

Evaluering – symbol og metode

Øyunn Syrstad Høydal

OSLOMET

Avhandling ph.d. sosialt arbeid og sosialpolitikk
Fakultet for samfunnsvitenskap
OsloMet – storbyuniversitetet

Vår 2019

CC-BY-SA OsloMet – storbyuniversitetet

OsloMet Avhandling 2019 nr 9

ISSN 2535-471X (trykt)

ISSN 2535-5454 (online)

ISBN 978-82-8364-157-8 (trykt)

ISBN 978-82-8364-291-9 (online)

OsloMet – storbyuniversitetet

Universitetsbiblioteket

Skriftserien

St. Olavs plass 4,

0130 Oslo,

Telefon (47) 64 84 90 00

Postadresse:

Postboks 4, St. Olavs plass

0130 Oslo

Trykket hos Byråservice

Trykket på Scandia 2000 white, 80 gram på materiesider/200 gram på coveret

Takk

Livet er fullt av tilfeldigheter. Små og store hendelser, som umerkelig drar oss i den ene eller andre retningen. Takk til dere som på ulike måter har ført meg fram til denne doktorgraden. Det har vært fine år. En spesiell takk til dere som har vist interesse for avhandlingens framdrift og skjebne. Det har betydd mye for meg.

Det sies at det skal en hel landsby til for å oppdra et barn. Det er ingenting. To universiteter, to institutter, et doktorgradsprogram, tre forskergrupper, seks direktorater, flere hundre byråkrater, to veiledere, kollegaer, venner og familie har bidratt til å gjøre denne avhandlingen til en realitet.

Tusen takk – alle sammen.

En spesiell takk til hovedveileder Sissel Hovik og biveileder Marit Haldar. Stor takk også til medforfatter Anne Grete Tøge, som bidro med nye faglige perspektiver og entusiasme på et kritisk tidspunkt. Takk til Simon Innvær, som har bidratt med kloke innspill over en lang periode og til andre medlemmer av forskergruppa *Samfunn, velferd og sosialpolitikk* (SWAP), som har delt av sin kunnskap. Takk også til forskergruppa *Profesjon, profesjonsutøvelse og styring* (PPS) og Silje Tellmann, som ga viktig og konstruktiv kritikk før siste innspurt.

Thanks to The University of Edinburgh and the research group SKAPE and Philip Cook for inviting me as a visiting student. Thanks Steve Yearley for letting me audit your truly inspiring course *Science, Knowledge and Expertise* and thank you Richard Freeman for giving me such a warm welcome and for your time and interest in my project.

Thank you Margaret for making the colonies feel like home.

I arbeidsdagen min på Institutt for sosialfag har jeg vært heldig å være omgitt av rausse, morsomme og kunnskapsrike kollegaer. En spesiell takk til Anja, Asbjørn, Carolina, Chris, Edda, Einar, Helle, Håvard, Knut, Kristian, Krisztina, Leah, Magne, Maja, Margrete, Mari, May-Britt, Siri, Steinar, Therese og Åsmund. Takk også til Steinar på NOVA for innspill og støtte underveis.

Stor takk til Endre Syrstad Høydal og Guro Syrstad Høydal, som hjalp med henholdsvis nødvendig nødhjelp til rotete referanseliste og effektiv punching av ødelagt spørreskjema. Takk til mamma og far - for alt mulig.

Til slutt:

Takk til Ingvald, som ikke er det ringeste opptatt av evaluering, men likevel har vært min klippe disse årene. Og takk til Anne og Einar. Dere er en daglig påminnelse om at livet er større enn alt dette.

Øyunn Syrstad Høydal

Oslo, mars 2019

Sammendrag

Til tross for at norsk statsforvaltning er pålagt å evaluere og har drevet utstrakt evalueringsevne i mer enn 30 år, finnes det liten kunnskap om konsekvensene av denne virksomheten. Denne avhandlingen er et bidrag til mer kunnskap om slike evalueringers konsekvenser eller innflytelse. Avhandlingen er avgrenset til evaluering i regi av direktorater innenfor det sosialpolitiske feltet. Direktoratene er viktige i evalueringssammenheng fordi de er faglige rådgivere for departementene, men også bidrar med kunnskapsoverføring til underliggende etater, media og allmenheten. I tillegg har direktoratene en sentral rolle i oppfølgingen av statlige tiltak og satsinger.

Evaluering er tradisjonelt betraktet som en metode eller et verktøy for vurdering. I denne avhandlingen kombineres dette perspektivet med et sosialkonstruktivistisk syn på evaluering. Det vil si at evaluering anerkjennes som en potensielt samfunnsnyttig metode, men også betraktes som en aktivitet med stor symbolsk betydning. Gjennom evaluering viser norske myndigheter hvilke verdier og idealer som er viktige i samfunnet. Sammen bidrar disse to perspektivene til å forklare evalueringenes rolle i forvaltningen og evalueringenes innflytelse.

Avhandlingen bygger på intervjuer med byråkrater, på en spørreundersøkelse med deltakere fra seks direktorater, og på sekundærdata om Norges flerårige deltakelse i PISA.

Avhandlingen består av tre artikler som undersøker evalueringers innflytelse fra ulike perspektiver. Artikkelen én undersøker den brede samfunnsmessige og politiske innflytelsen fra Norges mangeårige deltakelse i PISA (Programme for International Student Assessment). Konsekvensene av denne deltakelsen tolkes som et resultat av PISA som et verktøy eller en metode som bidrar med kunnskap som får klare konsekvenser for norsk politikk og samfunnsnivå. Samtidig sees PISAs innflytelse også som et resultat av norske myndigheters ønske om å framstå som endringsvillige og handlekraftige. Artikkelen to undersøker den instrumentelle eller direkte innflytelsen fra direktoratenes evalueringer på politisk-administrative beslutninger. Funnene viser at evaluering har direkte påvirkning på slike beslutninger. Dette forklares ut fra byråkraternes deltakelse i evalueringsprosessen og den praktisk nyttige kunnskapen dette resulterer i. Artikkelen tre undersøker betydningen av evalueringenes bidrag til ny kunnskap i form av nye ideer eller perspektiver, videre om hvilke typer data som bidrar til slik læring, og om dette er data i overensstemmelse med evidensbevegelsens metodologiske idealer. Funnene viser at evaluering er en viktig

kunnskapskilde for byråkratiet og at kvalitative evalueringsmetoder i størst grad bidrar til nye ideer og perspektiver.

Til sammen bidrar artiklene med ny kunnskap om hvordan byråkratene selv opplever evalueringenes innflytelse, og samtidig med innsikt i de langsiktige og komplekse virkningene av den internasjonale PISA-evalueringen. Samlet sett bidrar avhandlingen med ny kunnskap om konsekvensene av evalueringene som gjennomføres i regi av de sosialpolitiske direktoratene. Avhandlingen viser også at evalueringer utført av norske myndigheter ikke bare bør forstås ut fra deres praktiske funksjoner, men også i lys av deres symbolske betydning.

Summary

Despite the fact that Norwegian ministries and agencies have been evaluating for more than 30 years, little is known about the consequences of this activity. This dissertation is a contribution to increasing the knowledge of the impact or *influence* of these evaluations. The dissertation is limited to agencies within the field of social policy. The agencies play an important role in today's knowledge-based policy, and social policy is one of the areas where Norwegian authorities evaluate most extensively.

Evaluation may be viewed as a method for planning and developing social improvements. Evaluation could also be seen as a societal phenomenon, as an omnipresent, powerful and inescapable trend. These two perspectives represent different academic traditions. In this dissertation, evaluation is both acknowledged as a potentially useful method and a societal phenomenon and symbol of the values of society. Together, these two perspectives help explain the role of evaluations in Norwegian public administration and their influence.

The dissertation consists of three articles that examine the influence of evaluations from different perspectives. Article one examines the broad societal and political influence of Norway's long-term participation in the International Student Assessment (PISA) programme. The massive influence of the PISA evaluation is explained as a consequence of the evaluation being both a method – with results affecting Norwegian policy – and a result of the symbolic role of PISA, which reflects the values of society. Article 2 examines the instrumental or direct influence from evaluations in political-administrative decisions. In Article 3, the relationship between the ideal EBP methodological approaches and policymakers' perception of the impact of evaluations is investigated. The aim is twofold. First, to assess policymakers' belief in evaluations as a source of new ideas and policies. Second, to examine whether this belief depends on methodological ideals from the evidence hierarchy. The civil servants report strong belief in evaluations' impact on policy and development of new ideas – however, this is more or less independent of the type of methodological approach used. Exceptions are workshops, conferences and similar events, evaluation using observational data, and evaluations conducted with the intention of obtaining new knowledge.

The articles are based on interviews with Norwegian civil servants and a survey of participants from six agencies within the field of social policy.

Overall the dissertation reveals that the evaluative activity in Norwegian agencies does influence both policy and society. The study contributes new knowledge about how the bureaucrats themselves experience the influence of evaluations and reveals how a double perspective on evaluations could lead to a better understanding of the influence of the international PISA evaluation and evaluation in general.

Innhold

1. Innledning.....	4
2. Evaluering i statlig kontekst.....	8
2.1 Kunnskapsbasert politikkutforming.....	8
2.2 Evaluering i norsk statsforvaltning.....	9
2.3 Direktoratenes rolle i den kunnskapsbaserte politikkutforming.....	11
2.4 Evaluering i direktoratene.....	11
3. To perspektiver på evaluering.....	13
3.1 Modernistiske tilnærminger til evaluering.....	13
3.2 Sosialkonstruktivistiske perspektiver på evaluering.....	15
3.3 Evaluerings innflytelse og et dobbelt perspektiv.....	16
Denne avhandlingen tar imidlertid utgangspunkt i <i>evalueringsinnflytelse</i> (Kirkhart 2000).	17
4 Forskningsfeltet.....	19
4.1 Innflytelsen fra vitenskapelig basert kunnskap.....	19
4.2 Forklaringer på bruk av vitenskapelig basert kunnskap.....	21
4.2.1 Hvilke faktorer fører til bruk av vitenskapelig basert kunnskap?.....	21
4.2.2 Kontekstuelle forklaringer.....	23
4.3 Oppsummert.....	23
Denne gjennomgangen.....	23
5. Teoretisk tilnærming.....	24
5.1. Pragmatisk institusjonalisme.....	24
5.2. Et dobbelt blikk på evalueringsbølgen.....	27
5.3. Et pragmatisk perspektiv på norsk evalueringstradisjon.....	29
5.4. Evaluering i dagens forvaltning.....	31
6. Metode, data og egen rolle.....	35

6.1 En spørreundersøkelse blir til.....	36
6.1.1 Utgangspunkt og metodevalg.....	36
6.1.2 Testing, utvalg og deltakere	37
6.2 Spørreundersøkelsens struktur og operasjonalisering	41
6.2.1 Struktur.....	41
6.2.2 Operasjonalisering av instrumentell bruk.....	43
6.3 Metode og data – utfordringer og begrensninger	45
6.3.1 Data basert på spørreundersøkelsen	45
6.3.2 Intervjudata og andre data basert på kvalitative tilnærminger	47
6.3.3 Sekundærdata	48
7. Forskningsetiske betraktninger	50
8. Oppsummering av artiklene i avhandlingen.....	52
9. Avsluttende drøfting.....	55
9.1. Innflytelse gjennom instrumentell bruk	57
9.2. Innflytelse gjennom opplysende bruk	58
9.3. Innflytelse – i bred forstand.....	59
9.4. Oppsummert.....	60
10. Avhandlingens bidrag	62
11. Framtidig forskning.....	63
Referanser.....	64

Vedlegg:

Vedlegg 1: Godkjenning av NSD

Vedlegg 2: Forespørsel om deltakelse i undersøkelse om evaluering

Vedlegg 3: E-postinvitasjon

Vedlegg 4: Påminnelse om undersøkelse

Vedlegg 5: Spørreundersøkelse

Artikler:

Artikkel 1: Evaluation influence – method and societal phenomenon
Forfatter: Øyunn Syrstad Høydal
Språk: Engelsk
Status: Publisert i *Evaluation*

Artikkel 2: Tracking the Instrumental Use and Motives for Evaluations
Forfatter: Øyunn Syrstad Høydal
Språk: Engelsk
Status: Under vurdering i *American Journal of Evaluation*

Artikkel 3: Evaluating social policies – do methodological approaches determine the policy impact?
Forfatter: Øyunn Syrstad Høydal og Anne Grete Tøge
Språk: Engelsk
Status: Innsendt til *Evidence & Policy*

1. Innledning

Evaluering har vært en vanlig aktivitet i norsk statsforvaltning i cirka 30 år, og evalueringene spiller en stadig større rolle i den pågående kunnskapsbaserte politikktutforming. Evaluering omtales både som «en viktig kilde til læring og utvikling» og «virkemidler for å få et mer langsiktig perspektiv i styringen» (Direktoratet for økonomistyring 2011:2). Statlige etater er pålagt å evaluere gjennom både Regelverk for økonomistyring i staten og utredningsinstruksen¹.

Det samme samfunnet som i dag tilsynelatende har stor tro på kunnskapsproduksjon via evaluering, vet imidlertid lite om konsekvensene av den evalueringsvirksomheten som faktisk finner sted (Askim, Døving og Johnsen 2013, Halvorsen og Lier Madsen 2013, Sverdrup 2014). Derfor er det blitt etterlyst mer forskning og en systematisk kunnskapsbygging på dette området (Halvorsen, Lier Madsen og Jentoft 2013:8). Internasjonalt har det også blitt etterlyst flere studier av de faktiske konsekvensene av den utstrakte evalueringen av offentlig virksomhet (Coryn med flere 2017, Henry og Mark 2003, Pattyn med flere 2018, Saunders 2012). Formålet med denne avhandlingen er å bidra med ny kunnskap om evalueringers betydning i en norsk kontekst.

Avhandlingens tema er evaluering i regi av norske direktorater innenfor det sosialpolitiske feltet. Vidt definert omfatter sosialpolitikken utdanning, helse, oppvekst, innvandring, bomiljø, integrering, sysselsetting og sosiale ytelser. Sosialpolitikken legger grunnleggende premisser for befolkningens livskvalitet og utgjør dermed selve bærebjelken i den norske velferdsstaten. Dette feltet er også blant de politiske områdene statsforvaltningen evaluerer i størst utstrekning (Askim, Døving og Johnsen 2011, Tornes 2012). Dermed er det sosialpolitiske feltet interessant i evalueringssammenheng både i kraft av sin store samfunnsbetydning og på grunn av evalueringsomfanget. Direktoratene er særlig viktige i denne sammenhengen fordi de har en sentral rolle i dagens kunnskapsbaserte politikktutforming både som bestillere og potensielle brukere av evalueringer.

¹ Regelverk for økonomistyring i staten:

<https://www.regjeringen.no/no/tema/okonomi-og-budsjett/statlig-okonomistyring/reglement-for-okonomistyring-i-staten/id438887/>

Utredningsinstruksen:

<https://www.regjeringen.no/no/dokumenter/instruks-om-utredning-av-statlige-tiltak-utredningsinstruksen/id2476518/>

Dette utgangspunktet har resultert i følgende overordnede problemstilling for avhandlingen:

Hvilken innflytelse har evalueringene i regi av norske direktorater på politikk og samfunnsliv, og hvordan kan denne innflytelsen forstås som symbol og metode?

Innflytelse er i denne sammenhengen forstått i tråd med Kirkharts (2000) forsøk på å samle alle former for potensielle virkninger av evaluering i ett og samme begrep. Problemstillingen² er avgrenset til direktorater innenfor det sosialpolitiske feltet – i denne avhandlingen gjelder det Helsedirektoratet, Utlendingsdirektoratet (UDI), Integrerings- og mangfoldsdirektoratet (IMDi), Husbanken, Barne-, ungdoms- og familiedirektoratet (Bufdir) og Utdanningsdirektoratet (Udir).

For å besvare den overordnede problemstillingen er det nødvendig å klargjøre hvordan jeg forstår evaluering, og dermed også hvilket utgangspunkt jeg legger til grunn for denne avhandlingen. I årenes løp er det gjort utallige forsøk på å definere evaluering. I norsk sammenheng er den svenske evalueringsforskeren Evert Vedungs definisjon en av de mest kjente: «Noggrann efterhandsbedömning av utfall, slutprestationer eller förvaltning i offentlig verksamhet, vilken avses spela en roll i praktiska beslutsituationer» (1988:20). I denne definisjonen avgrenses evaluering til offentlig sektor. I tillegg ekskluderes både forhåndsanalyser og prosess- eller underveisevalueringer. En mye brukt og mer inkluderende tilnærming er Schrivens beskrivelse av evaluering som «the process of determining the merit, worth, or value of something, or the product of that process» (1991:139). Felles for disse to definisjonene er forståelsen av evaluering som en form for metode eller et verktøy som kan brukes når noe skal vurderes. Denne forståelsen har vært det dominerende perspektivet på evaluering siden den første internasjonale evalueringsbølgen oppstod i etterkrigstidens USA.

Evaluering kan imidlertid også betraktes som et kultur- eller samfunnsfenomen. Et ritual som ikke først og fremst gjennomføres for å skape bedre politiske løsninger, men som handler om å speile samfunnets overordnede normer og verdier (Dahler-Larsen 2012, Power 1997, Schwandt 2009). Gjennom evaluering viser samfunnet hva som er viktig og verdifullt. Slike sosiologiske eller sosialkonstruktivistiske perspektiver på hva evaluering er og betyr har vært langt mindre utbredt enn de modernistiske tilnærmingene. Det er imidlertid viktig å understreke at jeg her tegner opp evalueringskartet med relativ bred pensel. Selv om det

² Problemstillingen omhandler kun evalueringer utført av eksterne utøvere og omfatter ikke forvaltningsrevisjon.

finnes en skillelinje mellom sosialkonstruktivistiske og modernistiske tilnærminger både i evalueringsforskningen og annen samfunnsforskning (Røvik 2007), finnes det også mellomposisjoner. Guba og Lincolns (1989) forsøker gjennom *Fourth generation evaluation* å utvikle en ny og bedre evalueringsmetodikk basert på et konstruktivistisk kunnskapssyn.

Om forskere definerer evaluering som en potensielt nyttig metode eller som et kulturelt fenomen, har betydning for hvilke aspekter ved samfunnets evalueringsaktivitet som undersøkes, og hvordan en tolker konsekvensene av denne aktiviteten. Innenfor tradisjonell evalueringsforskning har ikke noe tema fått mer oppmerksomhet enn hvordan evalueringresultatene brukes og kommer til nytte (Alkin og King 2017, Højlund 2014, Patton 2008). Klare resultater har imidlertid latt vente på seg, og følge Højlund (2014) har dette resultert i en omfattende interesse for *hvilke faktorer* som bidrar til bruk av evalueringresultater (se for eksempel Cousin og Leithwood 1986, Johnson mfl. 2009). Evalueringsforskningen dreier seg også om hvordan riktige evalueringmetoder kan bidra til bedre evalueringer, tydeligere gevinster av evalueringene og dermed flere evalueringer (Henry og Mark 2003, Kirkhart 2000, Mark og Henry 2004, Pawson 2013). Forskningen som betrakter evaluering som et kulturelt fenomen, har derimot fokusert på hvilke verdier og ideer evaluering gjenspeiler, og hvilken rolle evaluering fyller i samfunnet, ut over å være et rent vurderingsverktøy (Dahler-Larsen 2012, Power 1997, Schwandt 2009).

Denne avhandlingen bygger på en tosidig tilnærming til evaluering. Jeg forsøker å kombinere en modernistisk og en sosiologisk eller sosialkonstruktivistisk tradisjon i en form for *pragmatisk institusjonalisme* (Røvik 1997). Denne tilnærmingen innebærer enkelt sagt at evaluering både anerkjennes som metode eller verktøy og som et sosialt eller kulturelt fenomen. Perspektivet er egnet både for å finne faktiske konsekvenser av evaluering, for å undersøke hvilke ideer og verdier som preger direktoratenes evalueringsaktivitet, og for å studere hva evalueringens symbolske rolle innebærer.

I denne avhandlingen undersøker jeg om evalueringene som gjennomføres av sosialpolitiske direktorater har innflytelse på norsk politikk og samfunnsniv. Dette er et stort spørsmål å besvare, og min løsning har vært å undersøke dette spørsmålet fra tre ulike perspektiver: innflytelse i instrumentell forstand, innflytelse i form av opplysning (Weiss 1977) og innflytelse i vid betydning (Kirkhart 2000). For å fange bredden i den potensielle innflytelsen av direktoratenes evaluering, bygger denne avhandlingen både på primær- og sekundærdata. Primærdataene består av resultatene av en spørreundersøkelse om evaluering blant ansatte i de

sosialpolitiske direktoratene, samt intervjuer med et utvalg byråkrater. Sekundærdataene omfatter dokumenter og tidligere forskning knyttet til Norges flerårige deltakelse i PISA (Programme for International Student Assessment). Til sammen kan dette datagrunnlaget 1) bidra til innsikt på aktørnivå om byråkratenes opplevelse av evalueringenes umiddelbare innflytelse, 2) bidra med kunnskap om evalueringers innflytelse på ulike samfunnsnivåer over tid.

I artikkel én: «Evaluation – Method and societal phenomenon» undersøkes innflytelsen fra PISA på norsk skolepolitikk og samfunnsliv i bred forstand. Her viser jeg også hvordan PISA både kan forstås som metode og symbol, og hvordan dette doble perspektivet kan bidra til en bedre forståelse av evaluering og evalueringers innflytelse. Dette perspektivet er mer implisitt til stede i de to andre artiklene, som begge tar utgangspunkt i begreper som gjenspeiler evalueringsforskningens modernistiske utgangspunkt.

I artikkel to undersøker jeg om byråkratene mener at evalueringer får *instrumentell* innflytelse på beslutninger i det politisk-administrative systemet og dermed også for norsk politikktutforming. Jeg tar utgangspunkt i evaluering som et administrativt verktøy som primært tas i bruk for å skaffe til veie praktisk nyttig kunnskap for byråkratiet (Dahler-Larsen 2000). Basert på denne forståelsen utvikler jeg en modell som beskriver forholdet mellom motiver for evaluering, evalueringsprosess, resultater og form for bruk. I motsetning til evalueringsforskning som gjerne tar utgangspunkt i et skille mellom utførere og brukere av evalueringsresultater, betrakter jeg evalueringsprosessen som en *samproduksjon* («co-production», Jasanoff 2004) av kunnskap – som en prosess der byråkratene er aktive bidragsytere og dermed sikrer at evalueringene resulterer i praktisk nyttig kunnskap som blir brukt.

Utgangspunktet for artikkel tre er evalueringenes bidrag til innflytelse gjennom *opplysning* (Weiss 1977) i direktoratene. Sammen med medforfatter Anne Grete Tøge, undersøker jeg betydningen av evalueringenes bidrag til ny kunnskap i form av nye ideer eller perspektiver. Vi ser på hvilke typer data som bidrar til slik læring, og om dette er data i overensstemmelse med evidensbevegelsens metodologiske idealer.

2. Evaluering i statlig kontekst

For å forstå betydningen av direktoratenes evalueringsvirksomhet er det nødvendig med innsikt i kunnskapsregimet denne virksomheten er en del av. Det vil si de strukturelle og institusjonelle rammene som preger denne kunnskapsproduksjonen (Campbell og Pedersen 2014, Christensen, Gornitzka og Holst 2017). I dette kapitlet vil jeg derfor redegjøre for evaluering i statlig forvaltning, for direktoratenes bruk av evaluering og for deres rolle i det politiske systemet. Framstillingen tar utgangspunkt i Norge som et kunnskapssamfunn preget av det internasjonale idealet om kunnskapsbasert politikkutforming.

2.1 Kunnskapsbasert politikkutforming

Norge omtales som et kunnskapssamfunn, eller et kunnskapsdemokrati (Christensen, Gornitzka og Holst 2017). Dette er samfunn der teoretisk kunnskap har betydning og autoritet i alle deler av samfunnet, og der kunnskap inngår som en helt sentral del av politisk maktkamp og styring (Stehr 2015). Norsk statsforvaltning har i hele perioden etter andre verdenskrig vært opptatt av å basere politiske beslutninger på vitenskapelig kunnskap (Christensen, Gornitzka og Holst 2017, Tranøy mfl. 2009). Dette idealet har tilsynelatende også fått ny vind i seilene de siste tiårene gjennom framveksten av EBP («evidence based policy»), eller kunnskapsbasert politikkutforming, som det kalles på norsk (Fretheim 2013).

I Europa knyttet evidensbevegelsen gjerne til Tony Blair og New Labours maktovertakelse i Storbritannia på slutten av 1990-tallet (Kisby 2011 Smith 2013). Det reformerte Labourpartiet stod fram som et alternativ til fortidas verdi- og tradisjonsbaserte politiske løsninger. Den nye politikken skulle baseres på vitenskap, og samfunnsvitere skulle bidra ved å informere politikerne om «...what works and why and what types of policy initiatives are likely to be most effective» (Blunkett 2000 i Kisby 2011:109). Tradisjonelle samfunnsvitenskapelige metoder som intervjuer, observasjoner og case-studier ble imidlertid ikke ansett som gode nok. Den nye kunnskapen skulle baseres på sikrere og mer objektive medisinske eller naturvitenskapelige idealer. Og øverst i det såkalte evidenshierarkiet stod eksperimenter (Randomised Controlled Trials, RCTs) og kunnskapsoversikter (Hammersley 2013).

Evidensbevegelsen fikk raskt stor internasjonal utbredelse, blant annet gjennom organisasjoner som EU, OECD og Verdens helseorganisasjon (WHO). Dette medførte også en fornyet interesse for og bruk av evaluering i europeisk forvaltning (Pattyn mfl. 2018).

Evidensbevegelsen har senere blitt gjenstand for betydelig kritikk og de opprinnelige metodologiske idealene er i noen grad moderert (Hammersley 2013, Smith 2013, Smith og Haux 2017). Verdens helseorganisasjon snakker ikke lenger om *evidence based policy*, men om *evidence-informed policy*,³ og i Storbritannia har den største kunnskapsoptimismen fra Blairs 90-tall bleknet. Til tross for dette og nye fenomener som *post truth* er det en økende internasjonal interesse for å bruke vitenskapelig basert kunnskap som grunnlag for politiske beslutninger (Smith og Haux 2017).

