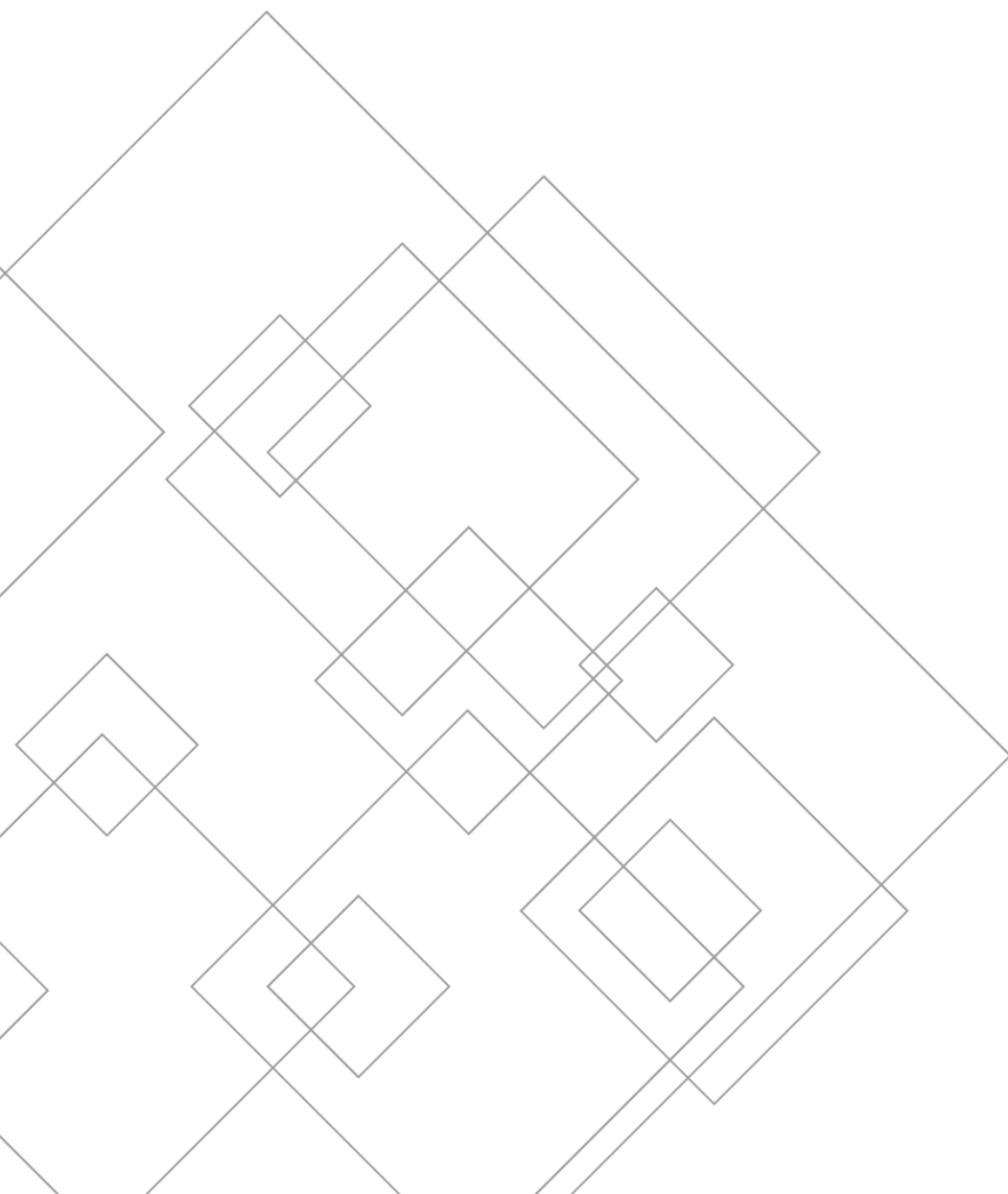


Effektevalueringer på integreringsfeltet

Hva vil det si, hvorfor bør det gjøres og hvordan kan man få det til?



IMDi
Integrerings- og
mangfoldsdirektoratet

Innhold

Sammendrag	2
1 Innledning	4
1.1 Hva er egentlig effekt og hvorfor er det så vanskelig å måle?	5
2 Hvordan kan vi finne ut om tiltakene våre virker?	7
2.1 Randomiserte, kontrollerte forsøk (RCT)	8
2.1.1 Valg i design av forsøk: Hvem, hva og hvordan?	8
2.2 Vanlige innvendinger mot randomiserte forsøk	10
2.1.1 Vet man ikke allerede hva som virker?.....	10
2.1.3 Er det ikke veldig dyrt?.....	11
2.1.4 Er det lov å gjennomføre slike forsøk?	11
2.1.5 Men det er jo allerede bestemt at tiltaket skal innføres?	12
2.2 Kvasiekperimentelle metoder	12
2.2.1 Eksempel: Regelverksendring	13
2.2.2 Eksempel: Klare objektive kriterier	14
3 Hva har blitt gjort i Norge? Erfaringer fra velferdsfeltet	15
3.1 Arbeids- og velferdspolitikken	15
3.2 Utdanningsfeltet	17
3.3 Valgforskningen.....	19
3.4 Diskriminering og holdninger	20
4 Hva skal til for å lykkes med effektevalueringer?	22
4.1 Hvordan få det til? Viktige læringspunkter	22
4.3 Nyankomne innvandrere som målgruppe: særskilte utfordringer?	26
4.4 Oppsummering.....	27
Kilder	29

Sammendrag

Det er behov for sikrere kunnskap om virkning av integreringstiltak. Integreringsmyndighetene bør i større grad legge til rette for forskning og evalueringer som kan gi troverdige mål på effekter.

I denne rapporten gjennomgår vi erfaringer med effektevalueringer fra velferdsfeltet. Hva går metodene ut på og hvilke erfaringer er gjort med dette på velferdsfeltet i Norge? Hensikten er å lære av disse erfaringene, for å kunne tilrettelegge for flere utprøvinger av integreringstiltak basert på metoder for effektevaluering.

I **kapittel 1** beskriver vi hva som egentlig menes med effekt og hvorfor det kan være vanskelig å måle. Den vanlige måten å gjennomføre evalueringer på integreringsfeltet har vært å observere resultatene av et tiltak i etterkant: Har tiltaket oppnådd de forventede resultatene? Dette gjør det imidlertid vanskelig å avgjøre om det var det nye tiltaket som ga resultater, eller andre forhold. Eller om resultatet kanskje ville vært det samme uten tiltaket. Disse problemene oppstår fordi man ikke har en kontrollgruppe å sammenligne med som *ikke* har fått tiltaket.

I **kapittel 2** gjennomgår vi metoder som kan gi robuste mål på effekter av tiltak. Å måle effekter kan gjøres både gjennom randomiserte kontrollerte studier (RCT-er) og med ulike kvasiekperimentelle design. RCT-er regnes som gullstandarden innen effektstudier, fordi det er det metodiske designet som gir de sikreste målene. Alle problemstillinger lar seg imidlertid ikke besvare med en RCT, og det kan også være etiske, politiske og økonomiske grunner til at man må velge andre metoder.

I Norge er vi fortsatt i startgropen når det gjelder å gjennomføre RCT-er i velferdspolitikken, men dette er i ferd med å endre seg. Det er etter hvert gjennomført flere RCT-er på velferdsfeltet, også i Norge.

I **kapittel 3** trekker vi fram noen eksempler på pågående og gjennomførte RCT-er fra velferdsfeltet som vi kan trekke lærdom av til gjennomføring av lignende eksperimenter på integreringsfeltet.

Det er mye man skal tenke på i utviklingen av en RCT, og det kan være komplisert og kostbart. I **kapittel 4** trekker vi frem noen sentrale grep man kan gjøre med sikte på å oppnå vellykkede prosjekter, også der nyankomne innvandrere er målgruppe:

- En RCT bør gjennomføres i flere faser. Evalueringen bør planlegges samtidig som prosjektet utvikles, og være basert på foreliggende kunnskap og en programteori. I første fase utvikler man modellen og forankrer prosjektet, deretter piloterer det, og modellen justeres før en (eventuell) oppskalering

og systematisk utprøving, med prosessevaluering.

- I mange tilfeller vil man være avhengig av tett samarbeid mellom forskere, byråkrater og praktikere for å lykkes med å designe og gjennomføre en god RCT.
- For å få mer kunnskap om virkning av integreringstiltak, bør de som utvikler tiltak samtidig planlegge hvordan det skal evalueres. Dersom det er ønskelig å måle effekt av tiltaket, bør man ikke starte implementeringen før det er laget et metodisk design for evaluering. På den måten holder man muligheten åpen for å teste ut tiltaket gjennom en RCT, som hvis det er gjennomført på en god måte er den sikreste metoden for å få vite om tiltaket har effekt.
- Forsøk der nyankomne innvandrere er målgruppe, kan innebære noen særskilte utfordringer som for eksempel språk, økt risiko for frafall og at populasjonen ikke er så stor. Når målgruppen regnes som en sårbar gruppe, er det også særlig viktig å vurdere etisk forsvarlighet.

1 Innledning

Kunnskapsbasert politikk er å utforme politiske programmer og tiltak basert på best mulig kunnskap om virkningene av de ulike virkemidlene. Kunnskapsbasert politikk skal være i tråd med forskning om hva som virker, for hvem og under hvilke betingelser. Det er et mål at politikkkutviklingen skal være mest mulig kunnskapsbasert¹.

Men er integreringspolitikken kunnskapsbasert? Vet vi nok om hva som virker av integreringspolitiske virkemidler?

Det er gjennomført både mange utprøvinger og evalueringer av integreringstiltak i Norge. En svakhet med utprøvingene er at de som regel er gjennomført på måter som ikke gjør det mulig å si om det har hatt effekt, altså om tiltaket virker eller ikke.

Den vanlige måten å gjennomføre evalueringer på har vært å observere resultatene av et tiltak i etterkant: Har tiltaket oppnådd de forventede resultatene?² Dette gjør det imidlertid vanskelig å avgjøre om det var det nye tiltaket som ga resultater, eller andre forhold. Eller om resultatet kanskje ville vært det samme uten tiltaket. Disse problemene oppstår fordi man ikke har en kontrollgruppe å sammenligne med som ikke har fått tiltaket.

