfra og komme med et mer objektivt syn på lederens atferd. Lindtein er imidlertid mer skeptisk til lønnsomheten av 360-gradersevalueringer for å måle utbyttet av ledelsesutviklingsprogram. Hans erfaring er at resultatene sjelden gir signifikante svar, og er så kostbare at de ikke lønner seg. Vi har ikke funnet forskning som støtter denne påstanden, men Lindteins erfaringer tilfører et økonomisk nytteperspektiv som er vel så viktig i spørsmålet om hvem som skal svare på evalueringen.

Læring kan også vise seg i kognitive endringer hos lederen, som f.eks. evnen til selvrefleksjon. Det er imidlertid svært vanskelig å evaluere objektivt, ifølge Arnulf:

> *Det vanskeligste med lederutvikling* [...] *er å evaluere prosesskompetanse eller evnen til å være i en refleksjonsprosess, f.eks. lære ting om å kjenne seg selv, f.eks. personlighet og å reflektere over personlighet, 360, gruppeprosesser, og det er klart at det er jo mye vanskeligere, men det kan man jo f.eks. ha med seg spørsmål om hvorvidt man har hatt et emosjonelt engasjement,* [...] *hvorvidt man føler at folk faktisk har åpnet seg og sluppet prosessen innpå seg* [...]. *Så kan man mot slutten av programmet spørre «har dette gitt en mening i forhold til det som var programmets intensjon". Altså, har de nå kommet i gang med prosesser som er knyttet til virkelige ledersituasjoner?*

Grøn presiserer imidlertid at selvrefleksjon ikke nødvendigvis fører til at lederen gjør en bedre jobb:

> [...] *Medarbejderen oplever jo ikke at lederen har tænkt nogen dybe tanker om sig selv – de oplever det først når lederen faktisk ændrer atfærd. Så jeg synes at det giver rigtig god mening at have elementer af selvreflektion, men vi ved også godt at the talk is cheap, ikke? Og jeg kan godt sidde og tænke at jeg kan skrive det i min reflektionsblok, men det er jo ikke nødvendigvis det samme som at jeg går ud og gør. Og derfor så er det et skridt på vejen for at genere en eller anden form for adfærdsændring, men det er bestemt ikke givet at man ændrer sit adfærd fordi man har lavet en reflektion over eget ledelsespraksis.*

Informantene våre er enige om at refleksjonsprosesser er viktige for lederens utvikling, men hvorvidt det har en verdi i seg selv er det ikke enighet om. Det er også utfordrende å finne objektive indikatorer for dette, som Arnulf påpeker. Likevel kan selvrefleksjon ifølge våre informanter være starten på en prosess, som resulterer i at lederen gjør ting annerledes. Dette kan ses i sammenheng med Argyris og Schöns (1996) dobbelkretslæring, hvor lederen reflekterer over tidligere handlinger og hvorfor de har gjort som de har gjort, og deretter endrer sine handlingsmønstre for å fylle sin rolle som leder på en bedre måte. Arnulf kommer med følgende eksempel på dette, hvor en leder aktivt gikk inn for å sanksjonere atferd som ikke var ønsket i bedriften. Siden fikk de inn en ny leder som istedenfor å bruke straff gikk inn og undersøkte årsaken til at den uønskede atferden oppsto. Da var det ikke lenger behov for å straffe, fordi de evnet å avverge den uønskede atferden.

> *Det funker ikke alltid å straffe folk. Poenget er at det er ikke alltid en opplagt sammenheng mellom problem og løsning. Kompliserte systemer som bedrifter er, det er ikke alltid at du skal se på den direkte effekten av atferd, men heller se på det sosiale systemet.*

Ifølge Sandvik kan det å måle endring på individnivå noen ganger være tilstrekkelig, men han presiserer at dersom en ønsker å oppnå resultater forankret i forskning, er det nødvendig å måle hvilken effekt denne endringen har i organisasjonen i tillegg. Dette kan imidlertid være utfordrende, ifølge våre informanter. Slemdal har følgende å si om dette: "[...] det er utfordrende for bedrifter å vie tid til å delta i evalueringen, men det er helt avgjørende for å kunne måle det organisatoriske. [...] Man kan måle eksempelvis endringer i turnover, vikarbruk og sykefravær. [...]". Slemdal nevner senere i intervjuet også indikatorene kommunikasjon, samarbeid og økonomistyring. Arnulf fremhever konfliktnivå som en viktig indikator på hvordan det går i organisasjonen. I tillegg til de mer kvantifiserbare indikatorene er det viktig å også gå mer kvalitativt til verks, ifølge Slemdal: "også dreier det seg om den kvalitative delen [...] Det er ofte like viktig hvordan menneskene selv opplever at effekten har vært i organisasjonen". Effekten av ledelsesutviklingsprogram kan med andre ord undersøkes på ulike måter, som også utfyller hverandre. Dette kommer vi tilbake til senere i analysen, når vi tar for oss kvalitative og kvantitative metoder for evaluering.

4.4 Kvantitative og kvalitative metoder i evalueringen

Evalueringer kan gjennomføres både ved bruk av kvalitative og kvantitative metoder, og det kan være flere faktorer som spiller inn på når én metode er å foretrekke fremfor en annen. Slemdal vektlegger betydningen av å samle inn kvalitative data, som kan si noe om deltakernes opplevelser av ledelsesutviklingsprogrammet. Han mener de kvalitative dataene er vel så viktige som de kvantitative:

> [...] *det er til syvende og sist menneskene som jobber der som er det viktigste, og da må man også ha med deres opplevelser. Det er noe med å se at organisasjonen består av alle disse menneskene, og da er deres opplevelser vel så viktig som de harde tallene, om ikke viktigere i enkelte tilfeller.*

Slemdal poengterer hvordan kvalitative metoder fungerer godt for å evaluere endringstiltak på individnivå, men vil være mindre hensiktsmessig om man ønsker å få klarhet i systemomfattende, organisatoriske endringer. For å måle hvordan organisatoriske endringer forekommer på bakgrunn av et ledersutviklingsprogram, vil det ifølge ham være mest relevant å ta i bruk en kvantitativ metode som enklere kan oppdage utvikling og forandring i de store linjene i organisasjonen.

Fordelen med kvantitative målinger for å oppdage endringer over tid vektla også Phillips da han utviklet sin «Learning Evaluation Model». Han fokuserte på at det var viktig å isolere betydningen av organisasjonsutviklingstiltak slik at man kunne fastslå at det var en korrelasjon mellom tiltaket og resultatet. For å få til dette argumenterte han for at det bør gjennomføres kvantitative målinger som tar utgangspunkt i konkrete variabler som kan knyttes til tiltaket, og at det bør skje målinger både før og etter organisasjonstiltaket er iverksatt, slik at dataene kan sammenlignes. En utfordring med kvalitative undersøkelser er nettopp at det er vanskelig å knytte resultatene direkte til ledelsesutviklingsprogrammet, så i tilfeller hvor man er opptatt av å isolere effekten kan kvantitative metoder være mer hensiktsmessig. Samtidig er det viktig å presisere at det generelt sett er vanskelig å fastslå en årsakssammenheng mellom tiltak og resultat, selv ved bruk av kvantitative metoder, da det er vanskelig å utelukke alle andre mulige faktorer som kan ha en innvirkning på resultatet.

Lindtein argumenterer for at fra kundeperspektivet er kvantitative undersøkelser i utgangspunktet beste metode for å måle korrelasjonen mellom ledelsesutviklingsprogrammet og lederens atferdsendring. Samtidig anerkjenner han at denne tilnærmingen er så ressurskrevende, at det i praksis ikke blir gjennomførbart:

> [...] *vi* [ser] *at kostnadene for å forsøke å lage kvantitative data her* [atferdsnivået]*, den blir så høy og at det output de får er så lite signifikant, at de gir opp, det går ikke, og det er ikke bare vi som har prøvd, mange har prøvd – man får det ikke til.*

Han viser derfor til at mange, i mangel på noe bedre, i stedet for benytter seg av kvalitativ tilnærming når de måler atferdsendring:

> [... i stedet for] *bruker* [man] *en kvalitativ tilnærming og spør de som har deltatt: er det bedre? Føler du at du har levert sånn og sånn? Og så kan du spørre noen som jobber sammen* [med kursdeltakeren] [...] *Så bygger du på dette, og skaper et bilde av at* [programmet] *var verdifullt og kjempebra.*

Dette er, som Lindtein påpeker, nyttig for å danne et bilde av om lederen selv opplever at kurset har gitt effekt, og om kollegaene rundt merker atferdsendring hos lederen. Det gir imidlertid ingen objektive, signifikante og harde data på at kurset *faktisk* har gitt utbytte. I en avveining mellom hva som er ideelt og hva som er praktisk mulig, argumenterer han likevel for at kvalitative målinger bør benyttes når kvantitative innsamlinger blir for ressurskrevende. Mange har forsøkt å måle effekter ved bruk av Kirkpatricks modell, uten å klare å dokumentere resultater og/eller verdi av ledelsesutvikling, forteller Lindtein. Derfor mener han det er legitimt å spørre seg om ideen om en lineær prosess fra læring til verdi er brukbar, eller om lederutvikling som konsept i det hele tatt er verdt investeringen.

I teorien vektlegger de i stor grad ulike datainnsamlingsmetoder, avhengig av hva slags tematikk man ønsker å få mer innsikt i. Da Brinkerhoff utviklet sin evalueringsmodell var han opptatt av å undersøke individuelle forskjeller i hvem tiltaket fungerte for, og hvem det ikke fungerte for (Singh & Ramdeo, 2020:411). Derfor utviklet han en evalueringsmetode som baserte seg på kvalitative undersøkelser, som nettopp ville plukke opp de individuelle opplevelsene. Det er naturlig å anta at ikke alle ledere som deltar på et lederutviklingsprogram

vil oppleve det samme utbyttet, og for å finne ut bakgrunnen for forskjellene i utbyttet, kan kvalitative metoder være hensiktsmessige. I likhet med Brinkerhoff, argumenterer også King og Nesbit (2015) for at økt bruk av kvalitative analysemetoder kan gi en dypere forståelse av deltakernes erfaringer og endrede perspektiver over tid.

Gjennom intervjuene våre argumenteres det i stor grad for at en kombinasjon av kvalitative og kvantitative metoder er hensiktsmessig å ta i bruk, dersom det er mulig. På denne måten får man både en kvantitativ innsikt i de store trendene av resultater i organisasjonen i etterkant av ledelsesutviklingsprogrammet, samtidig som man får mulighet til å dypdykke inn i de individuelle opplevelsene til deltakerne. For å få til dette, trekkes det frem at spørreskjema kan være et nyttig virkemiddel. En mulighet for å samle inn både kvantitative og kvalitative data via skjema, ifølge Horgen, er ved å både legge opp til avkrysning på skala, men også å åpne opp for en mer utbroderende skriftlig tilbakemelding.

En annen måte å kombinere bruk av spørreskjema med en dypere kvalitativ undersøkelse, er å ta utgangspunkt i de kvantitative resultatene fra spørreskjemaet for å utforme spørsmål man stiller lederne, og eventuelt deres nærmeste kollegaer, i en mer dyptgående samtale. Denne metoden tok Grøn i bruk da hun utviklet en modell for hvordan offentlige ansatte kan vurdere sine ledere:

> [...] *det* [er] *kvantitative spørgeskemaundersøgelser, men så har vi materialer understøtter oppfølgende dialog – så vi lægger op til at lederen skal have en dialog med sin egen leder, lederen skal ha en dialog med sine medarbejdere og at lederen eventuelt også skal have en dialog med sine sideordnede leder, som opfølgning på det kvantitative, så der er et kvantitativt skridt og så er der et kvalitativt skridt bagefter.*

I en prosessuell evalueringsmetode, som Lingham et al. presenterte, vises det konkret til hvordan kvalitative og kvantitative undersøkelser kan kombineres for å både kunne dypdykke inn i individuelle opplevelser, men også kunne måle trender og endringer over tid (2006:334). Ved å la de kvalitative og kvantitative undersøkelsene spille sammen, og la resultatene fra den ene undersøkelsen påvirke utformelsen av den andre, opplever man å kunne få et godt helhetsinntrykk av hvordan et opplæringsprogram påvirker en gruppe. Denne kombinasjonsmetoden vil også være ressursbesparende, fordi man kan spisse del-

evalueringene inn på tematikk man finner ekstra spennende på bakgrunn av resultatene fra forrige undersøkelse, eller der man ser man trenger mer informasjon, i stedet for å måtte gjennomføre brede og omfattende undersøkelser hver del-evaluering. Ved å få til et godt samspill mellom kvalitative og kvantitative del-evalueringer, kan man få en evaluering som danner et både bredere og dypere bilde av resultatene fra programmet, sammenlignet om man kun hadde benyttet seg av én av de to metodene.