Kunnskapsbasert politikk- og fagutvikling har de siste årene fått stadig større betydning i norsk statsforvaltning, og evaluering er del av denne trenden (DFØ 2016:30). Ifølge Tornes (2012) innebærer dette at forvaltningen delvis er avhengig av et kontinuerlig tilfang av ny kunnskap for å utføre egne oppgaver, samtidig som den i stor grad må organisere og tilrettelegge for produksjon av denne kunnskapen. Dermed kan en anta at idealet om kunnskapsbasert politikkutforming får tydelige praktiske konsekvenser for norske byråkrater. En slik antakelse underbygges av at stadig flere byråkrater mener jobben deres ligner mest på forskerrollen (Christensen nfl. 2014).

2.2 Evaluering i norsk statsforvaltning

Regelverket for økonomistyring i staten pålegger all statlig virksomhet å sikre informasjon om effektivitet, måloppnåelse og resultater gjennom evaluering (paragraf 16). Det heter imidlertid også at «Frekvens og omfang av evalueringene skal bestemmes ut fra virksomhetens egenart, risiko og vesentlighet» (paragraf 16). I praksis er det derfor en betydelig variasjon mellom de ulike sektorenes og enhetenes evalueringsevne. Utdannings-, sosial- og velferdsfeltet er blant områdene som i særlig grad har vært gjenstand for evaluering, mens samferdsel og forsvar er politikkområder som i liten grad evalueres (Askim, Døving og Johnsen 2013, Baklien 1985). Disse forskjellene kan forklares ut fra myndighetenes behov for kontroll med desentraliserte policy-områder. Ulikhetene kan også handle om at politiske områder preget av stor uenighet og usikkerhet evalueres mer enn andre politiske felt. Ifølge Bogenschneider og Corbett (2010) er det sosialpolitiske feltet spesielt utsatt for debatt og uenighet både fordi det er preget av verdiladete fenomener som familie, utdanning og helse, og fordi det rommer en høy andel såkalte *wicked problems*. Det vil si normative problemstillinger der utfordringene i seg selv er problematiske å definere,

³ <http://www.who.int/evidence/en/>

omskiftelige og vanskelige å løse. En annen mulig årsak til evalueringsaktiviteten på det sosialpolitiske området er at dette politiske feltet er preget av mange tidsavgrensede tiltak og programmer som myndighetene har et behov for å følge opp å kontrollere (Askim, Døving og Johnsen 2013:86). Kritikere har imidlertid hevdet at evalueringens viktigste funksjon innenfor deler av statlig virksomhet er legitimering av store overføringer, og at evaluering bare har blitt et motebegrep for å rettfærdiggjøre bruk av offentlige midler (Baklien 1993a:263).

Direktoratet for økonomistyring (DFØ), som har overordnet fagansvar for statlig evalueringsvirksomhet, definerer evaluering slik: «Evaluering er en systematisk datainnsamling, analyse og vurdering av en planlagt, pågående eller avsluttet aktivitet, en virksomhet, et virkemiddel eller en sektor.» (DFØ 2011:5. Selv om DFØ på denne måten har definert evaluering i statlig sammenheng, innebærer ikke dette nødvendigvis at norske direktorater forholder seg til denne definisjonen eller deler DFØs oppfatning av hva evaluering innebærer. DFØs egen rapport (2016) viser tvert imot at det hersker en viss usikkerhet eller variasjon i de ulike forvaltningsenhetenes definisjon eller forståelse av evaluering (s. 7).

Det finnes også ett klart særtrekk ved den statlige evalueringsaktiviteten som ikke kommer fram i DFØs definisjon, nemlig den sterke forskningsmessige forankringen som preger norsk evalueringstradisjon. Den danske evalueringsforskeren og sosiologen Peter Dahler-Larsen (2013) hevder at norsk evalueringstradisjon skiller seg ut i internasjonal sammenheng fordi den både teoretisk og metodisk har klare fellestrekk med forskning. I årenes løp har det likevel vokst fram en underskog av evalueringer som har mindre til felles med forskning. Halvorsen (2013) argumenterer derfor for et tydeligere skille mellom *forskningsbaserte evalueringer og andre evalueringer*. Men hva betyr det egentlig at noe er forskningsbasert? Må evalueringen være utført av en forsker? Ja, svarer Halvorsen, og hevder at slike evalueringer må utføres av evaluatorene med en *vitenskapelig innstilling* (s. 241).

Resultatene fra min egen spørreundersøkelse viser at evalueringene innenfor det sosialpolitiske feltet fremdeles primært utføres av forskere. Forskningsinstitutter var involvert i 42 prosent av evalueringene, forskere fra universiteter og høyskoler i 14 prosent og konsulenter i 40 prosent. Dette resultatet står i motsetning til Askim, Døving og Johnsen (2013), som hevder at flertallet av de statlige evalueringene utføres av konsulenter (s. 82). Disse resultatene er imidlertid basert på opplysninger fra hele statsforvaltningen. De ulike resultatene kan derfor skyldes at sektorene foretrekker forskjellige evalueringsmiljøer.

Forskjellene kan også skyldes at Askim, Døving og Johnsen har hentet sine tall fra evalueringsportalen⁴. Portalen er basert på forvaltningens frivillige registrering av evalueringer, og det er usikkert om denne ordningen noen gang har fungert. Selv byråkrater som jobber spesielt med evaluering, har liten kunnskap om portalen (DFØ 2016:15).

2.3 Direktoratenes rolle i den kunnskapsbaserte politikktutformingen

Direktoratene har to hovedroller. Den *iverksettende rollen* handler om direktoratenes myndighetsutøvelse, det vil si forvaltning av regelverk og tilskudd og gjennomføring av prosjekter og handlingsplaner. Direktoratene har også en *faglig rolle*, og den er todelt. For det første fungerer direktoratene som faglige rådgivere for departementene. De bidrar med faktaopplysninger, rapporter og medforfatterskap på stortingsmeldinger. For det andre formidler direktoratene kunnskap innenfor sine respektive fagfelt til medier og allmennhet (Botheim og Solumsmoen 2009). Kunnskap fra evaluering er relevant for alle disse oppgavene. Som faglige rådgivere er direktoratene avhengige av relevant og oppdatert kunnskap for å gjøre jobben sin, men evaluering er også et sentralt verktøy i forbindelse med forvaltning av regelverk, tilskudd og ulike statlige prosjekter.

Direktoratene er relativt autonome enheter, og relasjonen mellom departement og direktorat innebærer en balansegang mellom direktoratenes faglige uavhengighet og departementenes ønske om styring. Det har vært debattert om direktoratene burde gjøres til rent iverksettende organer, mens uavhengige fagmiljøer burde overta oppgaven som faglige rådgivere (Botheim og Solumsmoen 2009). Likevel har direktoratenes rolle i planlegging og politikktutvikling i praksis blitt styrket de siste tiårene (Christensen, Egeberg, Lægreid mfl. 2014), og direktoratenes faglige innspill har til dels stor innflytelse på politikktutformingen i departementene (Botheim og Solumsmoen 2009, Christensen, Egeberg, Lægreid mfl., 2014). Med sin dialog både oppover til departementsnivået, utover til allmennheten og media, samt nedover i forvaltningen til underetater og regionale nivåer, er direktoratene et nav i den kunnskapsbaserte politikktutformingen.

2.4 Evaluering i direktoratene

Direktoratenes evalueringsvirksomhet er allsidig. Den spenner fra mindre, egeninitierte evalueringer av tiltak og programmer til flerårige internasjonale evalueringer, som

⁴ <https://evalueringsportalen.no/>

Utdanningsdirektoratets administrasjon av den norske PISA-deltakelsen. De fleste eksterne evalueringene i statlig sektor varer et sted mellom seks måneder og ett og et halvt år. Få evalueringer strekker seg over mer enn to år, og prismessig ligger hovedtyngden mellom en halv og en og en halv million kroner (DFØ 2016:9). Dermed er PISA, som til sammen har pågått i over 15 år og kostet penger både i form av deltakeravgift, administrasjon og faglig oppfølging i alle disse årene, et atypisk eksempel på direktoratenes evalueringsvirksomhet. Samtidig har slike internasjonale evalueringer blitt mer utbredt de siste tiårene (Elstad 2010). Norge deltar eller har deltatt i flere mindre og lignende evalueringer innenfor utdanningssektoren som TIMSS (Trends in International Mathematics and Science Study) og PIRLS (Progress in International Reading Literacy Study).

Evalueringsaktiviteten i direktoratene er et resultat av tildelingsbrev, tilskuddsreglement, instruks, forsøksvirksomhet og strategier/planer, mens evalueringene som iverksettes av departementene, er mer ad hoc-preget og styrt av politiske hendelser (DFØ 2016).

Direktoratenes rolle i evalueringssammenheng kommer tydelig fram i resultatene fra undersøkelsen jeg gjennomførte i forbindelse med denne avhandlingen. Direktoratene er blant initiativtakerne til mer enn 70 prosent av evalueringene de er involvert i. Eget departement er involvert i drøye 30 prosent, mens over 30 prosent av evalueringene er et resultat av flere initiativtakere. Direktoratene betrakter også forvaltningen som de viktigste brukerne av evalueringsresultatene. Hele 99 prosent av de aktuelle evalueringene skulle brukes av forvaltningen, 54 prosent av evalueringene skulle brukes av politisk ledelse, mens praksisfeltet var aktuelle brukere av 71 prosent av evalueringene. Det faktum at evaluering synes å være en integrert del av direktoratenes virksomhet, at direktoratene tar initiativ til de fleste evalueringene og ser på forvaltningen som den viktigste brukeren av evalueringsresultatene, viser direktoratenes betydning i evalueringssammenheng.

Når direktoratene både er ansvarlige for ulike tiltak, initierer evaluering av de samme tiltakene og ser på seg selv som viktige brukere av evalueringsresultatene, bør en kanskje i større grad diskutere hvordan alle disse rollene preger forvaltningens evalueringer.

3. To perspektiver på evaluering

Skillet mellom evaluering forstått som metode og evaluering forstått som sosialt fenomen gjenspeiler hva Røvik (2007) beskriver som et fundamentalt skille i samfunnsvitenskap og organisasjonsvitenskap, nemlig skillet mellom modernistiske og sosialkonstruktivistiske tilnærminger. Den førstnevnte retningen er preget av utviklingsoptimisme, troen på organisering og organisasjon, samt vitenskaps- og kunnskapsoptimisme.

Sosialkonstruktivistiske tilnærminger er først og fremst kjennetegnet ved *ideen om den sosialt konstruerte virkeligheten*, men de er også skeptiske til modernismens tro på organisering og organisasjon, samt kjennetegnet av hva Røvik omtaler som *redskapsskepsis* (s.48). Det vil si at de sosialkonstruktivistiske retningene er skeptiske til den modernistiske ideen om at organisasjoner utelukkende er redskaper for å nå mål.

Evaluering er i seg selv et resultat av modernismens tro på rasjonalitet og framskritt gjennom vitenskapelig basert kunnskap (Albæk 1988, Dahler-Larsen 2012). Den første internasjonale evalueringebølgen oppstod i USA på 1960-tallet. Vedung omtaler dette som «*the science driven wave*» (2010:265). Dette var velferdsstatens oppblomstringstid, både i USA og Europa, og den første evalueringebølgen var tett forbundet med gjennomføringen av etterkrigstidens sosiale reformer (Krogstrup 2011). Enorme budsjetter ga behov for kunnskap om effekt, og Lyndon B. Johnsons satsing på fattigdomsbekjempelse «The War on Poverty», markerte starten på de store statsfinansierte evalueringene (Weiss 1998:12). Gjennom omfattende samfunnsplanlegging, eksperimenter, vitenskapelige analyser og evaluering skulle samfunnet styres framover (Vedung 2010). Dette synet på evaluering har siden satt sitt preg både på samfunnets og evalueringsforskningens forståelse av evaluering.

3.1 Modernistiske tilnærminger til evaluering

Modernistiske tilnærminger *tror* på evaluering, og denne troen kommer tydelig til uttrykk i evalueringsforskningens interesse for det politisk-administrative systemets bruk av evaluering. Deler av det amerikanske evalueringsmiljøet uttrykker også fremdeles sterke forhåpninger på evalueringens vegne, der *Social betterment* i form av en ny og bedre stat betraktes som selve målet med evaluering (Henry and Mark, 2003; Mark and Henry, 2004, 2013, Mathison 2018).

Ifølge Cuba og Lincoln (1989) har evalueringsforskningen vært dominert av et positivistisk vitenskapssyn. Dette innebærer ideen om en objektiv virkelighet som forskere eller

evaluatorer kan studere nøytralt og at data fra kvantitative studier er blitt ansett som mer objektive og troverdige enn data fra kvalitativt orienterte studier. Dette korresponderer med ideen om at evaluering skal bidra med objektive og nøytrale fakta. I tillegg innebærer dette også at betydningen av kontekst kan tones ned, noe som igjen gjør det lettere å generalisere evalueringresultatene fra et program til tilsvarende programmer andre steder (Guba og Lincoln 1989).

Modernismens tro på organisasjon og organisering gjenspeiles i evalueringens tilnærming til politiske beslutningsprosesser. Slike prosesser betraktes gjerne som en kjede eller sirkel av sammenhengende, rasjonelle stadier (se Albæk 1998, Jann og Wegrich 2007, Lasswell 1956, Nutley 2007). Modellene tar utgangspunkt i et politisk definert problem, i neste stadium tas det en beslutning i forhold til videre handling, deretter settes denne politikken ut i livet, før den så eventuelt evalueres. Her er vitenskapelig basert kunnskap spesielt relevant når politikerne skal velge mellom ulike handlingsalternativer, eller når den valgte løsningen skal evalueres. Vitenskapelig basert kunnskap kan imidlertid også være med på å definere det opprinnelige problemet eller spille en rolle i forbindelse med implementeringen av den valgte politiske løsningen (Nutley 2007:93).

Etter hvert er har rasjonelle modellene blitt erstattet eller supplert av teorier der samfunnsstyring anerkjennes som komplekse og uoversiktlige prosesser preget av tilfeldigheter, konflikter og koalisjoner (Cohen, March og Olsen 1972, Kingdon 1995, Sabatier og Jenkins-Smith 1999). Likevel er samfunnets syn på politisk-administrative beslutningsprosesser fremdeles dominert av ideen om organisasjoner som målretta og rasjonelle (Albæk 1988, Boswell 2009, Stone 2012). Denne oppfatningen medfører også klare forventninger knyttet til kunnskapens rolle – at kunnskap skal være et direkte bidrag til bedre politiske beslutninger (Albæk 1988, Parkhurst og Abeyasinghe 2016). Den rasjonelle modellen bygger også på en forestilling om at vitenskapelig basert kunnskap er et grunnleggende gode for politisk-administrative beslutningsprosesser. Dette står i kontrast til forskning innenfor det såkalte ekspertisefeltet, der en nettopp problematiserer ekspertenes og teknokratiets makt og påvirkning og frykter at slike krefter svekker de demokratiske beslutningskanalene og vanlige borgeres mulighet til å påvirke samfunnsutformingen (se for eksempel Holst og Molander 2017).

I tråd med det modernistiske kunnskapssynet hviler forventningene til kunnskapens direkte nytte også på en klar arbeidsdeling mellom politikere og byråkrater på den ene siden og

forskere eller konsulenter på den andre (Guba og Lincoln 1989, Pawson og Tilley 1997). Politikere og byråkrater diskuterer verdier og fastsetter mål, mens kunnskapsprodusentene finner svar på samfunnets utfordringer i form av objektive fakta. Vitenskapens innflytelse på politisk-administrative prosesser krenker dermed hverken demokratiske grunnprinsipper knyttet til ideen om folkestyre (i motsetning til ekspertvelde) eller vitenskapens egne idealer om uavhengig og verdifri forskning (Albæk 1998:31).

3.2 Sosialkonstruktivistiske perspektiver på evaluering

Sosialkonstruktivistiske perspektiver deler ikke den modernistiske troen på evaluering som et middel mot et bedre samfunn. Slike tilnærminger er heller ikke opptatt av evaluering som et samfunnsnyttig verktøy, og er derfor heller ikke interessert i å undersøke evaluering fra dette perspektivet. Disse perspektivene er i stedet interessert i evaluering som samfunnsfenomen, som en trend med internasjonal utbredelse og et uttrykk for makt og verdier.

Sosiologiske eller sosialkonstruktivistiske perspektiver betrakter kunnskap som et resultat av sosiale prosesser og den kulturen kunnskapen inngår som en del av (Berger og Luckmann 1966). Nettopp fordi hele fundamentet bak disse tilnærmingene er ideen om at kunnskap konstrueres sosialt innenfor en gitt kulturell og historisk kontekst, er sosialkonstruktivistiske tilnærminger til evaluering opptatt av å forstå evaluering i lys av omgivelsene.

Sosialkonstruktivistiske eller sosiologiske perspektiver betrakter evaluering og lignende fenomener – som revisjon og økonomiske analyser – som symboler på samfunnets verdier. I *The Audit Society* (1997) beskriver Power hvordan revisjon har spredd seg fra økonomi til en rekke andre samfunnsområder, og han diskuterer hva dette betyr, og hvorfor det har skjedd. Power mener denne utviklingen skyldes politikernes behov for kontroll. Revisjon har blitt et ritual som gjennomføres for å skape en illusjon av trygghet og kontroll. Konsekvensene av denne virksomheten er at stadig flere deler av samfunnet må tilpasse seg for å bli gjenstand for revisjon. Dermed endres organisasjoner i tråd med verdiene i revisjonssystemene.

Schwandts (2009) begrep *evaluation imaginary* viser til samfunnets ideer om hvorfor det skal evaluere, og hva evaluering kan bidra med. Schwandt mener dagens evalueringsbølge bygger på et tosidig moralsk-politisk utgangspunkt. På den ene siden betraktes evaluering som noe myndigheter nærmest er moralsk forpliktet til å gjøre. På den andre siden forventes det at evaluering danner grunnlaget for politiske beslutninger og styring. Evaluering er et internasjonalt fenomen, som blant annet er spredd gjennom organisasjoner som EU og OECD.

Schwandt mener nasjonale ideer om hva evaluering er eller skal bidra til, står i et spenn mellom internasjonale organisasjoners ideer om evaluering og nasjonale verdier og interesser.

Dahler-Larsen bygger videre på Powers og Schwandts arbeider i boken *The evaluation society* (2012). Utgangspunktet er dagens evalueringssamfunn: et samfunn der evaluering gjennomgår alle aspekter ved livet, hvor evaluering tas for gitt og er vanskelig både å kritisere og å unnsnippe. Dahler-Larsen er opptatt av å finne verdiene og ideene bak evalueringens bølge. Derfor undersøker han formålet og meningen med evaluering i lys av hva han betegner som tre sosiohistoriske stadier av modernitet: det moderne, den reflekterende modernitet og revisjonssamfunnet (*the audit society*)⁵. Dahler-Larsen viser dermed hvordan evaluering har hatt ulik betydning gjennom de forskjellige periodene av moderniteten.

Modernistiske tilnærminger til evaluering bygger på et tydelig skille mellom evaluator og evalueringsobjekt, politikk og kunnskapsproduksjon. Det finnes imidlertid perspektiver som tar et helt annet utgangspunkt og i stedet hevder det skjer en gjensidig påvirkning mellom politikk og forskning (Hoppe 2010, Jasanoff 2004). Ifølge Jasanoff (2004:3) foregår det en *co-production* (samproduksjon) av kunnskap og politikk. Det vil si at disse to systemene produserer kunnskap sammen, og at politikken er en del av kunnskapsproduksjonen, på samme måte som kunnskapen er en del av politikken. Ideen om samproduksjon er en kritikk av «...the realist ideology that persistently separates the domain of nature, facts, objectivity, reason and policy from those of culture, value, subjectivity, emotion and politics» (Jasanoff 2004:3). Slike perspektiver har imidlertid vært lite brukt innenfor den tradisjonelle evalueringsforskningen, som tvert imot synes å ha vært opptatt av å distansere seg fra politikken verden og verdier ved å hevde at evaluering er en ekstern og objektiv vurdering av *andres* verdier og mål (Tornes 2013:33).

3.3 Evalueringers innflytelse og et dobbelt perspektiv

Forskning på forholdet mellom kunnskap og politikk, skiller gjerne mellom tre hovedformer for bruk av kunnskap: instrumentell, opplysende/konseptuell og symbolsk/strategisk (se Alkin og King 2017, Cain 2015, Innvær mfl. 2002). Enkelt sagt handler disse kategoriene henholdsvis om evalueringens resultater som påvirker beslutninger

⁵Begrepet *refleksiv modernitet* er hentet fra *Reflexive Modernization. Politics, Tradition and Aesthetics in the Modern Social Order* (Beck, Giddens og Lash 1994). *Revisjonssamfunnet/the audit society* viser til tittelen *The Audit Society. Rituals of Verification* (Power 1997).

direkte, evalueringer som bidrar med ny kunnskap som på sikt kan endre holdninger og politiske beslutninger, og evalueringer som får betydning som symbol og ikke gjennom praktisk bruk av resultatene⁶.

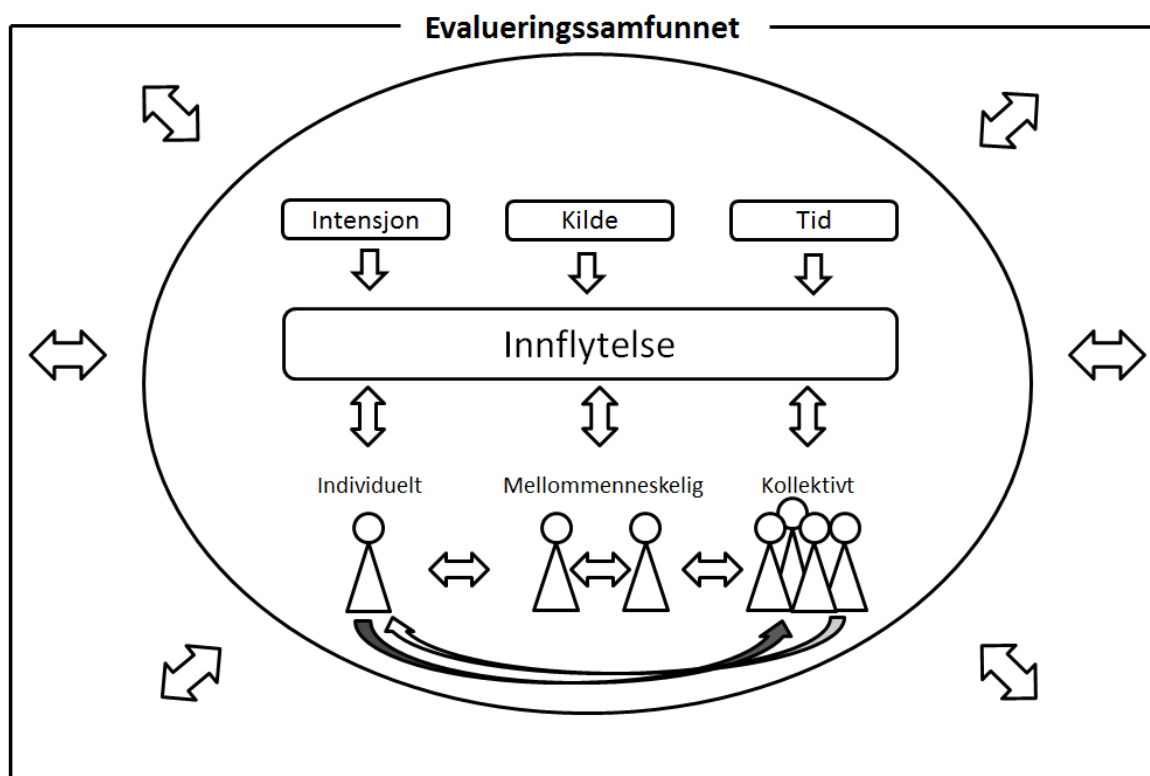
Denne avhandlingen tar imidlertid utgangspunkt i *evalueringssinnflytelse* (Kirkhart 2000). Dette begrepet søker å romme alle former for konsekvenser av evaluering, også de dimensjonen som undersøkes gjennom tidligere nevnte brukskategorier. Utgangspunktet for Kirkharts begrep er blant annet at evalueringer har langt mer sammensatte og vidtrekkende virkninger enn hva de tradisjonelle endimensjonale begrepene fanger opp hver for seg.

Basert på innflytelsesbegrepet, utvikler Kirkhart (2000) en modell der hun vektlegger tre dimensjoner av evalueringers innflytelse: evalueringers virkning over tid, intenderte og uintenderte konsekvenser av evalueringer, samt den delen av evalueringene som skaper endring, det vil si evalueringresultatene eller evalueringprosessen (Kirkhart 2000:17). Henry og Mark (2003) bygger videre på dette arbeidet og utvider det til en modell med tre ulike nivåer som påvirkes av evaluering: det individuelle, det mellommenneskelige og det kollektive nivået (s. 298).

I denne avhandlingen benytter jeg *evalueringssinnflytelse* i tråd med Kirkharts forsøk på å samle alle former for virkninger av evaluering i ett begrep. Jeg benytter også Kirkharts/Henry og Marks utvidede modell i kombinasjon med Dahler-Larsens (2012) begrep *evalueringssamfunnet* og utvikler et nytt analytisk rammeverk som benyttes i artikkel én. Modellen er imidlertid også illustrerende for den tosidige tilnærmingen til evaluering som hele denne avhandlingen bygger på.

⁶ For en utdypende diskusjon av forskjellene mellom evalueringers innflytelse og bruk, se artikkel én: Evaluation – method and societal phenomenon.

MODELL 1:



Evalueringssinnflytelse (Kirkhart 2000/Henry og Mark 2003) i kombinasjon med *evalueringssamfunnet* (Dahler-Larsen 2012).

Dette rammeverket kan fungere som et nyttig utgangspunkt for undersøkelser av evalueringers innflytelse. Modellen rommer innflytelse langs flere akser på ulike samfunnsnivåer. I tillegg viser modellen hvordan det skjer en gjensidig påvirkning mellom evalueringer og det samfunnet der evalueringene gjennomføres, nemlig *evalueringssamfunnet* (Dahler-Larsen 2012). Fordi modellen kan bidra til å konkretisere evalueringssinnflytelse, er den også et godt utgangspunkt for å diskutere evaluering som symbol.