I en nylig publisert artikkel som gjennomgår en rekke studier og evalueringer av integreringspolitiske tiltak i ulike EU og OECD-land, konkluderes det med at man vet for lite om hvorvidt integreringspolitiske tiltak i ulike land faktisk virker³. En kunnskapsoversikt⁴ fra 2017 viser at det er svært begrenset kunnskap, både nasjonalt og internasjonalt, om hva som er de mest effektive tiltakene for å lykkes med integreringen av voksne innvandrere med lav kompetanse. Forskerne mener det er behov for flere systematiske utprøvinger av tiltak og effektevalueringer med robuste design. Dette er også en av de tydeligste anbefalingene fra Brochmann-utvalget⁵. Utvalget påpeker at uten effektstudier av høy vitenskapelig kvalitet vil det ikke være mulig å bedre effektiviteten i integreringsarbeid fremover. Ifølge utvalget trengs det særlig mer kunnskap om hva som virker innenfor introduksjonsprogrammet. I Fafos evaluering av introduksjonsprogrammet og norskopplæringen⁶ anbefaler forskerne systematisk utprøving av metoder i norskopplæringen for nyankomne innvandrere.

¹ Se bl.a. DFØ (2018). Her står det blant annet at det i større grad bør brukes kunnskapsbasert design for utforming av nye tiltak der det er aktuelt.

² Finansdepartementet (2005).

³ Geribay & De Cuyper (2018).

⁴ Wollscheid m.fl. (2017).

⁵ NOU 2017:2.

⁶ Djuve m.fl. (2017).

I denne rapporten vil vi gjennomgå erfaringer med effektevalueringer fra velferdsfeltet. Hva går metodene ut på og hvilke erfaringer er gjort med dette på velferdsfeltet i Norge? Hensikten er å lære av disse erfaringene for å kunne tilrettelegge for flere utprøvinger av integreringstiltak basert på metoder for effektevaluering.

1.1 Hva er egentlig effekt og hvorfor er det så vanskelig å måle?

Effekt betyr kort sagt at noe påvirker noe annet. Det er en årsakssammenheng mellom tiltak (a) og resultat (b). Dette høres ikke så komplisert ut: Tenk deg en reklamekampanje på TV mot røyking. Dersom det er færre som røyker etter at kampanjen er gjennomført, har den hatt en effekt på folks røykevaner.

Eller et mer nærliggende eksempel: Det prøves ut en ny pedagogisk metode i norskundervisningen for voksne innvandrere, med mål om at flere deltakere skal bestå norskprøven ved kursets slutt. Dersom det er flere deltakere enn tidligere som består prøven etter at ny pedagogisk metode er prøvd ut, har endringen hatt en effekt på norskprøveresultatene. De nye pedagogiske metodene fungerer åpenbart bedre enn de gamle, og den innføres for alle norskkursdeltakere.

Men er det egentlig så enkelt? Hvordan kan vi vite at det faktisk er kampanjen som fikk folk til å slutte å røyke? Og kan det hende flere besto norskprøven av helt andre årsaker enn endret pedagogikk?

I de to eksemplene vi har nevnt her, kan vi ikke vite om tiltaket faktisk har hatt en effekt på resultatet. Dette handler om forskjellen på **korrelasjon** (samvariasjon) og **kausaltet** (årsakssammenheng).

Korrelasjon handler om at to eller flere fenomener opptrer samtidig. For å holde oss til et av eksemplene over: Innføringen av ny pedagogisk metode (a) og bedre norskprøveresultater (b). Dette kan forlede oss til å tro at vi er vitne til en kausalsammenheng, nemlig at det er a som har *påvirket* b. Det kan imidlertid være snakk om en *bakenforliggende variabel* som påvirker resultatet, men som man ikke har tatt høyde for. En bakenforliggende variabel i dette tilfellet kan for eksempel ha vært at det i samme periode ble lansert en ny app for språklæring, som ble bredt markedsført og mange av deltakerne tok i bruk.

For å kunne påstå at et tiltak (a) har effekt på resultatene (b) må man være så sikker som mulig på at det ikke er andre forhold som har påvirket resultatene. Man må altså kunne *isolere* effekten av det man prøver ut (a).

I utviklingen av ny politikk innfører man ofte et tiltak som forsøk, altså at det prøves ut noen steder før man eventuelt innfører det i stor skala. Gjennom å prøve ut et tiltak som forsøk, der noen grupper får tiltaket først, er allikevel ikke nok dersom man ønsker å måle effekt. Problemet er såkalte **seleksjonsmekanismer**: Personene eller enhetene (eks. kommuner, bydeler) som melder seg eller plukkes ut til å prøve ut nye tiltak, kan ha systematisk andre egenskaper enn dem som ikke deltar. For eksempel kan de være mer motiverte og mottagelige enn de som ikke mottar tiltaket.

Tiltaket «Jobbsjansen» er et eksempel på politikk som først ble innført som forsøk men der vi ikke kan måle effekt:

«Jobbsjansen» er en statlig tilskuddsordning som forvaltes av IMDi. Ordningens hovedmål er at man gjennom å tilby kvalifisering skal øke sysselsettingen blant innvandrere som står langt fra arbeidslivet. «Ny sjanse» var en forsøksordning i et utvalg kommuner og bydeler hvor formålet var å prøve ut ulike metoder for å få innvandrere over i arbeid eller utdanning, som ble gjort om til en permanent ordning etter gode resultater for målgruppen.

Det er godt mulig at «Jobbsjansen» har hatt god effekt for deltakerne. De har ikke fått andre tiltak samtidig som kan tenkes å påvirke resultatene. Men vi kan ikke vite om de gode resultatene skyldes tiltaket, eller at deltakerne hadde stor motivasjon. Det kan for eksempel tenkes at innenfor den bredere målgruppen for Jobbsjansen, var det personene som uansett hadde størst sannsynlighet for å komme i jobb som ble rekruttert inn i tiltaket. De kan for eksempel ha vært mer motiverte og mottagelige for tiltaket, eller de kan ha hatt mer uobserverte ressurser enn dem som ikke ble rekruttert. Dette gjør at det blir vanskelig å slå fast i hvor stor grad et utfall skyldes selve tiltaket, og ikke seleksjonsmekanismer.

Dette betyr ikke at «Jobbsjansen» ikke er nyttig. Det har vært en stadig økning i andel deltakere som kommer over i arbeid eller utdanning etter at ordningen ble innført. Vi vet allikevel ikke hvordan det ville gått med deltakerne uten tilbudet eller dersom de hadde fått et annet tilbud.

Det finnes imidlertid metoder som gir robuste mål på effekter av tiltak, der man kan avdekke kausalsammenhenger og unngå seleksjonseffekter. Noen slike metoder vil vi gjennomgå i neste kapittel.

2 Hvordan kan vi finne ut om tiltakene våre virker?

En effektevaluering undersøker virkningen av et tiltak eller et program (intervensjon). Dette innebærer mer enn å undersøke hvorvidt man har nådd målene man satte seg før intervensjonen. Begrepet effektevaluering innebærer også en ambisjon om å kunne avdekke hvorvidt, og i hvor stor grad, resultatene kan tilskrives selve intervensjonen. For å kunne si noe om årsak og virkning (kausaltitet), omfatter effektevalueringer en kontrafaktisk analyse, der man også stiller seg spørsmålet: hvordan hadde resultatene sett ut i fravær av intervensjonen?

Det finnes ulike metoder for å evaluere effekten av et tiltak, men ikke alle gir like sikre mål på effekt. Det finnes få eksempler, både nasjonalt og internasjonalt, på at integreringspolitiske tiltak effektevalueres. I Norge er de fleste evalueringene som er gjennomført av integreringstiltak registerbaserte analyser, der man har forsøkt å isolere effekter av et tiltak ved å kontrollere for andre forhold som kan tenkes å påvirke resultatene. Dette kan kalles kontrollvariabelmetoden⁷. Problemet med metoden er at man kun kan kontrollere for egenskaper ved individene/enhetene som man har opplysninger om. Det kan for eksempel være kjønn, alder, utdanning og innvandringsbakgrunn, hvis enhetene som sammenlignes er flyktninger. Hvis det er kommuner som sammenlignes, kan kjente egenskaper for eksempel være antall innbyggere, politisk lederskap og sentralitet. Det betyr imidlertid ikke at en eventuell effekt man finner kan tilskrives tiltaket man evaluerer, fordi det kan være egenskaper ved individene/enhetene man ikke vet noe om som egentlig forklarer resultatet. Eksempler på slike uobserverte faktorer kan være egenskaper ved individet som motivasjon og holdninger, dersom enhetene er flyktninger. Dersom enhetene er kommuner, kan det for eksempel være personlige egenskaper ved ledere eller trivsel hos de ansatte som ikke så lett lar seg måle.

I utviklingen av kunnskapsbaserte tiltak er det fornuftig å skille mellom grader eller nivåer av evidens («bevis»), etter hvor sikker kunnskap forskningen gir oss⁸. Med bakgrunn i svakhetene nevnt over, regnes ikke kontrollvariabelmetoden som en sikker metode for å måle effekt.

⁷ Ugreninov og Birkelund (2013).

Her skal vi gjennomgå noen typer effektstudier med mer robuste design, som kan bidra til å gi oss mer sikker kunnskap om hva som virker: Randomisert kontrollert forsøk (RCT: «randomized controlled trial») og kvasieksperimentelle metoder.

2.1 Randomiserte, kontrollerte forsøk (RCT)

RCT-er eller randomiserte, kontrollerte forsøk regnes som gullstandard innen forskningsmetoder når det gjelder å vurdere effekten av et tiltak, det vil si hvorvidt det *virker*. Metoden har vært mest vanlig innen medisinsk forskning, der den brukes til å undersøke og sammenligne effekter av ulike medikamenter og behandlingsformer. I senere år har metoden imidlertid blitt stadig mer utbredt også innen andre fagfelt, blant annet for å vurdere effekter av velferdspolitiske tiltak.