En annen faktor som kan være med på å påvirke hvilke metoder det er mest hensiktsmessig å benytte, er om evalueringen foretas av en intern eller ekstern evaluator. Meisingset mener at en ideell evalueringsmodell jobber med kvantitative data over tid, slik at de kan måle utvikling på lang sikt. Han understreker imidlertid at dette er en ressurskrevende og stor prosess, som kan være krevende å gjennomføre: «Men dette er en ideell modell som jeg ikke vet om andre enn Posten som har gjennomført [etter programgjennomføring, fordi data om lederprestasjoner var tilgjengelige]. Dette er en så stor, komplisert og ressurskrevende prosess [...]» Dersom evalueringen av et ledelsesutviklingsprogram foretas av en intern i bedriften som kjøpte programmet, kan utfordringene knyttet til å innhente kvantitative data antas å være mindre enn om det er en ekstern som gjennomfører evalueringen. Å gjennomføre kvantitative undersøkelser i etterkant av programmet for å følge utviklingen, er for det første lettere å gjennomføre om man har direktelinje ut til egne ansatte, samt kan pålegge egne ansatte å bruke av sin arbeidstid for å gjennomføre undersøkelsene. For det andre sitter bedrifter ofte på en del data om trender i bedriften, f.eks. turnovertall, sykefraværsstatistikk, faktureringsgrad, vikarbruk ol. som de kan følge utviklingen til over tid. Man kan anta at det vil være mer utfordrende for en ekstern evaluator å få innsyn i slike data, da evaluatoren selv må be om, og få innvilget, innsyn i disse dataene, og at man som evaluator er avhengig i at bedriften man undersøker samarbeider om å dele disse.

4.5 Før-, underveis- og ettermåling

Det er stor enighet om at det er hensiktsmessig med før- og ettermålinger. Dette gjør det mulig å følge noen fastsatte parametre over tid, og sammenligne disse. De fleste informantene har tatt i bruk kvantitative målinger før kursstart, som de sammenligner med en lik, eller tilsvarende, undersøkelse i etterkant av kurset. “Jeg har heller ikke troen på at folk husker hva de har krysset av på hvis man har et spørreskjema”, sier Sandvik, som mener det er uproblematisk å benytte seg av samme kvantitative spørreskjema før og etter kursstart.

Flere av de intervjuede fremhever at før- og ettermålinger er et godt alternativ til kontrollgruppe. F.eks. sier Madsen at før- og etterundersøkelser er den beste måten å sikre en sammenheng mellom program og læringsutbytte. Fordelen med å benytte seg av før- og ettermålinger i stedet for kontrollgruppe, er at måling gjennom kontrollgrupper, ofte fører med seg noen metodiske utfordringer angående hvem som skal være i gruppa, og at det krever mye ressurser. Dette beskriver blant andre Madsen:

> *Kontrollgruppe er liksom gullstandarden det, ikke sant. Da har du virkelig gjort det bra. Problemet er jo bare det at, hvem skal være kontrollgruppen. Ideelt sett skulle du da plukket ut en bunke med ledere som var på samme alder, som dreiv omtrent i samme bransje, og på samme nivå med de som gikk der. Skulle spurt de samme før og etter. Du får ikke tak i den kontrollgruppa, tenker jeg. Altså det er mulig, men det er fryktelig* [vanskelig].

Det argumenteres av flere for at man også bør måle underveis, slik at man kan justere programmet til det bedre mens det pågår. Arnulf er klokkeklar på at dette er den beste løsningen: "Nå har vi hatt 70 år med manglende fremdrift i evaluering av lederutvikling, og løsningen ligger rett foran oss: se underveis om programmet følger det man hadde tenkt." Han får støtte fra blant andre Horgen: "Så har vi jo også evaluering underveis, som er [...] svært viktig for å få tilbakemeldinger underveis for justering, eksempelvis på et program som er modulbasert, som går over et år."

Også Slemdal mener det er nyttig med en underveismåling for å kunne justere og tilpasse programmet:

> *Man må ha en nullpunktsmåling, gjerne kvantitativt før de starter. Så er det en* [måling] *underveis, det gir oss muligheten til å hjelpe oppdragsgiver til å endre eller gjøre forbedringer ut fra funnene vi har, slik at vi også er med på å heve kvaliteten på utdanningen.*

Madsen mener imidlertid at underveismålinger kan være forstyrrende og plagsomt for deltakerne på ledelsesutviklingskurset, og mener at det ikke er hensiktsmessig å gjennomføre underveismålinger på programmer som kun varer ett år:

Hvis du skal intervjue veldig mange, også skal du ha fokusgruppe i ledelsesprogrammet, så kan det fort blir veldig forstyrrende for hva de egentlig er der for. Derfor er det å lage en før og etterundersøkelse på en overkommelig måte noe jeg har stor tro på, for det kan gi dem ganske bra forståelse. Man kan også lage en forenklet midtveis-undersøkelse hvis dette går over litt lenger tid, men hvis det bare er ett år ville jeg ikke plaget folk med det.

De gjentagende anbefalingene fra informantene om å ty til underveismåling, underbygger Lingham sin prosessuelle evalueringsmetode der fokuset er at ledelsesutviklingsprogrammet løpende skal forbedres, tilpasses og oppdateres (Lingham et al., 2006:334). Ved å utforme en formativ evaluering som dette, sørger man for en dynamisk utvikling av ledelsesutviklingskurset, der programmet løpende kan utvikles og forbedres på bakgrunn av deltakernes tilbakemeldinger. I tillegg får man mulighet til å fange opp mulighetene deltakerne har for å implementere det de lærer på kurset i organisasjonen, eller om det er noen organisatoriske barrierer man må ta hensyn til. Slik unngår man å måtte vente til programmet er over før man får sett resultatene av kurset.

4.5.1 Viktigheten av underveismåling grunnet læring

Som skrevet, er det stor enighet om at det er hensiktsmessig med underveismåling grunnet muligheten det gir til å tilpasse og justere programmet. I tillegg til dette, tilsier våre funn at underveismåling er viktig sett fra et læringsperspektiv. Naturlig nok kan det sies at programansvarlige lærer når de får vite deltakernes meninger og refleksjoner knyttet til ledelsesutviklingsprogrammet, men informantene våre mener at også deltakerne lærer under denne prosessen. Grunnen til det er at evaluering underveis, i form av selvrefleksjon av aktuelle deltakere, kan bidra til at de blir mer bevisste over hvilken kunnskap de har tilegnet seg så langt i programmet, hvilke utfordringer de eventuelt trenger å jobbe mer med, hva som tilsynelatende har funket bra/dårlig for dem og andre refleksjoner som kan være nyttige for egen læring. Blant andre er Madsen enig i viktigheten av å ta hensyn til læring under evaluering:

Å få kunnskap så du eventuelt kan forbedre deg. Så kan du jo se på de forskjellige typene formål med evaluering [...] *om du bare vil at det skal legitimere eller bare vise at du gjør de rette tingene. Men jeg mener det er for læringsaspektet, for å kunne bli flinkere.*

4.5.2 Når bør etter-målingen være?

Når det kommer til tidspunkt for når ettermålingen bør gjennomføres, trekker flere av informantene frem samme utfordring; en ettermåling rett etter programslutt kan ikke si noe om langtidseffektene til kurset. Denne utfordringen utelukker imidlertid ikke at man skal gjennomføre en måling umiddelbart etter at kurset er ferdig, men det taler derimot for at man bør gjøre *flere* målinger etter at kurset er ferdig. Én som kan si noe om de umiddelbare opplevelsene tilknyttet kurset, og én (eller flere) som kan si noe om langtidseffektene fra kurset. Meisingset sier f.eks.: "Da tenker jeg også at det kan være lurt at AFF ikke evaluerer bare rett etter programslutt, men at man har evalueringspunkter etter et halvt år og ett år." Dette underbygger også Sandvik:

> *Den ene målingen etterpå er viktig for å få en selvrapport. Men jeg tror effekten i organisasjonen bør være litt etterpå. Jeg har ingen gode data for tidshorisont på hva som er det ideelle, men jeg tror mange er enige om at det bør gå noen måneder. 3-6 måneder vil jeg tippe er bra.*

Horgen er også enig i at det er nødvendig både med en umiddelbar måling etter kurset er over, som kan si noe om førsteinntrykket deltakerne sitter igjen med etter kurset:

> *Mtp tidspunkt så er det flere tidspunkter her. Hvis jeg skal tenke rasjonelt, så vil det være nyttig å gjennomføre en first impression måling, altså rett i etterkant. Det tenker jeg alltid vil være nyttig, ikke minst for de som gjennomfører det. Det å få en tilbakemelding på justeringer før neste runde med en annen kunde eller samme kunden.*

Deretter mener Horgen at det bør gjennomføres en kvalitativ og mer dyptgående måling etter at det har gått en stund siden programmet var over:

> *Jeg tenker at det da kan gjøres i form av et kundemøte eller et møte i etterkant. La oss si 2 måneder etter lederutviklingsprogrammet er gjennomført, gjennomføre en samtale med ansvarlig i virksomheten. Sette sammen en liten gruppe av de som har vært rundt ledere og/eller medarbeidere til de som har vært på lederutviklingsprogram og få mere kvalitativ informasjon om hvordan det her har virket.*

Videre sier hun:

> *Det er jo en kjensgjerning at det er veldig lett å få gode tilbakemeldinger på et lederutviklingsprogram program, og derfor så er det også viktig å la de få litt avstand, så at det ikke sitter i en slags eufori av fornøydhet og så levere ut et evalueringsskjema.* [...] *Man er jo ute etter å finne ut om det her har noe effekt i hverdagen,*

Grøn argumenterer for at tidspunktet for ettermålingen bør avhenge av hva man har lært. I programmer hvor man løpende tar i bruk det man lærer i egen arbeidshverdag, kan det være mulig å måle effekt umiddelbart når programmet avsluttes. Det kan imidlertid være lurt å vente i rundt 6 måneder før man måler, dersom man vil se at det de har lært blir tatt i bruk over tid, og ikke blir glemt igjen umiddelbart etter programslutt:

> *Fordi hvis man arbejder med at man løbende omsætter det man har lært i sin egen praksis, så tænker jeg at det må være en effekt lige når programmet slutter. Men jeg vil måske i virkeligheden også sige* [...] *at man faktisk skal vente 6 måneder med at måle.* [...] *Det er lidt ligesom når man spiser en pose vingummier, så stiger blodsukkeret meget hurtig, ikke? Og så falder det igen bagefter, og det samme kan godt være risikoen, at hvis man så stimulerer lederen til noget bestemt, så kan det godt være at de gør det i 14 dage, og så glemmer de det.* [...] *måske skal man i virkeligheden vente 6 måneder, for det er først der vi faktisk ser om det er en del af deres rutiner og adfærd* [er] *ændret.*

Som vi kan se, er det noe uenighet blant informantene om akkurat *når* den siste ettermålingene bør finne sted. Noen nevner 2 måneder, noen nevner et halvt år, mens andre synes det bør gå ett år eller to. Det viktigste synes å være at det skal ha gått lang nok tid til at man skal kunne måle om deltakerne faktisk har hatt en atferdsendring på bakgrunn av programmet.

Videre er det, som vi har sett i læringsteorien, en vesentlig forskjell på om en ansatt har lært noe individuelt og på om organisasjonen som helhet endrer seg. Dersom første måling viser at den ansatte har opplevd en atferdsendring på bakgrunn av kurset, kreves det dermed en egen måling for å kunne fastslå om også organisasjonen har opplevd endringer i etterkant av kurset. Organisatorisk læring innebærer at organisasjonens medlemmer må endre måten de

samhandler med hverandre på, og normer, verdier og grunnleggende antakelser i organisasjonen må justeres for å få dette til. Det kan derfor være langt mer krevende å oppnå organisatorisk atferdsendring enn individuell atferdsendring, og det vil naturligvis også ta lenger tid før en slik organisatorisk endring blir synlig.

Denne nyansen tar CIRO-evalueringsmetoden hensyn til i forbindelse med evaluering av ledelsesutviklingsprogram. Den legger opp til tre målinger etter kursslutt, med en umiddelbar evaluering så snart kurset er over, en som ser på hvordan læringen overføres av den enkelte tilbake på arbeidsplassen, og én som ser på langtidseffektene på avdelings- og organisasjonsnivå. Dette er en ressurskrevende evalueringsmetode, men som til gjengjeld vil gi et godt innsyn i hvordan implementeringen av kunnskap foregår over tid, og om den umiddelbare kunnskapen kursdeltakerne sitter med ved kursslutt vil anvendes på arbeidsplassen, eller om den vil glemmes og legges til side. Både informantene og teorien taler imidlertid samstemt for at dersom man ønsker å oppdage enten individuelle eller organisatoriske effekter på lang sikt etter et ledelsesutviklingskurs, bør man vente lenge nok med å måle til at en atferds- og rutineendring har funnet sted.