4 Forskningsfeltet

Forholdet mellom evaluering og politikktutforming er gjenstand for forskning innen ulike fagfelt. På den ene siden den rene evalueringsforskningen, som fra et tverrfaglig utgangspunkt studerer virksomhet definert som evaluering. På den andre siden samfunnsfaglige disipliner som fra ulike perspektiver er interessert i forholdet mellom politikk og forskning i bredere forstand. Denne avhandlingen er et bidrag til evalueringsforskningen og bygger primært på begreper og teorier fra dette feltet. Arbeidet er imidlertid også påvirket av bredere samfunnsfaglige perspektiver.

Evidensbevegelsen og dens metodologiske idealer har de siste tiårene preget både forskning og evaluering (Parkhurst og Abeysinghe 2016, Deterding og Solmeyer 2018, Smith 2017, Stockard og Wood 2017). Dermed har også forskningen på disse to feltene nærmet seg hverandre. Gitt disse forholdene og den sterke forskningsmessige forankringen i norsk evalueringstradisjon definerer jeg ingen skarpe grenser mellom forskning og evaluering i denne avhandlingen. Jeg vil også bruke begrepet *vitenskapelig basert kunnskap* som en fellesbetegnelse for forskning og evaluering. I tråd med Radaelli (1995) bruker jeg *kunnskap* som en samlebetegnelse om alt fra informasjon til forskning.

I arbeidet med denne avhandlingen er det spesielt tre forskningstemaer som har vært sentrale. Det første temaet handler hva evaluering er og hvordan evaluering kan forstås. Dette diskuterte jeg i forrige kapittel og vil også komme tilbake til i kapittel fem. De andre to temaene handler om a) i hvilken grad vitenskapelig basert kunnskap har innflytelse i politisk-administrative beslutningsprosesser b) hvordan slik innflytelse, eller mangel på dette er forsøkt forklart.

4.1 Innflytelsen fra vitenskapelig basert kunnskap

Forskningen på sammenhengen mellom evaluering og politikktutforming har pågått i et halvt århundre og opptatt utallige sider i vitenskapelige tidsskrifter og bøker. Derfor er det overraskende at dette arbeidet tilsynelatende har resultert i få tydelige svar. Mange av de mest populære begrepene og perspektivene som fremdeles brukes, stammer fra 1970- og tidlig 1980-tall, eller hva Henry og Mark (2003) omtaler som «The golden age of research on evaluation use». I denne perioden ble det foretatt en rekke relativt omfattende empiriske studier som senere har fått store følger (Caplan 1979, Weiss 1977, Knorr 1977). Senere har mye av forskningen foregått på et teoretisk plan (Coryn mfl. 2017, Mark 2008, Pattyn

mfl.2018), og deler av den empiriske forskningen har vært anklaget for å bygge på mindre case-studier av dårlig vitenskapelig kvalitet (Herbert 2014). Pattyn med flere (2018) går så langt som til å hevde at det meste av litteraturen som omhandler effekten av evalueringer «in practice is anecdotal» (s. 590).

Direkte sammenheng mellom evalueringsresultater og politiske avgjørelser er ifølge en rekke internasjonale forskere lite utbredt (Bogensneider og Corbett 2011, Højlund 2014, Lindgren 2014, Picciotto 2016, Saunders 2012, Vedung 2010b). På samme måte sås det tvil om den direkte nytten av forskning (Boswell 2009, Daviter 2015, Maybin 2016).

Resultatet av empiriske studier er imidlertid noe mer sprikende. Eberli (2016) konkluderer med at sveitsiske politikere i noen grad benytter evalueringsresultater som en kilde til egen læring, samtidig som resultatene også får instrumentell og symbolsk betydning. Bruken avhenger imidlertid av det politiske konfliktnivået. Evalueringsresultatene brukes i størst utstrekning når konfliktnivået er moderat. Boswell (2009) hevder at byråkratiets umettelige kunnskapsbehov skyldes behovet for å legitimere det eksisterende politiske systemet eller bygge autoritet. Disse symbolske funksjonene er ifølge Boswell betydelig viktigere enn kunnskapens instrumentelle betydning. Maybin (2016) hevder på sin side at britiske helsebyråkrater ser egne kollegaer som langt viktigere kilder til kunnskap enn både forskning og evaluering.

Blant enkeltstudier som undersøker sammenhengen mellom evaluering og politikktutforming, er skandinaviske studier spesielt interessante i denne sammenhengen, siden politikk- og samfunnsforhold i Norge, Sverige og Danmark er nokså sammenlignbare (Christensen, Gornitzka og Holst 2017). Resultatene av disse studiene peker i noe ulike retninger. Petersén og Olsson (2014) har undersøkt bruk av evalueringer i forbindelse med sosialt arbeid i Sverige, og konkluderer med at evalueringene primært fungerer som et ritual og har liten betydning som kilde til økt kunnskap eller endring. Tvert imot mener forfatterne at evalueringene først og fremst kommer *utførerne* til gode i form av ny kunnskap som kan resultere i artikler, nye oppdrag og økt personlig kompetanse. Breidahl med flere (2017), derimot, sammenligner konsekvensene av evalueringen av Nav og den danske kommunereformen og hevder at begge evalueringene ble brukt og kom til nytte. Den danske evalueringen fikk primært instrumentelle konsekvenser i form av justeringer av reformen. Nav-evalueringen fungerte i større grad som en kilde til ny kunnskap, og dette var også en uttalt hensikt med denne evalueringen (Breidahl mfl. 2017:12). Dermed avdekker studien

både at evalueringene brukes på ulike måter, og at disse forskjellige bruksformene i noen grad gjenspeiler intensjonene bak evalueringene.

Både den norske DFØ-rapporten *Bruk av evaluering i statlig styring* (2016) og *Surviving the routinization of evaluation: the administrative use of evaluations in Danish municipalities* (Dahler-Larsen 2000) konkluderer med at evalueringer kan ha instrumentelle konsekvenser. Dahler-Larsens studie er imidlertid cirka 20 år gammel. I tillegg er den gjennomført blant danske kommuneledere, en gruppe som i kraft av sin posisjon kanskje er mer positive til nytten av evaluering enn medarbeidere nedover i organisasjonshierarkiet. Deltakerne i den norske studien er personer som har vært oppdragsgivere for konkrete evalueringer, samt ledere og personer med et spesielt ansvar for evaluering (DFØ 2016:12). Dette utvalget gjør det vanskelig å trekke noen generelle slutninger om forvaltningens evalueringsvirksomhet basert på denne studien.

Til tross for at det i litteraturen er populært å vise til kunnskapens potensielle *opplysende* (Weiss 1977) virkning, er dette et begrep som i liten grad har vært inngående diskutert eller studert empirisk (Daviter 2015, Huckel og Blyth 2017, Smith 2013). Smith (2013) tar imidlertid utgangspunkt i enlightenment-begrepet i sin studie av forskningsbaserte ideers påvirkning på britisk helsepolitikk.

Få empiriske studier har også tatt utgangspunkt i Kirkharts (2000) innflytelsesbegrep. Herbert (2014) hevder at forskningen som har vært gjennomført, har vært preget av små studier der formålet hovedsakelig har vært å vise seg fram som profesjonelle brukere eller utførere av evalueringer. Riktignok ble en utvidet modell av innflytelsesbegrepet (Henry og Mark 2003) prøvd ut av Weiss med flere (2005), men forsøket strandet blant annet fordi modellen ble for komplisert å bruke (s. 26).

4.2 Forklaringer på bruk av vitenskapelig basert kunnskap

4.2.1 Hvilke faktorer fører til bruk av vitenskapelig basert kunnskap?

Den såkalte *factors affecting*-litteraturen undersøker *hva som skal til* for at evalueringsresultater eller forskning brukes og får politiske konsekvenser. Denne litteraturen tar gjerne eksplisitt eller implisitt utgangspunkt i Caplans (1979) tese om at forskning og politikk representerer to ulike samfunn eller kulturer, og at det er forskjellene mellom disse to kulturene som skaper utfordringer når forskningsbasert kunnskap skal overføres og implementeres i politikken verden. Dette perspektivet videreføres blant annet i den nyere

knowledge brokering-tradisjonen, der en vektlegger betydningen av aktører som kan overføre eller tolke kunnskap mellom produsenter og brukere (Olejniczak 2017).

De mest omfattende studiene av evaluering innenfor *factors affecting*-tradisjonen er gjennomført av Cousin og Leithwood (1986) og Johnson med flere (2009). Forskningen omfatter til sammen 106 empiriske studier av bruk av evaluering fra 1971 til 2005. Funnene viser at et felles engasjement, godt samarbeid og kommunikasjon mellom evalueringsmiljø og oppdragsgiver synes å være de viktigste elementene som fremmer bruk. Betydningen av relasjonen mellom forskere og policy-makere trekkes også fram av Oliver med flere (2014) i deres gjennomgang av 145 studier av bruk av vitenskapelig basert kunnskap i forbindelse med politikkkutforming. Andre studier har trukket fram betydningen av resultatenes praktiske nytteverdi (Contandriopoulos og Brousselle 2010, Innvær mfl. 2002, Nutley, Walter og Davies 2007, Weiss og Bucuvalas, 1980), forskningsmessig kvalitet, språklig tilgjengelighet og tidsmessig relevans (Bogenschneider og Corbett 2011, Innvær mfl. 2002, Oliver mfl. 2014), eller det faktum at resultatene stemmer overens med gjeldende politikk eller oppfatninger (Innvær mfl. 2002, Nutley, Walter og Davies 2007, Smith 2013, Weiss 1998).

De siste tiårene har både evaluering og samfunnsforskning vært sterkt preget av evidensbevegelsen og dens forventninger om instrumentell bruk av kunnskap basert på eksperimenter, metastudier og kvantitativ metode (Deterding og Solmeyer 2018, Parkhurst og Abeysinghe 2016, Smith 2017). Det er imidlertid i liten grad undersøkt hvilke former for kunnskap politikere eller byråkrater faktisk vurderer som nyttige (Azzam og Jacobsen 2015). I den grad *factors affecting*-litteraturen har behandlet denne tematikken, har spørsmålet blitt redusert til å handle om forskningsmessig kvalitet (Oliver mfl. 2014). En av få studier som faktisk undersøker byråkraters holdninger til evidensbevegelsens metodologi, er Petticrew med flere (2004). Byråkratene som intervjues i studien, er i liten grad opptatt av metodologiske trender eller hierarkier, men etterlyser i stedet kunnskap som kan brukes til å løse praktiske utfordringer og behov. Uavhengig av hva forskere måtte mene om metodene evidensbevegelsen bygger på, mangler det altså studier som undersøker hvilke former for kunnskap som oppfattes som nyttige og relevante blant de potensielle brukerne i det politisk-administrative systemet.

4.2.2 Kontekstuelle forklaringer

Den dominerende modernistiske tilnærmingen til evaluering har medført en begrenset forskningsmessig interesse for betydningen av samfunnsmessig kontekst i evalueringssammenheng. Vo og Christie (2015) hevder likevel at evalueringsforskningen de senere årene i økende grad er blitt seg bevisst betydningen av kontekstuelle forhold som institusjonelle rammer, politiske systemer, verdier og interesser. Denne forskningen synes imidlertid å være dominert av teoretiske studier (Contandriopoulos og Brousselle 2012, Daviter 2015, Højlund 2014, Picottio 2016, Saunders 2012). Og Vo og Christie (2015) mener at organisasjonsmessige faktorer og kontekst fremdeles får for lite forskningsmessig oppmerksomhet, og at slike faktorer bør spille en større rolle når politisk-administrativ bruk av evalueringer skal forklares (se også Alkin og King 2017, Højlund 2014). Kanskje har de siste årenes metodologiske påvirkning fra evidensbevegelsen bidratt til å dempe interessen for slike kontekstuelle faktorer (Parkhurst og Abeysinghe 2016).

4.3 Oppsummert

Denne gjennomgangen viser at det er behov for flere empiriske studier av evalueringers innflytelse på samfunnsliv og politikkutforming. Enten dette dreier seg om studier som fokuserer på ulike former for bruk eller evalueringers innflytelse på et bredere plan. Kunnskapsbehovet finnes både på nasjonalt, skandinavisk og internasjonalt nivå. Den eksisterende evalueringsforskningen er i stor grad preget av en modernistisk tilnærming til evaluering, der evaluering forstås som et nyttig verktøy, og der forholdet mellom politikk og kunnskap framstår som ryddig og avgrenset. Andre perspektiver på dette forholdet kan bidra til en bedre forståelse av dagens evalueringsaktivitet. Det er også behov for flere studier som vektlegger betydningen av kontekstuelle og organisasjonsmessige faktorer i forsøket på å forstå og forklare evalueringers konsekvenser.

5. Teoretisk tilnærming

I denne avhandlingen kombinerer jeg en modernistisk og en sosialkonstruktivistisk tilnærming til evaluering. Dette overordnede teoretiske utgangspunktet for avhandlingen vil jeg i tråd med Røvik (2007) omtale som pragmatisk institusjonalisme. Dette perspektivet har betydning for hvordan jeg undersøker og forstår direktoratenes bruk av evaluering. I tråd med sosialkonstruktivistiske eller sosiologiske perspektiver på kunnskap mener jeg at kunnskap må forstås ut fra historisk og kulturell kontekst. På lik linje med at dagens evalueringspraksis kan forstås fra et tosidig perspektiv, kan også evalueringshistorien betraktes fra to sider.

5.1. Pragmatisk institusjonalisme

Ideen om å kombinere to perspektiver på evaluering oppstod da jeg skrev den første artikkelen min, som handler om PISAs innflytelse. Jeg oppdaget at den norske PISA-deltakelsen både handler om å bruke evalueringsresultatene instrumentelt, som et grunnlag for justering av norsk utdanningspolitikk, og om å framstille Norge og norsk skolepolitikk i et bestemt lys. For eksempel kan en spørre seg om PISA-testene, som ikke er basert på nasjonale læreplaner, egentlig er den beste metoden for å undersøke elevprestasjoner over tid, eller om PISA-deltakelsen like gjerne kan forstås som et symbol på tilslutning til OECDs visjoner og målsettinger og et behov for å vise kontroll og oversikt over skolesektoren (Høydal 2018). Senere, da jeg leste *Trender og translasjoner. Ideer som former det 21. århundrets organisasjon*, oppdaget jeg at forfatteren Kjell Arne Røvik studerer tilbud, overføring og mottak av organisasjonsideer fra et tilsvarende tosidig perspektiv og kaller sin posisjon for *pragmatisk institusjonalisme*. Evaluering kan på samme måte som oppskrifter på organisasjonsstyring studeres som en ide eller trend, det vil si et fenomen «... som i en periode preger diskurser og aktiviteter på mange arenaer» (Røvik 2007: 123). Jeg mener derfor det er fruktbart å overføre Røviks pragmatiske institusjonalisme til evalueringsfeltet.

På samme måte som kombinasjonen av modernistiske og sosialkonstruktivistiske tilnærminger bidrar til å fange tvetydigheten og nyansene som preger spredning og mottak av internasjonale organisasjonstrender, kan en slik tosidig tilnærming bidra til å belyse dualiteten på evalueringsfeltet. Mens en modernistisk tilnærming vil ta utgangspunkt i evaluering som et nyttig og effektivt verktøy, vil et konstruktivistisk perspektiv forstå evaluering som et *symbol* på effektivitet, kontroll, god ledelse og lignende. En pragmatisk tilnærming innebærer at en anerkjenner at evaluering kan være begge deler. Ifølge Røvik (2007) skiller pragmatismen

ikke mellom ideer og praksis. Det vil si at ideen om hva evaluering er og bør være, ikke kan skilles fra den praktiske evalueringsvirksomheten som finner sted. Dermed unngår en de to ytterpunktene som modernistiske og sosialkonstruktivistiske perspektiver representerer. En pragmatisk tilnærming vil derfor anerkjenne at offentlig sektor evaluerer for å løse reelle utfordringer, samtidig som evalueringsvirksomheten også forstås som et forsøk på å framstå i et bestemt lys. Røvik hevder at det nettopp er modernismens vellykkede fortellinger om organisasjonsideenes nytteverdi som legger grunnlaget for den symbolske betydningen av de samme ideene (Røvik 2007, 53-54). Denne tolkningen synes også relevant for evalueringsfeltet. *Fortellingen* om evaluering som et praktisk nyttig og nødvendig verktøy fører til at evaluering også blir stående som et sterkt symbol på de samme egenskapene. Kanskje er det i realiteten også vanskelig å avgjøre hvor grensene mellom evaluerings symbolske kraft og praktiske virkning kan trekkes, eller om disse to sidene fungerer som en gjensidig forsterkning av hverandre.

Den tosidige tilnærmingen medfører også at en unngår noen av fallgruvene som kan prege rent modernistiske eller sosialkonstruktivistiske perspektiver. En sosialkonstruktivistisk tilnærming til evaluering kan for eksempel medføre at evaluering reduseres til et symbol, der potensialet for praktisk nytte hverken anerkjennes eller undersøkes. På den andre siden kan modernistiske perspektiver kritiseres for sin overdrevne tro på evaluerings potensiale i politiske beslutningsprosesser, og for et kunnskapssyn som i liten grad problematiserer de forutsetningene som preger den kunnskapen evaluering bidrar med. I tabellen under (tabell 3) viser jeg hvordan et modernistisk, sosialkonstruktivistisk og pragmatisk perspektiv vil betrakte evaluering, årsaker til evaluering og evalueringenens resultater:

Tabell 3: Tre tilnærminger til evaluering:

	modernistisk	sosialkonstruktivistisk	pragmatisk
Hva er evaluering?	rasjonelle verktøy	symboler	redskaper med instrumentelle effekter sosiale konstruksjoner
Hvorfor evaluere?	direktoratene evaluerer for å løse reelle utfordringer	evaluering gjennomføres for å framstå i et bestemt lys	virksomheten anser seg selv som en rasjonell aktør som vil løse reelle problemer virksomheter adopterer bestemte ideer for å framstå som en moderne organisasjon
Hva medfører evaluering?	bedre beslutninger	framstår som rasjonelle	bedre beslutninger framstår som rasjonelle

Samtidig kan en tosidig tilnærming til evaluering kritiseres for å bli tannløs fordi den spriker for bredt og rommer for mye. Etter min oppfatning er imidlertid realiteten så kompleks at et ensidig perspektiv ikke klarer å fange de nødvendige nyansene. En tosidig tilnærming er også et godt tilsvarende til påstander om at evalueringsforskningen og forskningen på forholdet mellom kunnskap og politikktutforming er unyansert og normativ (Huberman 1987, Maybin 2016, Mathison 2018, Vallin mfl. 2015). Ifølge Huberman (1987) skyldes mangelen på nyanser at denne forskningen behandler grunnleggende vitenskapsteoretiske spørsmål, og dermed reflekterer vitenskapsteoretiske skillelinjer knyttet til aktørers rasjonalitet, kunnskapens objektivitet og verdifrihet. Dette fører til at forskerne underkommuniserer nyanser i sine egne funn, eller ikke er interessert i å finne slike nyanser fordi de ikke passer inn i et todelt vitenskapsteoretisk verdensbilde (1987:590). Et teoretisk utgangspunkt som anerkjenner at evaluering på den ene siden kan bidra med kunnskap som brukes instrumentelt, og på den andre siden inkluderer evaluering som et symbol som gjenspeiler samfunnets verdier, rommer nettopp det spekteret av nyanser som Huberman etterlyser en forskningsmessig åpenhet og interesse for.

5.2. Et dobbelt blikk på evalueringsbølgen

Evalueringshistorien har i flere tilfeller blitt framstilt som en serie av internasjonale bølger (Dahler-Larsen 2012, Krogstrup 2011, Vedung 2010) der sedimenter fra hver bølge blir med i neste fase av den historiske utviklingen. I denne framstillingen tar jeg utgangspunkt A) Vedung (2010), som beskriver evalueringshistorien med utgangspunkt i evaluering som et nyttig verktøy eller metode B) Dahler-Larsen (2012), som med utgangspunkt i ideen om evalueringssamfunnet beskriver den samme historien om evaluering som et sosiokulturelt fenomen. Vedungs framstilling bygger på fire historiske perioder, mens Dahler-Larsens handler om hvordan evaluering kan forstås i tre ulike stadier av moderniteten. Dermed er det ikke noe direkte sammenligningsforhold mellom alle fasene. Det dreier seg vel også i realiteten mer om glidende overganger.

Som tidligere nevnt er evaluering et resultat av modernitetens tro på kunnskap og utvikling (Vedung 2010). I moderniteten erstattes religion og tradisjon med rasjonalitet, og evaluering er ikke bare et verktøy som manifesterer denne tankegangen, men også et symbol på «framskritt gjennom kunnskap» (Dahler-Larsen 2012:100). Utover 1970-tallet oppstod det imidlertid en voksende skepsis til ideen om at samfunnet kunne styres via naturvitenskapelig inspirerte eksperimenter og effektstudier. I stedet ble både forsknings- og evalueringsfeltet preget av nye strømninger som inkluderte brukermedvirkning, dialog, konsensus og demokratisk legitimitet, eller det Vedung omtaler som «*den dialogbaserte bølgen*» (2010:270). Den tredje evalueringsbølgen knytter Vedung til den internasjonale framveksten av New Public Management (NPM) og reformer i offentlig sektor på 1980- og 1990-tallet som ble påvirket av denne retningen. Reformene medførte et nytt fokus på mål- og resultatstyring og påfølgende evaluering. Evaluering ble på denne måten en integrert del av den ordinære driften av offentlig virksomhet og handlet dermed mer om offentlig styring enn om vitenskap.

Dahler-Larsen beskriver også en overgang fra etterkrigstidens vitenskapsoptimisme, via voksende tvil om vitenskapens samfunnsendrende kraft, til usikkerhet og mistanke og *den refleksive moderniteten*. I denne epoken av moderniteten sliter samfunnet med modernitetens bivirkninger, som forurensning, stress og velferdsavhengighet. Det offentlige må vise at disse problemene tas på alvor – ved å evaluere (2012: 144). I en tid med usikkerhet blir evaluering først og fremst et *symbol* på god ledelse, kunnskapsbasert beslutningstaking og refleksjon. Selv om evalueringene oppfattes som viktige, er det ikke lenger resultatene og konsekvensene som har størst betydning, det viktigste er at evalueringene gjennomføres (Dahler-Larsen

2012:145-147). Usikkerheten knyttet til konsekvensene av evalueringsvirksomheten medfører imidlertid at evaluering gradvis mister legitimitet, og dette får konsekvenser for evalueringen i *revisjonssamfunnet*.

Revisjonssamfunnet håndterer samfunnets utfordringer gjennom kontroll. Evaluering gir en illusjon av at problemer lar seg kontrollere og håndtere, og slik skapes det trygghet. I revisjonssamfunnet er evaluering standardisert, rutinebasert og integrert i andre organisatoriske prosedyrer. Betegnelsen *evalueringsmaskin* (Dahler-Larsen 2012:176) brukes som en analytisk metafor for evalueringsaktiviteten i revisjonssamfunnet.

Evalueringsmaskinene vektlegger risikostyring, kvantifisering og objektive standarder, og virkeligheten må justeres eller justeres for å passe til de nye evalueringsprosedyrene.

Den standardiserte og kvantitativt baserte evalueringen som finner sted i revisjonssamfunnet, har likhetstrekk med det Vedung (2010) omtaler som den fjerde bølgen, nemlig *evidensbølgen*. Her inngår evaluering som en del av idealet om kunnskapsbasert politikktutforming (EBP). Basert på et evidenshierarki rangeres evaluering etter metode, og i likhet med den første evalueringsbølgen er det eksperimenter og store kvantitative studier som representerer gullstandarden. Dette nye metodefokus gjør at evaluering går fra å være en ren politisk-administrativ aktivitet til igjen å nærme seg forskningsfeltet (Vedung 2010).

Denne gjennomgangen viser hvordan evaluering både har vært tett forbundet med samfunnsstyring og politikk på den ene siden og med forskning og vitenskap på den andre. Evalueringsfeltet har derfor både blitt påvirket av skiftende vitenskapelige idealer så vel som av endrede styringsidealene. Mens endringen mellom den første vitenskapsoptimistiske evalueringsbølgen og den dialogbaserte bølgen gjenspeiler et vitenskapelig paradigmeskifte, er overgangen til den tredje bølgen av en annen karakter. Evaluering går fra å være en vitenskap til å bli et institusjonalisert verktøy knyttet til drift og styring av offentlig sektor. NPM kan dermed beskrives som en styringsideologi der evaluering blir et kraftfullt symbol på god ledelse, og der evalueringsmetodikken gjenspeiler de nye praktiske behovene ideologien etterspør. Evidensbevegelsen er derimot både en styrings- og en forskningsideologi. Politiske beslutninger skal baseres på evidens, men evidens er knyttet til et bestemt metodesyn, og dette metodesynet får store konsekvenser for den praktiske gjennomføringen av evalueringer og hva slags kunnskap som *framstår* som god nok.

5.3. Et pragmatisk perspektiv på norsk evalueringstradisjon

Norge kan i tråd med Dahler-Larsen (2012) beskrives som et evalueringssamfunn. Evaluering har blitt et ritual – med eller uten praktiske konsekvenser, men like fullt et forventet aspekt ved en rekke sider av livet. Offentlige instanser som skoler og barnehager ber oss evaluere tilbudet deres. På flyplassen vil de vite om toalettene er rene, i butikken spør de om servicen holder mål, om du finner fram i hyllene og ellers hadde «en god kjøpsopplevelse». Som arbeidstaker kan du evaluere lederen din, arbeidsmiljøet, kantinetilbudet og innklimaet. Som student blir du bedt om å rangere forelesere, kurstilbud og fysiske fasiliteter. Summen av alle disse opplevelsene er at evaluering nærmest oppleves som en integrert del av det offentlige tjenestetilbudet – og som noe som hører sammen med rollene som forbruker, arbeidstaker, student, leder og borger. Selv om temaet for denne avhandlingen er evaluering i statlige direktorater, er det ikke likegyldig i denne sammenhengen at evalueringssamfunnet også preger hver og en av oss i dagliglivet. Ved at vi er omgitt av evaluering i et utall roller og situasjoner, blir evaluering noe vi tar for gitt og forventer. Hver og en av oss sosialiseres inn i rollen som evaluator – evaluering blir en del av mentaliteten vår, av våre forestillinger om hvordan verden skal fungere (Dahler-Larsen 2012).

