

Guide til evalueringsdesign med forsøgs- og kontrolklasser til evalueringer af uddannelseseksperimenter eller andet forsøgs- og udviklingsarbejde

Målrettet koordinatore og kvalitetsmedarbejdere

Udarbejdet i regi af Det erhvervsrettede uddannelseslaboratorium



Guide til evalueringsdesign med forsøgs- og kontrolklasser til evalueringer af uddannelseseksperimenter eller andet forsøgs- og udviklingsarbejde

For information on obtaining additional copies, permission to reprint or translate this work, and all other correspondence, please contact:

Det erhvervsrettede uddannelseslaboratorium /v Teknisk Erhvervsskole Center og Professionshøjskolen Metropol.

TEC

Nordre Fasanvej 27
2000 Frederiksberg C

Metropol

Tagensvej 18
2200 København N

Projektchef Dorrit Sørensen,
doso@phmetropol.dk
+45 72 48 70 09

Det erhvervsrettede uddannelseslaboratorium er bevilliget af Den europæiske socialfond og Region Hovedstaden.

Udarbejdet af:

Mette Foss Andersen, Udviklingskonsulent i Det erhvervsrettede uddannelseslaboratorium
Kristina Birch, Kvalitets- og udviklingskonsulent TEC

Indhold

Formål.....	4
Hvorfor evalueringværktøjer?	4
Læsevejledning til guiden	5
1. Hvad er et Uddannelseseksperiment?	5
2. Introduktion til evalueringsdesign med forsøgs – og kontrolklasser	9
3. Betydningen af forsøgs- og kontrolgrupper i uddannelseseksperimenter	9

Formål

Dette er en guide til evalueringsdesign med forsøgs- og kontrolklasser til evalueringer af uddannelseseksperimenter eller andet forsøgs- og udviklingsarbejde med fokus på kontrolklasser og sporbarhed til udviklingsinitiativer. Dette evalueringsdesign med forsøgs- og kontrolklasser er én af fire nye evalueringsformer, der er udviklet under Det Erhvervsrettede Uddannelseslaboratorium. Evalueringsmetoden indgår som en del af Uddannelseslaboratoriets samlede serie af værktøjer, der har til formål at understøtte den eksperimenterende metode i uddannelseskontekster. Dette evalueringsværktøj er samtidig et supplement til "Håndbog for evaluering og forandringsteorier for uddannelseseksperimenter"¹, som ligeledes har til formål at bidrage til samt understøtte uddannelsesinstitutioner i at evaluere uddannelseseksperimenter og udviklingsinitiativer, så det i højere grad kan føre til forbedret praksis for flere.

Vejledningen er udarbejdet af Det Erhvervsrettede Uddannelseslaboratorium i samarbejde med Kvalitets- og Udviklingsafdelingen på TEC med sparring fra kvalitetskonsulenter fra SOSU C og KEA. Guiden er udarbejdet på baggrund af en analyse af forsøgs- og kontrolhold på TEC og SOSU C², samt praksiserfaringer i forbindelse med uddannelseseksperimenter de sidste tre år.

Hensigten med *evalueringdesign med forsøgs- og kontrolklasser* er at inspirere samt vejlede uddannelsesinstitutioner til at give kvalificerede vurderinger af, hvorvidt udviklingsinitiativer skal forankres og implementeres i praksis, og dermed udvikle og løfte uddannelsernes kvalitet til gavn for hele organisationen.

Guidens målgruppe er dels koordinatore eller kvalitetsmedarbejdere på den pågældende

1

<http://uddannelseslaboratoriet.dk/metoder/evaluering-og-forandringsteorier/>

² <http://uddannelseslaboratoriet.dk/forskning-og-udvikling/udvikling/nye-evalueringsformer/>

uddannelsesinstitution, samt andre der er interesserede i en vejledning i, hvordan den eksperimenterende metode kan anvendes i praksis. Der er desuden udarbejdet en vejledning og brugervenlig trinmodel målrettet involverede undervisere og eksperimentansvarlige. Denne, samt andre værktøjer til den eksperimenterende metode kan downloades på <http://uddannelseslaboratoriet.dk/forskning-og-udvikling/udvikling/nye-evalueringsformer/>.

Hvorfor evalueringsværktøjer?

Det kan i dag være en udfordring for kvalitetssikringen og udviklingen i uddannelsessektoren, at der er mangel på velargumenteret viden om hvilke indsatser der er *værdiskabende* og kan bidrage til ny forbedret praksis. Kvalitetsarbejdet på uddannelsesinstitutioner udgøres som oftest af generelle evalueringer og tilfredshedsmålinger, der ikke nødvendigvis kan sikre sporbarheden til specifikke udviklingsinitiativer.