4.6 Evaluering i teorien vs. praktisk gjennomførbar modell

Løpende gjennom prosessen i dette utredningsprosjektet har vi stilt spørsmålene: "Finnes det en generisk modell som passer til alle former for evaluering?", og "Er teorien mulig å gjennomføre i praksis?"

Gjennom datainnsamlingsprosessen fremkom det at kontekst ofte er et sentralt element når man skal planlegge og gjennomføre en evaluering, slik som vi tidligere har argumentert og redegjort for i delkapittel 4.1. Dette belyses også av Slemdal: "Det er ikke nødvendigvis alltid lurt å ha en generisk modell. Det er noe kontekst og faktorer som man må ha med, for å kunne ha en mest mulig presis evaluering." Madsen argumenterer for at det er den teoretiske verktøykassen som er det viktigste når man skal lage, skape kontekst for og gjennomføre en evaluering:

> *Men som sagt teorier, teorier og teorier - det er det som er verktøykassa. Det er den beste verktøykassa du har.* [...] *dere må ha noen gode teorier. Ikke evalueringsteorier, men gode ledelsesteorier som gjør at dere sitter med de batteriene av spørsmålene som kan brukes, som er testet før og som holder mål.*

Madsen begrunnet dette med et eksempel fra egen erfaring med å gjennomføre en evaluering:

> *Vi gjorde det da vi evaluerte skattefunn. Da lagde vi spørsmål som var teoribaserte ut ifra bedriftsøkonomisk teori og entreprenørskap og sånn. Også prøvde vi å måle om det var noen endringer i de variablene når de var ferdige med prosjektet* [...]

Dette kan gi oss grunnlag for å hevde at dersom man legger til rette for tilpasning til kontekst og relevant teori, vil det være mulig å lage en anvendelig, generisk modell eller et generisk rammeverk for evaluering. Dette bringer oss videre til spørsmålet om hvorvidt evalueringsteorien er mulig å gjennomføre i praksis. Møtet mellom teori og praksis var også et emne for refleksjon under våre intervju. Meisingset sa bl.a.: "Evalueringsteorien er fin, men i praksis så synes jeg det er krevende når den møter praksis.". Lindtein poengterte at teori og praksis er to forskjellige felt, og at det alltid vil være et gap mellom de to feltene. Likevel eksisterer det en positiv holdning blant informantene til en løsning om å integrere teori og praksis. Informantene våre poengterer at det handler om å oversette modellene og teorien og gjøre dem mer praksisnære - en løsning der både teorien og praksisfeltet kommer til sin rett. Slemdal argumenterer for at konsulenter/praktikere i mindre grad er låst av teorien enn forskere og teoretikere:

> *Det er særlig modellene vi bruker mest, så oversetter vi modellene og gjør dem praksisnære. Konsulenter er ofte mer praktiske enn for eksempel forskere, og har ikke noen store utfordringer med det. Som forsker har man den enorme teoriforankringen, men vi* [praktikere] *kan la oss inspirere av teori og modeller, og så kan vi være kreative og finne egne løsninger ut ifra disse. I enkelte prosjekter følger vi teorier til punkt og prikke.* [...] *Hvis vi ser behov for det,* [...] *finner* [vi] *måter vi kan bruke teorier på i praksis. Vi er på en måte ikke så låst.*

Sandvik hevder imidlertid at forskere også kan klare å finne et kompromiss mellom forskerens teoretiske verden og den praktiske løsningen utenfor rammene av akademia:

> *Vi* [forskere] *er nok ikke så rigide som vi sier heller. Jeg kan som forsker tenke to ting på en gang, der jeg gjør noe enklere fordi selskapene vil ha informasjon om det. Også*

tar jeg heller den ene tingen jeg vil se på som forsker litt mer omfattende. Så det går an å finne løsninger slik at begge parter overlever. Jeg tror tilgangen til data, personer, kilder er en viktig faktor som må forhandles om.

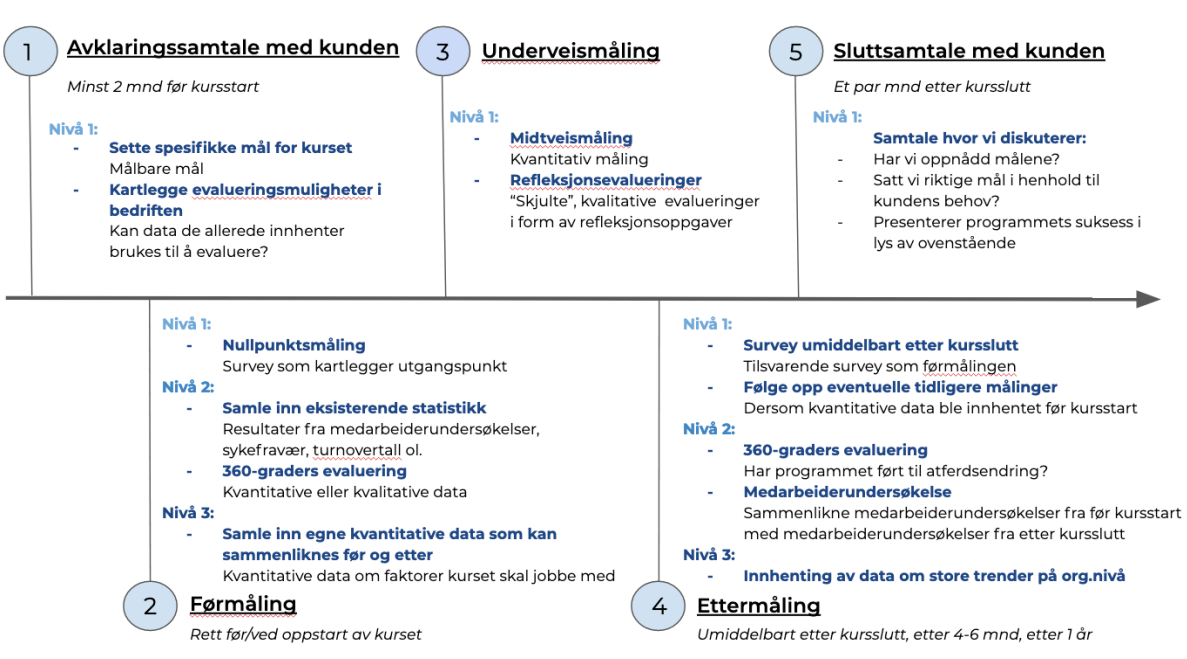
5.0 Bakgrunnen for utformingen av modellen

Vi har i oppgavens analysedel blitt oppmerksomme på utfordringene med å forene evalueringspraksis og evalueringsteori, og utfordringene med å lage en generisk evalueringsmodell som kan benyttes til å evaluere ledelsesutviklingsprogram uavhengig av faktorer som lengde, målgruppe og innhold. Informantene våre har imidlertid argumentert for at det er mulig å finne en gylden middelvei, der man forsøker å balansere ulike hensyn for å komme frem til en modell der anvendelighet og brukervennlighet også inkluderes. Dette på tross av at det i utgangspunktet bryter med hva som ifølge teorien er å anse som "absolutt beste" evalueringspraksis.

Av hensyn til gjennomførbarheten, og av AFF sitt ønske om å få en anvendbar modell de kan benytte seg av for å evaluere, har vi kommet frem til at vi må utvikle et verktøy som finner en balanse mellom beste praksis og gjennomførbarhet. Dersom man skulle ha tatt utgangspunkt i "beste praksis", hadde evalueringsmodellen blitt så omfattende, ressurskrevende og spesifikk, at vi vurderer det som lite sannsynlig at AFF ville ha tatt den i bruk. Vi har imidlertid forsøkt å ta slike avveininger med ambisjon om at det skal gå så lite som mulig ut over kvaliteten på evalueringen.

Vi vil i dette delkapittelet presentere modellen vi har utviklet til AFF på bakgrunn av oppgavens teori- og analysedel, og diskutere et utvalg avveininger vi har tatt mens vi har jobbet med denne.

5.1 Modell



På bakgrunn av funnene i analysen, har vi utviklet et generisk rammeverk som AFF kan ta i bruk når de skal utvikle sine ledelsesutviklingsprogram (se Vedlegg 1). Dette rammeverket er presentert som en femstegsmodell, der de ulike stegene er plassert på en tidslinje ut ifra når de skal gjennomføres. Til hvert steg er det beskrevet noen aktiviteter som er kategorisert fra "Nivå 1" til "Nivå 3". Til tross for at vi gjerne skulle sett at AFF tok seg tid til å gjennomføre omstendelige evalueringer i forbindelse med hvert ledelsesutviklingsprogram, har vi forståelse for at det ikke alltid vil være like stor kapasitet til å gjøre dette. Nivå-inndelingen er derfor et verktøy som presenterer hvilke aktiviteter vi anser som absolutte minstekrav for å gjennomføre en evaluering (nivå 1), hvilke aktiviteter vi anbefaler for å få en mer dyptgående innsikt i de individuelle opplevelsene av programmet, samt noe innsikt i læring på organisasjonsnivå (nivå 2), og hvilke aktiviteter vi foreslår for å få en enda dypere forståelse av effektene av programmet på organisasjonsnivå (nivå 3).

Som vi kan se, kombinerer modellen kvalitative og kvantitative aktiviteter. Dette er en avgjørelse tatt på bakgrunn av en felles anbefaling fra informantene, der vi har sett hvordan metodene utfyller hverandre. Surveyen i før- og ettermålingen anbefaler vi å utvikle både med avkrysningsspørsmål på skala, og med mer åpne og selvreflekterende spørsmål. Vi anbefaler å kombinere bruken av skalaspørsmål og åpne spørsmål, for å både kunne kvantifisere svarene med objektive og målbare indikatorer, samt å få dypere innsikt i holdninger og verdier i

organisasjonen. Eksempler på åpne spørsmål kan være "tenk på en gang du…" (Watkins et al., 2011:216), eller "Hvordan opplever du...". Ved hjelp av disse refleksjonene kan de grunnleggende verdiene og antagelsene i organisasjonen utfordres og endres som en del av læringsprosessene (Argyris & Schön, 1996:21-22).

Samtidig er bruken av selvrefleksjon for å undersøke utbytte som tidligere nevnt ikke uproblematisk, da lederens egen vurdering ikke er objektiv. Vi anser det likevel som et akseptabelt og gjennomførbart alternativ til 360-graders evaluering. I situasjoner der en imidlertid har ressurser til å måle på nivå 2, vil det også være hensiktsmessig å inkludere en kvalitativ eller kvantitativ 360-graders evaluering i før- og ettermålingene. Basert på det teoretiske rammeverket kan en sammenligning av resultatene fra en 360-graders evaluering før og etter programslutt, bidra til å avdekke endringer i deltakernes bruksteori. Ettersom det er kildene i lederens omgivelser som vurderer lederens jobbrelaterte atferd, kan dette si noe om den faktiske atferden i organisasjonen, og reflektere handlingsmønsteret hos organisasjonsmedlemmene (Argyris & Schön, 1996:13-14).

En utfordring ved bruk av survey er at det er begrenset hvor dyp innsikt en får, selv ved bruk av åpne spørsmål. Dersom AFF ønsker gode, kvalitative betraktninger, er individuelle samtaler i utgangspunktet den beste datagenereringsmetoden, der det er mulig å stille oppfølgingsspørsmål og få deltakeren til å utdype interessante betraktninger. I tilfeller der det imidlertid er for ressurskrevende med individuelle dybdesamtaler, vurderer vi åpne og selvreflekterende spørsmål i en survey til å være et hensiktsmessig alternativ. Dette er årsaken til at dette er foreslått som metode i vår modell.

En midtveismåling (steg 3) er lagt inn etter klar anbefaling fra informantene våre, da dette er et nyttig verktøy for å evaluere programmet underveis slik at programmet kan justeres fortløpende dersom det er nødvendig. Her er det spesielt nyttig å avdekke om deltakerne opplever at det er rom for å implementere det de har lært på kurset i arbeidshverdagen, slik at det kan gjøres tiltak dersom dette ikke er tilfellet. Vi kommer tilbake til steget om midtveisevaluering i kapittel 5.5.

Vi har valgt å inkludere en måling om hvorvidt deltakerne er tilfredse med kurset som en del av ettermålingen (steg 4). Dette på bakgrunn av at det i analysen poengteres at tilfredshet er hensiktsmessig å måle, da det gir en indikasjon på om kurset har gitt mersmak hos deltakerne. Vi anbefaler å måle tilfredshet ved bruk av åpne spørsmål, da dette vil gi en mer utdypende

beskrivelse av hva deltakerne har opplevd som bra og mindre bra ved kurset, enn ved skalaspørsmål.