Slik har det ikke alltid vært. Norge regnes som en *latecomer* i evalueringssammenheng (Baklien 1993a, Dahler-Larsen 2013, Tornes 2012). Mens den første evalueringebølgen spredte seg fra USA til land som Storbritannia, Sverige og Danmark ble evaluering vanlig i Norge først på 1980-90-tallet (Furubo, Rist og Sandahl 2002). Dette skjedde i forbindelse med innføringen av NPM-pregede reformer i offentlig sektor. Reformene innebar blant annet desentralisering av statlige oppgaver, økt fokus på mål- og resultatstyring og et behov for oppfølging og kontroll gjennom evaluering (Kvitastein 2002). Eide (1999) hevder at det også skjedde et brudd med den tidligere nordiske styringsmodellen på 1990-tallet. Mens politiske reformer tidligere var et resultat av langvarige høringsrunder, utredninger, dialog og kompromisser, skjedde mye av den nye styringen i etterkant av reformene, gjennom evaluering og eventuell kursjustering (Eide 1999:41).

I perioden mellom 1994 og 1996 fikk norske departementer gjennomført mer enn 200 evalueringssjekker til en prislapp på cirka 180 millioner kroner (Brofoss 1997, i Øvrelid 2000:135), og evaluering ble et nytt trendbegrep i norsk sentralforvaltning (Øvrelid 2000:134). Utover 1990- og 2000-tallet skjedde det en ytterligere delegering av oppgaver, og behovet for rapportering og tilsyn ble forsterket (Trygstad, Lorentzen og Løken 2005:8).

Norsk forvaltning var imidlertid ikke preget av det samme eksterne evalueringstrykket som EU, etter at landet stemte nei til EU-medlemskap i 1994 (Øvrelid og Bastøe 2002).

Til tross for at Norge tilsynelatende ikke ble truffet av Vedungs (2010) to første evalueringebølger, er det ingen tvil om at norske myndigheter har vært preget av den samme forskningsoptimistiske og rasjonalistiske bølgen som Vedung (2010) og Dahler-Larsen (2012) beskriver som karakteristisk for evalueringsfeltet i tiårene etter andre verdenskrig (Eide 1999, Tranøy, Christensen, Nuland og Hernes 2009). I denne perioden opprettet norske myndigheter blant annet flere frittstående forskningsinstitutter som skulle bidra med forskningsbasert kunnskap til det politisk-administrative systemet. Det ble også etablert planleggingsavdelinger i de fleste departementene med nære kontakter til eksterne samfunnsforskere (Øvrelid 2000, Eide 1999). Og på samme måte som det internasjonale evalueringsfeltet utover 1970-tallet gikk gjennom betydelige endringer knyttet til både kunnskapssyn og metodebruk, oppstod lignende trender innenfor den norske oppdragsforskningen. Ifølge Baklien (1993a) var mye av datidens aksjonsforskning i realiteten det samme som evaluering. På samme måte hevder Kvitastein (2002) at det i praksis er lite som skiller 1990-tallets evalueringer fra statlige rapporter, som tidligere ble kalt utredninger. Dersom en tar slike forhold i betraktning, blir det vanskelig å tidfeste når norsk forvaltning faktisk begynte å evaluere, og i hvilket omfang det tidligere har blitt evaluert. Forsknings- og utviklingsarbeidet har også vært ulikt organisert i sentralforvaltningen. Utenriksdepartementet etablerte egen evalueringsenhet allerede i 1978 (Øvrelid og Bastøe 2002). Utdanningssektoren drev også uttalt evalueringsvirksomhet på et tidlig tidspunkt, og Frøyland hevdet allerede i 1969 at evaluering hadde «... glidd inn i klisjeenes meningstomme rekke» (1969:91).

Denne gjennomgangen viser hvordan norsk statsforvaltning gjennom hele etterkrigsperioden har vært preget av modernitetens rasjonalisme og kunnskapstro. Selv om metodebruk og kunnskapssyn har variert, har vitenskapelig basert kunnskap blitt oppfattet som en ressurs og et bidrag til bedre politikkutforming. Deler av den forskningen norske myndigheter har bestilt i tiårene etter andre verdenskrig, har også i praksis vært evalueringer, selv om de ferdige

rapportene har gått under andre navn. Evalueringsbegrepet har også hatt ulik utbredelse innenfor ulike statlige sektorer⁷.

Likevel skjedde den en markant økning i antallet evalueringer på 1980- og 90-tallet, selv om deler av denne økningen mest sannsynlig kan tilskrives ny begrepsbruk (Øvrelid og Bastøe 2002). New Public Management er ifølge Røvik en form for internasjonal styringstrend som det er vanskelig å unngå, både fordi slike trender tilsynelatende representerer løsninger de fleste organisasjoner og stater kan ha nytte av, og fordi det oppstår et stort påtrykk for å adoptere slike ideer. Både EU og OECD har vært ivrige pådrivere for spredningen av NPM-inspirerte styringsidealer og bruk av evaluering. OECD etterlyste for eksempel evalueringssystemer knyttet opp mot norsk skolesektor på 1980-tallet (Monsen og Haug 2004). De NPM-inspirerte reformene og den påfølgende evalueringebølgen kan forklares ut fra et *behov* for å effektivisere og begrense veksten i offentlige kostnader (Olsen 1986 i Øvrelid og Bastøe 2002). Svekket tillit til at velferdsstaten ville klare å løse samfunnets stadig mer komplekse utfordringer, kan også ha bidratt til sterkere interesse for styringsrelevant kunnskap (Tranøy, Christensen, Nuland og Hernes 2009). Norske myndigheters tilslutning til NPM-ideologien og dens tilhørende verktøy- og begrepsapparat kan også forstås som et forsøk på å vise at de var mottakelige for internasjonale innspill og trender og villige til å ta nødvendige og fremtidsrettede grep om offentlig forvaltning.

5.4. Evaluering i dagens forvaltning

Lindgren (2014) betegner evaluering som et «forvaltningspolitisk hovedinstrument» (s.12). Selv om hun beskriver svenske forhold, synes det som om denne beskrivelsen også kan passe for norske forhold. Undersøkelsen i de sosialpolitiske direktoratene viser i hvert fall at det store flertallet av byråkratene oppfatter evaluering som et nyttig og nødvendig verktøy.

Dagens institusjonaliserte evalueringspraksis bærer tydelige spor av NPM-ideologiens fokus på økonomi, effektivitet og resultater. Dette kommer blant annet til uttrykk gjennom organiseringen av statlig evalueringsevne: Det overordnede faglige ansvaret er plassert hos DFØ, et direktorat under Finansdepartementet. Koblingen mellom evaluering og økonomi

⁷ Dermed er evaluering et fenomen med en lengre historie i norsk forvaltning enn hva det for eksempel gis inntrykk av gjennom forsøket på en samlet evalueringshistorisk framstilling i *International Atlas of Evaluation* (Furubo, Rist og Sandahl 2002).

viser seg også gjennom det faktum at etatene er pålagt å evaluere gjennom *Regelverket for økonomistyring i staten*, og gjennom ordlyden i dette regelverket. Her slås det fast at evaluering er et *styringsverktøy*, og det heter at direktoratene skal evaluere for å undersøke «effektivitet, måloppnåelse og resultater» (paragraf 16). Regelverket avgrensers dermed evalueringens rolle som styringsverktøy til å handle om en helt spesiell type kvalitetsindikatorer. I tråd med Røvik (2007) kan en hevde at økonomiregelverket gjennom sin beskrivelse av hva evaluering er og kan bidra med, også gjør evaluering til et symbol på god styring, økt effektivitet og større måloppnåelse. Dermed blir evaluering synonymt med for eksempel effektivitet, selv når evalueringsresultatene ikke medfører slike praktiske konsekvenser.

Statlig evalueringsvirksomhet handler imidlertid om mer enn styring og økonomi. Byråkratene som deltok i undersøkelsen mente at den viktigste hensikten med evaluering var å skaffe nye kunnskap. Evaluering inngår også som en del av det nye fokuset på kunnskapsbasert politikkutforming (DFØ 2016) og diskursen som omgir dette fenomenet. I denne sammenhengen handler evaluering om å undersøke *what works*, og om å bidra med evidensbasert kunnskap til politikere og forvaltning. Disse to evalueringsidealene gjenspeiler hoveddimensjonene i statlig politikk, som handler om å sørge for gode løsninger og godt innhold i de politiske beslutningene som vedtas, og samtidig sikre en sunn økonomi (Øvrelied og Bastøe 2002).

I Norge har evidensbevegelsen blant annet resultert i opprettelsen av kunnskapssentre for henholdsvis utdanning og helse. Kunnskapssenteret for helse, som nå er integrert i Folkehelseinstituttet, har i praksis hatt et bredt sosialpolitisk perspektiv (Elvbakken og Foss Hansen 2018). Sentrene overvåker relevante forskningsfelt, syntetiserer forskningskunnskap og driver med forskningsformidling. De samarbeider med internasjonale representanter for evidensbevegelsen, som Campbell Collaboration og What Works Clearinghouse⁸. Evidensbevegelsens idealer uttrykkes også gjennom økt bruk av randomiserte kontrollerte studier (RCT) i norsk utdanningsforskning (Pontoppidan mfl. 2018). Det kommer også til uttrykk gjennom Norges deltakelse i flere internasjonale komparative utdanningsstudier, gjennom PISAs betydning for norsk skole, og gjennom den betydningen de kvantitative dataene fra slike undersøkelser har fått. Byråkratene jeg intervjuet i forbindelse med denne

⁸ https://www.forskningsradet.no/prognetskunnskapssenter/Om_Kunnskapssenteret/1247146832396?lang=no

avhandlingen, ga også på ulike måter uttrykk for at de var påvirket av evidensbevegelsens metodologiske idealer. En fortalte at direktoratet han arbeidet i nærmest ble presset til å få eksterne eksperter til å utarbeide en kunnskapsoversikt innenfor et bestemt fagfelt.

Begrunnelsen var at «alle andre» gjorde det samme. En annen snakket om de metodiske utfordringene knyttet til gjennomføringen av RCT-er og hvordan dette hadde medført flere dårlige undersøkelser. En tredje var opptatt av politikernes interesse for effekten av ulike tiltak, og hvordan slike effekter i praksis ofte var vanskelige å måle. Dermed ble det ofte snakk om «å sannsynliggjøre en effekt» for å tilfredsstille forventningene.

Kunnskap fra eksperimenter, forskningssynteser og internasjonale skolestudier kan være nyttige framgangsmåter for å skaffe til veie deler av den kunnskapen som er nødvendig for å forstå komplekse sosialpolitiske utfordringer og tiltak. Resultatene fra spørreundersøkelsen viser imidlertid at bare 22 prosent av deltakerne mener evalueringer basert på kvantitativ metode bidrar til mer pålitelig informasjon enn andre evalueringer. Byråkratene betrakter dessuten kvalitative data som observasjoner, og møter, konferanser og workshops som de mest nyttige bidragene til nye ideer og perspektiver. Dermed kan det synes som om evidensbevegelsens viktigste funksjon er å være en form for symbolsk ledestjerne, og som om de metodologiske idealene har noe begrenset betydning ute i de enkelte direktoratene.

Vedung (2010) viser at det er betydelige likhetstrekk mellom den første internasjonale evalueringsbølgen og evidensbølgen. De deler metodologiske idealer og sterk tiltro til kunnskapens samfunnsmessige potensial. Dagens kunnskapsbaserte politikktutforming gir likevel uttrykk for mer beskjedne målsettinger enn den første evalueringsbølgen. Vitenskapen skal ikke lenger reformere velferdsstaten eller forandre verden, men representerer i stedet et system der samfunnet må tilpasse seg vitenskapens kategorier slik at det blir mulig å evaluere (Dahler-Larsen 2012).

Evalueringshistorien illustrerer hvordan det i praksis er vanskelig å trekke klare og allmenngyldige grenser mellom evaluering og andre former for kunnskapsproduksjon. Evaluering har både i Norge og internasjonalt vært assosiert med forskjellige metoder, mål og verdier og bør derfor forstås i en sosial og historisk kontekst. Samtidig illustrerer dette den fleksibiliteten som Power (1997) hevder er evalueringskonseptets styrke, og som har gjort det mulig å eksportere og tilpasse evaluering på tvers av landegrenser og institusjoner. Evalueringens dualitet kan også være et bidrag til den samme populariteten. Dersom evalueringer ikke får praktiske implikasjoner eller brukes overhodet, kan

evalueringsvirksomheten i seg selv likevel framstå som et symbol på rasjonalitet, ledelse og kontroll (Dahler-Larsen 2012).

Selv om evaluering har vist seg å være en tilpasningsdyktig internasjonal trend, er det viktig å forstå evaluering i lys av det samfunnet og den konteksten den inngår i. For eksempel er det vanskelig å sammenligne det vi kan kalle en norsk, forskningsforankret evalueringstradisjon med den profesjonsorienterte tradisjonen i USA eller den britiske evalueringstradisjonen påvirket av metoder og idealer fra økonomifaget (Duffy 2017). Forskjellene mellom evalueringsidealer fra henholdsvis økonomi og samfunnsforskning kan ha betydning for valg av evalueringsobjekt, evalueringskriterier, metode og perspektiver i gjennomføringen av evalueringer. Dermed har slike perspektiver også betydning for resultatene og hvilke konsekvenser evalueringsrapportene faktisk får. Forskjellene handler også om hvilke ambisjoner en har for evaluering. Den amerikanske ambisjonen om at evaluering skal lede til samfunnsendring (social betterment) er i hvert fall i liten grad uttalt når norske myndigheter evaluerer.

6. Metode, data og egen rolle

Selv om det finnes relativt lite norsk forskning om konsekvensene av forvaltningens evalueringsevne⁹, så finnes det erfaring og fortellinger. Både fra dem som gjennomfører evalueringene, bestillerne, brukerne og fra dem som selv blir evaluert. Det er fortellinger om rapporter som havnet i første og beste skuff da det plutselig ble et regjeringsskifte. Det er anekdoter fra departementsansatte om hvilke triks de bruker for å hindre oppmerksomhet rundt negative evalueringresultater. Og det er frustrerte fagfolk som mener de blir målt på resultater som har liten betydning, men er lette å måle. Disse erfaringene og fortellingene har jeg tatt med meg fra tidligere arbeidserfaring, men også fra venner og kjente i og utenfor statlig sektor.

Jeg fikk selv et slags evalueringssjokk i mitt første møte med forvaltningen etter seks år i privat næringsliv. Jeg hadde ingen erfaring med evaluering i jobbsammenheng og kom rett inn i en organisasjon der evaluering var det nye vidundermiddelet. Evaluering skulle ikke bare bidra til bedre tiltak og informerte beslutninger, men også medføre en generell kompetanseheving i organisasjonen, som blant annet betydde at byråkratene skulle lære å jobbe på nye måter, drive bedre prosesser og tenke mer helhetlig. Evaluering skulle transformere hele organisasjonen fra et «forvaltningsorgan» til en «kompetanseenhet». Som nyansatt var jeg litt avventende. Dels fordi jeg ikke kjente organisasjonen og det byråkratiske systemet, men også fordi jeg var ansatt som kommunikasjonsrådgiver og i praksis ikke var direkte involvert i noen av de pågående prosessene. Etter hvert ble det imidlertid min oppgave å skrive om evaluering. Dette opplevde jeg som en krevende øvelse. Fra et kommunikasjonsfaglig perspektiv var det mange store ord og planer, prosesser og strategier, men lite håndfast å ta tak i. Samtidig fikk jeg likevel inntrykk av at evalueringssjokkene og resultatene av dem hadde stor innflytelse i interne diskusjoner og beslutninger. Disse

⁹ Et utvalg tidligere norsk evalueringslitteratur:

- Almås, Reidar (1990), *Evaluering på norsk. Ei innføring i vurdering av prosjektarbeid og handlingsretta forskning*. Oslo: Universitetsforlaget
- Baklien, Bergljot (1993), Evalueringssjokk i Norge. *Tidsskrift for samfunnsforskning*, 34: 261–374
- Baklien, Bergljot (1985), Evalueringssjokkene mellom barken og veden. *Tidsskrift for samfunnsforskning*, 26: 457–475
- Sverdrup, Sidsel (2002), *Evaluering – Faser, design og gjennomføring*. Bergen: Fagbokforlaget

erfaringene gjorde meg kritisk interessert og undrende til evalueringsbølgen i norsk offentlig sektor.

6.1 En spørreundersøkelse blir til

6.1.1 Utgangspunkt og metodevalg

Da jeg begynte arbeidet med denne avhandlingen, var det nylig utgitt tre norske bøker om evaluering (Halvorsen, Lier-Madsen & Jentoft 2013, Sverdrup 2014, Tornes 2013). Bøkene belyser ulike sider ved evalueringsfeltet, men ingen tar utgangspunkt i byråkratiets perspektiver, og samtlige etterlyser mer forskning på statlig evalueringspraksis. Dette gjorde at jeg bestemte meg for å undersøke statens evalueringsvirksomhet fra byråkratiets ståsted. Den generelle mangelen på forskning gjorde at jeg ønsket å gjennomføre en spørreundersøkelse blant byråkrater innenfor et gitt politisk område for å få breddekunnskap om evaluering innenfor et spesifikt politisk felt.

Det sosialpolitiske feltet ble valgt både på grunn av feltets allmenne betydning og fordi det er innenfor dette feltet staten evaluerer i størst utstrekning. I tillegg er evaluering også internasjonalt forbundet med dette feltet både historisk og i dag. Valget falt på direktorater på grunn av deres sentrale rolle i den kunnskapsbaserte politikktutformingen. I denne sammenhengen er de sosialpolitiske direktoratene forstått som Helsedirektoratet, Utlendingsdirektoratet (UDI), Integrerings- og mangfoldsdirektoratet (IMDi), Husbanken, Barne-, ungdoms- og familiedirektoratet (Bufdir) og Utdanningsdirektoratet (Udir). Nav ble utelatt fordi organisasjonen relativt nylig har vært igjennom enorme omlegginger med påfølgende store evalueringer og dermed representerer en atypisk case blant direktoratene.

Mangelen på eksisterende forskning og kunnskap om byråkratiets bruk av evaluering gjorde at jeg innledningsvis både var nysgjerrig på alle de praktiske sidene ved direktoratenes evalueringsvirksomhet, samtidig som jeg var opptatt av byråkratenes holdninger til evaluering og i hvilken grad de mente dette var et nyttig kunnskapstilfang. For å forstå mitt eget forskningsfelt godt nok til å stille gode og relevante spørsmål, startet jeg hele prosjektet med å foreta orienteringsintervjuer (Titlestad 1982) med fire personer med evalueringserfaring fra tre ulike direktorater og et departement. Intervjuene ble gjennomført både for å lære mer om evalueringsaktiviteten i norske direktorater, og for å få bedre kunnskap om arbeidet i det politisk-administrative systemet og om hva politikere og byråkrater faktisk gjør når de utformer politikk (Maybin 2016).

Det ble ikke brukt noen standard intervjuguide i forbindelse med disse orienteringssamtalene, men noen temaer gikk igjen i hver samtale. Dette var spørsmål knyttet til selve bestillingsprosessen, eventuell uenighet og konflikt, kvaliteten på rapportene, og sist, men ikke minst i hvilken grad byråkratene oppfattet evaluering som nyttig, og om resultatene fikk konsekvenser. Informantenes innspill ble skrevet ned for hånd og deretter renskrevet som tematiske utsagn. Disse utgjorde senere en idebank som ble brukt i det videre arbeidet med spørreskjemaet.

Byråkratene var overraskende positive til evalueringsaktiviteten i egen organisasjon. Etter samtalene satt jeg igjen med et inntrykk av at evaluering bidrar til nyttig og nødvendig kunnskap som det ellers er vanskelig å få tilgang til. Dette stod i kontrast til en del av den litteraturen jeg hadde lest, som i begrenset grad ga inntrykk av evaluering som et praktisk nyttig verktøy (Maybin 2016, Petersén og Olsson 2014). Som en konsekvens av dette tilsynelatende misforholdet mellom påstander i litteraturen og norsk praksis, ble jeg stadig mer nysgjerrig på de faktiske følgene av direktoratenes evalueringsvirksomhet. Dette gjenspeiles i det ferdige spørreskjemaet.

6.1.2 Testing, utvalg og deltakere

Kvaliteten på en spørreundersøkelse avhenger av spørsmålenes presisjon. Det vil si i hvilken grad respondentene forstår hva spørsmålet dreier seg om, om de har en felles forståelse av ordlyden, og om denne forståelsen stemmer overens med spørsmålsstillerens forståelse (Haraldsen 1999:126). Testing av spørreskjemaer er derfor viktig for validiteten av senere svar. Det aller første utkastet mitt til spørreskjema ble diskutert i forbindelse med et metodekurs med kursholdere blant annet fra Statistisk sentralbyrå (SSB). Dette resulterte i nyttige innspill knyttet til skjemaets struktur og spørsmålsoppbygging. Tilbakemeldingene medførte blant annet at jeg 1) la inn filterspørsmål for å unngå at deltakere uten relevant kunnskap skulle besvare noen av de mest detaljerte spørsmålene, 2) utformet spørsmål der deltakerne aktivt måtte ta stilling til de ulike svaralternativene ved å svare JA eller NEI for hvert alternativ, 3) ryddet opp i spørsmålsrekkefølgen i skjemaet, og 4) inkluderte svaralternativer jeg ikke hadde kommet fram til på egen hånd. Blant annet ble workshops/konferanser/møter lagt inn som en mulig metode for datainnsamling som følge av dette (se vedlegg 5, spørsmål nr. 5).

Skjemaet ble testet til sammen sju ganger av til sammen fem personer. To av disse testpersonene hadde også deltatt i de innledende orienteringsintervjuene. I utgangspunktet hadde jeg planlagt at ingen av de tidligere informantene skulle teste spørreskjemaet, men da det viste seg å være svært tidkrevende å rekruttere nye testpersoner, benyttet jeg denne muligheten likevel. Jeg opplevde også at selv om enkelte testet skjemaet flere ganger og noen hadde deltatt som informanter i orienteringsintervjuene, bidro hver test med konstruktive innspill.

De første testene fungerte som samtaler der jeg sammen med testpersonene gikk gjennom en papirbasert variant av spørreskjemaet. Samtidig som testpersonene besvarte spørsmålene i skjemaet, bidro de også med mer generelle synspunkter og erfaringer knyttet til evaluering. Testingen medførte alt fra språklige endringer til nye svaralternativer. For eksempel tydeliggjorde testingen at én og samme evaluering kan innebære bruk av flere metoder, flere evalueringssubjekter og ulike hensikter. Som en følge av dette endret jeg skjemaet slik at respondentene i mindre grad ble presset til å velge eller prioritere ett bestemt svaralternativ, men i stedet kunne velge flere ulike svar. Testpersonene var på den ene siden langt på vei enige om nytten av evaluering, på den andre siden betraktet de også evalueringsevne som noe som var utenfor hver sin rolle og hvert sitt ståsted, det vil si fra ulike kontekstuelle posisjoner innenfor den samme statsforvaltningen. Testingen medførte derfor et stadig fornyet tilfang av interessante perspektiver og nyttige spesifiseringer. Dermed resulterte også denne prosessen i en kontinuerlig avveining mellom hvilke temaer og spørsmål som burde være med, og hvor langt skjemaet kunne være før framtidige respondenter ville vegre seg for å delta eller bryte underveis. Utover i testprosessen ble tilbakemeldingene stadig mindre kritiske. De tre siste testene fungerte som reelle undersøkelsessituasjoner, der de tre deltakerne svarte på den nettbaserte versjonen av undersøkelsen i programvaren Questback. I etterkant av disse testene ble det kun foretatt mindre språklige justeringer, og jeg anså derfor undersøkelsen som klar for utsendelse.

Arbeidet med utformingen av spørreundersøkelsen pågikk fra august 2015 til januar 2016. Selve undersøkelsen startet opp med første direktorat i januar 2016. Deretter gjennomførte jeg nye utsendinger etter hvert som nye direktorater sa seg villige til å delta. Siste undersøkelse ble avsluttet i juni 2016. Jeg opplevde at hele prosessen knyttet til rekruttering av deltakerorganisasjoner, testpersoner og informanter var vanskelig å styre eller planlegge fordi jeg hele tiden var avhengig av andre menneskers tid, velvilje og interesse. Målet var å

rekruttere så mange relevante deltakere til undersøkelsen som mulig fra hvert direktorat. Det vil si deltakere som jobbet innenfor de delene av organisasjonen som på ulike måter er i direkte berøring med evalueringer. Ved å inkludere alle ansatte ønsket jeg å kartlegge evalueringsevne og implikasjoner for ulike nivåer og fagfelt i organisasjonen. Det ville også gi anledning til å kartlegge holdninger til evaluering blant ulike grupper ansatte. Risikoen ved en slik bred tilnærming er at ansatte som i liten grad kjenner seg berørt av evalueringsskjemaene, unnlater å svare. For å motvirke et slikt frafall la jeg opp til at byråkrater med liten erfaring skulle få færre spørsmål. Dette kom fram i e-posten med lenke til undersøkelsen (Haraldsen 1999:208 og 269).

I praksis ble imidlertid deltakelsen i stor grad et resultat av det enkelte direktorats rutiner, kapasitet og engasjement (se tabell 1). Der det ene direktoratet brukte seks måneder på å bestemme seg for å delta, brukte et annet én time. Udir var den eneste organisasjonen som behandlet spørsmålet om deltakelse på direktørmøtenivå, og som deretter forankret deltakelsen i organisasjonen. Dette forklarer trolig den høye deltakelsen. IMDi og UDI definerte selv mindre, interne utvalg som undersøkelsen kunne være spesielt relevant for. De tre siste direktoratene sendte undersøkelsen til henholdsvis hele organisasjonen (Husbanken og Bufdir) og et større tilfeldig utvalg (Helsedirektoratet) – uten kjent oppfølging fra ledelse eller fagavdelinger. Som vist i tabell 1 ga dette svært varierende utslag. Både Husbanken og Helsedirektoratet har lav svarprosent, mens 59 prosent av utvalget fra Bufdir deltok i undersøkelsen. Høy deltakelse fra Bufdir kan skyldes at alle avdelingsdirektørene til slutt ble informert om undersøkelsen per SMS fordi jeg ikke nådde fram via andre kanaler. Det faktum at alle lederne i organisasjonen dermed kjente til undersøkelsen, kan ha ført til interne diskusjoner eller informasjonsutveksling som jeg ikke har kjennskap til. Helsedirektoratet var den eneste organisasjonen som ga inntrykk av å ha klare rutiner knyttet til henvendelser om deltakelse i undersøkelser. Dette kan tyde på hyppige forespørsler og kanskje en viss undersøkelsestrøtthet i organisasjonen. I hvert fall var det svært få fra Helsedirektoratets utvalg som ønsket å svare på mine spørsmål.