RCT er den beste måten å få kunnskap om hva som ville skjedd dersom tiltaket *ikke* ble innført. Evalueringen er designet som et forsøk: I forkant av forsøket defineres to grupper av individer eller enheter: en gruppe som mottar tiltaket og en kontrollgruppe som ikke mottar tiltaket, eller som mottar et annet tiltak. Det kan også være flere grupper med i forsøket. For eksempel kan to grupper få ulike tiltak, mens en tredje ikke mottar noe tiltak. Det viktige er at rekrutteringen til de ulike gruppene skjer gjennom *randomisering*: Det skal være helt tilfeldig hvem som havner i tiltaks- og kontrollgruppe. Det kan for eksempel trekkes lodd som bestemmer hvilken gruppe man havner i. Slik unngår man seleksjonsmekanismer i rekrutteringen av deltakere. Med unntak av denne sentrale forskjellen (at noen mottar tiltaket og andre ikke), er hensikten at de ulike gruppene skal være mest mulig like hverandre.

Tilfeldig trekning (randomisering) sikrer at det ikke finnes systematiske forskjeller som kan påvirke resultatene mellom personene i tiltaksgruppen og personene i kontrollgruppen, gitt at gruppene er store nok⁹. På denne måten kan man tolke forskjeller i utfall mellom gruppene i etterkant av tiltaket som effekter av intervensjonen.

2.1.1 Valg i design av forsøk: Hvem, hva og hvordan?

Randomisering er altså det viktigste som skiller en RCT fra andre utprøvinger. Det er allikevel mange måter å gjennomføre en RCT på. Hvordan evalueringer designes avhenger både av hva man vil finne ut av og hva som er gjennomførbart. Under redegjør vi for noen av de valgene man står overfor hvis man vil prøve ut et tiltak som et systematisk forsøk.

⁹ Antallet deltakere i et forsøk må være av en viss størrelse for å kunne avdekke effekt. Jo flere deltakere, desto høyere grad av presisjon hvis man vil påvise effekter med rimelig grad av sikkerhet. I planleggingen av en RCT er det derfor nødvendig å gjøre et estimat av antall/andel deltakere som trengs for å kunne måle effekt (såkalt styrkeberegning) (Malmberg-Heimonen m.fl. 2016).

Hvem skal randomiseres? En viktig vurdering gjelder hva som skal være «grunnheten» i studien. Det er ofte mest nærliggende å betrakte individer som grunnheten, og randomisere enkeltpersoner inn i tiltaks- og kontrollgruppe. En fordel som følger med dette er at man har gode forutsetninger for å et nokså høyt antall enheter (personer) i hver gruppe.

Man kan ikke alltid randomisere på individnivå. I mange sammenhenger ønsker man heller å randomisere på gruppenivå, slik at det er for eksempel skoler eller skoleklasser, kommuner eller NAV-kontorer som blir grunnheten. Hvis vi for eksempel skal undersøke en ny undervisningsmetode i en skoleklasse, må hele klassen undervises med samme metode. Klassen blir da en «klynge».

Hva skal testes ut? Noen RCT-er handler om effekten av kun å gi folk et lite «dytt» (nudge). Det er snakk om små virkemidler for å påvirke adferd i en eller annen retning, for eksempel en SMS eller et brev for å få folk til å stemme på valgdagen.

Andre eksperimenter kan undersøke mer omfattende tiltak og ta sikte på mer varig endring av adferd. Hvis man for eksempel ønsker at lærere skal bruke en annen metode i undervisningen, krever dette langsiktig implementering av den nye metoden. Hvis forsøket skal ha overføringsverdi, er det ikke nok at man finner en effekt av tiltaket. Man må vite *hva* som har hatt en effekt og *hvorfor*. Det krever at tiltaket man tester ut er presist beskrevet. Det bør være forankret i forskning som viser at det er grunn til å ha tro på tiltaket, og som sier noe om hvilke mekanismer man antar vil virke. Tiltaket bør også være pilotert før forsøk med randomisering iverksettes. Vi skriver mer om dette i kapittel 4.

Vil man teste ut ett eller flere tiltak? Når man skal forklare hva en RCT er, så er det vanlig å tenke seg en situasjon der noen får et tiltak, og andre ikke. Det er imidlertid også mulig å teste ulike tiltak opp mot hverandre, for å kunne sammenligne effekt av forskjellige tiltak, eventuelt også opp mot en kontrollgruppe som ikke mottar noe tiltak. Hvis man for eksempel ønsker å måle effekten av økt fysisk aktivitet på barns skolerestater, kan man teste ut ulike aktiviteter. Man randomiserer hvilke skoler som gjennomfører hvilke aktiviteter. På denne måten kan man finne ut ikke bare om økt fysisk aktivitet har en effekt, men også om noen aktiviteter er mer gunstige for trivsel og skolerestater enn andre.

Hva skal være utfallsmålene i prosjektet? Tenk deg for eksempel at man prøver ut nye metoder i norskopplæringen for nyankomne flyktninger, for å se om dette har effekt. Men effekt på hva? Et kortsiktig mål kan for eksempel være at deltakerne får bedre resultater på norskundervisningen. Men hva med på lang sikt? Et langsiktig mål kan være at deltakerne får en mer varig tilknytning til arbeidslivet. En analyse av langtidseffekter bør også planlegges for, og tas inn i prosjektet på et tidlig tidspunkt.

2.2 Vanlige innvendinger mot randomiserte forsøk¹⁰

2.1.1 Vet man ikke allerede hva som virker?

Politikere og praktikere mener ofte at de har en ganske god forståelse av hvilke tiltak som vil virke, og de lener seg på disse oppfatningene når de foreslår nye tiltak og programmer.

Selv om det kan være gode grunner for å tro at et nytt tiltak vil virke, kan det likevel være bryet verdt å gjennomføre en RCT før det innføres i storskala. I noen tilfeller har RCT-er nemlig vist at tradisjonelle oppfatninger om hva som virker, ikke stemmer. Det en trodde var effektive tiltak, kan vise seg å ikke ha noen effekt – eller i verste fall virke mot sin hensikt.

Et velkjent eksempel på dette er tiltaket «Scared Straight», et program for å redusere ungdomskriminalitet, som startet i New Jersey på 1970-tallet. Unge lovbrøtere og ungdommer i faresonen møtte tunge lovbrøtere som lærte dem om livet i fengsel. Målet var å skremme ungdommene bort fra en kriminell løpebane. Tall viste at programmet var svært vellykket – nesten ingen av ungdommene begikk kriminalitet tre måneder etter deltakelse. En Oscarvinnende dokumentar fra 1978 viste en suksessrate på 94 prosent, og programmet ble senere spredt til 30 stater i USA.

Men virket egentlig programmet? I 1982 ble det gjennomført et kontrollert eksperiment der deltakere ble valgt ut tilfeldig til eksperiment- og kontrollgrupper. Konklusjonen var at Scared Straight faktisk bidro til *mer* kriminalitet¹¹. Senere studier har bekreftet dette – programmet har bidratt til opp mot 30 prosent økning i kriminalitet blant deltakerne¹². Allikevel eksisterer programmet fortsatt.

2.1.2 Er det ikke uetisk?

Enkelte er motstandere av randomiserte forsøk i utformingen av offentlige tjenester fordi de mener at det er uetisk å holde tilbake nye tiltak fra personer som kunne ha nytte godt av dem. Dette gjelder særlig når det blir satt inn ekstra ressurser i det aktuelle tiltaket.

Det kan åpenbart være uetisk med eksperimenter, særlig overfor sårbare grupper. Det er for eksempel uetisk å utsette noen for et tiltak man har grunner til å tro kan ha negativ effekt. Det er også uetisk å bruke ressurser på å prøve ut noe man allerede vet virker.

¹⁰ Hentet fra Haynes et. al (2012).

¹¹ Årsakene til dette vet man ikke

¹² Petrosino m.fl. (2013).

Samtidig settes svært ofte politiske tiltak i verk uten at vi vet om de virker. Dette gjør ikke tiltakene mer etiske. Tvert imot kan det være mer uetisk å utsette for eksempel alle deltakere i introduksjonsprogram for noe vi ikke aner effekten av, enn å prøve ut lovende tiltak på et mindre utvalg på en systematisk måte. Det er dessuten mye billigere.

Nye offentlige tjenester rulles dessuten ofte ut gradvis, og noen geografiske områder får det før andre, uten at dette blir sett på som uetisk.

2.1.3 Er det ikke veldig dyrt?

Kostnadene avhenger av hvilket tiltak det er snakk om, og hvilket design man legger opp til. I kapittel 3 og 4 vil vi vise eksempler på at RCT-er ikke nødvendigvis trenger å være så dyre. Effektevalueringer av offentlige tiltak og tjenester kan også benytte seg av at tiltaket eller virkemiddelet allerede blir tilbudt, dersom de er innført på en måte som gjør at man kan konstruere naturlige kontrollgrupper i ettertid (naturlige eksperimenter)¹³.

Selv om en RCT kan bli nokså kostbar, kan det være verdt å bruke ekstra ressurser på å få klarhet i hvorvidt et tiltak virkelig gir ønsket effekt – og dermed ønsket utbytte for pengene tiltaket koster. Man kan hevde at det er store skjulte kostnader ved å ikke gjennomføre RCT-er, gjennom at det brukes svært mye offentlige penger på tiltak vi ikke kan vite om faktisk virker. Dersom kunnskapen om hva som virker øker, vil også offentlig pengebruk kunne effektiviseres.

2.1.4 Er det lov å gjennomføre slike forsøk?

I utgangspunktet ja. Folketrygdloven § 25-13¹⁴ åpner for forsøksvirksomhet i en del tilfeller:

«Departementet kan samtykke i avvik fra bestemmelsene i denne loven i forbindelse med forsøksvirksomhet. Forsøksvirksomheten må være faglig og økonomisk vel underbygd. -- Avviket må ikke innebære at enkeltpersoner får innskrenket sine rettigheter eller blir pålagt større plikter enn det som følger av loven».

Det er dessuten særskilt åpnet opp for forsøk i forbindelse med gjennomføring av introduksjonsprogram eller opplæring i norsk og samfunnskunnskap for voksne innvandrere. Introduksjonsloven¹⁵ ble endret fra 1.1.2017 med en ny § 27 om forsøksvirksomhet:

¹³ Mer om naturlige eksperimenter i avsnitt 2.2

¹⁴ Lov om folketrygd (folketrygdloven) (1997).

¹⁵ Lov om introduksjonsordning og norskopplæring for nyankomne innvandrere (introduksjonsloven) (2003)

«Departementet kan etter søknad fra kommunen godkjenne unntak fra loven og forskriftene etter loven i forbindelse med tidsavgrensede pedagogiske eller organisatoriske forsøk.»