Hensigten med dette evalueringsværktøj er at understøtte uddannelsesinstitutioner i at udarbejde praksisnære evalueringer, der gør det muligt at få mere viden om den effekt som udviklingsinitiativer genererer med henblik på at kunne tage en beslutning om implementering af udviklingstiltag, der bidrager til fornyet og forbedret praksis på et mere oplyst grundlag end hvad tilfældet ofte er i dag.

Tilgangen er inspireret af den *eksperimenterende metode*. Udgangspunktet er en undersøgende og eksperimenterende tilgang, hvor udviklingsinitiativer systematisk iværksættes, evalueres og eventuelt implementeres. Dette for at sikre at implementering og forankring af nye indsatser bygger på positive erfaringer. Kendetegnende for *evalueringdesign med forsøgs- og kontrolklasser* er oprettelse af *forsøgsklasser i sociale kontekster*.

Evalueringdesign kan udformes på mange mulige måder. På baggrund af erfaringer fra Uddannelseslaboratoriet oplever vi, at det stærkeste evalueringdesign indeholder et sammenligningsgrundlag, hvilket derfor vil være i fokus i denne guide.

Et sammenligningsgrundlag kan både være et kontrolhold og/eller et erfaringsgrundlag. Et *erfaringsgrundlag* kan være en formåling/baselinstudie, der skaber et udgangspunkt for indsatsens evaluering. Et *kontrolhold* kan være et hold, hvori der ikke interveres, men som har tilnærmelsesvis samme forudsætninger som forsøgsholdet.

Evalueringsdesign der indeholder et sammenligningsgrundlag giver større sikkerhed om indsatsers resultater og kan i højere grad sikre sporbarhed mellem udviklingstiltag og resultater.

Læsevejledning til guiden

- Kort introduktion til Uddannelseseksperimenter og Eksperimenthjulet
- Kort introduktion til evalueringsdesign med forsøgs – og kontrolklasser
- Betydning af kontrolhold i uddannelseseksperimenter
- Betydning af erfaringsgrundlag i uddannelseseksperimenter
- Overblik og gennemgang af den eksperimenterende evalueringsproces

1. Hvad er et Uddannelseseksperiment?

Et uddannelseseksperiment definerer vi som *et systematisk gennemført forsøg, der har til formål at afprøve for eksempel nye metoder i en konkret uddannelseskontekst*. Ved at eksperimenterer kan vi arbejde sammen om at forny og forbedre den daglige praksis på uddannelserne.

Uddannelseseksperimenter er kendetegnet ved at være:

- *Fokuserede på at forandre praksis*
Uddannelseseksperimenter tager afsæt i et formuleret behov for at forandre den nuværende praksis og arbejder målrettet med at skabe den ønskede forandring. Alle eksperimenter har en indbygget forandringsteori, det vil sige en teori om, hvad der kan skabe den ønskede forandring fra behov, til en ny forbedret praksis.
- *Hurtige og afgrænsede*
Uddannelseseksperimenter gør det muligt at afprøve noget nyt i en afgrænset periode – både tidsmæssigt og organisatorisk. Derved bliver det muligt at studere eksperimentets resultater, inden der tages stilling til hvordan, og om det skal udbredes til resten af organisationen.
- *Systematiske og viden-baserede*
Uddannelseseksperimenter tager afsæt i nyeste viden om, hvilke udfordringer, der aktuelt præger uddannelserne og arbejder systematisk med at få viden om, hvad der giver værdi.
- *Innovative og nyskabende*
Uddannelseseksperimenter sigter mod at skabe nye løsninger og metoder, som giver svar på konkrete udfordringer, og har værdi for praksis. Derfor er forudsætningen for at eksperimenterer også, at man som uddannelsesinstitution *tør tænke om*, og udvikle nye visioner for sin uddannelse.

Eksperimenthjulet

I dette afsnit introducerer vi Eksperimenthjulet, som er vores bud på en systematisk model til at strukturere arbejdet med uddannelseseksperimenter, se figur 1 og 2. Eksperimenthjulet bliver herved omdrejningspunkt for den forandringsproces, der skal skabes i de enkelte uddannelseseksperimenter.