Vi anbefaler også AFF å gjennomføre en samtale med kunden både før og etter kursstart, der det først settes konkrete og målbare mål som både kunden og AFF er enige om (steg 1), og der det i ettertid diskuteres i samråd med kunden hvorvidt disse målene var riktige med tanke på hva bedriften ønsket å oppnå (steg 5). Dette kan både være med på å skape bevissthet til fremtidige kurs om mål og kursinnhold bør spesifiseres bedre i fremtiden, og det kan være en god arena for AFF til å presentere funn fra evalueringen sin til kunden for å underbygge effekten av kurset.

I førsamtalen anbefaler vi også å kartlegge eventuelle evalueringsmuligheter i bedriften, slik at AFF kan dra nytte av disse. Dersom AFF kan legge inn noen spørsmål i bedriftens medarbeiderundersøkelse for å måle effekt av programmet, kan dette være en enkel og ressursbesparende måte å hente inn data. Dersom bedriften gjennomfører medarbeidersamtaler, kan det undersøkes hvorvidt kollegaene har opplevd endring hos lederen som var på kurs, som kan rapporteres tilbake til AFF. Ved å se på endringer i organisatoriske trender, som f.eks. sykefravær og turnover, kan organisatoriske effekter kartlegges. Endringer i disse organisatoriske trendene kan videre synliggjøre hvorvidt deltakerne har rekonstruert organisasjonens bruksteori, som har bidratt til dobbeltkretslæring, og dermed organisasjonslæring (Filstad, 2016:46).

For å benytte seg av disse eksisterende evalueringsplattformene, kreves det imidlertid en samtale rundt dette, der interesse og ønske om innsyn uttrykkes overfor bedriften. Under vil vi diskutere de største avveiningene i utvikling av evalueringsmodellen. En mer konkret og detaljert beskrivelse av modellen og innholdet i de ulike stegene finnes som vedlegg.

5.2 Modellen beskriver et rammeverk som behøver tilpasninger

Vi har sett at kontekst er svært viktig for hvordan en evaluering skal bygges opp, og at ledelsesutviklingsprogrammets mål må være styrende for hvilke indikatorer man benytter seg av i evalueringen. Det er derfor vanskelig å skulle utvikle en modell som er generisk, og som kan benyttes i alle evalueringssituasjoner uten å måtte tilpasses. Modellen vår er derfor utviklet som et rammeverk, og forutsetter at AFF tar noen pragmatiske og teoretiske avveininger underveis når evalueringen planlegges, slik at evalueringsopplegget blir best mulig tilpasset det programmet som evalueres. Som forutsetning for at evalueringen skal lykkes, er det viktig

at AFF tar seg tid til å tilpasse evalueringsmetoden til sitt program, og at hensyn til evaluering blir inkludert allerede på første møte med kunden. Å sette konkrete og målbare mål, og å tilpasse del-evalueringene etter de målene som er satt, er helt avgjørende for å sikre at man måler det man vil måle, og dermed at evalueringen gir noe utbytte overhodet.

Som vedlegg ligger et konkret eksempel på hvordan evalueringsmodellen kan tas i bruk i forbindelse med et av AFFs ledelsesutviklingskurs (se Vedlegg 2). Alle valg som ble tatt i utformingen av dette eksemplet er i tråd med rammeverket som er blitt gjort rede for i kapittel 5. Vi oppfordrer AFF til å se på hvordan dette evalueringseksempelet er tilpasset, og bruke det som inspirasjon når det skal utvikles nye evalueringer til andre utviklingsprogram. Videre anbefaler vi AFF å utvikle en "evalueringsbank" fordelt på teoretiske temaer, basert på tidligere evalueringer. Der kan det legges inn spørsmålsformuleringer, indikatorer og mål, som kan hentes frem igjen når det trengs.

5.3 Indikatorer

Informantene våre presenterte flere mulige indikatorer for å måle utbyttet av ledelsesutviklingsprogram: Turnover, sykefravær, vikarbruk, konfliktnivå, kommunikasjon, samarbeid og økonomistyring. Vi anser disse indikatorene som gode, men det er viktig å presisere at ikke alle indikatorer vil være hensiktsmessige i alle kontekster. Som vi har stadfestet tidligere i rapporten, er det viktig å kunne knytte indikatorene opp mot de konkrete målene som er satt opp i programmet. Det må derfor avgjøres om indikatorene som brukes er hensiktsmessige for å måle resultatet av de aktivitetene som har foregått under programmet. Dersom målet med programmet f.eks. er økt strategisk kompetanse hos lederne, kan det være hensiktsmessig å undersøke om de bruker tiden sin mer effektivt, f.eks. ved å se på timebruk. Det kan også være hensiktsmessig å se på salgstall. Dersom målet med programmet er styrkede relasjoner og samarbeid innad i organisasjonen, kan det være mer hensiktsmessig å spørre medarbeidere om de opplever at det er bedre kommunikasjon og samarbeid internt. Disse indikatorene er av natur mer kvalitativt orienterte, men kan f.eks. besvares ved hjelp av en rangert skala.

5.4 Bruk av 360-graders evaluering

Basert på analyse av innhentet data kan det sies at brorparten av våre informanter er for inkludering av 360-graders evaluering i modellen. Som nevnt i analysedelen av rapporten,

mener informantene våre at muligheten til å innhente deltagernes omgivelser sitt perspektiv på førstnevnte sin lederatferd gjør dette til en god evalueringsmetode. På den annen side, fant vi det nødvendig å ta hensyn til vårt mandat som ber oss utforme en praktisk anvendbar evalueringsmodell. Ettersom gjennomføring av 360-graders evaluering er ressurskrevende i form av tid, penger, tilgang til organisasjon og mer, har vi valgt å inkludere denne som valgfri på nivå 2 i før- og ettermåling (se Vedlegg 1).

Vi har valgt å inkludere 360-graders evaluering i "steg 2" (førmåling) og "steg 4" (ettermåling), som en aktivitet på nivå 2. I lys av læringsteori kan dette gi et mer korrekt bilde av om det er samsvar mellom de uttrykte handlingene som har blitt beskrevet gjennom selvrefleksjon, og handlingene som faktisk foretas (Argyris & Schön, 1996). Dermed gjør vi det klart for AFF at bruken av 360-graders modellen er høyst anbefalt, men kan utelates dersom tilgjengelige ressurser ikke tillater gjennomføring av dette. Hvis ressursene er til stede, men i begrenset kapasitet, foreslår vi bruken av en modifisert versjon av 360-graders evaluering hvor AFF eksempelvis gjennomfører evalueringen kun blant de antatt viktigste aktørene istedenfor på alle nivåer.

Det er også verdt å presisere at en 360-graders evaluering i praksis kan gjennomføres både ved hjelp av en kvalitativ og kvantitativ metode. Dette kan forklares nærmere ved hjelp av et eksempel; man kan innhente svar fra deltakernes omgivelser både ved hjelp av dybdeintervjuer og survey, hvorav førstnevnte er en kvalitativ metode mens survey er en kvantitativ metode. En vurdering av hvilken av disse som vil være mest hensiktsmessig å ta i bruk må vi overlate til AFF, da dette igjen avhenger av ressursene de vil ha tilgjengelig i det gitte prosjektet.

5.5 Bruk av midtveismålinger

Som illustrert i modellen, anbefaler vi i «steg 3» (underveismåling) bruken av både midtveismåling i form av en kvantitativ metode, og kvalitative refleksjonsevalueringer (se Vedlegg 1). Med refleksjonsevalueringer menes hyppige, gjerne ved hver samling, men korte refleksjonsoppgaver som tildeles programmets deltakere muntlig. Disse refleksjonsoppgavene skal bestå av spørsmål som stilles av programansvarlig til deltakerne knyttet til deres opplevelse av programmet så langt, deres meninger om hva som har virket positivt og hva som eventuelt kan forbedres. I diskusjoner med AFF kom det frem at dette allerede er en praksis blant deres programansvarlige. Vi anbefaler dem imidlertid å også skrive ned deltakernes svar

slik at de har materiale å jobbe å ta i bruk i etterkant og justere programmet basert på tilbakemeldingene. På den måten forbedrer de prosessens utbytte fra et evalueringsperspektiv.

Ideelt sett gjennomføres begge de nevnte prosessene for evaluering, i henhold til våre funn, men vi vil likevel argumentere for at en kvantitativ midtveisevaluering er den viktigste å gjennomføre av disse to metodene. Her får deltakerne mulighet til å fortelle om sine opplevelser midtveis i programmet, samt uttrykke hvilken læring de opplever at de har tilegnet seg. Ved å systematisk samle inn disse kvantitative dataene får programansvarlig tilgang til informasjon som kan indikere hvorvidt programmet er på rett spor, og hvilke justeringer som eventuelt bør gjøres.

Fordelen med en refleksjonsevaluering, er at den kan gjennomføres som en "skjult" evaluering. Det vil si at det ikke nødvendigvis må formidles som evaluering til deltakerne, da hensikten til sistnevnte bør være å lære av denne øvelsen. Dette ved å reflektere over det de har gjort og kunnskapen de har tilegnet seg i løpet av ledelsesutviklingsprogrammet. Programansvarlig, på sin side, bør være opptatt av å behandle denne informasjonen som evalueringsdata, og ta denne i bruk for å forbedre og justere programmet. På denne måten vil evalueringen inngå som en naturlig del av programmet, og ikke forstyrre det faglige innholdet. Da denne formen for evalueringsmetode også er et nyttig virkemiddel for læring, er det lett å presentere refleksjonsaktivitetene som læringselementer og ikke som evalueringsaktiviteter overfor deltakerne. Samtidig vil AFF unngå en form for enkeltkretslæring blant deltakerne, som gjerne kjennetegnes av mangel på refleksjon over årsaken til en konsekvens eller at en situasjon har oppstått (Filstad, 2016:45). Å få deltakerne til å reflektere over egen læring kan føre til positiv endring i atferd også etter at programmet er over, da det vil hjelpe dem med å identifisere knyttede misforhold og feil (Filstad, 2016:46).

6.0 Avslutning

Formålet med denne rapporten har vært å levere en anvendelig modell og fremgangsmåte til oppdragsgiver, AFF, som de kan benytte i evalueringen av sine bedriftsspesifikke ledelsesutviklingsprogram. I lys av dette har vi ønsket å besvare problemstillingen: *Hvordan bør AFF, med bakgrunn i oppdatert forskning og ekspertkompetanse, evaluere sine ledelsesutviklingsprogram?*

6.1 Konklusjon

Vi har blitt oppmerksomme på utfordringene ved å lage en generisk modell som er forenlig med teori og praksis. Modellen forutsetter derfor at AFF tar noen bedriftsspesifikke hensyn når evalueringen planlegges, slik evalueringen blir tilpasset det programmet som evalueres. Med bakgrunn i de mest fremtredende funnene i analysen, har vi utviklet et generisk rammeverk som er presentert som en femstegsmodell.

Våre funn understreker betydningen av å sette konkrete mål i forkant av programmet som gjenspeiler innholdet i programmet. På denne måten kan en i etterkant vurdere hvorvidt disse målene ble oppnådd og ga ønsket effekt. Videre er det viktig at en kombinerer kvantitative og kvalitative undersøkelser, slik at en både får innsikt i de store trendene på organisasjonsnivå og de individuelle opplevelsene til deltakerne. Våre funn vektlegger også betydningen av midtveisevalueringer som vurderer hvorvidt programmet er på rett spor, og som tilrettelegger for fortløpende justeringer. Avslutningsvis må innholdet og spørsmålene i evalueringen tilpasses etter kontekstuelle forhold. Det er også hensiktsmessig å påpeke at vi på bakgrunn av prosjektets omfang ikke hadde anledning til å teste ut modellen i praksis. Det kan dermed være nødvendig med justeringer når AFF tar modellen i bruk.

6.2 Videre forskning

I etterkant av prosjektet ser vi at det kunne vært aktuelt å undersøke om våre resultater gjelder for de åpne programmene i AFF, eventuelt en lignende organisasjon i samme bransje, eller om det kun gjelder for de bedriftsspesifikke programmene. I lys av dette, kunne det vært relevant å kombinere vår studie med et kvantitativ studie for å få et mer generaliserbart resultat. Det kunne eksempelvis vært interessant å undersøke hvilke indikatorer som kan brukes til å måle ulike målsetninger.

Til tross for at det ikke er mulig å generalisere funn i kvalitativ forskning, tror vi likevel at innsikten vi har fått om evaluering av ledelsesutviklingsprogram har en generaliseringsverdi, da den i kombinasjon med tidligere forskning kan øke forståelsen av evalueringsprosesser i organisasjoner.

7.0 Litteraturliste

AFF. (u.å.). Om AFF. Hentet fra: https://aff.no/om-aff/

Anderson, V. (2007). *The Value of learning: A new model of value and evaluation*. London: Chartered Institute of Personnel and Development.