2.1.5 Men det er jo allerede bestemt at tiltaket skal innføres?

I mange tilfeller er det allerede politisk vedtatt at et tiltak skal iverksettes og rulles ut nasjonalt. I slike tilfeller kan det allikevel være mulig å effektevaluere tiltaket, gjennom at det rulles ut gradvis, og ikke likt på alle steder. Ofte rulles politiske programmer ut gradvis uansett, blant annet med tanke på kapasitets- og kompetanseoppbygging. Hvorfor ikke designe utrulling slik at det kan effektevalueres?

2.2 Kvasieksperimentelle metoder

I mange tilfeller er det ikke mulig å gjennomføre eksperimenter (RCTer) i rendyrket form. For eksempel kan politikerne ha vedtatt en reform som omfatter hele befolkningen, og som gjøres gjeldende for hele befolkningen samtidig. Da er forutsetningene for å gjennomføre en RCT allerede brutt.

I slike tilfeller, der randomisering til tiltaks- og kontrollgruppe ikke er mulig, kan man imidlertid bruke analysestrategier som gjør det mulig å gi gode estimater av hvorvidt reformen har hatt effekt. Samlebetegnelsen for slike analyseteknikker kalles *kvasieksperimentelle metoder*. Slike metoder baseres på ulike teknikker for å etablere tiltaks- og kontrollgruppe, som imiterer den tilfeldige fordelingen i en RCT.

Et eksempel på at man kan bruke kvasieksperimentelle metoder til å estimere effekt, er når man kan identifisere et såkalt *naturlig eksperiment*. Man har en tilnærmet tilfeldig fordeling av personer i forsøks- og kontrollgruppe, og som er forårsaket av en «naturlig» oppstått hendelse, i den forstand at det er forhold utenfor forskernes kontroll som forårsaker eksperimentet.

En viktig forutsetning for å kunne benytte en reform eller regelendring som et naturlig eksperiment er at reformen blir implementert slik at individer ikke kan påvirke hvorvidt de omfattes av reformen eller ikke. Naturlige eksperimenter kan ses som et «lykketreff» for forskere, fordi randomisering av forsøkspersoner allerede er gjort, uten at forskerne selv må designe et forsøk¹⁶.

I virkeligheten er det imidlertid vanskelig å finne naturlige eksperimenter som fyller kravet til fullstendig tilfeldig fordeling mellom forsøkspersoner og kontrollgruppe. Forskerne studerer tross alt fenomenet etter at en endring er satt i verk. Hvis det allikevel er grunn til å tro at de to gruppene er like nok, kan forskerne ved hjelp av ulike

¹⁶ Les mer om naturlige eksperimenter i Ugreninov og Birkelund (2013).

kvasieksperimentelle analyseteknikker kontrollere for eventuelle kjente forskjeller mellom gruppene.

Kvasieksperimentelle metoder kan altså gjøre det mulig å si noe om kausalsammenhenger selv om et tiltak allerede er bestemt og kanskje iverksatt uten at man har randomisert deltakerne. Slik kan man også observere langsiktige effekter av tiltak som ble gjennomført for lenge siden.

Eksempler på kjente kvasieksperimentelle metoder er Regression discontinuity design, Differences-in-Differences og instrumentvariabelmetoden. Vi skal ikke gå nærmere inn på de ulike kvasieksperimentelle metodene her¹⁷, men vi vil beskrive et par eksempler på situasjoner der man kan estimere effekt uten at forutsetningene for en RCT er til stede.

2.2.1 Eksempel: Regelverksendring

Fra arbeids- og velferdspolitikken finner vi et eksempel på et prosjekt som utnytter et naturlig eksperiment til å undersøke effekten av ordinær utdanning som arbeidsmarkedstiltak for unge med nedsatte arbeidsevner, sammenlignet med andre tiltak¹⁸.

Blant unge som trenger bistand for å komme i jobb, tilbyr NAV ulike tiltak: korte, yrkesrettede kurs, eller støtte til å ta vanlig utdanning. Det har imidlertid manglet sikker kunnskap om hvilke tiltak som er best egnet til å få unge med nedsatt arbeidsevne inn i det ordinære arbeidslivet.

En utfordring når en skal undersøke effekter av arbeidsmarkedstiltak, er at unge med nedsatt arbeidsevne selv er med på å bestemme hvilke tiltak de skal få, i samarbeid med saksbehandleren i NAV. Det er med andre ord ikke tilfeldig hvem som får støtte til å ta vanlig utdanning, og hvem som får tilbud om andre arbeidsrettede tiltak. For å ta høyde for seleksjonsproblemer, har forskerne benyttet ulike kvasieksperimentelle analyseteknikker. De har utnyttet en endring i regelverket fra 1. januar 2004, da aldersgrensen for støtte til ordinær utdanning som arbeidsmarkedstiltak ble hevet fra 22 til 26 år.

Hevingen av aldersgrensen førte til at sammenlignbare grupper unge med nedsatt arbeidsevne ble vurdert under ulike regelverk. Forskerne har sammenlignet personer mellom 22 og 25 år halvåret før regelverksendringen med personer i samme aldersgruppe halvåret etter regelverksendringen. Man har altså sammenlignet en gruppe som har hatt ordinær utdanning som et mulig arbeidsmarkedstiltak, med en

¹⁷ Les mer i Angrist and Pischke (2009).

¹⁸ Prosjektet ble gjennomført av NIFU – Nordisk Institutt for studier av innovasjon, forskning og utdanning.

gruppe som i hovedregel¹⁹ ikke lenger hadde denne muligheten. Utfallet er tilknytning til arbeidslivet 9 år etter. Er det noen forskjell mellom gruppene?

Etter endringen i regelverket fikk altså færre under 26 år støtte til ordinær utdanning, og ifølge rapporten ble de i hovedsak tildelt tiltak i skjermet eller ordinær bedrift. Hovedkonklusjonen til forskerne er at unge som får støtte til å ta vanlig utdanning som tiltak, verken øker eller minker jobbsjansene sine på lang sikt²⁰.

2.2.2 Eksempel: Klare objektive kriterier

Et annet eksempel er fra utdanningsfeltet. I 2012 bevilget Stortinget 1,5 millioner kroner over fire år for å øke lærertettheten på ungdomstrinnet blant skoler med dårlige resultater. Vurderingen av hvilke skoler som fikk tildelt ekstra ressurser ble gjort etter klare objektive kriterier: Skolene måtte blant annet ha grunnskolepoeng under snittet for landet og mer enn 20 elever per lærer på ungdomstrinnet. Dermed kunne forskerne som evaluerte satsningen²¹ sammenligne skoler som (så vidt) kvalifiserte for deltakelse med skoler som (så vidt) ikke kvalifiserte²².

I dette konkrete tilfellet bygget en slik sammenligning på at skoler som på et gitt tidspunkt har nesten like gjennomsnittlige grunnskolepoeng - eller nesten lik gruppestørrelse - også på andre områder kan forventes å være (nesten) like. Disse i utgangspunktet like skolene henholdsvis får og ikke får ekstra ressurser. Etter at tilskuddet har fått virke en tid, vil det kunne oppstå en forskjell i resultater mellom skoler som er like over/under grensen. Basert på antakelsen om at skolene i utgangspunktet var like, og ville forblitt like i fravær av de ekstra ressursene, kan man tolke en eventuell forskjell som effekt av de ekstra ressursene.

Forskerne fant ingen effekt på elevenes læringsutbytte (eks. karakterer eller fravær) eller skolemiljøet ved de skolene som fikk ekstra lærere²³.

¹⁹ En del ble i praksis omfattet av en unntaksbestemmelse, og hadde fortsatt denne muligheten

²⁰ Salvanes, Kari Vea og Rune Borgan Reiling (2017).

²¹ Satsningen ble evaluert av Statistisk Sentralbyrå, Frischsenteret og Fafo.

²² Ved bruk av såkalt Regression Continuity Design

²³ Kirkebøen et.al (2017).

3 Hva har blitt gjort i Norge? Erfaringer fra velferdsfeltet

Nye medikamenter og medisinske behandlinger må testes grundig før de blir tatt i bruk, mens politiske satsinger som regel ikke blir gjenstand for samme granskning på forhånd. Ofte finnes det ingen bevis for at de faktisk vil virke før de innføres.

I Storbritannia har det regjeringsinitierte forskningsmiljøet Behavioural Insights Team (BIT) tatt til orde for at alle nye offentlige tiltak og virkemidler bør effektevalueres gjennom randomiserte, kontrollerte forsøk før de settes ut i livet. Fra opprettelsen i 2010 og frem til i dag har de gjennomført flere enn 300 randomiserte forsøk, også på politikkområder hvor en tidligere antok at det ville være svært vanskelig, som i sosialpolitikken²⁴.

Her hjemme er vi fortsatt i startgropen når det gjelder å gjennomføre RCT-er i velferdspolitikken, men dette er i ferd med å endre seg. Det er etter hvert gjennomført flere RCT-er på velferdsfeltet, også i Norge. I det følgende vil vi trekke fram noen eksempler fra velferdsfeltet som vi kan trekke lærdom av til gjennomføring av lignende eksperimenter på integreringsfeltet.

3.1 Arbeids- og velferdspolitikken

Arbeids- og velferdsdirektoratet (AVdir) har et ansvar for å sikre at tjenestene og virkemidlene i arbeids- og velferdsforvaltningen i størst mulig grad er kunnskapsbaserte. NAVs kunnskapsstrategi 2013-2020²⁵ har blant annet som formål å utvikle tjenester og virkemidler basert på kunnskap om hva som virker. Det er nå flere gode eksempler på vel utførte studier der NAV er oppdragsgiver, og vi nevner noen her som er interessante for integreringsfeltet.

Uni Research Helse har gjennomført en **effektevaluering av individuell jobbstøtte** (IPS: Individual Placement Support) for personer med moderate til alvorlige psykiske lidelser i perioden 2013-2016²⁶. Evalueringen ble gjennomført som en RCT, og supplert med kvalitative metoder som intervjuer og dokumentanalyse. Behandlingsgruppen fikk tilbud om oppfølgingsmetodikken IPS, som besto av tett, individuell oppfølging, koordinerte tjenester fra arbeid- og helsesektorene og bruk av vanlige arbeidsplasser som treningsarena. Kontrollgruppen fikk oppfølging uten bruk av IPS.