Tabell 1.**Direktorat, intern rekruttering, utvalg, deltakere & svarprosent.**

direktorat	Intern rekruttering	utvalg	deltakere	Svar (%)
Udir	Forankret i direktørmøte. Sendt til alle med potensiell kjennskap	201	128	64
Bufdir	Bred utsendelse uten annen kjent informasjon	200	78	59
Husbanken	Bred utsendelse uten annen informasjon	330	26	8
Helsedirektoratet	Tilfeldig utvalg truffet av egen organisasjon. Ingen intern oppfordring/forankring om deltakelse	230	30	13
IMDi	Sendt til internt definert relevant utvalg med intern oppfordring om deltakelse fra ledernivå	53	20	38
UDI	Sendt til internt definert relevant utvalg med intern oppfordring om deltakelse fra ledernivå	78	29	37
Totalt		1092	311	28

Alle direktoratene fikk tre purringer. Fra studier av responsrater i forbindelse med nettbaserte spørreundersøkelser vet en at hver purring gir uttelling i form av nye deltakere (Fan og Yan 2010). Selv om hver purring ga en viss uttelling i form av nye deltakere, var det likevel interessen etter første utsendelse som ble førende for det totale deltakerantallet fra hver organisasjon.

6.2 Spørreundersøkelsens struktur og operasjonalisering

6.2.1 Struktur

Spørreundersøkelsen består av tre deler. Del én handler om deltakernes siste avsluttede evalueringserfaring. Del to rommer spørsmål knyttet til deltakernes totale evalueringserfaring. Del tre er en ren personaliadel. 311 respondenter besvarte undersøkelsen, men i praksis ble antallet deltakere betydelig redusert som følge av filtrering etter første spørsmål: «Har du deltatt i en eller flere evalueringsprosesser, brukt evalueringsresultater, fått evalueringsresultater presentert eller lest om dem?» For å unngå svar fra deltakere som i realiteten ikke hadde relevant evalueringserfaring, ble deltakerne som svarte NEI på dette innledende spørsmålet, filtrert direkte til undersøkelsens personaliadel. Denne filtreringen førte imidlertid til at ca. 1/3 av respondentene i praksis forsvant ut av undersøkelsen. Dette var et overraskende høyt tall. Basert på egen arbeidserfaring fra direktorat forventet jeg at de fleste byråkratene i det minste ville ment at de hadde vært eksponert for evalueringsresultater.

Forklaringen kan være at spørsmålet er langt og kronglete. Samtidig vet en at respondentene oppfatter større grad av detaljer i spørsmålsutformingen tidlig i en undersøkelse (Elstad 2010). Det er rimelig å anta at det første spørsmålet i et skjema leses relativt grundig og oppfattes av de fleste. Svarfordelingen kan også skyldes ulik forståelse av selve evalueringbegrepet. I testprosessen kom det fram at informantene mine hadde noe ulik forståelse av evalueringbegrepet. De var også usikre på grenseoppgangen mellom evalueringer, utredninger og andre former for oppdragsforskning. Samtidig fikk jeg likevel et inntrykk av at denne usikkerheten snarere resulterte i en inkluderende enn en ekskluderende forståelse av evaluering. Resultatet kan derfor snarere tyde på at flere byråkrater enn jeg hadde forventet, opplever at de har liten befatning med statlig evalueringspraksis. En bekjent i et av de aktuelle direktoratene mente at det i praksis bare var mellom 30 og 50 stykker i vedkomendes organisasjon som hadde noen særlig befatning med evalueringer. Til tross for spørsmålets inkluderende formulering kan det derfor hende at en del av respondentene ikke kjente seg kvalifisert til å svare. Samtidig har de allerede ved å åpne undersøkelsen vist at evaluering ikke er et ukjent fenomen.

Datamaterialet gir mulighet til å undersøke om det finnes en sammenheng mellom deltakernes arbeidsområde og oppgitt evalueringserfaring (tabell 2). Ettersom for eksempel ledere og ansatte innen utredning/forskning kunne antas å ha større kjennskap til evalueringene som

direktoratene til enhver tid er involvert i enn for eksempel saksbehandlere, kan funn som bekreftet en slik tendens tyde på at det første filterspørsmålet fungerte etter hensikten.

Tabell 2. sammenheng mellom arbeidsområde og evalueringserfaring (prosent)

		Ledelse	Adm./øk/ styring	Utredning/ forskning	Saks- behandling	Annet
Evalueringserfaring	JA	90	60	77	60	59
	NEI	10	41	23	40	41

N=311

90 prosent av lederne oppgir at de har evalueringserfaring, og nesten 80 prosent av byråkratene innenfor forsknings-/utredningsfeltet oppgir det samme. Dette er høye tall, men likevel synes jeg det er overraskende at over 20 prosent av dem som jobber innen forskning/utredning, mener at de ikke kjenner til én eneste av evalueringssprosessene deres eget direktorat har vært involvert i. Dette kan tyde på at den høye NEI-andelen i noen grad skyldes spørsmålsutformingen.

Den første delen av undersøkelsen tar sikte på å kartlegge en del generelle trekk ved direktoratenes evalueringsspraksis. Deltakerne svarer ut fra siste evalueringserfaring og besvarer spørsmål om hva som ble evaluert, hvilken form for evaluering som ble gjennomført, hvilken metode som ble brukt, med mer (se vedlegg 5). Deretter filtreres deltakerne på nytt etter spørsmål 13. Det skilles mellom deltakere som antas å ha god kjennskap til den aktuelle evalueringssprosessen, og deltakere med mer perifer kjennskap. Den første gruppa besvarte deretter mer detaljerte spørsmål knyttet til hensikten med evaluering, potensielle brukere av resultatene, organisering, metodevalg, anbudsrunder, samarbeid og konflikter. Disse deltakerne besvarte også spørsmål om evalueringens konsekvenser. Denne filtreringen er gjennomført både for å spare respondenter med mer perifer evalueringserfaring for unødig detaljerte

spørsmål, og for å øke svarenes validitet. De to deltakergruppene slås deretter sammen i spørsmål 47, som handler om å vurdere den samlede kvaliteten på siste avsluttede evaluering.

I del to av undersøkelsen svarer deltakerne ut fra sin totale erfaring med eksterne evalueringer. I spørreundersøkelser går en gjerne fra spørsmål om det generelle til det spesielle, men i denne undersøkelsen valgte jeg å la tidsrekkefølgen være styrende. Tanken var at respondentene først skulle uttale seg om noe konkret, noe de hadde i friskt minne, og dermed få en introduksjon til tematikken før de svarte mer generelt. Del to kartlegger byråkratenes holdninger til evaluering. Holdningsspørsmål er kompliserte fordi spørsmålene kan presse respondenter til å gjøre seg opp en mening om fenomener de ikke har tenkt over tidligere. Dersom respondentene får spørsmål om temaer de kjenner godt fra sitt eget liv, øker sjansene for at de også har gjennomtenkte holdninger (Haraldsen 1999:176). Evaluering er en del av arbeidshverdagen til byråkratene, og jeg antar derfor at evaluering er noe de har en relativt stabil mening om. Gjennom såkalte sammensatte holdningsspørsmål, det vil si spørsmål som fra ulike vinkler kan sies å måle samme grunnholdning, har jeg dessuten forsøkt å få en sikrere indikasjon på byråkratenes holdning til evaluering (Haraldsen 1999:186). Eksempler på dette er spørsmål 66, 67 og 77 (se vedlegg 5) som alle består av utsagn som kan sies å indikere samme grunnleggende syn på evaluering:

Spørsmål 66:

«Evalueringer gir viktig kunnskap om offentlig sektor, som vi ellers ikke ville hatt tilgang til»

Spørsmål 67:

«Jeg mener evaluering bidrar til bedre styring av offentlig sektor»

Spørsmål 77:

«Evalueringer gir et nødvendig eksternt blikk på offentlig sektor»

6.2.2 Operasjonalisering av instrumentell bruk

Ved siden av å kartlegge generelle trekk ved direktoratenes evalueringsaktivitet, samt byråkratenes holdninger til evaluering, ønsket jeg å undersøke i hvilken grad evalueringene fikk konsekvenser. Et av temaene jeg gradvis ble mer opptatt av i denne sammenhengen, var spørsmålet knyttet til evalueringers eventuelle instrumentelle konsekvenser.

Brofoss og Wiig (2006) viser hvordan operasjonalisering av bruksbegrepet påvirker resultatene av statistiske undersøkelser, og at det er nødvendig med en flerdimensjonal forståelse for å fange dette komplekse fenomenet. Jeg tok utgangspunkt i Dahler-Larsens (2000) studie *Surviving the routinization of evaluation: the administrative use of evaluations in Danish municipalities*) og operasjonaliserte *direkte/instrumentell bruk* på fire ulike måter: 1) Evalueringsobjektet (EO) ble satt på den politiske dagsordenen, 2) EO ble justert, 3) EO ble avsluttet/fjernet, 4) EO ble utvidet/videreført. Gjennom ordlyden i spørsmålet understreket jeg at det her handlet om evalueringens eventuelle *direkte/instrumentelle* betydning og konsekvenser for evalueringsobjektet.

I utgangspunktet ønsket jeg å sammenligne egne resultater med resultatene av Dahler-Larsens (2000) studie av bruk av evaluering i danske kommuner. Dette syntes å være det mest relevante sammenligningsgrunnlaget på det aktuelle tidspunktet¹⁰ Testingen av spørreskjemaet førte imidlertid til flere endringer i forhold til det opprinnelige spørsmålsoppsettet. I den danske undersøkelsen *måtte* deltakerne svare JA på ett av svaralternativene *justert, avsluttet eller videreført*. Med andre ord presses deltakerne til å bekrefte en eller annen form for konsekvens. Mine deltakere fikk i stedet fire svaralternativer knyttet til alle formene for bruk, og i tillegg ble spørsmålet om evalueringenes dagsordenfunksjon lagt til (se spørsmål nr 42, vedlegg 5). Noen vil kanskje hevde at evalueringenes dagsordenfunksjon primært handler om symbolsk bruk av evalueringresultater, men i tråd med Saunders' definisjon og informantenes innspill mener jeg at dette er et utfall som «... endrer evalueringsobjektet på en eller annen måte» (Saunders 2012) og derfor hører hjemme under instrumentell bruk.

Jeg mener validiteten av mine funn styrkes av samsvaret mellom kvalitative og kvantitative data og også av de sammenlignbare tendensene i rapporten fra DFØ (2016) og Dahler-Larsens (2000) undersøkelse. Likevel er det vanskelig å trekke helt skarpe skillelinjer mellom ulike former for bruk av evalueringer, og det er også disse begrepene kritisert for gjennom tidligere forskning (Albæk 1988, Henry og Mark 2003, Kirkhart 2000).

¹⁰ DFØs Rapport *Bruk av evaluering i statlig styring* var på det tidspunktet ikke utgitt.

6.3 Metode og data – utfordringer og begrensninger

Her diskuterer jeg begrensninger og utfordringer ved mine valg av metode og ved de dataene metodevalget resulterte i. I denne avhandlingen innebærer dette en diskusjon om både innhenting, behandling og tolking av primærdata i form av spørreskjema og orienteringsintervjuer, samt sekundærdata i form av foreliggende forskning omkring Norges deltakelse i PISA, samt mediedekningen i forbindelse med denne deltakelsen og informasjon fra OECD og norske myndigheter.

6.3.1 Data basert på spørreundersøkelsen

Maybin (2016) kritiserer bruken av survey-undersøkelser og semistrukturerte intervjuer i studier av bruk av forskning i politisk-administrative prosesser. Hun hevder at disse metodene primært bidrar til å fortelle at byråkrater i liten grad gjør nytte av forskning, og dermed ikke klarer å belyse hvorfor politiske dokumenter og prosesser tydelig er bygget på relevant kunnskap. For å forstå hvordan politiske prosesser foregår og beslutninger faktisk fattes, argumenterer Maybin derfor for mer utstrakt bruk av ulike former for dybdestudier (2016:2). Etter min oppfatning er det ikke slik at den ene metoden utelukker den andre. Innenfor evalueringsfeltet, som både i norsk og internasjonal sammenheng framstår med hvite flekker på kartet, burde det være rom for et mangfold av metodiske tilnærminger basert på spesifikke problemstillinger i de ulike studiene. I arbeidet med avhandlingen min mener jeg at spørreundersøkelsen var en velegnet metode for mine spesifikke forskningsinteresser. For det første har undersøkelsen bidratt med en generell oversiktskunnskap om direktoratenes evalueringspraksis. For det andre mener jeg undersøkelsen gir viktig informasjon om evalueringenes innflytelse på norsk politikkutforming.

Det er en fare for at deltakerne i evalueringsprosesser ønsker å rasjonalisere erfaringene sine etter at selve prosessen er avsluttet, og at survey-svarene derfor ikke gjenspeiler en reell situasjon, men i stedet et slags ideal (Albæk 1988, Dahler-Larsen 2000). Samsvaret mellom kvalitative og kvantitative data indikerer imidlertid at jeg både måler det jeg ønsker å måle, og at jeg har dekning for de konklusjonene jeg trekker. Samtidig kan det diskuteres om operasjonaliseringen av *bruk av evalueringsresultater* i undersøkelsen er god nok, og om jeg klarer å skille mellom instrumentell og symbolsk bruk. Også her mener jeg kvaliteten av mine funn styrkes av at evalueringenes direkte konsekvenser også forstås ut fra kvalitative data. Disse dataene viser for eksempel at evalueringer kan få instrumentelle konsekvenser når

direktoratene lurer på hvordan et spesifikt tiltak virker, og om dette tiltaket bør videreføres i samme form eller endres.

Undersøkelsen er basert på deltakelse fra byråkrater i seks ulike direktorater. Direktoratene bestemte selv hvem som skulle få tilbud om å delta, og utvalgene er ikke sammenlignbare. I tillegg er det en relativ overvekt av ledere blant respondentene (17 prosent). Studien er likevel den mest omfattende undersøkelsen av evaluering i statsforvaltningen i senere år og derfor et viktig kunnskapsbidrag på tross av disse svakhetene.

Jeg har prøvd å begrense lengden på undersøkelsen. Skjemaet har maks 79 spørsmål for deltakere som besvarer den lengste versjonen, og maks 69 spørsmål for dem som besvarer den kortere varianten.¹¹ Undersøkelsen tar cirka ti minutter å besvare. Ifølge tidligere forskning er dette en lengde som ikke bør avskrekke deltakerne underveis (Galesic og Bosnjak 2009, Guo mfl. 2016). Tematikken, som er relevant for deltakernes jobbsituasjon, kan også virke positivt både på deltakelse og fullføringsgrad (Haraldsen 1999). På den andre siden kan det hende at deltakerne er lei av å delta på ulike undersøkelser i jobbsammenheng. *Evalueringssamfunnet* er tydelig til stede i norsk arbeidsliv, der alt fra møter og seminarer til arbeidsmiljø fortløpende skal vurderes av de ansatte. Svartaten fra Helsedirektoratet kan peke i en slik retning. Enkelte respondenter har også oppgitt at de er «lei av denne typen undersøkelser» i spørreskjemaet.

I norsk forvaltning praktiseres evaluering ulikt innenfor ulike sektorer, og funnene fra undersøkelsen min er derfor ikke uten videre generaliserbare til hele statsforvaltningen. Samtidig peker funnene i samme retning som resultatene av DFØs undersøkelse, og ettersom denne studien har deltakere fra 9 ulike departementer og 16 underliggende enheter, kan en anta at mine funn også har relevans ut over det sosialpolitiske feltet (DFØ 2016:15). Det kan imidlertid hende at byråkratene i min undersøkelse var spesielt positive til nytten av evaluering fordi de tilhører enheter der evaluering spiller en stor rolle, og fordi evaluering er utbredt innenfor sosialpolitikk. Det er alt i alt vanskelig å avgjøre hva slike forhold har å si for validiteten av min undersøkelse.

Wilensky (1997) hevder at det innenfor korporative demokratier, som de skandinaviske, er større sannsynlighet for at forskningsresultater brukes og påvirker politikktutformingen enn i

¹¹ Totallengden vil variere på grunn av filtrering i forbindelse med oppfølgingsspørsmål.

fragmenterte og desentraliserte stater som USA (s. 1241). Slike forklaringsfaktorer, og det faktum at mine egne resultater og Dahler-Larsens funn peker i samme retning, kan legge grunnlag for hypoteser om lignende sammenhenger og resultater innenfor Norden.

6.3.2 Intervjudata og andre data basert på kvalitative tilnærminger

Metoder er redskaper for å finne svar, skriver sosiologen Karin Widerberg (2001). I løpet av et doktorgradsprosjekt kan både spørsmål og metode endres. I utgangspunktet gjennomførte jeg intervjuer for å bli bedre kjent med et felt jeg manglet kunnskap om, og som også var lite beskrevet i litteraturen. Intervjuene ble dermed kun gjennomført for min egen opplysning. Da jeg analyserte resultatene fra spørreundersøkelsen i forbindelse med arbeidet med artikkel to, oppdaget jeg imidlertid hvordan intervjudataene utfylte resultatene fra spørreundersøkelsen, og at kvalitative og kvantitative data sammen ga den beste beskrivelsen av direktoratenes evalueringsevne. Blant annet fordi intervjudataene bidro til å eksemplifisere og tydeliggjøre resultatene i spørreundersøkelsen. Informantene ga for eksempel klare eksempler på hvordan evalueringresultater ble brukt og satt pris på. Informantene var dessuten opptatt av ulike sider ved forvaltningens evalueringsevne og bidro dermed til et bredere fokus. Mens én var spesielt opptatt av bestillingsprosessen og fortalte mye om denne, dreide en annen samtale seg mye om forholdet mellom bestiller og utføremiljøer.

Informantene ble rekruttert via bekjente og eget nettverk. Dette var en tidsbesparende metode. I tillegg tror jeg denne framgangsmåten bidro til at informantene la mindre bånd på seg for å framstå som korrekte i en forskningssammenheng. Dermed ga disse samtalene det jeg oppfatter som ærlige og tydelige framstillinger av evalueringsevnen i egen organisasjon. Jeg tror også at deres visshet om min begrensede kunnskap gjorde at de i liten grad var opptatt av å tilpasse sin egen framstilling til mine oppfatninger (Jacobsen 2010). Relasjonene mellom oss hadde derfor, slik jeg ser det, ingen klare negative konsekvenser for validiteten av informasjonen som ble gitt.

I tillegg til orienteringsintervjuene består det kvalitative datamaterialet av e-postkontakt med to skoleforskere ved Universitetet i Oslo. Henvendelsene dreide seg henholdsvis om Norges motiver for å delta i PISA-undersøkelsen, og om tips til forskning som kunne bidra med positive argumenter i forbindelse med norsk PISA-deltakelse.

6.3.3 Sekundærdata

Kirkharts (2000) innflytelsesbegrep danner utgangspunktet for min første artikkel. For å belyse dette begrepet ønsket jeg å benytte en case som kunne illustrere betydningen av alle de ulike dimensjonene som knyttes til konseptet. I den sammenhengen var jeg spesielt opptatt av tidsdimensjonen, og dermed falt valget på PISA. Denne casen ga dessuten rikelig tilgang til sekundærdata som kunne illustrere teoretiske poenger.

Artikkelen bygger primært på to bøker som fra ulike sider belyser konsekvensene av PISA-deltakelsen i norsk sammenheng, samt relevant mediedekning og i noen grad informasjon fra norske skolemyndigheter og OECD. Bøkene det er snakk om, er antologien *PISA – sannheten om skolen?* (Elstad og Sivesind, red. 2010) og *Kunnskapsbløffen. Skoler som jukser, barn som gruer seg* (2011) av journalist og forfatter Magnus Marsdal. PISA-antologien er skrevet av noen av Norges fremste skoleforskere og dekker et bredt utvalg perspektiver på norsk PISA-deltakelse, samtidig som den inkluderer mange sentrale stemmer. Marsdals bok bidrar med dybdekunnskap gjennom intervjuer med lærere og ledere i norsk skole, som forteller om sine erfaringer med skolen etter PISA. Selv om bøkene til sammen bidrar med et bredt bilde av PISA-deltakelsens konsekvenser, kan begge bøkene beskrives som overveiende PISA-kritiske. Generelle litteratursøk avdekket også en overvekt av PISA-kritisk litteratur. Ettersom mitt formål var å vise konsekvensene av PISA i bred forstand, tok jeg derfor personlig kontakt med en skoleforsker ved Universitetet i Oslo og etterspurte tips om litteratur som var mer positiv til PISA. De mest relevante henvisningene ble benyttet.

Aviser og andre medier er unike kilder til forståelse av debatter og oppfatninger som preger både samtid og fortid (Østbye mfl. 2013:56). Min bruk av artikler, leserinnlegg og kronikker som datamateriale tok nettopp utgangspunkt i behovet for å illustrere PISA-debatten og holdningene i befolkningen. Samtidig ga mediedekningen også tilgang til andre former for kunnskap, for eksempel i form av uttalelser om Norges PISA-deltakelse fra OECDs ledelse (Helljesen, Grønli og Omvik 2013) Denne typen informasjon illustrerer hvordan media er en viktig kilde i studier av pågående prosesser og aktuelle fenomener. Bruk av artikler og reportasjer som forskningsdata kan imidlertid by på utfordringer fordi innholdet i liten grad kvalitetssikres (Østbye mfl. 2013:56). Jeg har forsøkt å ha et kritisk forhold til mitt mediebaserte datamateriale, blant annet gjennom en samlet vurdering av mediet, avsenderne bak kronikker og leserinnlegg og journalistene som har skrevet saker.

Evalueringers innflytelse er et komplekst fenomen. Gjennom ulike tilnæringsmetoder, data og perspektiver mener jeg denne avhandlingen avdekker deler av denne kompleksiteten. Gjennom å bruke norsk PISA-deltakelse som eksempel, samt eksisterende forskning og dokumenter, viser avhandlingen hvordan evaluering virker over ulike tidsløp og på ulike samfunnsnivåer. Resultatene fra spørreundersøkelsen bidrar på sin side med å gi et oversiktsbilde over direktoratenes evalueringsaktivitet, deres bruk av evalueringresultatene og de holdningene som finnes til evaluering blant byråkratene. Intervjuene supplerer og styrker troverdigheten av survey-dataene gjennom eksempler og utfyllende informasjon.

7. Forskningsetiske betraktninger

Alle forskningsprosjekter må forholde seg til de generelle retningslinjene som finnes innenfor feltet. I dette tilfellet er prosjektet omfattet av NSDs (Norsk senter for forskningsdata) regelverk, og prosjektet er godkjent i tråd med dette (se vedlegg nr. 1). Forskningsetikk dreier seg imidlertid ikke bare om nedskrevne kjøreregler, men også om situasjoner der forskeren selv må finne løsninger på utfordringer som oppstår underveis i prosessen.

Jeg har forsøkt å ha en ryddig og klar dialog med alle som har bidratt i prosjektet mitt – og med dem jeg prøvde å rekruttere, men ikke fikk med meg. Det er viktig både for samfunnets generelle tillit og holdninger til forskning og for tilliten til hvert enkelt forskningsprosjekt at prosessene er ryddige og redelige. I mitt tilfelle medførte det enkelte dilemmaer knyttet til hvor mange ganger en kan spørre enkeltindivider og institusjoner om deltakelse i et forskningsprosjekt før forespørselen blir oppfattet som utidig mas. Jeg opplevde for eksempel at et direktorat aldri svarte på henvendelsene mine før jeg til slutt sendte SMS direkte til alle medlemmene av direktørgruppa. Jeg opplevde dette selv som en litt påtrengende og lite *forskeraktig* måte å ta kontakt på, men samtidig medførte denne strategien både svar og deltakelse. Selv om noen kanskje opplevde denne henvendelsen som masete eller useriøs, fikk jeg ingen negative reaksjoner i etterkant.

Som hovedregel skal forskning foregå ut fra et frivillig, informert samtykke. Hensikten med informert samtykke er først og fremst å beskytte informantene (NESH 2016, s. 14-15). Informantene som deltok i orienteringssamtalene, hadde opprinnelig fått beskjed om at deres innspill kun skulle bidra til å øke min kunnskap om statlig evalueringsvirksomhet. Da jeg gradvis oppdaget at disse intervjuene representerte viktig kunnskap, oppstod dermed en vanskelig situasjon. For det første ønsket jeg hverken å bryte med forskningsmessige idealer eller opptre uryddig overfor egne informanter. For det andre ble jeg usikker på hvordan jeg rent praktisk skulle gå fram for å bruke kvalitative data i artikkelsammenheng. Hvordan skulle jeg informere informantene, og hvordan skulle jeg omtale dette skiftet i min egen doktoravhandling? Jeg kom til at jeg ønsket å bruke sitater fra tre ulike informanter, og sendte e-post for å spørre disse om tillatelse. Jeg fikk en aktiv tillatelse til å bruke intervjuene. Denne avgjørelsen ble basert på intervjuenes jobbrelaterte, faglige karakter. Dersom intervjuene hadde dreid seg om personlige forhold eller sensitiv tematikk, hadde en slik framgangsmåte vært vanskelig både å forsvare og å gjennomføre. Informantene som brukes i artikkelen, ble

rekruttert direkte via meg og kan ikke spores av andre informanter gjennom rekrutteringsprosessen. I tillegg er de fullstendig anonymisert i artikkelen.

8. Oppsummering av artiklene i avhandlingen

Artikkel 1: Evaluation influence – method and societal phenomenon

Forfatter: Øyunn Syrstad Høydal

Språk: Engelsk

Status: Publisert i *Evaluation*

I denne artikkelen undersøker jeg hvordan et tosidig perspektiv på evaluering kan bidra til å forstå Norges deltakelse i PISA og til å forklare de omfattende konsekvensene av denne evalueringen. Dette perspektivet innebærer at jeg på den ene siden anerkjenner PISA som en potensielt nyttig metode med konkrete resultater som får praktiske følger. På den andre siden betrakter jeg PISA som et sosialt fenomen som også spiller en betydelig rolle som symbol på kontroll, styring og vilje til å satse på utdanning.