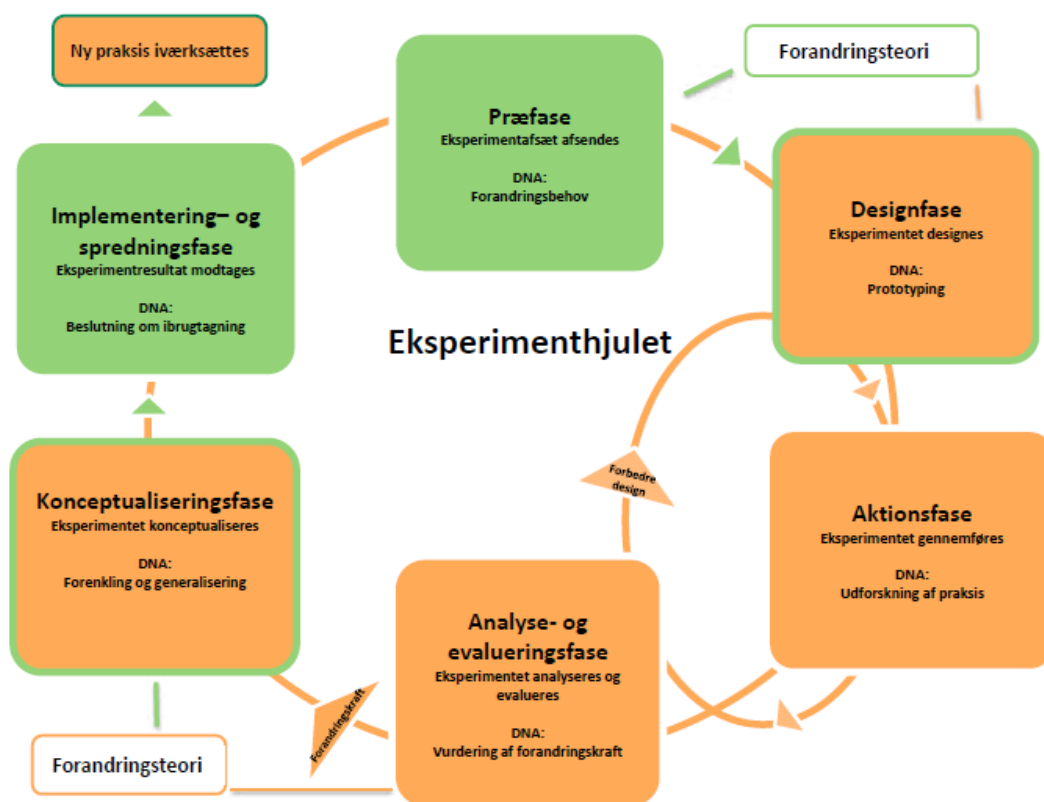
I udviklingen af modellen har vi løbende justeret Eksperimenthjulet som følge af tilbagemeldinger fra lærere og ledere om løbende erfaringer, der er fulgt med det enkelte eksperiment.

Eksperimenthjulet består af seks faser: Præfasen, designfasen, aktionsfasen, analyse- og evalueringsfasen, konceptualiseringsfasen og endelig implementerings- og spredningsfasen. Af figur 1 fremgår det centrale DNA i hver fase: Forandringsbehov – Prototyping – Udforskning – Vurdering – Forenkling – Ibrugtagning. Den sidste pil i Eksperimenthjulet

peger på 'Ny praksis iværksættes'. Hjulet illustrerer på den ene side det flow, som bliver skabt i eksperimentet, hvor faserne bygger ovenpå hinanden. På den anden side bliver faserne holdt adskilt, så der er sikret en systematik i den måde, eksperimentet bliver gennemført på.

Eksperimenthjulet har to farver.

- Den grønne farve er knyttet til 1) det arbejde, der skal danne afsæt for de enkelte uddannelseseksperimenter og 2) det arbejde, der har at gøre med forankring og implementering af de forandringer, som eksperimenterne har afstedkommet. Det er disse faser der er i fokus i denne guide.
- Den orange farve er knyttet til det udførende arbejde med de enkelte eksperimenter. Det er i disse faser, man konkret arbejder med at designe, afprøve og evaluere nye måder at gøre uddannelse på.



FIGUR 1: UDDANNELSESLABORATORIETS EKSPERIMENTHJUL - DNA

Overgangen mellem præfase og designfase, samt konceptualiseringsfase og implementerings- og spredningsfase illustreres ved, at farven grøn omkredser den orange fase.