²⁴ Les mer her: <https://www.bi.team/>

²⁵ NAV (2012)

²⁶ Uni Research Helse/Uni Research Rokkansenteret (2016). Les mer: <https://www.nav.no/no/NAV+og+samfunn/Kunnskap/FoU-midler/Relatert+informasjon/effektevaluering-av-individuell-jobbst%C3%B8tte-ips>

Forskerne fant en tydelig effekt i favør av IPS når det gjaldt overgang til arbeid, men også når det gjaldt livskvalitet og selvopplevd helse. Samlet sett ga dette et solid grunnlag for å anbefale videreføring av IPS-tilbudet.

IPS er en metode som i utgangspunktet har til hensikt å få psykisk syke personer inn i ordinært arbeidsliv. Metoden prøves nå også ut for andre grupper som oppfølging av prosjektet nevnt over, blant annet flyktninger i introduksjonsprogrammet. Uni Research Helse er nå i gang med å gjennomføre en **effektevaluering av «Raskt i jobb» for flyktninger i Bergen kommune**²⁷. «Raskt i jobb» har som hensikt å hjelpe flyktninger i introduksjonsprogrammet i kommunen ut i ordinært arbeidsliv ved hjelp av IPS-metodikken. Ved å gjennomføre en RCT vil forskerne sammenligne deltakere i «Raskt i jobb» med en kontrollgruppe, og gi et svar på hvorvidt satsningen faktisk gir bedre resultater i form av arbeidsdeltakelse, enn det dagens praksis gjør.

Dette prosjektet er et eksempel på at et tiltak (bruk av IPS) som har vist seg å ha effekt for en gruppe, prøves ut på en annen. Slike prosjekter kan gi viktig lærdom inn i utviklingen av RCT-er på integreringsfeltet: Det er mulig å ta utgangspunkt i tiltak man vet virker på andre grupper, og undersøke hvorvidt de også har effekt på våre målgrupper (flyktninger og nyankomne innvandrere).

Et annet eksempel er et pågående **forsøk med Supported Employment (SE)**²⁸, som undersøker om individuell oppfølging basert på kvalitetskravene i metoden Supported Employment fører til at flere personer med nedsatt arbeidsevne får og beholder arbeid i ordinært arbeidsliv, sammenlignet med oppfølgingstiltak hvor det ikke er satt tilsvarende krav til at denne metodikken benyttes. Personer som har nedsatt arbeidsevne og behov for et oppfølgingstiltak for å komme i jobb, blir trukket ut til deltakelse enten i SE-forsøket eller i et annet oppfølgingstiltak. De som arbeider med å bistå personer ut i arbeid i SE-forsøket er underlagt mer detaljerte og manualbaserte krav til arbeidsmåter fra NAV enn det som gjelder for andre oppfølgingstiltak.

Et tredje pågående forsøk i NAV er **Helhetlig oppfølging av lavinntektsfamilier (HOLF)**²⁹, som prøver ut en modell for helhetlig oppfølging av lavinntektsfamilier. Forskningen gjennomføres som en randomisert kontrollert studie, hvor NAV-kontorene som er med i undersøkelsen fordeles på en tiltaksgruppe og kontrollgruppe. Forsøket er altså grupperandomisert. Målgruppen, målområdene og ressurstilfør-

²⁷ Prosjektet varer i perioden 2017-2020

²⁸ Prosjektet varer i perioden 2014-2018, og AVdir har ansvaret for å koordinere forsøket (NAV 2018a)

²⁹ NAV (2018b). Evalueringen gjennomføres av forskere ved OsloMet, og det har kommet to underveisrapporter: (Malmberg-Heimonen m.fl. 2016, Malmberg-Heimonen m.fl. 2018)

selen er den samme for begge kontorgruppene. Alle kontorene skal tilsette to familiekoordinatorer som skal følge opp familiene. Kontorene i tiltaksgruppen forplikter seg til å følge en bestemt oppfølgingsmetodikk i arbeidet, mens kontorene i kontrollgruppen velger organisering og arbeidsmetodikk selv.

Den kortsiktige målsettingen er å utvikle og prøve ut en oppfølgingsmodell som kan gi bedre resultater for familiene i målgruppen. Det overordnede og langsiktige målet er å forhindre at fattigdom går i arv ved å forebygge at barn i lavinntektsfamilier forblir fattige. Prosjektet har altså både kortsiktige og langsiktige utfallsmål.

Prosjektet gjennomføres som et samarbeid mellom brukere, praktikere og forskere, i to faser. I fase 1 (utviklingsfasen) ble metode og verktøy utviklet, og prosjektet ble pilotert i i tre NAV-kontorer. Prosjektmedarbeidere fra disse tre kontorene har bidratt i utvikling og utprøving av metoder og verktøy. I fase 2 (forskningsfasen) blir prosjektet prosess- og effektevaluert. Gjennom *effektevalueringen* ønsker man å finne ut om det er effekter av modellen som er utviklet, og om denne er bedre enn «business as usual». Gjennom en tilhørende *prosessevaluering*³⁰ vil man finne ut hvilke mekanismer ved bruk av modellen og metodikken som har størst betydning.

3.2 Utdanningsfeltet

I Kunnskapsdepartementet (KD) har man de seneste årene vært opptatt av at nye tiltak bør prøves ut og testes før de rulles ut i stor skala, gjennom «skrittvis reform». Tanken om skrittvis reform i KD innebærer først å innhente relevant kunnskap om tiltaket som skal prøves ut. Deretter ønsker man å sørge for en grundig modellutvikling og pilotering av tiltaket³¹. Deretter prøves tiltaket ut systematisk, helst som en RCT. Først etter en slik systematisk utprøving bør det tas en beslutning om en eventuell oppskalering av tiltaket.

Forskningsrådet fikk i 2017 i oppdrag å etablere ordningen **Forskning på effekter av tiltak for økt kvalitet i barnehage og grunnskole**, der kommuner og forskere sammen kan søke om støtte for å gjennomføre systematiske forsøk³². I ordningen vil det bli stilt strenge krav til design som gjør det mulig å måle effektene av tiltakene som prøves ut. Prosjektene skal også gi kunnskap om hva som forklarer hvorfor tiltakene har effekt og hvordan tiltakene ble lokalt forankret og implementert. Ordningen skal sørge for at resultatene fra utprøvingene formidles og kan tas i bruk av relevante aktører som skole- og barnehageeiere og lærere og barnehagelærere. For at kunnskapen fra utprøvingene av tiltakene skal bli tatt i bruk, må prosjektene vise

³⁰ Les mer om prosessevaluering i kapittel 4

³¹ Les mer om modellutvikling og pilotering i kapittel 4

³² Les mer på forskningsrådet.no https://www.forskningsradet.no/no/Utlysning/FINNUT/1254035549619&WT.mc_id=prgfor-utl-FINNUT

hvordan vellykkede tiltak kan spres og oppskaleres. Det vil bli stilt krav om at prosjektene gir en vurdering av realismen i en eventuell oppskalering av tiltakene, herunder tiltakenes kostnadseffektivitet.

KD vil foreslå bevilgninger til programmet gjennom de årlige statsbudsjettene. Innenfor programmet kan det tildeles midler både til forskningsaktiviteten og til å gjennomføre tiltakene i sektoren. For eksempel kan det avsettes midler til prosjektkoordinatorer på skoler/barnehager eller i fylker/kommuner. Skoleeierne/barnehageeierne kan også avsette egne midler til tiltakene. Det er foreløpig lagt få innholdsmessige føringer for hvilke tiltak som skal prøves ut innenfor ordningen, det er opp til forskere i samarbeid med kommunene å utforme tiltakene.

Konkrete eksempler på RCT-er på utdanningsfeltet:

Program for bedre gjennomføring (PBG) i regi av KD, er en satsning for å få flere ungdommer til å fullføre og bestå videregående opplæring. En del av programmet handler om å sette i gang systematiske forsøk, som ble startet opp høsten 2016.

To prosjekter prøver ut intensivopplæring i matematikk på ungdomstrinnet og i videregående, ett prosjekt prøver ut to programmer for bedre læringsmiljø med mål om økt gjennomføring og det siste prosjektet skal undersøke effektene på gjennomføring av at fire fylkeskommuner tar i bruk et system for identifisering og oppfølging av risikoelever i videregående. Det er for tidlig å si noe om resultatene av prosjektene.

I alle prosjektene gjennomføres utprøvingene i et tett samarbeid mellom forskere og fylkeskommunene. Det er fylkene og forskerne som har utviklet og som gjennomfører tiltakene. KD la seg ikke opp i tiltakenes innhold, men stilte krav om at de skulle designes slik at det var mulig å måle effekter³³.

Et annet eksempel er «**Et lag rundt eleven**», i regi av Utdanningsdirektoratet, som skal undersøke hvordan det er hensiktsmessig å bruke flerfaglig kompetanse i skolen³⁴. Dette kan handle om økt tverrfaglig samarbeid mellom aktører i og rundt skolen, styrking av utvalgte ressurser, eller kombinasjoner av dette.

I det ene prosjektet får noen skoler i 15 norske kommuner en ekstra helsesøsterressurs i en toårsperiode. Prosjektet er en RCT hvor fire skoler i hver av de 15 samarbeidskommunene er valgt ut. De øvrige skolene i den enkelte kommune følges opp som kontrollgruppe. Etter to år skal det undersøkes om det er endringer i lærings-

³³ Les mer på regjeringen.no <https://www.regjeringen.no/no/dep/kd/Bedre-gjennomforing-i-videregaende-/id2005356/>

³⁴ Leser mer på www.lagrundteleven.no

miljø, elevenes fravær og resultater på nasjonale prøver. I tillegg vil prosjektet undersøke og vurdere implementeringspraksis slik at det er mulig å spre tiltaket til andre skoler om det viser seg å være vellykket.³⁵

I det andre prosjektet skal den såkalte «LOG-modellen» prøves ut og effektevalueres fra og med skoleåret 2017/18³⁶. Modellen er et systematisk opplegg for ledelse, organisering og gjennomføring av en utviklingsprosess for å oppnå bedre flerfaglig samarbeid. Dette betyr at alle grunnskolene i deltagende kommuner deltar i evalueringen, men halvparten av skolene blir tilfeldig trukket ut til en tiltaksgruppe og halvparten til en kontrollgruppe. Undersøkelsen vil gi svar på om skoler som bruker LOG-modellen (tiltaksgruppen) forbedrer det flerfaglige arbeidet slik at læringsmiljø, læringsutbytte og om opplevd profesjonell kompetanse blir bedre enn i skoler som ikke bruker modellen (kontrollgruppen).