Jeg tar utgangspunkt i Kirkharts (2000) influence-begrep, som er et forsøk på å romme konsekvensene av evaluering i vid forstand. Kirkhart har en modell av begrepet som vektlegger tre dimensjoner av evalueringers innflytelse: evalueringers virkning over tid, intenderte og uintenderte konsekvenser av evaluering, samt den delen av evalueringen som skal skape endring (det vil si evalueringresultatene eller evalueringprosessen) (Kirkhart 2000:17). Henry og Mark (2003) bygger videre på Kirkharts arbeid og utvider modellen med tre ulike nivåer som påvirkes av evaluering: det individuelle, det mellommenneskelige og det kollektive nivået (Henry og Mark 2003:298). Basert på dette arbeidet utvikler jeg en modell eller et analytisk rammeverk som bidrar til å synliggjøre de omfattende følgene av PISA-evalueringen langs flere akser og på ulike samfunnsnivåer. Disse følgene diskuterer jeg ut fra Dahler-Larsens (2012) begrep *evalueringssamfunnet*, som viser til evaluering som et samfunnsfenomen og et symbol på samfunnets overordnede normer og verdier. Og jeg viser hvordan evaluering påvirker politiske beslutninger og samfunnsliv, men også at evaluering blir påvirket av samfunnet rundt.

Artikkelen bygger på sekundærdata i form av foreliggende forskning, mediedebatt og dokumenter.

Artikkel 2:

Tittel: Tracking the Instrumental Use and Motives for Evaluations
Forfatter: Øyunn Syrstad Høydal
Språk: Engelsk
Status: Under vurdering i *American Journal of Evaluation*

Hensikten med denne artikkelen er å undersøke om evalueringer har direkte politiske konsekvenser, og om slike virkninger kan forklares ut fra organisasjonsmessig kontekst og motivene bak evalueringsevirsomheten (Højlund 2014). Artikkelen er basert på data fra spørreundersøkelsen blant ansatte i seks norske direktorater, samt intervjuer. Jeg tar utgangspunkt i evaluering som et administrativt verktøy, motivert av byråkratiets behov for rask og praktisk nyttig kunnskap (Dahler-Larsen 2000) og betrakter evaluering som en form for samproduksjon, eller «co-production» (Jasanoff 2004), av kunnskap mellom byråkrater og forskere/konsulenter. Ut fra dette utvikler jeg en modell som illustrerer sammenhengen mellom organisasjoners motiver for evaluering, evalueringsprosessen, resultatene og bruken av disse.

Direktoratenes rolle skaper et behov for stadig tilfang av relevant og oppdatert kunnskap. Direktoratene ønsker reell kunnskap, og de ønsker evalueringsmiljøer som er engasjerte, kunnskapsrike og i *noen* grad utfordrer deres egne ideer og virkelighetsforståelse. Samtidig er evalueringsrapportene i deres øyne et arbeidsverktøy, og dermed ønsker de anbefalinger og konklusjoner som kan brukes og komme til nytte innenfor rammene av eksisterende politikk og budsjetter. Nettopp fordi de ønsker seg nyttige evalueringsrapporter, er byråkratene aktive deltakere i evalueringsprosessen gjennom utforming av oppdraget, innspill til case og metoder og oppfølging av utkast. Derfor gir det større mening å betrakte evalueringsprosessen som en form for samproduksjon av kunnskap mellom forskere og byråkrater, enn å betrakte dette fra et modernistisk perspektiv der kunnskapsbruker og kunnskapsprodusent er strengt adskilte og preges av ulike idealer og normer. Direktoratenes engasjement i evalueringsprosessen resulterer i kunnskap som er nyttig og relevant fra byråkratenes ståsted. Dermed blir evalueringresultatene også brukt og får instrumentelle konsekvenser.

Artikkel 3:

Tittel: Evaluating social policies – do methodological approaches determine the policy impact?
Forfatter: Øyunn Syrstad Høydal og Anne Grete Tøge
Språk: Engelsk
Status: Innsendt til *Evidence & Policy*

Evidensbevegelsens popularitet har medført et nokså ensidig fokus på instrumentell bruk av evalueringresultater framskaffet via eksperimenter, metastudier og kvantitative tilnærminger. Til tross for at evidensbevegelsens metodologiske idealer har vært gjenstand for betydelig kritikk de seneste årene, har en i svært liten grad vært opptatt av å undersøke hvilke metodiske tilnærminger som vurderes som mest hensiktsmessige fra byråkratenes perspektiv. En har heller ikke vært opptatt av å undersøke den opplysende virkningen av all kunnskapen som har blitt produsert (Weiss 1977).

Artikkelen bygger på data fra spørreundersøkelsen blant norske direktoratsansatte. Sammen med medforfatter Anne Grete Tøge, undersøker jeg hvordan byråkratenes kunnskapsbehov stemmer overens med evidensbevegelsens idealer. Resultatene viser at direktoratene primært evaluerer for å skaffe ny kunnskap. Det vil si at evalueringenes eventuelle instrumentelle konsekvenser er av mindre betydning. Byråkratene oppgir at evalueringer basert på kvalitative metoder i størst grad bidrar til nye perspektiver og ideer. Både observasjoner og workshops/møter/konferanser er viktige i så måte. Resultatene viser betydningen av den sosiale dimensjonen ved læring, og hvordan mennesker sammen skaper kunnskap.

De generelt positive holdningene til evalueringenes nytteverdi, forklares som en konsekvens av direktoratenes rolle i det politisk-administrative systemet og dermed kontinuerlige behov for ny kunnskap. Holdningene sees også som en mulig konsekvens av idealet om kunnskapsbasert politikktutforming og *evalueringssamfunnet* (Dahler-Larsen 2012).

9. Avsluttende drøfting

I denne avhandlingen undersøker jeg om evalueringer har betydning for norsk politikk og samfunnsliv, og hvordan denne innflytelsen kan forstås. Min forståelse av evaluering bygger på et tosidig perspektiv. På den ene siden mener jeg at evaluering er et verktøy med potensielt samfunnsnyttige konsekvenser, på den andre siden handler evaluering også om å framstille politikk og forvaltning i lys av samfunnets verdier. Når norske direktorater evaluerer, kan denne aktiviteten både forklares ut fra et behov for bedre styring og mer kunnskap, men også ut fra et ønske om å framstå som rasjonelle, effektive og kunnskapsbaserte. Det er rimelig å anta at disse to hensiktene vanskelig lar seg skille fra hverandre, og at denne dualiteten også bidrar til evalueringens sterke posisjon i forvaltningen og samfunnet for øvrig.

I tråd med sosialkonstruktivistiske eller sosiologiske kunnskapssyn forstås direktoratenes evaluering i en kulturell og historisk kontekst. Norsk statsforvaltning har i hele perioden etter andre verdenskrig vært preget av modernismens tro på vitenskapelig basert kunnskap som en ressurs i politisk sammenheng. Det er også rimelig å hevde at norske myndigheter i praksis har brukt evaluering som en kunnskapskilde i store deler av denne perioden. Dermed er dagens evalueringsaktivitet en videreføring av en langvarig forvaltningstradisjon.

Norske direktorater har en helt sentral posisjon i dagens kunnskapsbaserte politikkutforming, både som faglige rådgivere og som ansvarlige for oppfølging av statens politikk. Dersom direktoratene skal oppfylle disse oppgavene, er de avhengige av et kontinuerlig tilfang av ny og relevant kunnskap. Fordi kunnskapsbehovet ofte er knyttet til spesifikke satsinger eller tiltak, preget av kort tidshorison og bestemte perspektiver, er evalueringer og annen oppdragsforskning nødvendige bidrag (Coleman 1972). Selv om det finnes praktiske grunner for direktoratenes evalueringsvirksomhet, spiller også evalueringenes symbolske betydning inn. Og denne symbolske betydningen kan ikke forstås løsrevet fra den konteksten der evalueringene iverksettes, gjennomføres og brukes. Selv om evalueringshistorien viser hvordan evaluering som symbol kan forstås i lys av internasjonale epoker av modernitet, vil meningen med evaluering i et bestemt samfunn være et resultat av slike internasjonale strømninger og nasjonale eller lokale verdier (Schwandt 2009).

Ved å evaluere framstår direktoratene som veldrevne og effektive organisasjoner, i tråd med dagens styringsidealer. Evaluering handler også om tilslutning til evidensbevegelsen og ideen om kunnskapsbasert politikkutforming. Ved å delta i internasjonale studier som PISA viser

norske myndigheter at norsk skolepolitikk er basert på objektiv kunnskap, og at norske styresmakter har kontroll over dette viktige politiske området. Samtidig viser survey-dataene at norske byråkrater ikke deler denne entusiasmen for evidensbevegelsens metodologiske idealer. Tvert imot mener byråkratene at kunnskap basert på kvalitative metoder i størst grad bidrar til nye ideer og perspektiver (se artikkel 3).

Selv om idealet om kunnskapsbasert politikktutforming har preget norsk forvaltning i mange tiår, er evaluering noe annet i dag enn for 30 eller 50 år siden (Lindgren 2014).

Evalueringsvirksomheten både i offentlig sektor og i samfunnet for øvrig er mye mer omfattende, og det faktum at nær sagt alle sider av tilværelsen er gjenstand for evaluering, gjør det rimelig å omtale Norge som et evalueringssamfunn (Dahler-Larsen 2012). Det vil si et samfunn der evaluering er en selvfølge, en forventet rutine. Forventningen om evaluering kan i seg selv betraktes som en konsekvens av evaluering, og viser dermed også hvor omfattende innflytelsen fra evaluering kan være i siste instans. Men hva gjør slike forventninger med oss? Hva er resultatene av et slikt samfunn? Uavhengig av hvilke metodiske idealer og praktiske formål som ligger bak samfunnets evalueringsvirksomhet, handler evaluering om måling, vurdering og en sammenligning med et ideal. Noe er bedre enn noe annet, noe er dårligere. Samfunnet må gjøres målbart og tilpasses de kategoriene evalueringer tilbyr. De verdiene eller resultatene som ikke kan måles, står dermed i fare for å miste verdi og betydning. Samtidig kan evalueringssamfunnet også tolkes som et uttrykk for en grunnleggende mangel på tillit (Power 1997) og en mangel på tiltro til profesjonelle grupper kunnskap og erfaringer (Lindgren 2014). Et system basert på kontinuerlig kontroll er både arbeids- og kostnadskrevede, og dette er temaer som etter min oppfatning oftere burde diskuteres i forbindelse med offentlig evaluering.

I tråd med den pragmatiske institusjonalismens åpenhet for nyanser og kompleksitet har evalueringers innflytelse blitt undersøkt fra tre ulike perspektiver i artiklene i denne avhandlingen. I *Tracking the Instrumental Use and Motives for Evaluations* tar jeg utgangspunkt i den modernistiske forståelsen av evaluering som et verktøy med instrumentelle konsekvenser for politisk-administrative beslutningsprosesser. I *Evaluating social policies – do methodological approaches determine the policy impact?* Utvider Anne Grete Tøge og jeg perspektivet og undersøker i hvilken grad byråkratene opplever evalueringene som nyttige når evalueringene ikke bidrar til direkte endring etter evidensbevegelsens idealer, men i stedet fungerer som en kilde til opplysning eller kunnskap

(Weiss 1977). I tillegg undersøker vi om evidensbevegelsens metodologi er det beste utgangspunktet for kunnskap som skal brukes i det politisk-administrative systemet. I *Evaluation – method and societal phenomenon* undersøker jeg PISAs innflytelse på norsk skolepolitikk og samfunnsliv i bred forstand, fra en tosidig evalueringsperspektiv.

9.1. Innflytelse gjennom instrumentell bruk

Både evalueringstilhengere og -skeptikere har vært opptatt av de instrumentelle konsekvensene av evaluering. Det er denne formen for direkte nytte evalueringssiden hviler på, og det er slik bruk tilhengerne leter etter og skeptikerne betviler. Evidensbevegelsens popularitet har dessuten bidratt til et fornyet fokus på kunnskapens instrumentelle konsekvenser de siste årene.

Avhandlingen viser at byråkratene innenfor det sosialpolitiske feltet mener at evalueringresultatene får praktiske konsekvenser i politisk-administrative beslutningsprosesser. Jeg argumenterer for at disse funnene må forstås i lys av direktoratenes rolle og hensikten med evalueringene. Det er vanskelig å vite om satsinger på anti-mobbearbeid i skolen eller på bedre oppvekstmiljø og levekår i Groruddalen¹² fungerer som planlagt, uten at dette undersøkes skikkelig. Det er også komplisert å vite hva som er *sannheten* innenfor en gitt politisk debatt, når partene som deltar, presenterer rapporter der konklusjonene spriker i ulike retninger – i tråd med egne standpunkter. I slike tilfeller er det naturlig at direktoratene søker sin egen sannhet i form av evalueringsrapporter, og at disse rapportene brukes og får direkte konsekvenser i politisk-administrative beslutningsprosesser. Selv om resultatene får innflytelse på denne måten er det likevel usikkert om denne innvirkningen er av en veldig omfattende karakter. Snarere er det slik at kunnskap som passer inn med gjeldende politiske oppfatninger og budsjetter har de største mulighetene for instrumentell innflytelse (Weiss 1998).

Konklusjonene i en evalueringsrapport vil variere i tråd med hvilke rammer som legges for oppdraget, og dette er et viktig aspekt i evalueringssammenheng (Dahler-Larsen 2012, Mathison 2018). Mens det ut fra byråkratens perspektiv vil være nyttig å gi klare rammer for evalueringoppdraget, delta aktivt med forslag og råd underveis i evalueringprosessen, samt

¹² <https://www.regjeringen.no/no/aktuelt/103-skoler-reduerte-mobbing/id2617588/>
<https://www.regjeringen.no/no/tema/kommuner-og-regioner/by--og-stedsutvikling/omradeloft1/groruddalssatsingen/id2344376/>

følge opp tidlige rapportutkast, kan denne formen for samproduksjon (Jasanoff 2004) oppfattes som mer problematisk av forskerne eller konsulentene som gjennomfører evalueringoppdraget. Detaljerte oppdragsrammer kan sees som uelegante forsøk på å styre prosessen, mens innspill underveis oppleves som utidig innblanding. Slik den ferdige rapporten kan betraktes fra et maktperspektiv, kan også selve evalueringprosessen sees med slike briller (Dahler-Larsen 2006).

9.2. Innflytelse gjennom opplysende bruk

Evaluerings innflytelse gjennom deres mer generelle kunnskapsbidrag eller opplysende effekt (Weiss 1977) undersøkes i artikkelen *Evaluating social policies – do methodological approaches determine the policy impact?* Resultatene av spørreundersøkelsen viser at byråkratene verdsetter evalueringene som et bidrag til ny kunnskap, nye perspektiver og ideer. Våre funn viser dermed – stikk i strid med evidensbevegelsens instrumentelle fokus – at evaluering er en viktig kunnskapskilde i direktoratene også når de ikke brukes instrumentelt. Funnene indikerer dermed et system der vitenskapelig basert kunnskap ansees som en ressurs også i et mer generelt eller langsiktig perspektiv.

Opplysende bruk (Weiss 1977) av kunnskap handler på den ene siden om umiddelbar læring eller mottak av kunnskap og informasjon. På den andre siden viser begrepet til langvarig kunnskapsakkumulasjon, kollektiv holdningsendring og eventuelle politiske endringer som en følge av dette. I denne avhandlingen er det den mest umiddelbare formen for opplysning som undersøkes, eller hva Vedung (2009) omtaler som opplysning på første nivå.

Våre funn viser at evalueringer som gjennomføres for å skaffe ny kunnskap også i størst grad bidrar til nye ideer og perspektiver. Kanskje har evalueringer som gjennomføres med dette formålet for øye, et mer åpent og kritisk perspektiv enn evalueringer som gjennomføres for instrumentell bruk (Weiss 1977). Undersøkelsen viser også at respondentene mener observasjoner, workshops og konferanser er de metodene som i størst grad bidrar til slike nye ideer og perspektiver. I slike situasjoner, som i begrenset grad styres av forskere eller andre, skapes nye løsninger og uventede konklusjoner (Silverman 2013). Dette viser betydningen av sosiale dimensjoner når læring og kunnskapens rolle i politikken diskuteres (Freeman og Sturdy 2014).

I denne artikkelen undersøker vi ikke evalueringenes opplysende virkning over tid. Likevel er det rimelig å forvente at evalueringsrapportene spiller en betydelig rolle i den langsiktige

kunnskapsakkumulasjonen i direktoratene. Evalueringsrapportene er i stor grad skapt på direktoratenes premisser. Dermed representerer rapportene perspektiver, definisjoner og begreper som gjenspeiler direktoratene interesser. Rapportene inneholder dermed en form for internt anerkjent eller sikker kunnskap, som er av stor betydning i byråkraters arbeidshverdag (Maybin 2016). Kunnskap i skriftlig form har dessuten en helt sentral rolle i det politisk-administrative systemet fordi denne kunnskapen er stabil, enkel å spre og reproducere (Freeman og Sturdy 2014). Når evalueringsrapportene dessuten produseres i digital form, er evalueringresultatene bare et tastetrykk unna. Slik spres kunnskap og akkumuleres i systemet og bidrar til den kontinuerlige læringsprosessen som finner sted (Bemelmans-Videc og Eriksen 1994). På sikt kan kunnskap fra disse evalueringene og de ideene rapportene representerer, dermed bidra til å endre holdninger og politikk (Smith 2013).

9.3. Innflytelse – i bred forstand

For å vise hvorfor det er meningsfullt og viktig å tenke bredt om evalueringers innflytelse, bruker jeg konsekvensene av Norges flerårige PISA-deltakelse som eksempel i artikkel én. Selv om PISA både gjennom budsjetter, varighet og innflytelse representerer et ytterpunkt i direktoratenes evalueringsevne, er det et eksempel som kan bidra til å undersøke lignende og mindre åpenbare konsekvenser i kjølvannet av mindre evalueringer. For også andre evalueringer kan påvirke enkeltindivider og samfunnsgrupper lenge etter at politikere, myndigheter og forskere har mistet interessen. Det er videre rimelig å anta at ikke bare PISA har virkninger som hverken var planlagt eller forutsett. Og akkurat som PISA kan også andre evalueringer både forstås som et verktøy og et symbol på samfunnets overordnede verdier og idealer.

PISA har medført politiske avgjørelser som har resultert i konkrete endringer i norsk skole. Disse endringene innebærer både et testbasert skolesystem og strukturer som sørger for at dette systemet fungerer. PISA har også medført mindre håndfaste endringer, blant annet i form av en ny kollektiv diskurs om skole og utdanning (Elstad 2010). Gjennom PISA har også selve skoledagen til norske barn blitt endret. Mer testing, trening før testing og nye faglige prioriteringer påvirker mestringsfølelse og læringsutbytte og vil dermed følge disse barna resten av livet. Samtidig, uavhengig av bekymrede forskere og mistroiske foreldre, er den testbaserte skolen den eneste skolen disse barna kjenner. Denne erfaringen medfører kanskje at deres generasjon ser PISA og lignende tester i et annet lys enn dagens kritikere, som vokste opp med en annen skolehverdag. På samme måte som evaluering skaper nye

forståelser av sosiale fenomener, bidrar også samfunnet rundt til å endre effektene av evaluering (Dahler-Larsen 2012).

PISA har hatt en enorm definisjonsmakt. Da vi gikk og la oss, hadde vi verdens beste skole, da vi våknet, var vi middelmådige. På den ene siden kan det hevdes at PISA bidro til en realitetsorientering i det politisk-administrative systemet og samfunnet for øvrig. På den andre siden hadde ingenting endret seg over natten. Det eneste som endret seg, var vår oppfatning og vårt bilde av skolen. PISA er et kraftfullt symbol. Gjennom sin utbredelse, internasjonale organisering og eierskap, samt objektive grafer, representerer PISA nøytral og sikker viten, eller *sannheten om skolen* (Elstad og Sivesind 2010).

9.4. Oppsummert

Til tross for at evalueringsforskningen har lett etter bruk av evalueringsresultater i mer enn 60 år, er det få entydige svar om hvilken innflytelse evaluering har i samfunnet. En av årsakene kan være både forskere og forvaltnings store forventninger til evaluerings nytteverdi.

Dersom målet med evaluering er sosial forandring er fallhøyden stor og resultatene vanskelige å finne. I de sosialpolitiske direktoratene er forventningene mer moderate. Her er evaluering først og fremst et nyttig arbeidsverktøy, og avhandlingen viser at byråkratene mener disse evalueringene både bidrar til læring, nye ideer og politisk-administrative avgjørelser.

Forvaltningens interesse for praktisk nyttig kunnskap gjenspeiler en lang tradisjon, og ifølge Tornes (2012:81) har dette fokuset også preget norsk samfunnsforskning. Dermed er det to parter som drar i samme retning.

For å forklare evalueringens posisjon i byråkratiet og samfunnet for øvrig, er det også viktig å ta hensyn til evalueringens symbolske betydning. Når norske direktorater bestiller kunnskapsoversikter, skaffer de seg samtidig autoritet og kontroll over et bestemt politikkområde. Evalueringer som på den ene siden bidrar med praktisk nyttig kunnskap om konsekvensene av statlige tiltak, bidrar på den andre siden til å vise at norsk politikk er basert på kunnskap og ikke verdier. Men evaluering handler om verdier.

Verdiene kommer til syne gjennom direktoratenes valg av evalueringsobjekt og fortsetter videre gjennom ulike former for innspill, deltakelse og kontroll underveis i evalueringsprosessen. Det er dermed ikke sagt at evalueringsrapporter er diktert av oppdragsgiveren, at de er uredelige eller ikke kan bringe frem nye perspektiver. De

representerer et perspektiv på verden, og dette perspektivet er basert på direktoratenes, eller byråkreatiets, egne ønsker og behov.

Slik bidrar evalueringer i stor grad til å besvare det politisk-administrative systemets egne spørsmål, innenfor dets egne rammer. Dette resulterer i kunnskap som kan kritisere resultatene eller gjennomføringen av offentlige programmer, men som ikke stiller spørsmålsteget ved selve systemet (Almås 1990, Mathison 2018). Derfor er det også i realiteten vanskelig å avgjøre hvor evalueringenes innflytelse stopper, og hvor grensen går mellom kunnskap og verdier, evaluering og politikk. Når evalueringensresultater er i samsvar med det politiske regimets ideer og verdivurderinger, kan en imidlertid hevde at evalueringer bidrar med demokratisk kunnskap. Hadde det vært motsatt, hadde vi vært på vei mot ekspertveldet.

10. Avhandlingens bidrag

Denne avhandlingen bidrar først og fremst med ny kunnskap om evaluering blant norske direktorater innenfor det sosialpolitiske feltet, og om hvilken innflytelse disse evalueringene har. Evalueringenes innflytelse undersøkes fra tre ulike perspektiver, og avhandlingen viser at evalueringene brukes på ulike måter og har ulike former for konsekvenser.

I motsetning til hva som hevdes i en del litteratur, synes det som om direktoratenes evalueringer har instrumentelle følger. Byråkratene mener likevel at evalueringenes viktigste rolle er å bidra med ny kunnskap. Dette funnet viser at evidensbevegelsen og deler av evalueringforskningens ensidige fokus på instrumentell bruk av kunnskap bør utfordres og diskuteres. I artikkel tre undersøker også medforfatter Anne Grete Tøge og jeg hvordan byråkratene oppfatter nytten av ulike metodiske tilnærminger i evalueringssammenheng. Her finner vi at byråkratene ikke deler evidensbevegelsens oppfatning om at eksperimenter og kvantitativt baserte data nødvendigvis bidrar med den mest nyttige kunnskapen sett fra forvaltningens ståsted. Til tross for at hele evidensbevegelsen hviler på nettopp et helt bestemt metodologisk fundament, har en i liten grad undersøkt hvordan dette idealet oppleves i praksis. Artikkelen bidrar derfor med ny og viktig kunnskap på et felt der en bestemt metodologi nærmest er tatt for gitt, til tross for at evidensbevegelsen selv er basert på ideen om evidens i stedet for ideologi.

Avhandlingen viser også at det er behov for mer innsikt om evalueringers samlede innflytelse på ulike samfunnsnivåer over tid. Dette er et felt det finnes lite kunnskap om, samtidig som artikkel én i denne avhandlingen viser hvordan en evaluering potensielt sett kan ha betydelig innflytelse på både politikk og samfunnsliv. Gjennom et tosidig perspektiv på evaluering bidrar dessuten artikkelen til å vise at evalueringers innflytelse også er knyttet til deres betydning som samfunnsfenomen og sosiale symboler.

Avhandlingen bygger på et tosidig perspektiv på evaluering, som dels forstås som en metode og et potensielt nyttig verktøy, og dels som et sosialt fenomen. Dette perspektivet er i seg selv et bidrag til et forskningsfelt som har blitt kritisert for å være normativt og unyansert.

11. Framtidig forskning

Utgangspunktet for denne avhandlingen var mangelen på forskningsbasert kunnskap og oversikt over de fleste aspekter ved statlig evalueringsvirksomhet. Denne avhandlingen er et skritt på veien mot bedre innsikt i dette feltet. Samtidig gjenstår det mange områder å belyse og spørsmål å besvare.

Selv om avhandlingen avdekker en overveiende positiv holdning til evaluering, og et inntrykk av at evalueringsresultatene kommer til praktisk nytte i det politisk-administrative systemet, går den ikke i dybden av disse resultatene. Det er behov for forskning som går bak disse funnene og undersøker hvordan det politisk-administrative systemet i spesifikke tilfeller faktisk bruker evalueringsresultater, og hvilke konsekvenser dette får for norsk politikkkutforming. På samme måte er det behov for en dypere innsikt i hvordan kunnskap fra evalueringer spres i byråkratiet, og hvilke konsekvenser denne kunnskapsakkumuleringen får. Framtidig forskning bør også gå grundigere inn i evalueringsprosesser og undersøke premisene for kunnskapen som skapes, og det praktiske arbeidet og avveiningene som leder fram til en evalueringsrapport.