Argyris, C., & Schön, D. A. (1978) *Organizational learning.* Reading, MA: Addison-Wesley.

Argyris, C. & Schön, D. A. (1996). *Organizational Learning II: Theory, Method and Practice.* Reading, Massachusetts: Addison-Wesley.

Askim, J., Døving, E. & Johnsen, Å. (2013) Evalueringspraksis i den norske statsforvaltningen 2005–2011. I Anne Halvorsen, Nina Jentoft og Einar Lie Madsen (red): *Evaluering: Tradisjoner, praksis, mangfold*. (s. 72-88). Bergen: Fagbokforlaget

Bastøe, P. Ø. (2020). Fem Forutsetninger For Gode Evalueringer. Oslo: Gyldendal.

Bates, R. (2004). A critical analysis of evaluation practice: the Kirkpatrick model and the principle of beneficence. *Evaluation and Program Planning,* 27(3), 341-347. https://doi.org/10.1016/j.evalprogplan.2004.04.011

Boulmetis, J. & Dutwin, P. (2011). *The ABCs of Evaluation: Timeless techniques for program and project managers* (3. utg.). San Francisco: Jossey-Bass.

Boyatzis, R.E. (2008). Leadership development from a complexity perspective. *Consulting Psychology Journal: Practice and Research*, 60(4), 298-313. https://doi.org/10.1037/1065-9293.60.4.298

Brinkerhoff, R. (2003). *The Success Case Method*: *Find Out Quickly What's Working and What's Not.* San Francisco: Berrett-Koehler.

Chen, H. T. (2015). *Practical Program Evaluation: Theory-Driven Evaluation and the Integrated Evaluation Perspective*. Sage Publications.

Day, D. V. & Dragoni, L. (2015). Leadership development: An outcome-oriented review based on time and levels of analyses. *Annual Review of Organizational Psychology and Organizational Behavior.* (2), 133–156. https://doi.org/10.1146/annurev-orgpsych-032414-111328

Edwards, G., & Turnbull, S. (2013). A Cultural Approach to Evaluating Leadership Development. *Advances in Developing Human Resources*, 15(1), 46 –60. https://doi.org/10.1177/1523422312467144

Filstad, C. (2016). *Organisasjonslæring – fra kunnskap til kompetanse*. Bergen: Fagbokforlaget.

Fournier, D. M. (1995). Establishing evaluative conclusions: A distinction between general and working logic. *New Directions for Program Evaluation*, 68(Winter 1995), 15-32. https://doi.org/10.1002/ev.1017

Giangreco, A., Carugati, A., & Sebastiano, A. (2010). Are we doing the right thing? Food for thought on training evaluation and its context. *Personnel Review*, 39(2), 162-177. https://doi.org/10.1108/00483481011017390

Halvorsen, A. & Madsen, E. L. (2013). Styring og læring gjennom evaluering. I Anne Halvorsen, Nina Jentoft og Einar Lie Madsen (red): *Evaluering: Tradisjoner, praksis, mangfold*. (s. 13-27). Bergen: Fagbokforlaget

King, E., & Nesbit, P. (2015). Collusion with denial: Leadership development and its evaluation. *Journal of Management Development.* 34(2). https://doi.org/10.1108/JMD-02-2013-0023

Kirkpatrick, D.L. (1959). Techniques for Evaluating Training Programs. *Journal of American Society for Training and Development*, 13(11), 3-9.

Kirkpatrick, J. D., & Kirkpatrick, W. K. (2016). *Kirkpatrick's four levels of training evaluation*. Association for Talent Development.

Kuvaas, B., & Dysvik, A. (2016). *Lønnsomhet gjennom menneskelige ressurser: Evidensbasert HRM* (3.utg). Bergen: Fagbokforlaget.

Lai, L. (2013). *Strategisk kompetanseledelse* (3.utg). Bergen: Fagbokforlaget.

Lingham, T., Richley, B., & Rezania, D. (2006). An evaluation system for training programs: a case study using a four-phase approach. *Career development international*, 11(4), 334-351. https://doi.org/10.1108/13620430610672540

McDavid, J., & Hawthorn, L. (2006). *Program evaluation & performance measurement: An introduction to practice.* Thousand Oaks, Calif: Sage.

Mehlbye, J., Rieper, O., & Togeby, M. (1993). *Håndbog i evaluering*. AKF Forlaget.

Owen, J. M. (2007). *Program evaluation: Forms and approaches* (3. utg). New York: The Guilford Press.

Passmore, J. and Velez, M. (2012). SOAP-M: a training evaluation model for HR. *Industrial and Commercial Training*. 44(6), 315-325. https://doi.org/10.1108/00197851211254743

Phillips, J. (1996). *Accountability in Human Resource Management. Oxford:* Butterworth-Heinemann.

Singh, R. & Ramdeo, S. (2020). *Principles and Contextual Perspectives. Leading Organizational Development and Change*. Cham: Palgrave Macmillan.

Sverdrup, S. (2014). *Evaluering Tilnærminger, modeller og eksempler.* (1.utg). Oslo: Gyldendal Akademisk.

Tjora, A. (2017). *Kvalitative forskningsmetoder i praksis* (3. utg.). Oslo: Gyldendal Akademisk.

Topno, H. (2012). Evaluation of Training and Development: An Analysis of Various Models. *Journal of Business and Management*, 5(2), 16-22. https://doi.org/10.9790/487X-0521622

Ulrich, D. (2014). The future targets or outcomes of HR work: Individuals, organizations and leadership. *Human Resource Development International*, 17(1), 1-9. https://doi.org/10.1080/13678868.2013.825144

Wallace, D. M., Torres, E, M., Zaccaro, S, J. (2021). Just what do we think we are doing? Learning outcomes of leader and leadership development. *The Leadership quarterly*, 32(5), 1-13. https://doi.org/10.1016/j.leaqua.2020.101494

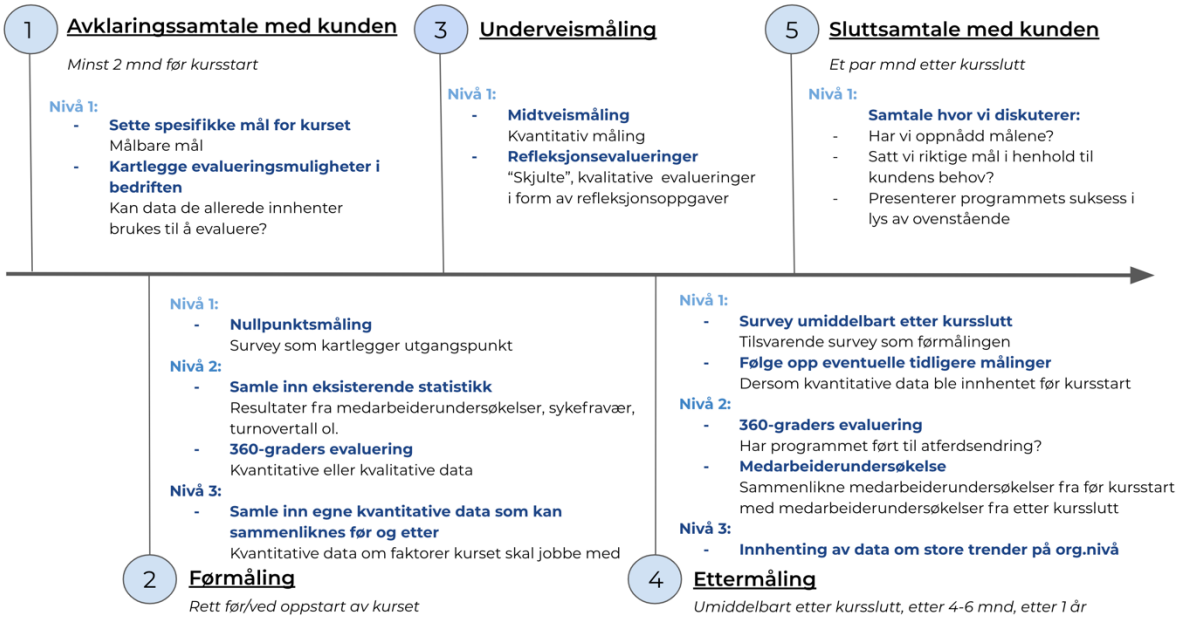
Warr, P., Bird, M., & Rackham, N. (1970). *Evaluation of management training: A practical framework, with cases, for evaluating training needs and results*. Gower Press.

Watkins, K. E., Lysø, I. H., & deMarrais, K. (2011). Evaluating executive leadership programs: A theory of change approach. *Advances in Developing Human Resources*, *13*(2), 208-239. https://doi.org/10.1177/1523422311415643

8.0 Vedlegg

Vedlegg 1: Generisk modell for evaluering

Modellen er i utgangspunktet laget som PowerPoint-fil, men er vist i bildeformat. AFF vil få tilgang til PowerPoint-filen.



1 Avklaringssamtale med kunden

Tidspunkt: *Minst to måneder før*

Hvem skal være med?

Nivå 1: Evaluator/kursleder og kunde (representant fra ledelsen eller HR-avdeling)

Nivå 2: Kursdeltakere og tillitsvalgt

Møtets innhold:

- **Fastsette mål for programmet**
  - Spesifikke, målbare, oppnåelige, relevante mål som er tidsfestet
  - Ikke for generelle
  - Målene må reflektere innholdet til kurset

- **Kartlegge tilgjengelige verktøy for evaluering i samarbeid med kunde**
  - Kartlegge evalueringsmuligheter hos kunden: medarbeiderundersøkelse, medarbeidersamtale, eksisterende statistikk og lignende
    - Eks.: Mulighetene for å gjennomføre en 360- graders evaluering 6 mnd. etter kursslutt?
    - Eks.: Mulig å diskutere lederens utvikling ifm. kurset i medarbeidersamtalen, som kan rapporteres til AFF?
    - Eks.: Muligheter for å legge inn et par spørsmål i bedriftens medarbeiderundersøkelse, som er relatert til kursets mål? Gjerne både i medarbeiderundersøkelsen som gjennomføres før og etter kurset, slik at det kan sammenliknes

2 Førmåling

Tidspunkt: *Så tett opp til kurset som mulig*

Hensikt: Samle inn data som kan sammenliknes med data fra etter kursslutt. Avdekke effekt og korrelasjon.

Nivå 1:

- **Nullpunktsmåling (Ideelt sett: skal ikke ha skjedd endringer mellom førmålingen og kursstart)**
  - I form av survey enten i forkant av kurset (eks. to uker før) eller første kursdag
    - Selvrefleksjon: nåværende status for å ha noe å sammenlikne med
    - Både skala og åpne spørsmål

Nivå 2:

- **Samle inn eksisterende statistikk (må kunne knyttes til målene)**
  - Eksempler på eksisterende statistikk:
    - Medarbeiderundersøkelser
    - Sykefravær
    - Turnover
    - Konfliktnivå
    - Vikarbruk
    - Fakturering stall
    - Salgstall
    - Intern VS. ekstern timeføring

- **360-graders evaluering**
  - Innhente vurderinger av den enkelte leders jobbrelaterte atferd, der en vektlegger ønsket utvikling fra flere kilder i lederens omgivelser.

Nivå 3:

- **Samle inn egne kvantitative data som kan sammenliknes før og etter**
  - Kvantitative data om faktorer kurset skal jobbe med
    - Eksempler på dette kan være de samme som på **Nivå 2:** , forskjellen er dog at evaluatoren må bidra til datagenereringen, og ikke bare innhente eksisterende data

3 Underveismåling

Hensikt: Å få tilbakemelding for justeringer - se om programmet følger det man har tenkt

Nivå 1:

Midtveismåling: En kvantitativ evaluering midt i for kursdeltakerne

- Eksempler til relevante spørsmål:
  - Hvor mye har du jobbet med x siden sist?
  - I hvilken grad opplever du at du har fått verktøy til x?
  - I hvilken grad oppleves innholdet i kurset som relevant for jobben din?

Nivå 1:

Refleksjonsevaluering: Dette er flere diskré evalueringer som mates inn i kurset:

- En refleksjonsevaluering kan for eksempelvis være samtale i plenum, dagbok, akvarium, mentimeter og/eller kahoot
- Formålet:
  - Evaluator får signaler om en har tatt med seg videre det en lærte sist, eller om læringen er blitt glemt.
  - Evaluator får signaler om det de lærer er mulig å implementere i virksomheten, eller hvilke utfordringer som eventuelt oppstår
- OBS: Viktig at det er en naturlig del av programmet, og ikke forstyrrende for deltakerne. Dette skal ikke fremstilles som en evaluering overfor deltakerne eller kunden.
  - Eks: Kursleder spør kursdeltakerne:
    - Hva lærte vi sist?
    - Hvordan har dere tatt i bruk det dere lærte?
    - Hva har vært utfordrende/utfordringer?
    - Tenk på en gang siden sist samling der du ...