Et eksperiment har til formål at skabe en forandring i praksis. For hvert eksperiment, designes en forandringsteori, som er central i den endelige evaluering af eksperimentet. For mere information om evaluering af uddannelseseksperimenter henvises til *Håndbog til evaluering og forandringsteori for uddannelseseksperimenter*.³

I det følgende er en kort beskrivelse af indholdet i hver enkelt fase. For en uddybning af eksperimenthullets faser se *Håndbog til eksperimenthullet – Sådan arbejder du med Uddannelseseksperimenter i praksis*.

1) Præfasen: I præfasen defineres udviklingsstrategiens afsæt, som er afsættet for det eksperimenterende arbejde i hele organisationen. Det strategiske niveau identificerer de udfordringer og forandringsbehov, der har fremmet et ønske om at eksperimenter for at udvikle ny praksis. De indsamler aktuel og relevant viden for at skabe et vidensgrundlag, som gør det muligt at tænke nyt og formulere nogle begrundede transformativt hypoteser om, hvad der vil kunne skabe den ønskede forandring af praksis. En central del af præfasen er desuden at forberede uddannelsesorganisationen på at eksperimenter. Blandt andet ved at etablere en forsøgszone, hvor det er muligt at bryde med rutinerne, og rum og tid til at kunne eksperimenter.

2) Designfasen: I denne fase designes det enkelte eksperiment af eksperimentteamet. Eksperimentteamet omsætter de identificerede udfordringer og forandringsbehov til konkrete eksperimenter, som kan iværksættes og afprøves i uddannelsens praksis. Eksperimentet planlægges og udtænkes – helt fra udarbejdelsen af den transformativt hypotese, der skal testes, til udformning af de konkrete aktioner, hvorigennem eksperimentet gennemføres. Aktionerne formål er at

³ <http://uddannelseslaboratoriet.dk/om-projektet/forandringsteorier/>

påvirke og forbedre praksis ved at afprøve nye handlingsforløb. I designfasen udarbejdes desuden en plan for dataindsamling og evaluering. I forbindelse med evalueringsdesignet, udvælger eksperimentteamet, hvilket sammenligningsgrundlag eksperimentets resultater skal holdes op imod. Eksperimentteamet udarbejder erfaringsgrundlag og tager valg om kontrolgruppe(r). Planen gør det muligt for eksperimentteamet løbende at følge op på, hvorvidt eksperimentet bidrager til de ønskede forandringer, eller om det bør justeres.

3) Aktionsfasen: I denne fase gennemfører eksperimentteamet eksperimentets aktioner og indsamler data. Aktionernes formål er at afprøve ny praksis med henblik på efterfølgende evaluering. Eksperimentteamet udarbejder beskrivelser af de gennemførte aktioner i "Aktionsloggen"⁴. Formålet med Aktionsloggen er at få beskrivelser, illustrationer og refleksioner over, hvad der er afprøvet i aktionerne, for at indsamle/dokumentere data herom. Desuden indsamles anden data, som benyttes til eksperimentets evaluering. Dette kan eksempelvis være via interviews, observationer, spørgeskemaer mv.

4) Analyse- og evalueringsfasen: I denne fase opsamles, evalueres og vurderes eksperimentet. Denne evalueringen kan opdeles i en kvalitativ og kvantitativ del, der varetages af henholdsvis eksperimentteam, og uddannelsesinstitutionens kvalitets- og udviklingsafdeling.

Eksperimentteamet analyserer kvalitativt eksperimentets resultater og processer med henblik på at finde indsigter, nye erkendelser eller overraskende "fund". Analysens fund kan dels relatere sig til den transformativt hypotese, og dels være mere utilsigtede følgevirkninger af aktionerne. Ved at eksperimentteamet systematiserer og organiserer disse fund, vil de kunne organisere fundene i mønstre, der repræsenterer den nye indsigt, som eksperimentet

⁴ Refleksionslogbog til aktionsfasen, samt andre værktøjer til udarbejdelse af eksperimenter, kan downloades på <http://uddannelseslaboratoriet.dk/metoder/eksperimenthullet/vaerktøjskasse/>

har frembragt. Formålet er at vurdere i hvilken grad og hvordan, eksperimentet har medvirket til at skabe ny tænkning eller nye praksisformer, der kan understøtte den ønskede forandring. Kan fundene og mønstrene tildeles tilstrækkelig værdi som pejlemærker for, hvordan fremtidig praksis kan planlægges og udføres?