Denne avhandlingen bygger på byråkratenes synspunkter. Det er også behov for mer kunnskap om forskernes rolle og praktiske innsats på dette feltet. Hvordan er det å være oppdragsforsker på jobb for norske myndigheter? Ifølge Furre og Horrigmo (2013), kan det være utfordrende å levere fra seg evalueringskonklusjoner som ikke er i tråd med oppdragsgivernes virkelighetsforståelse. Hvordan preger denne vissheten kvaliteten på norske evalueringsrapporter og forskernes stemmer i et marked der relasjoner og godt rykte kan forventes å ha en viss betydning for senere oppdrag?

Selv om evaluering både internasjonalt og i Norge har vært tett forbundet med det sosialpolitiske feltet, ville det også være interessant å undersøke hvordan evaluering brukes i andre deler av statsforvaltningen. Medfører en lavere evalueringsfrekvens innenfor disse politikkområdene at de evalueringene som gjennomføres, er mer omfattende og får mer tyngde? Og dersom disse evalueringene har større innflytelse, hvordan blir de til?

Sist, men ikke minst er det betimelig å spørre se hva den neste evalueringsbølgen blir. Dersom evidensbølgen er på retur, i tråd med Smith og Haux (2017) - Hva blir den femte bølgen?

Referanser

- Albæk, E. (1988). *Fra sandhed til information: Evalueringsforskning i USA - før og nu*. København: Akademisk forlag.
- Alkin, M. C., & King, J. A. (2017). Definitions of evaluation use and misuse, evaluation influence, and factors affecting use. *American Journal of Evaluation*, 38(3), 434-450.
- Alkin, M. C., & Taut, S. M. (2003). Unbundling evaluation use. *Studies in Educational Evaluation*, 29(1), 1-12.
- Almås, R.(1990). *Evaluering på norsk. Ei innføring i vurdering av prosjektarbeid og handlingsretta forskning*. Oslo: Universitetsforlaget.
- Askim, J., Døving, E. & Johnsen, Å. (2013). Evalueringspraksis i den norske statsforvaltningen 2005-2011. I A. Halvorsen, E. Lier-Madsen, & N. Jentoft (Red.), *Evaluering. Tradisjoner, praksis, mangfold* (s. 72-88). Bergen: Fagbokforlaget.
- Azzam, T. & Jacobsen, M. (2015). Reflections on the Future of Research on Evaluation. I P. R. Brandon (Red.), *Research on evaluation: New Directions for Evaluation*, 148, 103-116.
- Baklien, B. (1985). Evalueringsforskeren mellom barken og veden. *Tidsskrift for samfunnsforskning*, 26, 457-475.
- Baklien, B. (1993)a. Evalueringsforskningens historie og utvikling i Norge. *Politica*, 25(1), 56-62.
- Baklien, B. (1993)b. Evalueringsforskning i Norge. *Tidsskrift for samfunnsforskning*, 34, 261-374.
- Bemelmans-Videc, M. L., & Eriksen, B. G., Edie N. (1994). Facilitating Organizational Learning: Human Resource Management and Program Evaluation. I F. L. Leeuw, R. C. Rist, & R. C. Sonnichsen (Red.), *Can governments learn?: Comparative perspectives on evaluation & organizational learning*. New Brunswick: Transaction Publishers.

- Berger, P. og Luckmann T. (1976). *Den samfundsskabte virkelighed: en videnssociologisk afhandling*. København: Lindhardt & Ringhof.
- Boaz, A., Grayson, L., Levitt, R., & Solesbury, W. (2008). Does evidence-based policy work? Learning from the UK experience. *Evidence & Policy: A Journal of Research, Debate and Practice*, 4(2), 233-253.
- Bogenschneider, K., & Corbett, T. J. (2011). *Evidence-based policymaking: Insights from policy-minded researchers and research-minded policymakers*. New York: Routledge.
- Boswell, C. (2009). *The political uses of expert knowledge: Immigration policy and social research*. Cambridge: Cambridge University Press.
- Boswell, C. (2015). Cultures of Knowledge Use in Policymaking: The Functions of Research in German and UK Immigration Policy. I P. Scholten, H. Entzinger, R. Penninx & S. Verbeek (Red.), *Integrating Immigrants in Europe* (s. 19-38). Cham: Springer.
- Botheim, I., & Solumsmoen, D. (2009). Direktoratene: rolleblanding eller nyttig rollekombinasjon? *Stat & Styring*, 19(01), 8-9.
- Braeken, J. (2014, 27. oktober). Et nyansert syn på PISA. *Faculty of Educational Sciences, University of Oslo*. Hentet fra <https://issuu.com/shanecol/docs/forsknings-bro-web?e=8763060/9855158>
- Breidahl, K. N., Gjelstrup, G., Hansen, H. F., & Hansen, M. B. (2017). Evaluation of Large-Scale Public-Sector Reforms: A Comparative Analysis. *American Journal of Evaluation*, 38(2), 226-245.
- Bromark, M. (2014, 12. desember). Voksne ville blitt stresset av ungdoms skolehverdag. *Klikk.no*. Hentet fra <http://www.klikk.no/foreldre/skolebarn/ungdom-og-stress-1488938.ece>
- Brofoss, K. E., & Wiig, O. (2006). *Bruk av forskning: Fasettert begrep – mange forklaringsmodeller*. Oslo: NIFU STEP.
- Cain, T. (2015). Teachers' engagement with research texts: Beyond instrumental, conceptual or strategic use. *Journal of education for teaching*, 41(5), 478-492.

- Campbell, J. L., & Pedersen, O. K. (2014). *The national origins of policy ideas: Knowledge regimes in the United States, France, Germany, and Denmark*. Princeton: Princeton University Press.
- Caplan, N. (1979). The two-communities theory and knowledge utilization. *American behavioral scientist*, 22(3), 459-470.
- Carden, F. & Alkin, M.C. (2012). Evaluation roots: An international perspective. *Journal of MultiDisciplinary Evaluation* 8(17): 102–118.
- Christensen, J., Gornitzka, Å, & Holst, C. (2017). Knowledge regimes in the Nordic countries. I O. Knudsen (Red.), *The Nordic Models in Political Science: Challenged, But Still Viable?* (s. 239-252). Oslo: Fagbokforlaget.
- Christensen, T., Egeberg, M., Lægreid, P., & Aars, J. (2014). *Forvaltning og politikk* (4. utg.). Bergen: Universitetsforlaget.
- Cohen, M. D., March, J. G., & Olsen, J. P. (1972). A garbage can model of organizational choice. *Administrative science quarterly*, 1-25.
- Coleman, J. S. (1972). *Policy research in the social sciences*. Morristown: General Learning Press.
- Contandriopoulos, D., & Brousselle, A. (2012). Evaluation models and evaluation use. *Evaluation*, 18(1), 61-77.
- Coryn, C. L., Wilson, L. N., Westine, C. D., Hobson, K. A., Ozeki, S., Fiekowsky, E. L., & Schröter, D. C. (2017). A decade of research on evaluation: A systematic review of research on evaluation published between 2005 and 2014. *American Journal of Evaluation*, 38(3), 329-347.
- Cousins, J. B., & Leithwood, K. A. (1986). Current empirical research on evaluation utilization. *Review of educational research*, 56(3), 331-364.
- Dahler-Larsen, P. (2000). Surviving the routinization of evaluation: the administrative use of evaluations in Danish municipalities. *Administration & society*, 32(1), 70-92.

- Dahler-Larsen, P. (2006). *Evaluering og magt*. Århus: Magtutredningen.
- Dahler-Larsen, P. (2012). *The evaluation society*. Stanford, CA: Stanford University Press.
- Dahler-Larsen, P. (2013). Evaluering i Norge set udefra. I A. Halvorsen, E. Lier-Madsen, & N. Jentoft (Red.), *Evaluering. Tradisjoner, praksis, mangfold* (s. 43-57). Bergen: Fagbokforlaget.
- Davies, H. T., Nutley, S. M., & Smith, P. C. (2000). *What works?: Evidence-based policy and practice in public services*. Bristol: The Policy Press.
- Daviter, F. (2015). The political use of knowledge in the policy process. *Policy Sciences*, 48(4), 491-505.
- Deterding, N. M., & Solmeyer, A. R. (2018). Building Evidence in Challenging Contexts: Introduction to the Special Section. *American Journal of Evaluation*, 39(1), 24-26.
- Direktoratet for forvaltning og ikt (2013, 19. november). *Merverdi eller unødig omvei? Om direktoratenes rolle i gjennomføring av nasjonal politikk* (Difi-rapport 2013:11). Hentet fra <https://www.difi.no/rapport/2014/03/merverdi-eller-unodig-omvei-om-direktoratenes-rolle-i-gjennomforingen-av-nasjonal-politikk>
- Direktoratet for økonomistyring (2016, 28. september). *Bruk av evaluering i statlig styring* (Rapportnr.: R9392). Hentet fra <https://evalueringsportalen.no/evaluering/bruk-av-evaluering-i-statlig-styring/Rapport%20DF%C3%98%20Bruk%20av%20evaluering%20i%20statlig%20styring.pdf/@;inline>
- Direktoratet for økonomistyring (2018, 3. oktober). *Økonomiregelverket i staten*. Hentet fra <https://dfo.no/fagomrader/ekonomiregelverket>
- Duffy, D. N. (2017). *Evaluation and Governing in the 21st Century: Disciplinary Measures, Transformative Possibilities*. Hampshire: Palgrave Macmillan.
- Eberli, D. (2016). Tracing the use of evaluations in legislative processes in Swiss cantonal parliaments. *Evaluation and program planning*, 69, 139-147.
- Eide, K. (1999). *Hvor kommer utdanningspolitikken fra?* Oslo: Norsk institutt for studier av forskning og utdanning.

- Elstad, E. & Sivesind, K. (Red.). (2010). *PISA – sannheten om skolen?* Oslo: Universitetsforlaget.
- Elstad, E. (2010). PISA i norsk offentlighet: politisk teknologi for styring og bebreidelsesmanøvrering. I E. Elstad & K. Sivesind (Red.), *PISA – sannheten om skolen?* (s.100-121). Oslo: Universitetsforlaget.
- Elstad, Jon Ivar (2010). Spørreskjemaundersøkelsens fallgruber. I D. Album, M. N. Hansen, K. Widerberg (Red.), *Metodene våre. Eksempler fra samfunnsvitenskapelig forskning*. Kap. 9, s. 155-169. Oslo: Universitetsforlaget.
- Elvbakken, K. T., & Hansen, H. F. (2018). Evidence producing organizations: Organizational translation of travelling evaluation ideas. *Evaluation*. Hentet fra <https://doi.org/10.1177/1356389018803965>
- Fan, W., & Yan, Z. (2010). Factors affecting response rates of the web survey: A systematic review. *Computers in human behavior*, 26(2), 132-139.
- Freeman, R. & Sturdy, S. (Red.). (2014). *Knowledge in policy: Embodied, inscribed, enacted*. Bristol: Policy Press.
- Fretheim, A. (2013). Kunnskapsbasert politikkutforming. *Norsk Epidemiologi*, 23(2).
- Frøyland, E. (1969). *Utforskning og utvikling av utdanningssamfunnet*. Oslo: Universitetsforlaget.
- Furre, H. & Horrigmo, A. M. J. (2013). Nærhet og distanse. Om forholdet mellom bestiller og utøver av evalueringer. I A. Halvorsen, E. Lier-Madsen, & N. Jentoft (Red.), *Evaluering. Tradisjoner, praksis, mangfold* (s. 180-194). Bergen: Fagbokforlaget.
- Furubo, J. E., Rist, R. C., & Sandahl, R. (2002). *International atlas of evaluation*. New Brunswick: Transaction Publishers.
- Galesic, M., & Bosnjak, M. (2009). Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public opinion quarterly*, 73(2), 349-360.

- Guo, Y., Kopec, J. A., Cibere, J., Li, L. C., & Goldsmith, C. H. (2016). Population survey features and response rates: a randomized experiment. *American journal of public health, 106*(8), 1422-1426.
- Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Newbury Park, CA: Sage.
- Gustafsson, J.E. (2008). Effects of international comparative studies on educational quality on the quality of educational research. *European Educational Research Journal, 7*(1), 1-17.
- Halvorsen, A. (2013). Forskningsbaserte evalueringer og andre evalueringer. I A. Halvorsen, E. Lier-Madsen, & N. Jentoft (Red.), *Evaluering. Tradisjoner, praksis, mangfold* (s. 233-246). Bergen: Fagbokforlaget.
- Halvorsen, A., Lier-Madsen, E., og Jentoft, N. (Red.) (2013), *Evaluering. Tradisjoner. Praksis. Mangfold*. Bergen: Fagbokforlaget.
- Hammersley, M. (2013). *The myth of research-based policy and practice*. London: Sage Publications Ltd.
- Haraldsen, G. (1999). *Spørreskjemametodikk etter kokebokmetoden*. Oslo: Ad Notam Gyldendal.
- Helljesen, V., Grønli, H., og Omvik, O.R. (2013). PISA-sjefen mener norsk skole har et stort, skjult problem. *NRK*. Hentet fra: https://www.nrk.no/valg2013/pisa-sjef_-_norges-store-problem-1.11186771
- Henry, G. T., & Mark, M. M. (2003). Beyond use: Understanding evaluation's influence on attitudes and actions. *American Journal of Evaluation, 24*(3), 293–314.
- Herbert, J.L. (2014). Researching evaluation influence: A review of the literature. *Evaluation Review, 38*: 388–419.
- Holst, C., & Molander, A. (2017). Public deliberation and the fact of expertise: making experts accountable. *Social Epistemology, 31*(3), 235-250.

- Hoppe, R. (2010). From “knowledge use” towards “boundary work”: sketch of an emerging new agenda for inquiry into science-policy interaction. In *Knowledge Democracy* (pp. 169-186). Springer, Berlin, Heidelberg.
- Huberman, M. (1987). Steps toward an integrated model of research utilization. *Knowledge*, 8(4), 586-611.
- Huckel, C. S., & Blyth, F. (2017). Challenges of integrating evidence into health policy and planning: linking multiple disciplinary approaches. *Public health research & practice*, 27(2), 1-6.
- Højlund, S. (2014). Evaluation use in the organizational context – changing focus to improve theory. *Evaluation*, 20(1), 26-43.
- Højlund, S. (2014). Evaluation use in evaluation systems – the case of the European Commission. *Evaluation*, 20(4), 428-446.
- Innvær, S., Vist, G., Trommald, M. & Oxman, A. (2002). Health policy-makers' perceptions of their use of evidence: a systematic review. *Journal of health services research & policy*, 7(4), 239-244.
- Jacobsen, D. I. (2010). *Forståelse, beskrivelse og forklaring: innføring i metode for helse- og sosialfagene*. Oslo: Høyskoleforlaget.
- Jann, W. & Wegrich, K. (2007). Theories of the policy cycle. I Fischer, F., Miller, G. J., & Sidney, M. S. (Red.). *Handbook of Public Policy Analysis*, 43-62. Boca Raton: Taylor & Francis Group LLC.
- Jasanoff, S. (Red.). (2004). *States of knowledge: the co-production of science and the social order*. New York, NY: Routledge.
- Johnson, K., Greenseid, L. O., Toal, S. A., King, J. A., Lawrenz, F., & Volkov, B. (2009). Research on evaluation use: A review of the empirical literature from 1986 to 2005. *American Journal of Evaluation*, 30(3), 377-410.

- Kallemeyn, L. M., Hall, J., Friche, N., & McReynolds, C. (2015). Cross-continental reflections on evaluation practice: Methods, use, and valuing. *American Journal of Evaluation*, 36(3), 339-357.
- Kingdon, J. W. (1995). The policy window, and joining the streams. *Agendas, alternatives, and public policies*, 172-189.
- Kirkhart, K.E. (2000). Reconceptualizing evaluation use: An integrated theory of influence. *New Directions for Evaluation*, 2000: 5–23.
- Knorr K. (1977). Policy-makers use of social science knowledge: symbolic or instrumental, I Weiss, C. H. (Red.), *Using social research in public policy making* (Vol. 11). Lexington, MA: Lexington Books.
- Krogstrup, H. K. (2011). *Kampen om evidens: Resultatmåling, effektevaluering og evidens*. København: Hans Reitzels Forlag.
- Kvitastein, O.A. (2002). *Offentlige evalueringer som styringsinstrumenter: Kravspesifikasjoner og kontrollproblemer* (SNF-rapport nr.3 30/2002). Bergen: Stiftelsen for samfunns- og næringslivsforskning.
- Lasswell, H. D. (1956). The decision process. *Seven Categories of Functional Analysis*. College Park, MD: University of Maryland.
- Lindgren, L. (2014). *Nya utvärderingsmonstret: om kvalitetsmätning i den offentliga sektorn*. Lund: Studentlitteratur AB.
- Mark, M. M. (2008). Building a better evidence base for evaluation theory. I N.L. Smith & P. R. Brandon (Red.), *Fundamental issues in evaluation* (s. 111-134). New York, NY: The Guilford Press.
- Mark, M. M., & Henry, G. T. (2004). The mechanisms and outcomes of evaluation influence. *Evaluation*, 10(1), 35–57.

- Mark, M. M. & Henry, G.T. (2013). Multiple routes: Evaluation, assisted sensemaking, and pathways to betterment. I M. C. Alvin (Red.), *Evaluation Roots: A Wider Perspective of Theorists' Views and Influences* (s. 144-156). Thousand Oaks: Sage Publications Inc.
- Marsdal, M. E. (2011). *Kunnskapsbløffen*. Oslo: Forlaget Manifest.
- Mason, S., & Azzam, T. (2018). In Need of an Attitude Adjustment? The Role of Data Visualization in Attitude Change and Evaluation Influence. *American Journal of Evaluation*.
- Mathison, S. (2018). Does evaluation contribute to the public good?. *Evaluation*, 24(1), 113-119.
- Maybin, J. (2016). *Producing Health Policy: Knowledge and Knowing in Government Policy Work*. Hampshire: Palgrave Macmillan.
- Meyer, S. & Myklebust, S. (Red.). (2002). *Kunnskapsmakt*. Oslo: Gyldendal akademisk, Fagbokforlaget.
- NESH, Den nasjonale forskningsetiske komité for samfunnsvitenskap og humaniora. 2016. *Forskningsetiske retningslinjer for samfunnsvitenskap, humaniora, juss og teologi*. Oslo: De nasjonale forskningsetiske komitéene.
- Nutley, S. M., Walter, I., & Davies, H.T.O. (2007). *Using evidence: How research can inform public services*. Bristol: The Policy Press.
- Olejniczak, K. (2017). The Game of Knowledge Brokering: A New Method for Increasing Evaluation Use. *American Journal of Evaluation*, 38(4), 554-576.
- Oliver, K., Innvar, S., Lorenc, T., Woodman, J., & Thomas, J. (2014). A systematic review of barriers to and facilitators of the use of evidence by policymakers. *BMC health services research*, 14(1), 2.
- Parkhurst, J. O., & Abeyasinghe, S. (2016). What constitutes “good” evidence for public health and social policy-making? From hierarchies to appropriateness. *Social epistemology*, 30(5-6), 665-679.

- Patton, M. Q. (2008). *Utilization-focused evaluation*. Newbury Park, CA: Sage.
- Patton, M.Q. (1998). Discovering process use. *Evaluation*, 4(2), 225–233.
- Pattyn, V., Van Voorst, S., Mastenbroek, E. & Dunlop, C. (2018). Policy Evaluation in Europe. I E. Ongaro, & S. Van Thiel (Red.), *The Palgrave handbook of public administration and management in Europe* (s. 577-593). Hampshire: Palgrave Macmillan.
- Pawson, R. (2013). *The science of evaluation: a realist manifesto*. London: Sage.
- Pawson, R., Tilley, N., & Tilley, N. (1997). *Realistic evaluation*. London: Sage.
- Petersén, A. C., & Olsson, J. I. (2014). An evaluation paradox in social work? An empirical study of evaluation use in connection with temporary programmes in Swedish social work. *European Journal of Social Work*, 17(2), 175-191.
- Petticrew, M., Whitehead, M., Macintyre, S. J., Graham, H., & Egan, M. (2004). Evidence for public health policy on inequalities: 1: the reality according to policymakers. *Journal of Epidemiology & Community Health*, 58(10), 811-816.
- Picciotto, R. (2016). Evaluation and bureaucracy: the tricky rectangle. *Evaluation*, 22(4), 424-434.
- Ponterotto, J. G. (2006). Brief note on the origins, evolution, and meaning of the qualitative research concept thick description. *The Qualitative Report*, 11(3), 538-549.
- Power, M. (1997). *The Audit Society: Rituals of Verification*. Oxford: Oxford University Press.
- Radaelli, C. M. (1995). The role of knowledge in the policy process. *Journal of European public policy*, 2(2), 159-183.
- Regelverket for økonomistyring i staten, hentet fra:
https://www.regjeringen.no/globalassets/upload/fin/vedlegg/okstyring/reglement_for_ekonomistyring_i_staten.pdf

- Rey, O. (2010). The use of external assessments and the impact on education systems. I S. M. Stoney (Red.), *Beyond Lisbon 2010: Perspectives from Research and Development for Education Policy in Europe* (CIDREE Yearbook 2010) (s. 137–157). Slough: NFER.
- Røvik, K. A. (2007). *Trender og translasjoner. Ideer som former det 21. århundrets organisasjon*. Oslo: Universitetsforlaget.
- Sabatier, P. & Jenkins-Smith H. *Policy change and learning: An advocacy coalition approach*, (s. 1-9). Boulder: Westview Press.
- Sanderson, I. (2002). Evaluation, policy learning and evidence-based policy making. *Public administration*, 80(1), 1-22.
- Saunders, M. (2012). The use and usability of evaluation outputs: A social practice approach. *Evaluation*, 18(4), 421-436.
- Schwandt, T. A. (2009). Globalizing influences on the western evaluation imaginary. I K. Ryan & J. B. Cousins (Red.), *The Sage handbook of educational evaluation*, (19-36). Thousand Oaks, CA: SAGE Publications, Inc
- Scriven, M. (1991). *Evaluation thesaurus*. Newbury Park, CA: Sage.
- Silverman, D. (2013). *A very short, fairly interesting and reasonably cheap book about qualitative research*. London: Sage Publications Ltd.
- Smith, K. E. (2013). *Beyond evidence based policy in public health: The interplay of ideas*. Hampshire: Palgrave Macmillan.
- Smith, K. E., & Haux, T. (2017). Evidence-based policymaking (EBPM). I B. Greve (Red.), *Handbook of Social Policy Evaluation*. Cheltenham: Edward Elgar Publishing.
- Stehr, N. (2015). Knowledge society, history of. I J. Wright (Red.), *International Encyclopedia of the Social & Behavioral Sciences*, 2nd edition, (Volume 1). Amsterdam: Elsevier.
- Stockard, J., & Wood, T. W. (2017). The Threshold and Inclusive Approaches to Determining “Best Available Evidence” An Empirical Analysis. *American Journal of Evaluation*, 38(4), 471-492.

- Stone, D. A. (2012). *Policy paradox: The art of political decision making*. New York: W.W. Norton & Company.
- Sverdrup, S. (2014). *Evaluering: tilnærminger, modeller og eksempler*. Oslo: Gyldendal akademisk.
- Sverdrup, Sidsel (2002), *Evaluering – Faser, design og gjennomføring*. Bergen: Fagbokforlaget
- Titlestad, T. (1982). *Når folket fortel: ei handbok i intervjuteknikk og tradisjonsinnsamling*. Oslo: Universitetsforlaget.
- Tornes, K. (2012). Evalueringstradisjoner. I K. Tornes (Red.), *Evaluering i teori og praksis* (s. 71-97). Bergen: Fagbokforlaget
- Tornes, K. (2013). Om verdier og kriterier i evaluering. I A. Halvorsen, E. Lier-Madsen, & N. Jentoft (Red.), *Evaluering. Tradisjoner, praksis, mangfold* (s. 28-42). Bergen: Fagbokforlaget.
- Tranøy, B.S., Christensen, J., Nuland, B.R. & Hernes, G. (2009). Staten og velferdsforskningen. I B. R. Nuland, B. S. Tranøy & J. Christensen (Red.), *Hjernen er alene: institusjonalisering, kvalitet og relevans i norsk velferdsforskning* (s. 13-42). Oslo: Universitetsforlaget.
- Trygstad, S., Lorentzen, T., Løken, E., Moland, L., & Skalle, N. (2006). *Den nye staten. Hovedrapport. Omfang og effekter av omstillingene i staten 1990-2004* (Fafo-rapport 530). Oslo: Fafo.
- Vallin, L. M., Philippoff, J., Pierce, S., & Brandon, P. R. (2015). Research-on-Evaluation Articles Published in the American Journal of Evaluation, 1998–2014. *New Directions for Evaluation*, 2015(148), 7-15.
- Vanderkruik, R., & McPherson, M. E. (2017). A contextual factors framework to inform implementation and evaluation of public health initiatives. *American Journal of Evaluation*, 38(3), 348-359.
- Vedung, E. (2010)a. Four waves of evaluation diffusion. *Evaluation*, 16(3), 263-277.

- Vedung, E. (2010)b. Utvärdering och de sex användningarna. I Rombach, B., & Sahlin, K. (Red.), *Från sanningsökande till styrmedel: moderna utvärderingar i offentlig sektor* (s.25-52). Stockholm: Santérus Förlag
- Vedung, E. (2009). *Utvärdering i politik och förvaltning*. Lund: Studentlitteratur.
- Vo, A. T., & Christie, C.A. (2015). Advancing research on evaluation through the study of context. I P. R. Brandon (Red.), *Research on evaluation: New Directions for Evaluation*, 148, 43-55.
- Weiss, C. H. (1977). Research for Policy's Sake: The Enlightenment Function of Social Research. *Policy Analysis*, 3(4), 531-545.
- Weiss, C. H. (1978). Improving the linkage between social research and public policy. *Knowledge and policy: The uncertain connection*, 23-81.
- Weiss, C.H. (1979). The many meanings of research utilization. *Public administration review*, 39(5), 426-431.
- Weiss, C. H., & Bucuvalas, M. J. (1980). *Social science research and decision-making*. Columbia University Press.
- Weiss, C. H. (1998). Have we learned anything new about the use of evaluation? *American Journal of Evaluation*, 19(1), 21-33.
- Weiss, C.H., Murphy-Graham, E. & Birkeland, S. (2005). An alternate route to policy influence how evaluations affect DARE. *American Journal of Evaluation*, 26, 12–30.
- Widerberg, K. (2001). *Historien om et kvalitativt forskningsprosjekt: en alternativ lærebok*. Oslo: Universitetsforlaget.
- Wilensky, H. L. (1997). Social science and the public agenda: Reflections on the relation of knowledge to policy in the United States and abroad. *Journal of Health Politics, Policy and Law*, 22(5), 1241-1265.
- Østbye, H., Helland, K., Knapskog, K., Larsen, L. O., & Moe, H. (2013). *Metodebok for mediefag*. Oslo: Fagbokforlaget.