4 Ettermåling

Tidspunkt: *Umiddelbart etter kursslutt, etter 4-6 mnd, etter 1 år*

Hensikt: Se i sammenhengen med nullpunktsmålingen - har programmet hatt ønsket effekt? Er målene nådd?

Nivå 1:

- **Umiddelbar survey som matcher nullpunktssurvey, men med noen tilleggsspørsmål**
- **Følge opp eventuelle tidligere målinger**
  - Dersom man innhentet egne kvantitative data i forkant av kurset, må tilsvarende materiale samles inn på nytt for å kunne sammenliknes
  - Dersom man har lagt inn spørsmål i forbindelse med kursets mål i bedriftens medarbeiderundersøkelse, må dette følges opp

Nivå 2:

- **360-graders evaluering**
  - Når: *4-6 mnd etter kursslutt*
  - Formål: Hvordan har lederen implementert læring fra programmet i arbeidet?

- **Medarbeiderundersøkelse**
  - Relevant dersom man fikk innsyn i medarbeiderundersøkelser fra før kursstart og sammenlikne med medarbeiderundersøkelse ved kursslutt

Nivå 3:

- **Innhenting av data om store trender på organisasjonsnivå**
  - Når: *1 år etter kursslutt*
  - Eks: Følge opp og sykefraværsstatistikk

5 Sluttsamtale med kunden

Tidspunkt: *Etter at ettermålingene er gjennomført*

Nivå 1:

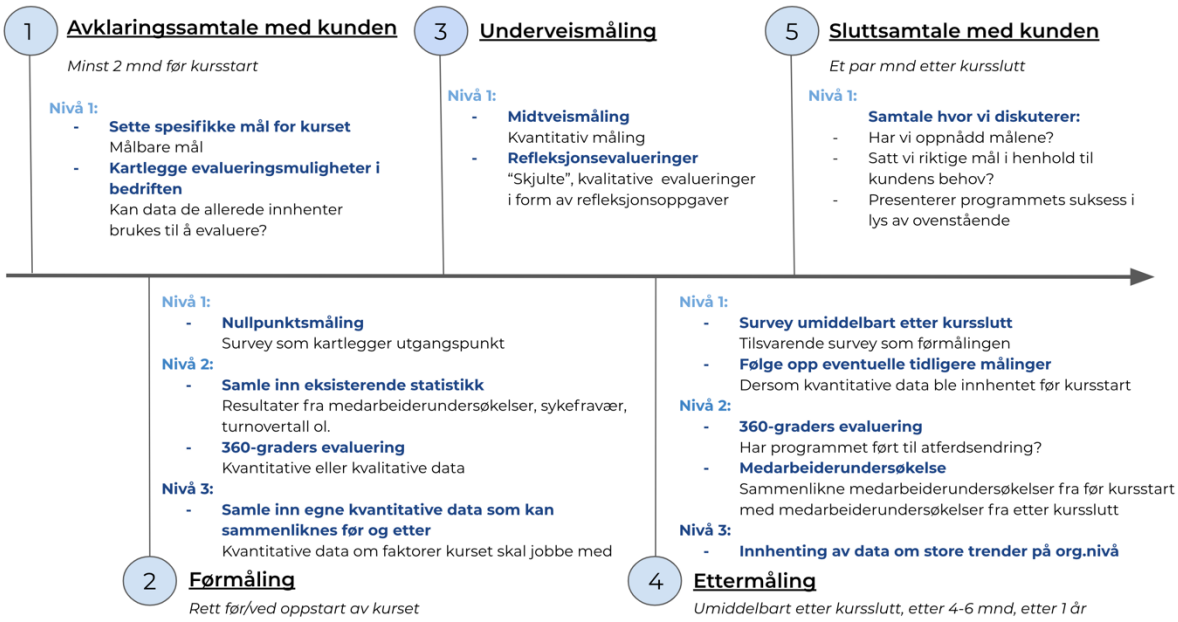
Innhold til diskusjon:

- Har vi oppnådd målene?
- Satt vi riktige mål i henhold til kundens behov?

  - Presentere programmets suksess i lys av de (to) ovenstående poengene

Vedlegg 2: Modell for evaluering med case-eksempler

Modellen er i utgangspunktet laget som PowerPoint-fil, men er vist i bildeformat. AFF vil få tilgang til PowerPoint-filen.



1 Avklaringssamtale med kunden

Tidspunkt: *Minst to måneder før*

Hvem skal være med?

Nivå 1: Evaluator/kursleder og kunde (representant fra ledelsen eller HR-avdeling)

Nivå 2: Kursdeltakere og tillitsvalgt

Møtets innhold:

- **Fastsette mål for programmet**
  - Spesifikke, målbare, oppnåelige, relevante mål som er tidsfestet
  - Ikke for generelle
  - Målene må reflektere innholdet til kurset

- **Kartlegge tilgjengelige verktøy for evaluering i samarbeid med kunde**
  - Kartlegge evalueringsmuligheter hos kunden: medarbeiderundersøkelse, medarbeidersamtale, eksisterende statistikk og lignende
    - Eks.: Mulighetene for å gjennomføre en 360- graders evaluering 6 mnd. etter kursslutt?
    - Eks.: Mulig å diskutere lederens utvikling ifm. kurset i medarbeidersamtalen, som kan rapporteres til AFF?
    - Eks.: Muligheter for å legge inn et par spørsmål i bedriftens medarbeiderundersøkelse, som er relatert til kursets mål? Gjerne både i medarbeiderundersøkelsen som gjennomføres før og etter kurset, slik at det kan sammenliknes

AFF

1 Fastsette mål for programmet

AFF sine nåværende mål:

- Økt strategisk kompetanse blant topplederne
- Styrket relasjoner og nettverk på tvers av organisasjonen (sosial kapital)

Hvordan gjøre disse målene målbare?

- Deltakerne skal etter programslutt i større grad samarbeide med/benytte seg av kompetansen til kollegaer på tvers av bedriftskontorer/lokasjoner (Gjerne tallfest "i større grad")
- Deltakerne skal benytte *dette* verktøyet oftere enn ved kursstart (f.eks. en intern kommunikasjonsplattform) (Gjerne tallfest "i større grad" - f.eks. 10% mer)
- Deltakerne skal i løpet av programmet utvikle nye bekjentskaper med kollegaer på tvers i landet
- Deltakerne skal bli bevisst hvordan barrierer og etablerte normer i bransjen hindrer nytenkning

Flere mål kan ytterligere spesifiseres ut ifra de spesifikke aktivitetene i kurset - hva de konkret øver på. Måle konkret om de blir bedre på de aktivitetene de øver på.

AFF

1 Kartlegge mulighet for måling i bedriften

- Er bedriften villig til å la AFF gjennomføre en 360 graders evaluering av et utvalg ledere i forkant og etterkant av programmet?
    - Eventuelt at bedriften gjør 360-evaluering selv, eller inkluderer oppfølgingsspørsmål fra kurset i en medarbeidersamtale?
- Har bedriften en medarbeiderundersøkelse? Hvor ofte gjennomfører de denne? Er det mulig å legge inn et par spørsmål i denne, som AFF kan få innsikt i utviklingen til?
    - Hvis ja: Formuler spørsmål som undersøker om det lederne lærer på kurs får ringvirkninger ut i bedriften. Eksempelvis:
        - I hvilken grad opplever du at det er fokus på ... i bedriften
        - I hvilken grad kjenner du til ...
        - I hvilken grad opplever du at lederen din inkluderer deg i ...
- Samler bedriften inn eksisterende statistikk det kan være interessant å få innsikt i?
    - Eksempelvis:
        - Hvor ofte prosjekter består av samarbeid mellom ansatte i ulike deler av Norge?
        - Bruken av dette verktøyet, som indikerer at de ansatte har jobbet med dette som kurset så på?

2 Førmåling

Tidspunkt: *Så tett opp til kurset som mulig*

Hensikt: Samle inn data som kan sammenliknes med data fra etter kursslutt. Avdekke effekt og korrelasjon.

Nivå 1:

- **Nullpunktsmåling (Ideelt sett: skal ikke ha skjedd endringer mellom førmålingen og kursstart)**
    - I form av survey enten i forkant av kurset (eks. to uker før) eller første kursdag
        - Selvrefleksjon: nåværende status for å ha noe å sammenlikne med
        - Både skala og åpne spørsmål

Nivå 2:

- **Samle inn eksisterende statistikk (må kunne knyttes til målene)**
    - Eksempler på eksisterende statistikk:
        - Medarbeiderundersøkelser
        - Sykefravær
        - Turnover
        - Konfliktnivå
        - Vikarbruk
        - Fakturering stall
        - Salgstall
        - Intern VS. ekstern timeføring
- **360-graders evaluering**
    - Innhente vurderinger av den enkelte leders jobbrelaterte atferd, der en vektlegger ønsket utvikling fra flere kilder i lederens omgivelser.

Nivå 3:

- **Samle inn egne kvantitative data som kan sammenliknes før og etter**
    - Kvantitative data om faktorer kurset skal jobbe med
        - Eksempler på dette kan være de samme som på **Nivå 2:** , forskjellen er dog at evaluatoren må bidra til datagenereringen, og ikke bare innhente eksisterende data

AFF

2 Nullpunktsmåling, survey

Nivå 1:

Første kursdag deles ut en survey, med spørsmål som kartlegger deltakernes utgangspunkt før kurset. Husk at tilsvarende undersøkelse skal besvares rett etter kursslutt for å måle utvikling!

- Skalaspørsmål:
  - I hvilken grad opplever du å ha kjennskap til bedriftens strategi i dag?
  - Anslagsvis hvor mange ganger har du tatt kontakt med kollegaer ved andre bedriftskontor enn ditt eget den siste måneden?
  - I anslagsvis hvor mange prosent av prosjektene dine er det inkludert ansatte fra andre bedriftskontor?

- Åpne spørsmål:
  - Hvordan opplever du at bedriftens strategi er synlig i bedriften i dag?
  - Tenk på en gang du jobbet på et prosjekt og ikke hadde kompetansen du behøvde i ditt team. Hvordan håndterte du det? Hvordan kunne denne situasjonen ha vært løst?

AFF

2 Nullpunktsmåling, samle inn ytterligere data (nivå 2 & 3)

Forskjellen på dataene fra survey-målingen og denne, er at det her samles inn data som ikke er basert på ledernes egne oppfatning, men objektive data fra statistikk, medarbeidere ol.

Eksisterende data (Nivå 2):
Dersom dere i førsamtalen fikk bekreftet at bedriften samler inn data det er aktuelt å få innsikt i for å sammenligne trender i bedriften før og etter kurset, samler dere inn denne nå.

Eksempelvis:
Hvor ofte samarbeider lederne med kollegaer ved andre bedriftskontor enn sitt eget?
- Eks.: Statistikk over prosjekter på tvers av lokasjoner

360-graders evaluering (Nivå 2):
Samle inn egne data gjennom kvalitative eller kvantitative 360-graders evalueringer.

Samle inn egne data (Nivå 3):
Dersom bedriften ikke har eksisterende data på området dere ønsker innsikt i, og de lar dere samle inn data på egenhånd, gjør dere dette nå.

3 Underveismåling

Hensikt: Å få tilbakemelding for justeringer - se om programmet følger det man har tenkt

Nivå 1:
Midtveismåling: En kvantitativ evaluering midt i for kursdeltakerne
- Eksempler til relevante spørsmål:
  - Hvor mye har du jobbet med x siden sist?
  - I hvilken grad opplever du at du har fått verktøy til x?
  - I hvilken grad oppleves innholdet i kurset som relevant for jobben din?

Nivå 1:
Refleksjonsevaluering: Dette er flere diskré evalueringer som mates inn i kurset:
- En refleksjonsevaluering kan for eksempelvis være samtale i plenum, dagbok, akvarium, mentimeter og/eller kahoot
- Formålet:
  - Evaluator får signaler om en har tatt med seg videre det en lærte sist, eller om læringen er blitt glemt.
  - Evaluator får signaler om det de lærer er mulig å implementere i virksomheten, eller hvilke utfordringer som eventuelt oppstår
- OBS: Viktig at det er en naturlig del av programmet, og ikke forstyrrende for deltakerne. Dette skal ikke fremstilles som en evaluering overfor deltakerne eller kunden.
  - Eks: Kursleder spør kursdeltakerne:
    - Hva lærte vi sist?
    - Hvordan har dere tatt i bruk det dere lærte?
    - Hva har vært utfordrende/utfordringer?
    - Tenk på en gang siden sist samling der du ...

AFF

3 Midtveismåling

Del ut et spørreskjema når programmet er ca. halvveis. Dette skjemaet skal ikke være identisk skjemaet som ble utdelt første og siste seminardag.