Eksperimentteamet overgiver den kvalitative analyse og eksperimentets kvantitative data til kvalitet- og udviklingsafdelingen. Ved at supplere eksperimentets kvantitative data fra eksperiment og kontrolgrupper med eksperimentteamets kvalitative analyser, udarbejder kvalitet- og udviklingsafdelingen en analyse af, hvorvidt eksperimentet kan skabe ny forbedret praksis på institutionen.

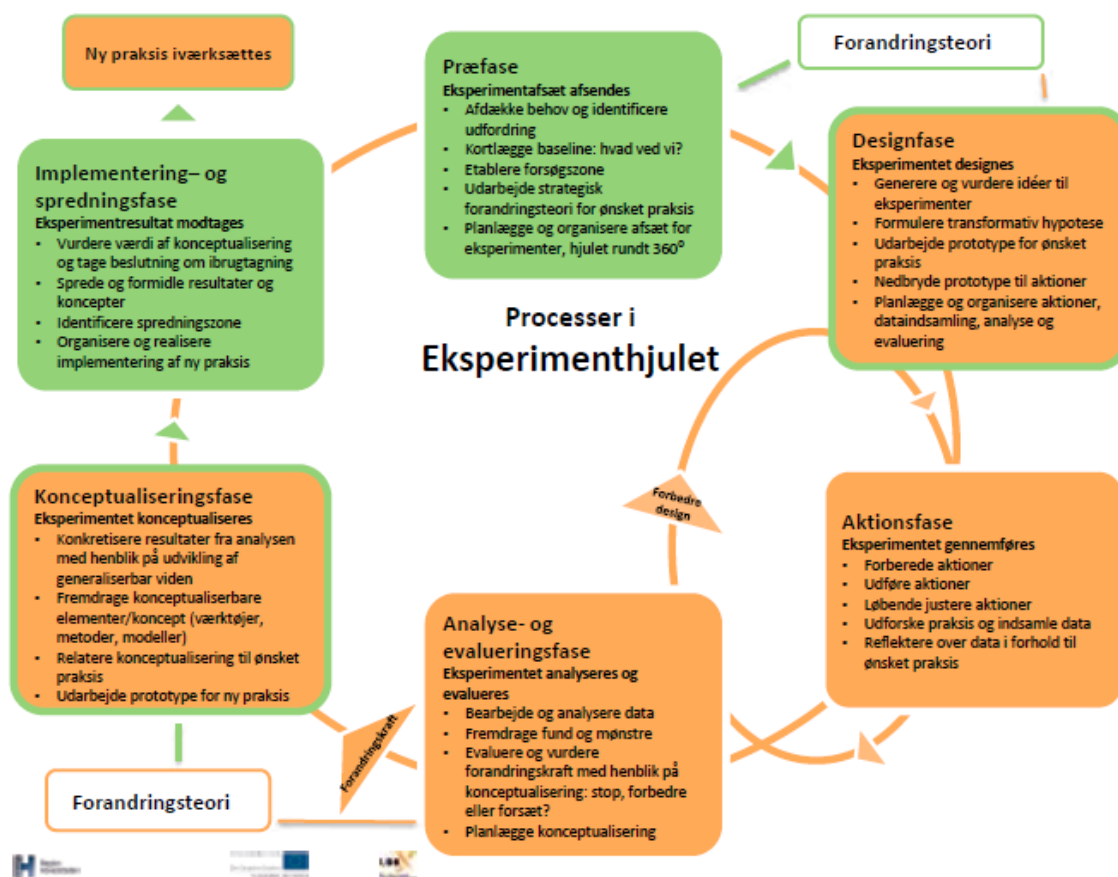
5) Konceptualiseringsfasen: I denne fase ser eksperimentteamet nærmere på de resultater der er fremkommet af analyse- og evalueringssfasen med henblik på at udlede generaliseret viden om eksperimentet er værdiskabende. Hvis eksperimentet

vurderes til at skabe ny forbedret praksis, omsættes og konkretiseres resultaterne med henblik på ibrugtagning af andre i andre kontekster.

Den generaliserbare viden sammenfattes i et åbent koncept, der kan tages i brug af andre. Et *koncept* er den samlede betegnelse for værktøjer, metoder og/eller modeller, der har været testet i en helhed, og vurderet som værdiskabende i forhold til at skabe ny forbedret praksis.

6) Implementerings- og spredningsfasen:

I denne fase skal uddannelsesinstitutionen på et strategisk og praktisk niveau arbejde med at udbrede og ibrugtage nye koncepter. Implementering og spredning handler om at skabe forbedringer og værdi for organisationen, ved at de gode resultater, der er udviklet i eksperimenterne bliver integreret og forankret som nye forbedrede praksisser i institutionen.