I **prosjektet «Fysisk aktivitet»** i regi av Utdanningsdirektoratet, skal forskere evaluere effekten av to modeller for økt fysisk aktivitet og kroppsøving på fysisk og psykisk helse, læring og læringsmiljø blant elever på 30 ungdomsskoler fra og med 2017³⁷. Noen av skolene kommer blant annet til å ha en ekstra kroppsøvingstime i uka, noen av undervisningstimene i fag vil være utendørs og det vil være flere aktive pauser fra stillesittende undervisning.³⁸

3.3 Valgforskningen

I USA er det gjennomført mange RCT-er som undersøker hvorvidt ulike tiltak (stimuli) rettet mot potensielle velgere har effekt på valgdeltakelsen. Det er kun få eksempler på lignende forskningseksperimenter i europeiske land. De senere årene er det imidlertid gjennomført to vellykkede eksperimenter i Norge, som også inkluderer egne tiltak rettet mot personer med innvandrerbakgrunn.

Institutt for Samfunnsforskning har både ved kommunevalget i 2015 og stortingsvalget i 2017 gjennomført **eksperimenter med tiltak for å øke valgdeltakelsen**, der eksperimentgruppen mottar en oppfordring om å stemme, mens kontrollgruppen ikke gjør det³⁹.

Eksperimentene handlet om å få grupper som ellers ikke ville ha stemt til å delta i politiske valg ved å gi dem en påminnelse, eller en liten dytt («nudge») i retning av å

³⁵ Prosjektet gjennomføres av Nordisk institutt for studier av innovasjon, forskning og utdanning (NIFU) i samarbeid med Læringsmiljøsentret ved Universitetet i Stavanger og Folkehelseinstituttet.

³⁶ Modellen implementeres og evalueres i samarbeid med forskere fra Høgskolen i Oslo og Akershus og Universitetet i Tromsø

³⁷ Kolle m.fl. (2016).

³⁸ Forsøket gjennomføres av forskere ved Norges idrettshøgskole, Universitet i Agder, Universitetet i Stavanger og Høgskulen på Vestlandet.

³⁹ Bergh, Johannes et.al. (2016), Bergh, Johannes et.al. (2018).

stemme. Ved begge valg ble det sendt ut SMS-er med en påminnelse og oppfordring om å stemme til tilfeldig utvalgte velgere. Velgere som ikke fikk SMS-ene, utgjorde kontrollgruppen i eksperimentet. I tillegg ble det sendt ut mer omfattende brev til velgere med innvandrerbakgrunn. Brevene inneholdt praktisk informasjon om stemmegivningen samt et avsnitt som varierte mellom tre ulike brev.

Resultatene fra eksperimentene viser at tiltaket med SMS hadde en effekt på valgdeltakelsen i 2015, og størst effekt blant unge og innvandrere.

Både i 2015 og 2017 var det brev til personer med innvandrerbakgrunn som hadde størst effekt. Ved lokale valg har utenlandske statsborgere stemmerett; ved stortingsvalg har bare norske statsborgere stemmerett. Gruppene av velgere med innvandrerbakgrunn er derfor ikke direkte sammenliknbare for de to valgene.

Generelt hadde de ulike tiltakene størst effekt i 2015. En sannsynlig forklaring er forskjellen i deltakelsesnivåene ved de to valgene. Ved stortingsvalg er deltakelsen høyere, og personene som mottar tiltak ville ha stemt uansett. Ved lokalvalget i 2015 var det nettopp i grupper med lav deltakelse, unge velgere og innvandrere (SMS), og utenlandske statsborgere (brev) at effekten var størst. Forskerne anbefaler derfor at det vil være mest effektivt å innføre slike tiltak i større skala ved lokalvalg.

Statlige myndigheter har ved mange valg gjennomført mobiliseringskampanjer for å få folk til å stemme. For eksempel har IMDi jobbet for å øke valgdeltakelsen i innvandrerbefolkningen, gjennom for eksempel postkort og valgbrosjyrer. Men man vet ikke om disse mobiliseringstiltakene faktisk har bidratt til økt valgdeltakelse. Disse eksperimentene gir oss økt kunnskap om hvorvidt tiltak for å øke valgdeltakelsen faktisk har en effekt, og for hvilke grupper.

Dette er et eksempel på at en RCT kan gjennomføres i stor skala uten at det trenger å være så dyrt og omfattende. Man utnytter at valggjennomføringen uansett skal skje, og tiltaket (sende ut SMS og brev) koster lite.

3.4 Diskriminering og holdninger

I Norge er det de senere år to ganger blitt gjennomført **randomiserte felteksperimenter for å måle diskriminering i arbeidslivet**⁴⁰. Det er en lang forskningstradisjon for slike eksperimenter internasjonalt, også i Sverige og Danmark.

⁴⁰ Midtbøen og Rogstad (2012), Birkelund m.fl (2014).

Forskerne har identifisert utlyste stillinger og sendt to søknader fra fiktive kandidater til hver stilling. «Søkerne» var kvalifiserte for stillingen de søkte på, og hadde lik utdanning og arbeidserfaring. Den eneste forskjellen var at en av søkerne hadde et typisk norsk navn, og den andre et typisk utenlandsk (pakistansk-muslimsk) navn.

Resultatene fra begge eksperimenter viser at jobbsøkere med utenlandsk navn blir diskriminert i arbeidslivet, og har mindre sannsynlighet for å bli innkalt til intervju.

Begge felteksperimentene er supplert med kvalitative intervjuer med arbeidsgivere. Det siste felteksperimentet har vært en del av et større forskningsprogram finansiert av forskningsrådet om diskriminering i arbeidslivet og er supplert med registerdata, samt intervjudata for å forstå hvordan arbeidsgiverne resonnerer når de bestemmer hvem som skal kalles inn til intervju. Slik får man også større innsikt i hvorfor og hvordan den påviste diskrimineringen forekommer.

Det er også gjennomført et felteksperiment for å undersøke **effekten av personlig kontakt med innvandrere på holdninger til innvandring og integrering**. Bakgrunnen for dette er det som kalles «kontaktteori»; er de som har mer kontakt med innvandrere også mer positive til innvandring og integrering, og i hvilke kontekster? Dette har det allerede blitt forsket en del på, men først og fremst gjennom spørreundersøkelser. Det har manglet et eksperiment som følger en gruppe personer over tid.

Soldater ved en brigade i Nord-Norge ble plassert tilfeldig sammen på ulike rom. Det ble randomisert hvem som skulle bo med en person med innvandrerbakgrunn og ikke. Soldatene ble intervjuet den første dagen, og helt på slutten av en åtte ukers rekruttperiode. Resultatene viste at soldatene som bodde sammen med en person med innvandrerbakgrunn, endret klart syn på innvandreres arbeidsmoral. De endret imidlertid ikke syn på om innvandrere bør ha samme rett til sosiale goder som nordmenn, og om de var positive eller negative til innvandring.⁴¹

De to sistnevnte prosjektene handler ikke om å prøve ut tiltak og undersøke om de har effekt, som de andre eksemplene vi tar for oss her. I disse prosjektene man benyttet randomisering som metode og konstruert eksperimenter for å *måle* forekomsten av et fenomen, altså om og i hvilken grad noe forekommer.

Diskriminering og holdninger er vanskelig å måle, blant annet fordi det ofte baserer seg på selvrapportering. Randomisering er en metode som kan bidra til å gi oss mer sikker kunnskap, som er viktig for å igangsette effektive tiltak for å forebygge eller begrense forekomsten av diskriminering.

⁴¹ Henning Finseraas m.fl (2017)

4 Hva skal til for å lykkes med effektevalueringer?

Det er altså gjennomført få systematiske forsøk og effektevalueringer på integreringsfeltet, både i Norge og internasjonalt. Samtidig etterspørres kunnskap om effekten av ulike integreringstiltak fra mange hold.

Gjennomgangen så langt har vist at RCT-er på velferdsfeltet er relativt nytt, særlig i Norge. Det er ikke bare på integreringsfeltet det er et problem med lite kunnskap om effekter av tiltak. Utfordringen gjelder for hele velferdsfeltet. Etter hvert har imidlertid ulike sektorer gjennomført flere forsøk, som eksemplene i kapittel 3 har vist.

Vi har innhentet erfaringer fra velferdssektorer som har erfaringer med å bestille og legge til rette for effektevalueringer. Vi har også hatt dialog med forskningsmiljøer som har erfaring med gjennomføring av RCT-er og kvasiekperimentelle studier⁴². Hva erfarer bestillere og forskere som avgjørende for å lykkes med effektevalueringer?

Det er mye man skal tenke på i utviklingen av en RCT, og det kan være komplisert og kostbart. Basert på erfaringene fra både bestillere og utførere av effektevalueringer på velferdsfeltet, vil vi her trekke frem noen grep man kan gjøre med sikte på å oppnå vellykkede prosjekter. Noen av grepene er det forskerne som må gjøre, mens andre er det særlig bestillere som må tenke på.

4.1 Hvordan få det til? Viktige læringspunkter

Start tidlig! Planlegg for evaluering mens tiltaket utvikles

Det er viktig å planlegge for evaluering samtidig som tiltak utvikles. Dette er en god regel for alle typer evalueringer, men er ekstra viktig for gjennomføring av en RCT. Hvis man starter implementeringen uten å ha designet for eksperimentet klart, kan man ha ødelagt muligheter eller gjort det vanskeligere å måle effekter. Man må sikre at tiltaket implementeres på en måte som gjør at det faktisk kan effektevalueres. De som utvikler tiltak og bestiller forskning, har sjelden den nødvendige kunnskapen og erfaringen for å designe en RCT. Derfor bør forskere med metodekompetanse involveres så tidlig som mulig når tiltak utvikles.

⁴² Dialogen har foregått gjennom anskaffelsesmetoden «dialog med markedet», der vi har gjennomført én-til-én-møter med interesserte leverandører etter en offentlig utlysning på Doffin.

Hvis man som bestiller av en evaluering ikke er sikker på om tiltaket som skal testes ut egner seg for å effektevalueres, kan det være lurt å involvere forskere i vurderingen. Det kan gjøres både i forkant av utlysning, gjennom dialog med markedet, eller man kan i utlysningen be tilbydere vurdere om tiltaket lar seg teste ut på en måte som kan gi mål på effekt.