Øvrelid, R. & Bastøe, P. (2002). Norway: Toward a Results-Orientated Government Administration? I J.E. Furubo, R. Rist & R. Sandahl (Red.), *International Atlas of Evaluation* (s. 209-225). New Brunswick: Transaction Publishers.

Øvrelid, R. (2000). Evaluering i Norge. *Nordisk administrativt tidsskrift*, 2, 131-47.



Harald Hårfages gate 29
N-5007 Bergen
Norway
Tel: +47-55 58 21 17
Fax: +47-55 58 96 50
nsd@nsd.uib.no
www.nsd.uib.no
Org.nr. 985 321 884

Øyunn Syrstad Høydal

Institutt for offentlig administrasjon og velferdsfag Høgskolen i Oslo og Akershus

Postboks 4 St. Olavs plass

0130 OSLO

Vår dato: 12.10.2015

Vår ref: 44958 / 3 / AMS

Deres dato:

Deres ref:

TILBAKEMELDING PÅ MELDING OM BEHANDLING AV PERSONOPPLYSNINGER

Vi viser til melding om behandling av personopplysninger, mottatt 01.10.2015. Meldingen gjelder prosjektet:

<i>44958</i>	<i>Evalueringsprosjekt i statlig sosialpolitikk</i>
<i>Behandlingsansvarlig</i>	<i>Høgskolen i Oslo og Akershus, ved institusjonens øverste leder</i>
<i>Daglig ansvarlig</i>	<i>Øyunn Syrstad Høydal</i>

Personvernombudet har vurdert prosjektet og finner at behandlingen av personopplysninger er meldepliktig i henhold til personopplysningsloven § 31. Behandlingen tilfredsstiller kravene i personopplysningsloven.

Personvernombudets vurdering forutsetter at prosjektet gjennomføres i tråd med opplysningene gitt i meldeskjemaet, korrespondanse med ombudet, ombudets kommentarer samt personopplysningsloven og helseregisterloven med forskrifter. Behandlingen av personopplysninger kan settes i gang.

Det gjøres oppmerksom på at det skal gis ny melding dersom behandlingen endres i forhold til de opplysninger som ligger til grunn for personvernombudets vurdering. Endringsmeldinger gis via et eget skjema, <http://www.nsd.uib.no/personvern/meldeplikt/skjema.html>. Det skal også gis melding etter tre år dersom prosjektet fortsatt pågår. Meldinger skal skje skriftlig til ombudet.

Personvernombudet har lagt ut opplysninger om prosjektet i en offentlig database, <http://pvo.nsd.no/prosjekt>.

Personvernombudet vil ved prosjektets avslutning, 01.05.2018, rette en henvendelse angående status for behandlingen av personopplysninger.

Vennlig hilsen

Katrine Utaaker Segadal

Anne-Mette Somby

Kontaktperson: Anne-Mette Somby tlf: 55 58 24 10

Vedlegg: Prosjektvurdering

Dokumentet er elektronisk produsert og godkjent ved NSDs rutiner for elektronisk godkjenning.

Avdelingskontorer / District Offices:

OSLO: NSD, Universitetet i Oslo, Postboks 1055 Blindern, 0316 Oslo. Tel: +47-22 85 52 11. nsd@uio.no

TRONDHEIM: NSD, Norges teknisk-naturvitenskapelige universitet, 7491 Trondheim. Tel: +47-73 59 19 07. kyrre.svarva@svt.ntnu.no

TROMSØ: NSD, SVF, Universitetet i Tromsø, 9037 Tromsø. Tel: +47-77 64 43 36. nsdmaa@sv.uit.no



Prosjektvurdering - Kommentar

Prosjektnr: 44958

Tilrådommen gjelder for spørreundersøkelsen. Intervjuene må meldes separat, se under.

INFORMASJON OG SAMTYKKE

Utvalget informeres skriftlig om prosjektet og samtykker til deltakelse. Informasjonsskrivet er godt utformet.

INFORMASJONSSIKKERHET

Personvernombudet legger til grunn at forsker etterfølger Høgskolen i Oslo og Akershus sine interne rutiner for datasikkerhet.

PROSJEKTSLUTT OG ANONYMISERING

Forventet prosjektslutt er 01.05.2018. Ifølge prosjektmeldingen skal innsamlede opplysninger da anonymiseres. Anonymisering innebærer å bearbeide datamaterialet slik at ingen enkeltpersoner kan gjenkjennes. Det gjøres ved å:

- slette direkte personopplysninger (som navn/koblingsnøkkel)
- slette/omskrive indirekte personopplysninger (identifiserende sammenstilling av bakgrunnsopplysninger som f.eks. bosted/arbeidssted, alder og kjønn)

INTERVJUER

Vi ber om at intervjudelen meldes til personvernombudet@nsd.uib.no med intervjuguide og informasjonsskriv i god tid før oppstart. Benytt endringsskjema som finnes på våre nettsider til å beskrive denne delen.

Vedlegg 2: Forespørsel om deltakelse i undersøkelse om evaluering:

Forespørsel om deltakelse i e-postbasert undersøkelse om evaluering

I forbindelse med mitt PhD-prosjekt ved Høgskolen i Oslo og Akershus, ønsker jeg å gjennomføre en e-postbasert spørreundersøkelse blant ansatte i et utvalg direktorater innen det sosialpolitiske feltet.

Målet med denne undersøkelsen er å bidra med ny kunnskap om bruk av og holdninger til evaluering i statlig regi.

I forbindelse med undersøkelsen ønsker jeg et bredt bilde av forvaltningens bruk av ekstern evaluering. Derfor håper jeg å sende undersøkelsen til alle deler av organisasjonen som denne tematikken kan være relevant for. Det vil si både til ledelse, til ansatte som arbeider med analyse og kunnskapsproduksjon til daglig og til ansatte som bare indirekte forholder seg til evaluering gjennom for eksempel ny kunnskap.

Undersøkelsen er lagt opp slik at ansatte med betydelig evalueringserfaring besvarer en lengre versjon enn ansatte med begrenset erfaring. Undersøkelsen vil dermed ta henholdsvis drøye ti minutter å besvare for dem med mye erfaring og knappe ti minutter for ansatte med lite erfaring. Jeg ønsker også å følge opp med intervjuer av rundt 15 deltakere i undersøkelsen. Disse vil velges ut på tvers av deltakerorganisasjonene.

Fra deltakerorganisasjonenes side vil jeg ha behov for assistanse i forbindelse med utsendelse av undersøkelsen. Både når det gjelder utvalg av deltakere og i forbindelse med selve utsendelsen. Selve undersøkelsen vil ligge som lenke i en e-post. Det er også fint om det kan informeres internt om undersøkelsen i forkant. Her vil jeg naturligvis kunne være behjelpelig.

Alle data behandles konfidensielt. Ingen enkeltpersoner vil være gjenkjennelige i artikler eller annet materiale.

Jeg presenterer gjerne resultatene av den ferdige studien.

Med vennlig hilsen

Øyunn Syrstad Høydal
stipendiat

41 21 71 48
oyhoy@hioa.no
Institutt for Offentlig administrasjon og velferd
Høgskolen i Oslo og Akershus

Personvern

Prosjektet er meldt til Personvernombudet for forskning, Norsk samfunnsvitenskapelig datatjeneste (SND). Alle data behandles konfidensielt. I prosjektperioden oppbevares datamaterialet avidentifisert. Ingen enkeltpersoner vil være gjenkjennelige i artikler eller annet materiale. Du kan trekke deg underveis i prosjektet, uten å begrunne dette nærmere. Dersom du trekker deg vil alle innsamlede data om deg bli slettet. Alle data slettes når prosjektet avsluttes 31.12.2018.

Vedlegg 3: E-postinvitasjon

Hei,

Høgskolen i Oslo og Akershus vil gjennomføre en undersøkelse om bruk av evaluering i statlig sektor. Denne e-posten sendes på vegne av dem til ansatte i xxxx

Har du deltatt i en eller flere evalueringsprosesser, brukt evalueringsresultater, fått evalueringsresultater presentert eller lest om dem? Gå inn på undersøkelsen:

<https://response.questback.com/yunnhydal/4k5sjmjwmj>

Målet med undersøkelsen er å bidra med ny kunnskap om bruk av evaluering i statlig sektor. Undersøkelsen er en del av et doktorgradsprosjekt ved Høgskolen i Oslo og Akershus.

Alle data behandles konfidensielt. Ingen enkeltpersoner vil være gjenkjennelige i artikler eller annet materiale.

Dersom du har spørsmål, ta gjerne kontakt med undertegnede.

Med vennlig hilsen

Øyunn Syrstad Høydal
Stipendiat

Institutt for offentlig administrasjon og velferdsfag Fakultet for samfunnsfag Høgskolen i Oslo og Akershus

E-post: oyhoy@hioa.no

Vedlegg 4: Påminnelse om undersøkelse

Hei, du har tidligere mottatt invitasjon til å delta i en spørreundersøkelse om evaluering.

Det finnes lite forskning på bruk av evaluering i statlig sektor. Målet med denne spørreundersøkelsen er å bidra med ny kunnskap om dette feltet. Jeg håper du tar deg tid til å delta.

Undersøkelsen er en del av et doktorgradsprosjekt ved Høgskolen i Oslo og Akershus. Alle data vil behandles konfidensielt. Ingen enkeltpersoner vil være gjenkjennelige i artikler eller annet materiale.

[Klikk her for å delta](#)

Med vennlig hilsen

Øyunn Syrstad Høydal
Stipendiat

Institutt for offentlig administrasjon og velferdsfag
Høgskolen i Oslo og Akershus
E-post: oyhoy@hioa.no

Vedlegg 5: Undersøkelse: Bruk av evaluering i statlig sektor

1. Har du deltatt i en eller flere evalueringsprosesser, brukt evalueringsresultater, fått evalueringsresultater presentert eller lest om dem?

- Ja
- Nei

2. Når ble denne evalueringen avsluttet?

Vær vennlig og svar ut fra den siste avsluttede evalueringen du har erfaring fra. Husk at evalueringen må være utført av eksterne miljøer/utførere.

- Siste 11 måneder
- Siste 1-2 år
- Siste 3-4 år
- Siste 5 år eller mer
- Vet ikke

3. Hva ble evaluert? (Flere svar er mulig)

- Tjenester eller tiltak
- Organisering eller styreform
- Pilotprosjekt/ forsøksordning
- En reform
- En lov, eller regelendring
- Annet

4. Hva slags form for evaluering var dette? (Flere svar er mulig)

- Førevaluering/ forhåndsanalyse
- Følgeevaluering/ underveisevaluering
- Sluttevaluering
- Målevaluering
- Effektevaluering
- Randomisert kontrollert studie
- Metaevaluering
- Vet ikke
- Annet

5. Hvilken form for datainnsamling ble benyttet i denne evalueringen? (Flere svar er mulig)

- Intervjuer
- Spørreundersøkelse
- Registerdata
- Observasjon
- Forsøk/ eksperiment med kontrollgruppe
- Dokument-/ policyanalyse
- Case-studie
- Workshops/ konferanser/ møter
- Vet ikke
- Annet

6. Har du:

- ...lest hele eller deler av evalueringsrapporten?
- ...deltatt på ett eller flere interne møter der erfaringer fra denne evalueringen ble delt?
- ...deltatt på ett eller flere eksterne møter der erfaringer fra denne evalueringen ble delt?
- ...lest om evalueringen i fagtidsskrift?
- ...lest/ hørt om evalueringen i media?

7. ...lest hele eller deler av evalueringsrapporten?

- Ja
- Nei
- Vet ikke

8. ...deltatt på ett eller flere interne møter der erfaringer fra denne evalueringen ble delt?

- Ja
- Nei
- Vet ikke

9. ...deltatt på ett eller flere eksterne møter der erfaringer fra denne evalueringen ble delt?

- Ja
- Nei
- Vet ikke

10. ...lest om evalueringen i fagtidsskrift?

- Ja
- Nei
- Vet ikke

11. ...lest/ hørt om evalueringen i media?

- Ja
- Nei
- Vet ikke

12. Hvem utførte denne evalueringen? (Flere svar er mulig)

- Forskningsinstitutt
- Universitet/ høyskole
- Konsulentfirma
- Vet ikke
- Annet

13. Hvordan kjenner du til den aktuelle evalueringsprosessen? (Flere svar er mulig)

- Jeg deltok i initierings- eller anbudsprosess.
- Jeg ble holdt løpende orientert underveis.
- Jeg var medlem av styringsgruppe, referansegruppe eller intern prosjektgruppe.
- Jeg har bidratt med kunnskapsdeling/ informasjonsarbeid.
- Jeg har på annen måte bidratt faglig eller administrativt inn i prosessen.
- Jeg er blitt informert/ har tilegnet meg informasjon i etterkant.
- Annet

14. Hvem tok initiativet til denne evalueringen? (Flere svar er mulig)

I denne delen av undersøkelsen får du spørsmål knyttet til oppstart og gjennomføring av den aktuelle evalueringen.

- Brukere/ klienter/ interessegrupper
- Politisk ledelse
- Stortinget
- Eget departement
- Eget direktorat
- Annen statlig etat
- Vet ikke
- Annet

15. Var evalueringen

- ...lovpålagt?
- ...pålagt etter Statens økonomiregelverk?

16. ...lovpålagt?

- Ja
- Nei
- Vet ikke

17. ...pålagt etter Statens økonomiregelverk?

- Ja
- Nei
- Vet ikke

18. Hvem skulle bruke resultatene av denne evalueringen?

- Politisk ledelse
- Forvaltningen
- Praksisfeltet

19. Politisk ledelse

- Ja
- Nei
- Vet ikke

20. Forvaltningen

- Ja
- Nei
- Vet ikke

21. Praksisfeltet

- Ja
- Nei
- Vet ikke

22. Ble evalueringen gjennomført for å

- ...undersøke virkning/ effekt
- ...undersøke måloppnåelse
- ...skaffe ny kunnskap
- ...bekrefte antakelser
- ...sikre videreføring av et tiltak eller lignende
- ...sette sak/ felt på dagsordenen
- ...bedre kontroll/ styring
- ...vurdere organisering

23. ...undersøke virkning/ effekt

- Ja
- Nei
- Vet ikke

24. ...undersøke måloppnåelse

- Ja
- Nei
- Vet ikke

25. ...skaffe ny kunnskap

- Ja
- Nei
- Vet ikke

26. ...bekrefte antakelser

- Ja
- Nei
- Vet ikke

27. ...sikre videreføring av et tiltak eller lignende

- Ja
- Nei
- Vet ikke

28. ...sette sak/ felt på dagsordenen

- Ja
- Nei
- Vet ikke

29. ...bedre kontroll/ styring

- Ja
- Nei
- Vet ikke

30. ...vurdere organisering

- Ja
- Nei
- Vet ikke

31. Fantest det i forbindelse med denne evalueringen

- ...en styringsgruppe?
- ...en referansegruppe?
- ...et formalisert samarbeid med brukere/ interessegrupper?

32. ...en styringsgruppe?

- Ja
- Nei
- Vet ikke

33. ...en referansegruppe?

- Ja
- Nei
- Vet ikke

34. ...et formalisert samarbeid med brukere/ interessegrupper?

- Ja
- Nei
- Vet ikke

35. I forbindelse med denne evalueringen var det intern uenighet i mitt direktorat om

- ...behovet for evalueringen
- ...valg av metode for datainnsamling
- ...tidspunktet for evalueringen
- ...hva som skulle inngå i evalueringsoppdraget

36. ...behovet for evalueringen

- Ja
- Nei
- Vet ikke

37. ...valg av metode for datainnsamling

- Ja
- Nei
- Vet ikke

38. ...tidspunktet for evalueringen

- Ja
- Nei
- Vet ikke

39. ...hva som skulle inngå i evalueringsoppdraget

- Ja
- Nei
- Vet ikke

40. Ble det noen form for uenighet mellom oppdragsgiver og andre parter i denne evalueringsprosessen?

(For eksempel evaluator/ utførermiljø eller referansegruppe/ styringsgruppe)

- Ja
- Nei
- Vet ikke

41. Hva handlet denne uenigheten om? (Flere svar er mulig)

- Eierskap til prosessen
- Valg av datainnsamlingsmetode
- Valg av evalueringsform
- Valg av informanter/ case
- Analyse/ resultater
- Kvaliteten på evalueringen
- Vet ikke
- Annet

42. Fikk evalueringen direkte konsekvenser for evalueringsobjektet (loven/ tiltaket/ reformen eller lignende)?

- Evalueringsobjektet ble satt på den politiske dagsordenen
- Evalueringsobjektet ble justert
- Evalueringsobjektet ble avsluttet/ fjernet
- Evalueringsobjektet ble utvidet/ videreført

43. Evalueringsobjektet ble satt på den politiske dagsordenen

- Ja
- Nei
- Vet ikke
- Uaktuelt

44. Evalueringsobjektet ble justert

- Ja
- Nei
- Vet ikke
- Uaktuelt

45. Evalueringsobjektet ble avsluttet/ fjernet

- Ja
- Nei
- Vet ikke
- Uaktuelt

46. Evalueringsobjektet ble utvidet/ videreført

- Ja
- Nei
- Vet ikke
- Uaktuelt

47. Hvordan vil du vurdere den samlede kvaliteten på evalueringen som ble gjennomført?

- Svært høy
- Høy
- Middels
- Lav
- Svært lav
- Vet ikke

48. Har du brukt resultater eller annen kunnskap fra evalueringen i eget arbeid?

- Ja
- Nei
- Vet ikke

49. Hvorfor har du ikke brukt resultater eller annen kunnskap fra denne evalueringen?
(Flere svar er mulig)

- Evalueringen har ikke vært relevant for min jobb
- Kvaliteten er ikke god nok
- Evalueringen ga meg ikke ny kunnskap
- Det kan hende jeg vil bruke evalueringsresultatene senere
- Det finnes andre og bedre kunnskapskilder på dette feltet
- Vet ikke
- Annet

50. Hvordan har du benyttet deg av kunnskap fra den aktuelle evalueringen?

- Som grunnlag for beslutninger i forbindelse med evalueringsprosessen.
- Som generelt kunnskapsgrunnlag for eget arbeid.
- Som bekreftelse på eksisterende kunnskap/ antakelser.
- I forbindelse med kunnskaps-/ informasjonsdeling.

51. Som grunnlag for beslutninger i forbindelse med evalueringsprosessen.

- Ja
- Nei
- Vet ikke

52. Som generelt kunnskapsgrunnlag for eget arbeid.

- Ja
- Nei
- Vet ikke

53. Som bekreftelse på eksisterende kunnskap/ antakelser.

- Ja
- Nei
- Vet ikke

54. I forbindelse med kunnskaps-/ informasjonsdeling.

- Ja
- Nei
- Vet ikke

55. Denne evalueringen ga nyttig informasjon/kunnskap.

Nå vil du presenteres for et antall utsagn om den aktuelle evalueringen. For hvert utsagn finnes det seks ulike svaralternativer.

- Helt uenig
- Delvis uenig
- Hverken enig/ uenig
- Delvis enig
- Helt enig
- Vet ikke

56. Evalueringsrapporten kom på et tidspunkt da resultatene var blitt uaktuelle.

- Helt uenig
- Delvis uenig
- Hverken uenig/ enig
- Delvis enig
- Helt enig
- Vet ikke

57. Evalueringsresultatene var i overensstemmelse med egne erfaringer/ antakelser.

- Helt uenig
- Delvis uenig
- Hverken uenig/ enig
- Delvis enig
- Helt enig
- Vet ikke

58. Evalueringen ble svekket på grunn av evaluatorens manglende kunnskap om feltet.

- Helt uenig
- Delvis uenig
- Hverken uenig/ enig
- Delvis enig
- Helt enig
- Vet ikke

59. Evalueringen ga nye ideer eller synspunkter.

- Helt uenig
- Delvis uenig
- Hverken uenig/ enig
- Delvis enig
- Helt enig
- Vet ikke

60. Evalueringen ble svekket av evaluators personlige oppfatninger om feltet.

- Helt uenig
- Delvis uenig
- Hverken uenig/ enig
- Delvis enig
- Helt enig
- Vet ikke

61. Jeg hadde tillit til den/ de som utførte evalueringen.

- Helt uenig
- Delvis uenig
- Hverken uenig/ enig
- Delvis enig
- Helt enig
- Vet ikke

62. Det har vært intern uenighet i mitt direktorat om evalueringresultatet.

- Helt uenig
- Delvis uenig
- Hverken uenig/ enig
- Delvis enig
- Helt enig
- Vet ikke

63. Dersom du ønsker å oppgi evalueringens navn eller tema, kan du gjøre dette her:

64. Angi hvor mange eksterne evalueringprosesser i statlig sektor du har erfaring med fra siste tre år.

Nå bes du om å svare på spørsmål om din samlede erfaring med eksterne evalueringer av tiltak, program, lover, reformer, satsinger og lignende i statlig regi. Forvaltningsrevisjon i regi av Riksrevisjonen er heller ikke et tema i denne delen av undersøkelsen.

- 0
- 1-3
- 4-6
- 7-9
- 10 eller flere
- Vet ikke

65. Hva er ditt samlede inntrykk av kvaliteten på evalueringene som gjennomføres i regi av din etat/ som din etat er involvert i

- Svært lav
- Lav
- Hverken eller
- Høy
- Svært høy
- Vet ikke

66. Evaluering gir viktig kunnskap om offentlig sektor, som vi ellers ikke ville hatt tilgang til.

Du bes fremdeles om å svare ut fra din samlede erfaring med ekstern evaluering i statlig sammenheng når du nå presenteres for nye utsagn om evaluering og evalueringsprosesser.

- Helt uenig
- Delvis uenig
- Hverken uenig/ enig
- Delvis enig
- Helt enig
- Vet ikke

67. Jeg mener evalueringer bidrar til bedre styring av offentlig sektor.

- Helt uenig
- Delvis uenig
- Hverken uenig/ enig
- Delvis enig
- Helt enig
- Vet ikke

68. Evalueringer spiller en stor rolle som grunnlag for politikkutforming.

- Helt uenig
- Delvis uenig
- Hverken uenig/ enig
- Delvis enig
- Helt enig
- Vet ikke

69. I min organisasjon legger vi til rette for evaluering når vi planlegger en ny satsing/ nye tiltak.

- Helt uenig
- Delvis uenig
- Hverken uenig/ enig
- Delvis enig
- Helt enig
- Vet ikke

70. Evalueringer basert på kvantitativ metode gir mer pålitelig informasjon enn andre evalueringer.

- Helt uenig
- Delvis uenig
- Hverken uenig/ enig
- Delvis enig
- Helt enig
- Vet ikke

71. I min organisasjon foretrekker vi evalueringer utført av forskere/ forskningsmiljøer framfor konsulentmiljøer.

- Helt uenig
- Delvis uenig
- Hverken uenig/ enig
- Delvis enig
- Helt enig
- Vet ikke

72. Min organisasjon initierer effektstudier i tilfeller der det er tvilsomt om det er mulig å måle effekt.

- Helt uenig
- Delvis uenig
- Hverken uenig/ enig
- Delvis enig
- Helt enig
- Vet ikke

73. Evalueringsresultater blir ofte gjenstand for intern uenighet i min organisasjon.

De siste påstandene handler om omfang og nytte av evalueringene som gjennomføres i din organisasjon.

- Helt uenig
- Delvis uenig
- Hverken uenig/ enig
- Delvis enig
- Helt enig
- Vet ikke

74. Evalueringsresultatene får sjelden praktiske konsekvenser.

- Helt uenig
- Delvis uenig
- Hverken uenig/ enig
- Delvis enig
- Helt enig
- Vet ikke

75. I min organisasjon evaluerer vi mindre enn vi er pålagt.

- Helt uenig
- Delvis uenig
- Hverken uenig/ enig
- Delvis enig
- Helt enig
- Vet ikke

76. Evaluering er et viktig virkemiddel for å sette et fagfelt på dagsordenen.

- Helt uenig
- Delvis uenig
- Hverken uenig/ enig
- Delvis enig
- Helt enig
- Vet ikke

77. Evalueringer gir et nødvendig eksternt blikk på offentlig sektor.

- Helt uenig
- Delvis uenig
- Hverken uenig/ enig
- Delvis enig
- Helt enig
- Vet ikke

78. Dersom du ønsker å tilføye en egen kommentar om evaluering, kan du gjøre dette her:

79. Hvilken utdanning har du?

Til slutt bes du om å besvare spørsmål om utdanning og arbeid.

- Samfunnsfaglig
- Juridisk
- Humanistisk
- Økonomisk
- Sosialfaglig
- Realfaglig
- Annet

80. Mitt nåværende arbeidsområde kan best beskrives som:

- Ledelse
- Administrasjon/ økonomi/styring
- Utredning/ forskning/ analyse
- Saksbehandling
- Kommunikasjon
- Annet

81. I etterkant av denne undersøkelsen vil det bli gjennomført intervjuer med et mindre antall respondenter. Hensikten med intervjuene er å få en grundigere forståelse av evalueringspraksis i statsforvaltningen. Både personer med mye og med lite evalueringserfaring vil bli intervjuet. Dersom du har lyst til å delta på et slikt intervju, kan du skrive e-posten din i feltet under. Deltakerne i intervjuene vil bli anonymisert i eventuelle artikler og i avhandlingen.

Øyunn Syrstad Høydal.

Høydal ØS. Evaluation influence: Method and societal phenomenon. *Evaluation*.

2019;25(3):381-398. DOI: <https://dx.doi.org/10.1177/1356389018804263>

[Article not included due to copyright]

Artikkel 2

Øyunn Syrstad Høydal. Tracking the Instrumental Use and Motives for Evaluations. *Under vurdering i American Journal of Evaluation.*

[Article not included due to copyright]

Artikkel 3

Øyunn Syrstad Høydal og Anne Grete Tøge. Evaluating social policies – do methodological approaches determine the policy impact? *Innsendt til Evidence & Policy*.

[Article not included due to copyright]