FIGUR 2 EKSPERIMENTHJULET – INDHOLD I FASER

2. Introduktion til evalueringsdesign med forsøgs – og kontrolklasser

I det følgende beskrives det *evalueringdesign med forsøgs – og kontrolklasser* i en form som kan benyttes af uddannelsesinstitutioner i forbindelse med uddannelseseksperimenter og andre udviklingsinitiativer.

Evalueringsdesign med forsøgs – og kontrolklasser – hvad vil I gerne vide noget om?

Allerede inden et eksperiment eller udviklingsindsats sættes i gang, er det nødvendigt at planlægge hvordan tiltaget evalueres.

Hvis en evaluering af en indsats skulle være så solid som muligt, skulle institutionerne tilfældigt udtrække elever eller studerende til deltagelse i udviklingsinitiativer/eksperimenter. Derved ville der være et *perfekt sammenligningsgrundlag* for eksperimentets resultater. Det vil dog sjældent være praktisk muligt. Det centrale er derfor at finde et alternativt sammenligningsgrundlag, som eksperimenternes resultater kan måles op imod.

Er der ikke et sammenligningsgrundlag er det nemt at lave fejlslutninger. Positive eller negative resultater fra et eksperiment kan skyldes en lang række andre faktorer – lige fra den enkelte undervisers pædagogiske handleberedskab til politiske reformer.

Uden et sammenligningsgrundlag, til at spejle indsatserne i, kan det være vanskeligt at vurdere om en indsats er værdiskabende og sammenholde resultater og erfaringer fra forskellige eksperimenter med hinanden.

Det centrale i *evalueringdesign med forsøgs – og kontrolklasser* er at fremme *sporbarheden* fra indsatsen til indsatsens resultat. Sporbarheden fremmes

- Dels ved, at resultaterne fra den forsøgsklasse der påvirkes af en indsats, sammenlignes med et erfaringsgrundlag, og/eller kontrolklasser.

- Dels ved, at aktørernes samt kontrolklassens hold og elev-id registreres, således at eventuelle effekter af en indsats kan følges og sammenlignes over tid.
- Dels ved, at der via en udforskende tilgang, stilles spørgsmål til resultaterne, og indsatsens forudsætninger

Når eksperimentteamet indsamler kvalitative og kvantitative data fra eksperimentet, kan informationerne bruges til at underbygge og evaluere eksperimentets *forandringsteori*. I det følgende beskrives kontrolgrupper i forbindelse med evalueringdesign med forsøgs – og kontrolklasser.

3. Betydningen af forsøgs- og kontrolgrupper i uddannelseseksperimenter

En *forsøgsklasse* er en klasse hvori der eksperimenteres med en aktivitet, eksempelvis en *aktion*. Et *kontrolhold* består af personer, der er sammenlignelige med udviklingsindsatsens deltagere, men som ikke deltager i aktionen. Er eksperimentets deltagere eksempelvis et hold af elever, kan en kontrolgruppe bestå af et parallelhold, der modtager samme undervisning. Formålet med at anvende kontrolklasser, er at opstille et sammenligningsgrundlag der til en hvis grad er ens for begge grupper.

Det er aldrig muligt at udvælge en "perfekt" sammenlignelig kontrolklasse, da der altid vil være forskel på elever og studerende. Det centrale, er at kontrolklasserne er så sammenlignelige, at det som udgangspunkt er tilfældigt, om den enkelte elev eller studerende går på forsøgsholdet eller parallelholdet (kontrolholdet). Er det *tilfældigt* om en elev går i den ene eller den anden klasse, vil influerende faktorer, udover indsatsen, som udgangspunkt *påvirke begge grupper på samme vis*.

Er influerende faktorer, som eksempelvis *mangel på praktikpladser*, eller *generelle initiativer til fremme af et sundt studiemiljø*, ens for begge grupper, og det er tilfældigt om en elev eller studerende befinder sig i

forsøgs- eller kontrolklasse, skabes større sikkerhed om indsatsens resultater.

Eksempel:

Et eksperimentteam eksperimenterer med en indsats der skal øge elevernes faglige niveau.

Først afdækker teamet det gennemsnitlige faglige niveau i forsøgs- og kontrolklasse (erfaringsgrundlag).

Derefter gennemfører de en *aktion*, eksempelvis en ny form for pædagogisk og didaktisk praksis i forsøgsklassen, men ikke i kontrolklassen.