Eksempler på spørsmål:
- Skalaspørsmål:
  - Hvor mye har du jobbet med kollegaer ved andre lokasjoner siden sist?
  - I hvilken grad opplever du at du har fått verktøy til samarbeide med kollegaer på tvers av organisasjonen?
  - I hvilken grad opplever du verdien av å samarbeide med kollegaer ved andre bedriftskontorer enn ditt eget?
  - I hvilken grad oppleves innholdet i kurset som relevant for jobben din?

- Åpne spørsmål:
  - Hvilke utfordringer har du opplevd i forbindelse med å samarbeide på tvers av organisasjonen i det daglige arbeidet ditt?
  - Hvordan kan dette kurset tilpasses enda bedre til å møte dine utfordringer?

AFF

3 Refleksjonsevalueringer

Arrangér aktiviteter der du får innsikt i om deltakerne har tatt med seg videre det de lærte på forrige samling.

<u>Eksempler på aktiviteter og spørsmål:</u>

Dagbok
- Notere når man blir bevisst på en situasjon der barrierer og etablerte normer i bransjen hindrer nytenkning

Akvarium
- Hvilke verktøy fra kurset har du benyttet deg av i arbeidshverdagen din?
- Tenk på en situasjon der hadde mer verdi å samarbeide enn å utføre oppgaven alene

Samtale/diskusjon mellom deltakerne
- Hva lærte vi sist?
- Hvordan har dere tatt i bruk det dere lærte?
- Hva har vært utfordrende/utfordringer?
- Tenk på en gang siden siste samling du samarbeidet på tvers av organisasjonen, der dette bidro positivt til prosjektet ditt
- Tenk på en gang siden siste samling du samarbeidet på tvers av organisasjonen, der dette bidro negativ

4 Ettermåling

Tidspunkt: *Umiddelbart etter kursslutt, etter 4-6 mnd, etter 1 år*

Hensikt: Se i sammenhengen med nullpunktsmålingen - har programmet hatt ønsket effekt? Er målene nådd?

Nivå 1:
- **Umiddelbar survey som matcher nullpunktssurvey, men med noen tilleggsspørsmål**
- **Følge opp eventuelle tidligere målinger**
  - Dersom man innhentet egne kvantitative data i forkant av kurset, må tilsvarende materiale samles inn på nytt for å kunne sammenliknes
  - Dersom man har lagt inn spørsmål i forbindelse med kursets mål i bedriftens medarbeiderundersøkelse, må dette følges opp

Nivå 2:
- **360-graders evaluering**
  - Når: *4-6 mnd etter kursslutt*
  - Formål: Hvordan har lederen implementert læring fra programmet i arbeidet?
- **Medarbeiderundersøkelse**
  - Relevant dersom man fikk innsyn i medarbeiderundersøkelser fra før kursstart og sammenlikne med medarbeiderundersøkelse ved kursslutt

Nivå 3:
- **Innhenting av data om store trender på organisasjonsnivå**
  - Når: *1 år etter kursslutt*
  - Eks: Følge opp og sykefraværsstatistikk

AFF

4 Survey ved kursslutt (Nivå 1)

Nivå 1:

Identisk som den som ble utdelt ved kursstart, men med noen åpne tilleggsspørsmål rundt tilfredsheten:

Eksemplevis:
- Hvilke deler av kurset opplevde du som mest relevant for din arbeidshverdag, og hvorfor?
- Hvilke deler av kurset opplevde du som minst relevant for din arbeidshverdag, og hvorfor?

AFF

4 Innhenting av ytterligere data (Nivå 1):

Dersom dette ble innhentet i steg 2, innhent samme data på nytt slik at dette kan sammenliknes.

Eksisterende data (Nivå 2):
Eksempelvis:
Hvor ofte samarbeider lederne med kollegaer ved andre bedriftskontor enn sitt eget?
- Eks.: Statistikk over prosjekter på tvers av lokasjoner

360 graders evaluering (Nivå 2):
Eksempelvis:
Kvantitativ formulering: I hvilken grad opplever du at lederen kommuniserer og samarbeider på tvers i organisasjonen?

Kvalitativ formulering: Opplever du at lederen kommuniserer og samarbeider på tvers i organisasjonen?
- Hvis ja, hvordan?
- Hvis nei, tenk på en situasjoner der det hadde vært gunstig med samarbeid på tvers av organisasjonen?

Samle inn egne data (Nivå 3):
Dersom bedriften ikke hadde dataene dere var ute etter i førmålingen, eller dersom bedriften ikke har oppdaterte tall, må dere samle inn egne data.

AFF

4 Ettermåling (Nivå 2 & 3)

Nivå 2:

- 360-graders evaluering
  - Formål: Hvordan har lederen implementert læring fra programmet i arbeidet?
  - Disse spørsmålene kan gjerne utformes pba. svarene gitt i underveismålingene
  - Eksempler på spørsmål:
    - Opplever du at kursdeltakeren initierer til samarbeid med kollega på tvers av organisasjonen mer enn før?
      - Hvis ja, på hvilken måte?
    - Opplever du at kursdeltakeren legger til rette for kreativitet og nytenkning?
      - Hvis ja, hvordan?

- Medarbeiderundersøkelse
  - Relevant dersom man fikk innsyn i medarbeiderundersøkelser fra før kursstart og sammenlikne med medarbeiderundersøkelse ved kursslutt

Nivå 3:

- Innhenting av data om store trender på organisasjonsnivå
  - Eks: Følge opp på antall prosjekter med samarbeid på tvers av organisasjonen

5 Sluttsamtale med kunden

Tidspunkt: *Etter at ettermålingene er gjennomført*

Nivå 1:

Innhold til diskusjon:

- Har vi oppnådd målene?
- Satt vi riktige mål i henhold til kundens behov?

  - Presentere programmets suksess i lys av de (to) ovenstående poengene

AFF

5 Sluttsamtale med kunden

Vi diskuterer i fellesskap:

- Har vi oppnådd målene?
- Satt vi riktige mål i henhold til kundens behov?
  - Var dette et godt mål for å oppnå det dere ønsket?
- Opplever kunden endring hos de ansatte?

- Presentere programmets suksess i lys av de ovenstående poengene

Vedlegg 3: Mandat fra AFF

AFF

Evaluering av ledelses- og organisasjonsutvikling

1. OM AFF

AFFs misjon er å bidra til verdiskapende utvikling av lederskap og organisasjoner i et komplekst og globalisert samfunn. Vi hjelper samfunnsbyggende virksomheter i privat-, ideell- og offentlig sektor med økt verdiskapning gjennom leder- og organisasjonsutvikling. Vi er godt forankret i norsk og skandinavisk tradisjon, samtidig som vi bygger på og samarbeider med de beste internasjonale miljøene innen ledelsesforskning.

Våre tjenesteområder dekker de fleste områder innen utvikling av organisasjoner, medarbeiderskap og ledelse. Vi er spesielt opptatt av noen strategiske kompetanseområder: endring og utvikling av organisasjoner, ledergruppeutvikling, innovasjonsprosesser og konflikthåndtering.

Flere av våre konsulenter er psykologer og organisasjonspsykologer med dyptgående kompetanse innen atferdspsykologi og tilgrensende fagområder. Mange har også vært ledere og kjenner lederes hverdag fra innsiden.

Kort om oss:

- Blant Norges største ledelse- og organisasjonsutviklingsmiljø med over 50 kompetente medarbeidere.
- Blant verdens 50 beste tilbydere av lederutviklingsprogram på Financial Times ranking.
- Etablert i 1952 med et klart samfunnsoppdrag.
- Partnerskap og samlokalisert med Norges Handelshøyskole i Bergen og i Oslo.

Vi er stolte av våre åpne program som supplerer lederutviklingsporteføljen for mange sentrale virksomheter. Gjennom programmene er vi tett på aktuelle lederutfordringer og sikrer at vi til enhver tid kan tilby relevant utvikling og faglig påfyll. Våre åpne programmer er: Solstrandprogrammet, Solstrandprogrammet Accelerate og AFF Samspill & Ledelse. Programmene rates årlig i Financial Times sin globale måling og sammen med NHH er vi rangert blant verdens 50 beste åpne program.

2. EVALUERING AV LEDERUTVIKLING

Hvert år gjennomfører vi mer en rekke utviklingsprosjekter i samarbeid med en bredt spekter av kunder i ulike bransjer og kontekster. Prosjektene varierer i omfang fra en kort workshop til lederutviklingsprogram som strekker seg over flere år. Ambisjonene kan variere fra å hjelpe en ledergruppe å bli mer effektiv, til å å endre kulturen i et stort konsern. Hvordan vi evaluerer prosjektene varierer også ut fra kundens behov og de involverte konsulentenes anbefaling. Oftest evaluerer vi gjennom deltakeres opplevde utbytte og grad av fornøydhet. I enkeltprosjekter utvikler vi mer omfattende evalueringsdesign med før- og etter målinger.

Vi vet at selvevaluering medfører en rekke metodiske problemer og begrenset innsikt i en intervensjons effekt. I tråd med vår verdi om å være faglige solide ønsker vi å utvikle en modell og fremgangsmåte for hvordan vi skal evaluere våre utviklingstiltak.

3. PROBLEMSTILLING

3.1 Kartlegging av forskning og beste praksis på evaluering

AFFs evalueringsmodell må være faglig forankret i oppdatert forskning og beste praksis. Vi ønsker derfor en kartlegging av forskning på evaluering av utviklingstiltak (herunder, lederutvikling og annen relevant ledelses- og organisasjonsutvikling) og en undersøkelse av «beste praksis». Vi ser for oss at sistnevnte kan gjennomføres gjennom intervjuer eller fokusgrupper med eksperter og praktikere.

Vi ønsker at studentene vurderer om det er for omfattende å ta for seg evaluering av alle typer utviklingstiltak innen ledelse- og organisasjonsutvikling, og om det er hensiktsmessig å avgrense mot evaluering av lederutviklingsprogram.

Vi ønsker at studentene inkluderer en vurdering av Kirkpatricks og Brinkerhoffs modeller for evaluering, samt at NHH-forsker Alexander Sandvik inkluderes i ekspert-undersøkelsene.

3.2 Utvikling av modell og fremgangsmåte for evaluering

Vi ønsker en modell og fremgangsmåte for evaluering som vi kan benytte i våre prosjekter. Modellen må balansere forskningsfaglig forankring med praktisk anvendbarhet.

Vi ønsker ikke å legge føringer for modellen og fremgangsmåten studentene utvikler, men for illustrasjon kan vi se for oss at modellen eksempelvis inkluderer (1) flere steg (før- og etter målinger), (2) beskrivelse av hvem som skal besvare målingene (selvrapport, medarbeidere, overordnet), (3) format på målinger (spørreundersøkelse, intervjuer, fokusgrupper), og (4) anbefalte indikatorer/skalaer.

Vi ønsker at studentene lager en så praktisk fremgangsmåte som mulig for våre konsulenter som skal implementere modellen og gjennomføre evalueringene. For eksempel ønsker vi en beskrivelse av hvor lenge før utviklingstiltaket en før-måling bør gjennomføres, hvilke spørreskjema/indikatorer som bør inkluderes, osv.

Vedlegg 4: Informasjonsskriv

Deltakelse i forskningsprosjekt knyttet til evalueringsmetoder av ledelsesutviklingsprogram

Vi ønsker med dette å invitere deg til å delta i et forskningsprosjekt hvor formålet er å finne beste praksis for å evaluere ledelsesutviklingsprogram. I dette skrivet får du informasjon om prosjektets mål og hva deltakelse vil innebære for deg.

Formål

Formålet med dette prosjektet er å gi AFF en anvendelig modell og fremgangsmåte for å evaluere egne utviklingstiltak. Modellen skal være forankret i oppdatert forskning og teori fra evalueringsfeltet. AFFs nåværende evalueringsrutiner består i all hovedsak av vurdering av deltakernes egne opplevde utbytte og grad av fornøydhet. Denne formen for evaluering medfører en rekke metodiske problemer, og gir ikke en reell og helhetlig måling av det som ønskes evaluert. En sentral del av formålet med dette prosjektet er dermed også å sikre at AFFs fremtidige evalueringsprosesser inkluderer en måling på andre nivåer enn utelukkende opplevd utbytte, som f.eks. måling av læring og endret atferd.

Dette prosjektet er en del av emnet Prosjektforum på masterprogrammet Organisasjon, ledelse og arbeid ved Universitetet i Oslo. Vi er en gruppe på seks studenter som har fått et prosjektmandat fra AFF. Prosjektet gjennomføres høsten 2021, og har et omfang på 20 studiepoeng.

Prosjektet vil resultere i en rapport, som blir tilgjengelig for følgende parter:

- AFF, til bruk i eget arbeid
- UiO, Institutt for sosiologi og samfunnsgeografi for vurdering

Hvem er ansvarlig for forskningsprosjektet?