Det er også et alternativ å gjennomføre en anskaffelse i flere faser, enten med separate forprosjekt om å utvikle design for evalueringen, eller med en konkurranseform der utvikling av design er første fase. En annen mulighet er å lyse ut en konkurranse om å utvikle og teste design for forsøket, med opsjoner om gjennomføring av selve forsøket⁴³.

IMDi har for eksempel satt i gang et prosjekt der en gruppe forskere skal utvikle modeller med bruk av økonomiske insentiver i introduksjonsprogrammet og opplæringen i norsk og samfunnskunnskap. Prosjektet er lyst ut med opsjoner om gjennomføring av effektevaluering (RCT) dersom man velger å prøve ut en eller flere av modellene.⁴⁴

Velg tiltak det er grunn til å tro på: Kunnskapsbasert utprøving

Et randomisert kontrollert eksperiment trenger ikke å være dyrt, randomisering i seg selv er gratis. Hvis man allikevel skal sende ut SMS til hele befolkningen med en oppfordring om å stemme på valgdagen, koster det ikke mye å eksperimentere med ulike SMS-er etter tilfeldig fordeling. Men utprøvinger av tiltak som handler om at mennesker skal endre atferd på en bestemt og varig måte, som for eksempel at lærere skal ta i bruk en ny metode for å undervise i norsk, er komplisert. Det tar lang tid å planlegge, lang tid å få resultater og er ofte kostbart. Ved utprøving av tiltak som krever mye ressurser, er det særlig viktig at man har tro på tiltaket. Test bare ut tiltak det er god grunn til å anta at har effekt.

Å ha begrunnet tro på tiltaket, handler også om forvaltning av ressurser. Å gjennomføre en RCT er som regel tidkrevende, og i mange tilfeller dyrt. Hvis utprøvingen handler om at noen skal få et annet tilbud enn de ellers ville ha fått, kan forsøket dessuten være uetisk dersom man ikke har god grunn til å tro at tiltaket som testet ut vil gi effekt.

Et godt startpunkt for å identifisere lovende tiltak, er en kunnskapsoppsummering som sammenfatter internasjonal evidensbasert forskning om problemstillingen man er ute etter å løse. På felt der man ikke har særlig med effektstudier fra før, som på integreringsfeltet, bør man ta utgangspunkt i god praksis fra Norge og andre land: Man kan systematisere kunnskap fra praksisfeltet, for å komme på sporet av noe

⁴³ Et eksempel på en slik anskaffelsesform er konkurransepreget dialog: <https://www.anskaffelser.no/avtaler-og-regelverk/anskaffelsesprosedyrer/konkurransepreget-dialog>

⁴⁴ <https://doffin.no/Notice/Details/2018-327043>

man har grunn til å anta at virker.

Hvorfor virker tiltaket? Viktigheten av programteori

Det holder ikke å tro at et tiltak har effekt. For at eksperimentet skal ha noen overføringsverdi må man vite *hvorfor*. Hvilke mekanismer er det man antar virker?

Et godt prosjekt er forankret i en programteori. Dette er kort sagt en omfattende beskrivelse av hvordan og hvorfor en ønsket endring er forventet å skje i en spesifikk kontekst. En slik beskrivelse må altså være basert på foreliggende kunnskap fra feltet (i tråd med foregående punkt), og være basert på en teoretisk forankring. Hvorfor tror vi et tiltak har effekt? Hvorfor bør det prøves ut?

En programteori er kort sagt ideer om hvordan og hvorfor en intervensjon fører til bestemte resultater. Programteorien bør inneholde beskrivelse av målgrupper, aktiviteter, virksomme mekanismer, delmål og slutt mål.⁴⁵

Sikre godt samarbeid: Forankring hos forskere og praktikere

Eierskap til prosjektet bør være nøye gjennomtenkt, og det er viktig med forpliktelse fra alle parter for å lykkes. En viktig del av prosessen er å legge til rette for dialog mellom forskere, praktikere og brukere.

Det kan for eksempel være tidkrevende å rekruttere aktører som deltagere i eksperimenter, og prosessen må være forankret på alle nivåer. Det er vel og bra om kommuneledelsen har tro på tiltaket, men hva om praktikerne er negative til den nye norskopplæringsmetoden? Da er det sannsynlig at metoden ikke implementeres på god nok måte. Det er også viktig å sikre forståelse i alle ledd for hva som skal til for å lykkes med en RCT, for eksempel trofasthet til det som skal testes ut.

Bruk tid på å «få alle med». Dette er ikke minst viktig i prosjekter som skal gå over lang tid, og der det er fare for at de involverte minster motivasjonen underveis.

Tiltaket bør piloteres

Det som skal testes ut bør altså være presist beskrevet og godt forankret. Tiltaket bør også piloteres og følgeevalueres, for å sikre om det i det hele tatt lar seg gjennomføre i praksis. Pilotering gir blant annet innsikt i praktiske utfordringer ved å gjennomføre tiltaket og om det kan være noen uforutsette (negative) effekter av tiltaket.

⁴⁵ Funnell & Rogers (2011), Malmberg-Heimonen m.fl. (2016)

Pilotering skjer typisk gjennom at tiltaket prøves ut på en mindre gruppe. Ved en eventuell oppskalering kan tiltaket justeres etter lærdommene man fikk under piloteringen.

Kvalitet i implementeringen er alfa og omega

Å gjennomføre en RCT krever svært god planlegging, også for å sikre god kvalitet i implementeringen av tiltaket. Dette gjelder særlig hvis tiltaket krever at personer langt nede i virkemiddelapparatet endrer måten de vanligvis gjør ting på, som for eksempel ved utprøving av en ny undervisningsmetode på skoler. Hvis tiltaket ikke er implementert på en tilfredsstillende måte, har dette store konsekvenser for hvordan man skal fortolke resultatene: Hvis man finner ut at tiltaket ikke hadde noen effekt: Skyldes det selve tiltaket, eller at det ikke ble implementert riktig? Det er mange ting som kan gå annerledes enn man har tenkt seg ved en systematisk utprøving.

God planlegging og forankring i praksisfeltet er avgjørende for god implementering. En prosessevaluering er viktig for å vurdere kvalitet på implementeringen, og kan også gjøre det mulig å gjennomføre noen nødvendige justeringer underveis.

Fidelity er her et sentralt begrep, og kan på norsk oversettes til «trofasthet». Det handler om hvor trofast utøverne er til bruken av metoden/modellen i en gitt utprøving. Fidelity kan også defineres som implementeringskvalitet; hvorvidt og i hvilken grad tiltak blir implementert etter hensikten.

Følg prosjektet underveis: Gjennomfør en god prosessevaluering

Det er viktig å planlegge godt og forankre prosjektet hos alle involverte, for å sikre god implementering. Men hvordan vet man egentlig om prosjektet har blitt implementert i henhold til plan? Og hvordan fanger man opp eventuelle avvik? Ved en god prosessevaluering.

For å bidra til å kaste lys over både *om* en intervensjon virker, *hva* som virker og *hvorfor* den virker, har man i økende grad begynt å kombinere RCT-er med omfattende prosessevalueringer⁴⁶. Relevante metoder man bruker i en prosessevaluering er for eksempel observasjon av hvordan tiltaket blir implementert, intervjuer, og tilstedeværelse på møter osv.

Randomiserte forsøk er altså den metoden som gir de sikreste målene på om et tiltak har effekt, men sier oss lite om hvorfor og hvordan. Randomisering alene er ikke tilstrekkelig for at forsøket skal gi oss den kunnskapen vi er ute etter.

⁴⁶ Malmberg-Heimonen m.fl. (2016)

Er det egentlig verdt det? Kost/nytte-analyse

Planlegging og gjennomføring av en RCT kan være dyrt og ressurskrevende. Kostnadene avhenger av tiltakets kvalitet, karakter og omfang. Det bør gjøres en kost/nytteanalyse av om det er verdt det i hvert enkelt tilfelle, eller om det kan være fornuftig å bruke andre design som kan sannsynliggjøre effekter, men som koster mindre i både tid og penger (kvasiekperimentelt design).

Men selv store kostnader til gjennomføring av en RCT kan være lønnsomt i et lengre perspektiv. Det bør derfor gjennomføres en kost-nytteanalyse også på lang sikt, før tiltaket foreslås innført i storskala. Kanskje hadde tiltaket effekt, men er nytten større enn kostnadene? Nytten av tiltaket bør altså vurderes å være verdt kostnadene på lang sikt.

Problemstilling først, deretter valg av metode

I denne rapporten har vi vist mange gode argumenter for å evaluere tiltak gjennom randomiserte kontrollerte forsøk. Men det er viktig å huske på at det er det metodiske designet som må være tilpasset forskningsspørsmålet, ikke omvendt. Det er ikke alle problemstillinger som best besvares med en RCT. Det er heller ikke alltid mulig å gjennomføre, for eksempel av etiske, politiske eller kostnadmessige hensyn.

Når man skal bestille evaluering av et eller flere tiltak, med ønske om å måle effekter, er det et fornuftig utgangspunkt å være åpen for også kvasiekperimentelle design.

4.3 Nyankomne innvandrere som målgruppe: særskilte utfordringer?

Eksemplene vi har oppsummert i kapittel 3 viser at det bør være mulig å få til systematiske utprøvinger også på integreringsfeltet. Samtidig er det viktig å understreke at erfaringene fra andre sektorer ikke nødvendigvis er direkte overførbare. For eksempel har Arbeids- og velferdsdirektoratet sterkere styringsverktøy overfor kommunene enn integreringsmyndighetene har, noe som kan gjøre det lettere å gjennomføre eksperimenter. Felles for de to feltene er at det kan være utfordringer knyttet til utførelse av eksperimenter med sårbare grupper. Ved gjennomføring av forsøk må man alltid gjøre en vurdering av om det er etisk forsvarlig. Etske vurderinger blir ekstra viktige når målgruppen for forsøket i utgangspunktet er en sårbar gruppe.

Det er også en del særskilte utfordringer når innvandrere med kort botid er målgruppen, for eksempel språk og høyere risiko for frafall. Det kan også være utfordrende

at populasjonen ikke er så stor, særlig hvis man vil teste ut effekt av tiltak på utvalgte grupper (etter for eksempel landbakgrunn og utdanningsnivå).