Efter aktionen tester teamet igen det faglige niveau i forsøgs- og kontrolhold.

Er det gennemsnitlige faglige niveau steget i forsøgsklassen, men ikke på kontrolholdet, er det et argument for at aktion(-erne) har øget elevernes faglige niveau i forsøgsklassen.

Er det faglige niveau fortsat ens i begge grupper, eller lavere på forsøgsklassen, er det en indikator på at aktionerne ikke har haft den effekt der har været *ønsket*.

Eksperimentet bør dog prøves af i flere klasser, for at skabe yderligere sikkerhed om værdien af indsatsen. Desuden bør testresultaterne holdes op imod de kvalitative evalueringer af eksperimentet.

Denne antagelse skal dog ses i lyset af forudsætning, og hvorvidt klasserne var *ens inden aktionerne*. Dette skal tages i betragtning i den endelige vurdering af resultaterne. Hvis forsøgsholdet eksempelvis på forhånd er fagligt stærkere end kontrolholdet, vil positive resultater fra en aktion være "skæve" – forsøgsholdet har fra start af haft en stærkere faglighed, og positive resultater i forhold til kontrolgruppen vil ikke kunne tilskrives aktionen, men bero på forsøgsholdets stærkere faglighed. Det er derfor vigtigt, at være opmærksom på hvilken

kontrolklasse resultaterne sammenlignes med, og hvorvidt det er muligt at følge de kvantitative data op med kvalitative data, der kan afdække andre nuancer i resultaterne.

Der vil altid være forskelle mellem forsøgs- og kontrolholdene, da det sjældent vil være *fuldstændigt tilfældigt* om en elev eller en studerende befinder sig på de ene eller det andet hold. I et forsøg på at imødekomme denne forudsætning for enhver sammenligning, kan kvantitative data suppleres med kvalitative metoder eksempelvis i form interview og deltagerobservation. Den kvalitative empiri kan, sammen med forandringsteorien, medvirke til at sandsynliggøre eller "af-sandsynliggøre", at der er sammenhæng mellem negative eller positive resultater af et tiltag og selve tiltaget.

Det er derfor vigtigt at være opmærksom på at forsøgshold og kontrolhold er umiddelbart ens (eksempelvis på køn, alder, gennemsnitligt karakterniveau, fravær mv). Det kan give skæve resultater, hvis et almindeligt hold, sammenlignes med en voksenklasse, specialklasse eller lignende.

Eksempel:

I eksperimentet "Eleven som aktør i praktikpladssøgningen" eksperimenteres der med aktioner der søger at fremme elevernes muligheder for at få en uddannelsesaftale med en virksomhed. Forsøgsholdet udgøres af elever der har specielt svage forudsætninger for at opnå en uddannelsesaftale.

I dette tilfælde skal et kontrolhold være en gruppe af elever med lige så svage forudsætninger for at få en uddannelsesaftale. Består kontrolholdet af elever med stærkere forudsætninger, vil en efterfølgende sammenligning af antallet af praktikpladser på de to hold være usikker.

Hvad hvis der ikke findes parallelhold til forsøgsholdet?

Er det ikke muligt at finde en parallelklasse i samme periode, kan et alternativ være et hold fra undervisningsforløb fra det foregående optag.

Bemærk, at elevsammensætningen ofte kan være forskellig afhængig af optagstidspunktet, eksempelvis kan der være forskelle blandt elever der optages i august, og elevgruppen der optages senere på året. Man bør derfor være opmærksom på tidspunktet for hvornår kontrolklassen er blevet optaget. Det kan være nødvendigt at udvælge en kontrolklasse et helt skoleår tilbage, hvis dette bedst repræsenterer eksperimentets elevgruppe/optag.

Vælger man en kontrolklasse fra det foregående skoleår, da skal dette tydeliggøres i evalueringen, da der kan ske meget på et år, der også kan påvirke resultaterne. Dette kan eksempelvis være politiske ændringer, organisationsændringer, ændring i uddannelsesplaner m.v., som kan betyde at eksperimentklassen og kontrolklassen ikke har haft samme forudsætninger.

Det er vigtigt at være opmærksom på, at selv om forsøgshold og kontrolhold umiddelbart synes éns, vil man aldrig med sikkerhed kunne tilskrive positive resultater som afkommet fra eksperimentets aktioner. Positive resultater kan udledes fra et utal af andre faktorer i og udenfor klasserummet. Alligevel er fordelene ved at benytte kontrolgrupper at der er en større sikkerhed for, samt et argument for at der er en sammenhæng mellem eksperimentets aktioner og eventuelle positive resultater. Argumentet kan styrkes når det holdes op imod forandringsteorien, samt eksperimentets kvantitative og kvalitative fund. Dette vil blive beskrevet nærmere i det følgende.