- Universitetet i Oslo

Hvorfor får du spørsmål om å delta?

Utvalget består av eksperter innenfor evalueringsfeltet og læringsfeltet. Vi vil intervjue 6-12 personer med ekspertkunnskap om prosjektets tematikk, og ønsker at du skal delta på bakgrunn av din kjennskap til temaet.

Hva innebærer det for deg å delta?

Du blir invitert til et intervju som varer mellom 45 og 60 minutter. Intervjuet vil bli tatt opp på en kyptert diktafon, og vil senere bli transkribert på en sikret lagringstjeneste godkjent av UiO. Deretter vil lydopptaket slettes. Du får selv anledning til å avgjøre om du ønsker å omtales med navn i oppgaven, eller om du vil bli anonymisert.

Det er frivillig å delta

Det er frivillig å delta i prosjektet. Hvis du velger å delta, kan du når som helst trekke ditt samtykke tilbake uten å oppgi noen grunn. Alle opplysninger om deg vil da bli slettet. Det vil ikke ha noen negative konsekvenser for deg hvis du ikke vil delta eller senere velger å trekke deg.

Ditt personvern – hvordan vi oppbevarer og bruker dine opplysninger

Vi vil bare bruke opplysningene om deg til formålene vi har fortalt om i dette skrivet. Vi behandler opplysningene konfidensielt og i samsvar med personvernregelverket. Det er kun studentgruppen og prosjektansvarlig som vil ha tilgang ved behandlingsansvarlig institusjon. Dersom ønskelig, vil opplysninger som kan identifisere deg som person anonymiseres under transkriberingen. Etter dette slettes lydopptaket.

Hva skjer med opplysningene dine når vi avslutter forskningsprosjektet?

Prosjektet skal etter planen avsluttes 17. november 2021. Ved prosjektslutt vil alle personopplysninger og opptak være slettet.

Dine rettigheter

Så lenge du kan identifiseres i datamaterialet, har du rett til:

- innsyn i hvilke personopplysninger som er registrert om deg,
- å få rettet personopplysninger om deg,
- få slettet personopplysninger om deg,
- få utlevert en kopi av dine personopplysninger (dataportabilitet), og
- å sende klage til personvernombudet eller Datatilsynet om behandlingen av dine personopplysninger.

Hva gir oss rett til å behandle personopplysninger om deg?
Vi behandler opplysninger om deg basert på ditt samtykke.

På oppdrag fra UiO, Institutt for sosiologi og samfunnsgeografi, har NSD – Norsk senter for forskningsdata AS vurdert at behandlingen av personopplysninger i dette prosjektet er i samsvar med personvernregelverket.

Hvor kan jeg finne ut mer?
Hvis du har spørsmål til studien, eller ønsker å benytte deg av dine rettigheter, ta kontakt med:

- Kontaktperson for prosjektgruppen, Solveig Vik Hofseth på epost: solveivh@uio.no eller telefon: 98 02 50 69.
- Emneansvarlig for Prosjektforum ved UiO, Beate Karlsen på epost: beate.karlsen@sosgeo.uio.no
- NSD – Norsk senter for forskningsdata AS, på epost (personverntjenester@nsd.no) eller telefon: 55 58 21 17.

Samtykkeerklæring

Jeg har mottatt og forstått informasjon om prosjektet ”Evaluering av ledelsesutviklingsprogrammer”, og har fått anledning til å stille spørsmål. Jeg samtykker til:

- å delta i intervju

Kryss av i **en** av boksene for å gi samtykke til bruk av begrensede personopplysninger (navn, jobbtittel, formalkompetanse) i publiserte resultater av studien:

☐ Jeg godtar at dere bruker mitt navn, jobbtittel og formalkompetanse i publiserte resultater av studien

☐ Jeg ønsker å være anonym og godtar derfor ikke bruk av navn, jobbtittel og formalkompetanse i publiserte resultater av studien.

Jeg samtykker til at mine opplysninger behandles frem til prosjektet er avsluttet.

(Signert av prosjektdeltaker, dato)

Vedlegg 5: Hoved-intervjuguide

Introduksjon:

- Kort introduksjon om formålet med prosjektet
    - Vi er en prosjektgruppe bestående av seks studenter fra masterstudiet Organisasjon, ledelse og arbeid ved UiO. Gjennom emnet Prosjektforum har vi fått utdelt et mandat fra Norges største ledelses- og organisasjonsutviklingsmiljø, AFF, der vi skal utvikle en generisk modell og fremgangsmåte for hvordan AFF bør evaluere sine ledelsesutviklingsprogram. Disse programmene er bedriftsspesifikke med varighet på ca. 1 år. Vi er derfor i utgangspunktet opptatt av å belyse spørsmålene våre ut ifra denne konteksten.
    - (For å besvare problemstillingen ønsker vi å intervjue eksperter og praktikere for å få innsikt i beste praksis på evaluering. Vi ønsker å intervjue deg ettersom du har X som ditt ekspertområde, og håper derfor at du kan bidra til å sette nytt lys på evalueringsfeltet.)
- Er det greit for deg at vi tar taleopptak av intervjuet? Behandlingen av opptaket vil foregå som beskrevet i samtykkeerklæringen og i tråd med NSD sine retningslinjer.
- Vi ønsker å avklare/bekrefte om du ønsker sitatsjekk eller anonymitet. Dersom du godtar bruk av begrensede opplysninger, vil vi bruke ditt navn, jobbtittel og formalkompetanse i publiserte resultater av studien.

Evaluering og ledelsesutviklingsprogram:

1. Hva er din bakgrunn fra evalueringsfeltet? (kunnskaper, ferdigheter eller erfaringer)
2. Hva tenker du er formålet med evaluering?
3. Hva opplever du er det viktigste å tenke på når man skal gjennomføre en evaluering?
4. Hvordan skiller evaluering av ledelsesutviklingsprogram seg fra andre evalueringer?
5. Hvordan ville du gått frem for å evaluere et ledelsesutviklingsprogram?
    a. Hvem bør utføre evalueringen?
    b. Hva er det som bør evalueres?
    c. Når bør man gjennomføre evaluering (tid for før- og ettermålinger)?
    d. Hva slags metode bør man benytte seg av?
        i. Observasjon, fokusgruppe, intervju, kvantitative spørreundersøkelser osv.?

    e. Hva er styrker og svakheter ved den evalueringsmetoden du har lagt frem?
6. Hva tenker du om å benytte kontrollgruppe i evalueringen?
    a. Har du forslag til et alternativ for å sikre at læringsutbyttet kan knyttes til det konkrete programmet, dersom bruk av kontrollgruppe er uaktuelt?
7. I hvilken grad opplever du at man kan benytte seg av samme evalueringsmetode til ulike programmer?
    a. Bør de tilpasses spesifikt til det de skal måle?
8. Mange teoretiske modeller inkluderer flere nivåer av evaluering, for eksempel på individ- og organisasjonsnivå. I praksis har vi inntrykk av at virksomheter ikke har tid og ressurser til å evaluere på flere nivåer. Hvilke problemer kan oppstå i møtet mellom teori og praksis?
    a. Hvordan kan de løses?
9. Har du noen tanker om brukervennligheten til evalueringsteori?
10. Hva tenker du om selvevaluering ved ledelsesutviklingsprogrammer?
11. Vi ønsker å intervjue deg ettersom du har skrevet om...
    a. Hvordan kan man på best mulig måte fange opp negative utfall av ledelsesutviklingsprogrammer i evaluering?
    b. Hvilke konsekvenser kan oppstå dersom en unngår å evaluere?

Læring:

12. Hva vil du si er læring, i forbindelse med ledelsesutviklingsprogram?
13. Hva mener du er den beste måten å vurdere læringsutbyttet av et ledelsesprogram?
14. Hva er en god indikator på at noen har lært av programmet?
15. Forskning viser et økende fokus på ledelse i organisasjonen, og ikke bare hos individer. I hvilken grad er organisasjonslæring ønskelig/nyttig å måle i evaluering i et ledelsesutviklingsprogram?
    a. Hvordan kan og bør dette i så fall måles?
16. AFF evaluerer stort sett kun tilfredshet og opplevd læringsutbytte. Hva tror du er grunnen til det?
    a. Hva tenker du om å bruke tilfredshet som et mål på læring?
    b. Hvordan kan man eventuelt få en mer helhetlig vurdering?
17. Hvordan kan man vurdere/måle atferdsendring i en evaluering?
    a. Hva tenker om sammenhengen mellom atferdsendring og læring?

b. Hvilke parametere tenker du kan måle atferdsendring?

Avrundingsspørsmål:

- Er det noe du ønsker å tilføye?
- Kan vi kontakte deg igjen hvis det blir aktuelt?
- Har du forslag til andre eksperter/praktikere som kan være interessante å intervjue?
- Takke for at informanten stilte opp på intervjuet

Vedlegg 6: Godkjenning fra NSD

Behandlingen av personopplysninger er vurdert av NSD. Vurderingen er:

Det er vår vurdering at behandlingen av personopplysninger i prosjektet vil være i samsvar med personvernlovgivningen så fremt den gjennomføres i tråd med det som er dokumentert i meldeskjemaet med vedlegg den 06.09.2021, samt i meldingsdialogen mellom innmelder og NSD. Behandlingen kan starte.

DEL PROSJEKTET MED PROSJEKTANSVARLIG

For studenter er det obligatorisk å dele prosjektet med prosjektansvarlig (veileder). Del ved å trykke på knappen «Del prosjekt» i menylinjen øverst i meldeskjemaet. Prosjektansvarlig bes akseptere invitasjonen innen en uke. Om invitasjonen utløper, må prosjektansvarlig inviteres på nytt.

TYPE OPPLYSNINGER OG VARIGHET

Prosjektet vil behandle alminnelige kategorier av personopplysninger frem til 17.11.2021.

LOVLIG GRUNNLAG

Prosjektet vil innhente samtykke fra de registrerte til behandlingen av personopplysninger. Vår vurdering er at prosjektet legger opp til et samtykke i samsvar med kravene i art. 4 og 7, ved at det er en frivillig, spesifikk, informert og utvetydig bekreftelse som kan dokumenteres, og som den registrerte kan trekke tilbake.

Lovlig grunnlag for behandlingen vil dermed være den registrertes samtykke, jf. personvernforordningen art. 6 nr. 1 bokstav a.

PERSONVERNPRINSIPPER

NSD vurderer at den planlagte behandlingen av personopplysninger vil følge prinsippene i personvernforordningen om:

· lovlighet, rettferdighet og åpenhet (art. 5.1 a), ved at de registrerte får tilfredsstillende informasjon om og samtykker til behandlingen

· formålsbegrensning (art. 5.1 b), ved at personopplysninger samles inn for spesifikke, uttrykkelig angitte og berettigede formål, og ikke behandles til nye, uforenlige formål

· dataminimering (art. 5.1 c), ved at det kun behandles opplysninger som er adekvate, relevante og nødvendige for formålet med prosjektet
· lagringsbegrensning (art. 5.1 e), ved at personopplysningene ikke lagres lengre enn nødvendig for å oppfylle formålet

DE REGISTRERTES RETTIGHETER
Så lenge de registrerte kan identifiseres i datamaterialet vil de ha følgende rettigheter: innsyn (art. 15), retting (art. 16), sletting (art. 17), begrensning (art. 18), og dataportabilitet (art. 20).

NSD vurderer at informasjonen om behandlingen som de registrerte vil motta oppfyller lovens krav til form og innhold, jf. art. 12.1 og art. 13.

Vi minner om at hvis en registrert tar kontakt om sine rettigheter, har behandlingsansvarlig institusjon plikt til å svare innen en måned.

FØLG DIN INSTITUSJONS RETNINGSLINJER
NSD legger til grunn at behandlingen oppfyller kravene i personvernforordningen om riktighet (art. 5.1 d), integritet og konfidensialitet (art. 5.1. f) og sikkerhet (art. 32).

For å forsikre dere om at kravene oppfylles, må dere følge interne retningslinjer og/eller rådføre dere med behandlingsansvarlig institusjon.

MELD VESENTLIGE ENDRINGER
Dersom det skjer vesentlige endringer i behandlingen av personopplysninger, kan det være nødvendig å melde dette til NSD ved å oppdatere meldeskjemaet. Før du melder inn en endring, oppfordrer vi deg til å lese om hvilke type endringer det er nødvendig å melde: nsd.no/personverntjenester/fylle-ut-meldeskjema-for-personopplysninger/melde-endringer-i-meldeskjema. Du må vente på svar fra NSD før endringen gjennomføres.

OPPFØLGING AV PROSJEKTET
NSD vil følge opp ved planlagt avslutning for å avklare om behandlingen av personopplysningene er avsluttet.

- Lykke til med prosjektet!