En annen utfordring er at målgruppen kan variere betydelig fra år til år, avhengig av asylankomster og innvilgede familiegjenforeninger. Det kan være stor variasjon både i antall og i sammensetting (f. eks. landbakgrunn, kjønn, alder, utdanning). Selv om innvandrere ikke har vært definert som målgruppe i eksemplene vi har beskrevet, er det flere av forsøkene som har innvandrere i målgruppen. I NAVs HOLF-prosjekt, er flertallet i målgruppen innvandrere. Erfaring fra dette og andre prosjekter med innvandrere i målgruppen, kan være særlige nyttige å se til ved utforming av forsøk rettet særskilt mot IMDIs målgruppe: nyankomne innvandrere. Erfaringer fra andre forskningsprosjekter med innvandrere som målgruppe, kan også være relevante⁴⁷.

4.4 Oppsummering

- Det er behov for sikrere kunnskap om virkning av integreringstiltak. Integreringsmyndighetene bør i større grad legge til rette for forskning og evalueringer som kan gi troverdige mål på effekter.
- Å måle effekter kan gjøres både gjennom randomiserte kontrollerte studier (RCTer) og med ulike kvasiexperimentelle design. RCTer regnes som gullstandard innen effektstudier, fordi det er det metodiske designet som gir de sikreste målene. Alle problemstillinger lar seg imidlertid ikke besvare med en RCT, og det kan også være etiske, politiske og økonomiske grunner til at man må velge andre metoder.

I denne teksten har vi hovedsakelig fokusert på RCTer, og mindre på andre metodiske design som også kan brukes for å måle effekter (kvasiexperimentelle design).

- Forskere og bestillere som har erfaring med å gjennomføre RCTer på velferdsfeltet i Norge har en del felles erfaringer om hva som vurderes som avgjørende for å lykkes: En RCT bør gjennomføres i flere faser. Evalueringen bør planlegges samtidig som prosjektet utvikles, og være basert på foreliggende kunnskap og en programteori. I første fase utvikler man modellen og forankrer prosjektet, deretter piloteres det, og modellen justeres før en (eventuell) oppskalering og systematisk utprøving, med prosessevaluering. I

⁴⁷ IMDI/Fafos (2010) veileder i brukerundersøkelser beskriver utfordringer som også kan være relevante for andre typer forskningsprosjekter med innvandrere som målgruppe.

mange tilfeller vil man være avhengig av tett samarbeid mellom forskere, byråkrater og praktikere for å lykkes med å designe og gjennomføre en god RCT.

- Ved utforming av RCTer på integreringsfeltet, bør disse erfaringene legges til grunn. Forsøk der nyankomne innvandrere er målgruppe, kan innebære noen særskilte utfordringer som for eksempel språk, økt risiko for frafall, og at populasjonen ikke er så stor. Når målgruppa regnes som en sårbar gruppe, er det også særlig viktig å vurdere etisk forsvarlighet.
- For å få mer kunnskap om virkning av integreringstiltak, bør de som utvikler tiltak samtidig planlegge hvordan det skal evalueres. Dersom det er ønskelig å måle effekt av tiltaket, bør man ikke starte implementeringen før det er laget et metodisk design for evaluering. På den måten holder man muligheten åpen for å teste ut tiltaket gjennom en RCT, som hvis det er gjennomført på en god måte er den sikreste metoden for å få vite om tiltaket har effekt.

Kilder

Angrist, Joshua D. and Jorn-Steffen Pischke (2009). *Mostly Harmless Econometrics. An Empiricist's Companion*. Princeton: Princeton University Press

Backe-Hansen, Elisabeth, Tonje Gundersen, Marie Hatlelid Føleide og Aina Winsvold (2015). "Forskningsoversikt om alvorlige atferdsvansker blant barn og unge". *Velferdsinstituttet NOVA*, November 2015.

Bergh, Johannes, Dag Arne Christensen og Richard E. (2016). "Getting Out the Vote. Experiments in voter mobilization among immigrants and natives in Norway". *Institutt for Samfunnsforskning*, Report 2016:12.

Bergh, Johannes, Dag Arne Christensen og Richard E. Matland (2018). "Voter Mobilisation in a High-Turnout Context. Get out the Vote Experiments in the 2017 Norwegian Parliamentary Election". *Institutt for samfunnsforskning*, Report 2018:7.

Birkelund, Gunn Elisabeth, Jon Rogstad, Kristian Hellebø, Tove Mogstad Aspøy og Heide Fischer Bjelland (2014). «Diskriminering i arbeidsliver – Resultater fra randomiserte felteksperiment i Oslo, Stavanger, Bergen og Trondheim». *Sosiologisk tidsskrift* 04/2014(22)

DFØ (2018). *Veileder til utredningsinstruksen. Instruks om utredning av statlige tiltak*. Oslo: Direktoratet for økonomistyring.

Djuve, Anne Britt, Hanne Cecilie Kavli, Erika Braanen Sterri og Beret Bråten. (2017). "Introduksjonsprogram og norskopplæring: Hva virker – for hvem?" *Fafo-rapport* 2017:31. Oslo: Fafo

Finansdepartementet (2005). *Veileder til gjennomføring av evalueringer*. Oslo: Finansdepartementet

Finseraas, Henning, Torbjørn Hanson, Åshild A. Johnsen, Andreas Kotsadam og Gaute Torsvik (2017). "Trust, Ethnic Diversity, and Personal Contact: Experimental Field Evidence". *Work in progress*.

Funnell, Sue C. and Patricia C. Rogers (2011). *Purposeful Program Theory: Effective Use of Theories of Change and Logic Models*.

Geribay, Montserrat Gonzalez, & Peter De Cuyper (2018). Is there an evidence basis for immigrant integration policies? A methodological enquiry. *Nordic Journal of Migration Research* 8:1.

Haynes, Service, Goldacre & Torgerson: Test, Learn, Adapt: Developing Public Policy with Randomised Controlled Trials. Cabinet Office/Behavioural Insights Team, 2012

IMDi/Fafo (2010). *Veileder i brukerundersøkelser når innvandrere er brukere*. Oslo: Integrerings- og mangfoldsdirektoratet.

Kirkebøen, Lars J., Andreas Kotsadam, Oddbjørn Raum, Silje Andresen og Jon Rogstad (2017). Effekter av satsing på økt lærertetthet. SSB rapport 2017/39. Oslo-Kongsvinger: Statistisk Sentralbyrå.

Kolle, Elin, Reidar Safvenbom, Ulf Ekelund, Runar Solberg, May Grydeland, Sigmund A. Anderssen, Jostein Steene-Johannessen. (2016). "Utprøving og evaluering av modeller for fysisk aktivitet for elever i ungdomsskolen. Kunnskapsoversikt". *Norges idrettshøgskole*, Juli 2016

Lov om folketrygd (folketrygdloven) (1997). *LOV-1997-02-28-19*.
<https://lovdata.no/dokument/NL/lov/1997-02-28-19>

Lov om introduksjonsordning og norskopplæring for nyankomne innvandrere (introduksjonsloven) (2003). *LOV-2003-07-04-80*. [https://lovdata.no/dokument/NL/lov/2003-07-04-80?q=Lov om introduksjonsordning og norskopplæring](https://lovdata.no/dokument/NL/lov/2003-07-04-80?q=Lov%20om%20introduksjonsordning%20og%20norskoppl%C3%A6ring)

Malmberg-Heimonen, Ira, Knut Fossetøl, Marianne Rugkåsa, Anne Grete Tøge, Krisztina Gyure og Erika Gubrium (2016). «Helhetlig oppfølging av lavinntektsfamilier. Første underveisrapport». *AFI Rapport* 2016:11.

Malmberg-Heimonen, Ira, Anne Grete Tøge, Tone Liodden, Marianne Rugkåsa, Krisztina Gyure, Knut Fossetøl og Berit Berghei, (2018). «Helhetlig oppfølging av lavinntektsfamilier. Andre underveisrapport». *OsloMet Rapport* 2018 nr 9.

Midtbøen, Arnfinn H. og Jon Rogstad (2012). Diskrimineringens omfang og årsaker. Etniske minoriteters tilgang til norsk arbeidsliv. Institutt for samfunnsforskning, rapport 2012:1.

NAV (2013). *Kunnskapsstrategi i NAV 2013-2020*. Oslo: Arbeids- og velferdsdirektoratet.

NAV (2018b). Helhetlig oppfølging av lavinntektsfamilier. Publisert 23.05.2016.

<https://www.nav.no/no/NAV+og+samfunn/Kunnskap/Forsok+i+regi+av+NAV/helhetlig-oppf%C3%B8lging-av-lavinntektsfamilier>

NAV 2018a). Supported Employment. Publisert 23.05.16.

<https://www.nav.no/no/NAV+og+samfunn/Kunnskap/Forsok+i+regi+av+NAV/fors%C3%B8k-med-supported-employment>

NOU 2017:2. *Integrasjon og tillit – langsiktige konsekvenser av høy innvandring*. Oslo: Justis- og beredsdepartementet

Petrosino, Anthony, Carolyn Turpin-Petrosino, Meghan E. Hollis-Peel and Julia G. Lavenberg (2013). "Scared Straight and Other Juvenile Awareness Programs for Preventing Juvenile Delinquency: A Systematic Review". *Campbell Systematic Reviews* 2013:5.

Salvanes, Kari Veia og Rune Borgan Reiling (2017). "Utdanning som arbeidsrettet tiltak for ungdom med redusert arbeidsevne". *NIFU-rapport* 2017:1.

Ugreninov, Elisabeth og Gunn Elisabeth Birkelund (2013). "Naturlige eksperiment". *Sosiologi i dag* vol 43, nr 3.

Uni Research Helse/Uni Research Rokkansenteret (2016). Evaluering av Individuell jobbstøtte (IPS): Sluttrapport, desember 2016. Les mer:

<https://www.nav.no/no/NAV+og+samfunn/Kunnskap/FoU-midler/Relatert+informasjon/effektevaluering-av-individuell-jobbst%C3%B8tte-ips>

Wollscheid, Sabine, Martin Flatø, Hanne Næss Hjetland og Ingrid Smette (2017). "Effekter av opplæringstilbud for tospråklige elever og kompetansehevnstiltak for voksne innvandrere: En kunnskapsoversikt". *NIFU-rapport* 2017:30.