Fordele ved at benytte kontrolgrupper

- *Stærkere evalueringdesign:* kontrolgrupper giver et sammenligningsgrundlag og dermed et belæg for at antage at der sammenhæng mellem eksperimentets aktioner og fund.
- *Større sikkerhed om resultater:* der er større sikkerhed for at positive resultater kan tilbageføres til eksperimentets aktioner, hvis de viser sig i forsøgsklassen og ikke i kontrolgruppen (under forudsætning af at eksperiment- og kontrolgruppe er sammenlignelige inden eksperimentets aktioner igangsættes).

4. Betydning af et erfaringsgrundlag i uddannelseseksperimenter

I *evalueringdesign med forsøgs – og kontrolklasser* er *erfaringsgrundlaget* central. Erfaringsgrundlaget kan fungere som en før-måling, og relatere sig til det udkom, som indsatsen måles på. Et erfaringsgrundlag kan baseres på observationer, resultater fra tilfredshedsundersøgelser, karaktergennemsnit, fravær og lignende.

For at undersøge om et indsats er værdiskabende, kan eksperimentteamet, eventuelt i samarbejde med en koordinator eller kvalitetsmedarbejder, udarbejde et erfaringsgrundlag for eksperiment – samt eventuelt kontrolhold, inden indsatsen igangsættes.

Det anbefales at erfaringsgrundlaget i nogen grad består af kvantitative data, suppleret af kvalitativ empiri, da de kvantitative data gør det lettere at sammenligne resultater fra indsatser med hinanden. Det følgende er eksempler på data som et erfaringsgrundlag kan indeholde.

Eksempler på kvantitative data:

- Elektronisk anonymiseret spørgeskema, der spørger til elevtrivsel
- Fraværsregistrering – gennemsnitligt fravær i eksperimental- og kontrolklasse
- Fagligt niveau – gennemsnitligt fagligt niveau måles i eksperimental- og kontrolklasse

Eksempler på supplerende kvalitative data:

- Deltagerobservation: eksperimentteamet beskriver hvordan praksis og undervisningsmiljøet er i forsøgsklassen på nuværende tidspunkt
- Korte interview med elever om undervisningspraksis på nuværende tidspunkt
- Tidligere resultater fra indsatser og udviklingsinitiativer

Jo flere oplysninger, desto bedre!

Om det er før-målinger, kontrolklasser, interview eller andre oplysninger, bidrager det til et sammenligningsgrundlag, og dermed et argument for at der kan være sammenhæng mellem eksperimentets aktioner og resultat

Som tommelfingerregel vil flest mulige datakilder styrke evalueringen af eksperimentets resultater. Desto flere datakilder der peger i samme retning, desto stærkere er evalueringen af eksperimentets forandringsteori og desto større er sandsynligheden for at eksperimentets udkom kan være afledt af indsatsen. Vel og mærke under forudsætning at eksperiment- og kontrolhold er sammenlignelige⁵ inden eksperimentets begyndelse.

⁵ Med "sammenlignelige" menes der at der ikke er skævheder på eksempelvis køn, alder og fagligt niveau i eksperiment- og kontrolhold inden eksperimentets begyndelse.

Variable i undersøgelsen

Undersøgelsens variable, som er de faktorer som erfaringsgrundlaget og den efterfølgende evaluering baserer sig på, indsamles af henholdsvis eksperimentteam og udviklingsafdeling.

Det vil typisk være eksperimentteamet der indsamler oplysninger om eksempelvis:

- Trivsel: information om trivsel kan indsamles ved hjælp af elektroniske spørgeskemaer, deltagerobservation og interview.
- Fravær: Gennemsnitlig fraværprocent indsamles hos undervisere i eksperimental- og kontrolklasser
- Fagligt niveau kan indsamles ved hjælp af faglig test og registreres som en gennemsnitlig testscore for eksperiment- og kontrolklasse.

Koordinator eller kvalitetsmedarbejder indsamler typisk:

- Frafald, afslutningskarakter, køn og alder indsamles hos elevadministration ved hjælp af registrering af holdnummer

Det overstående er ikke en udtømmende liste over mulige data. Data kan suppleres med andre datakilder, eksempelvis pædagogiske evalueringer og/eller individuelle oplysninger om elever og studerende på holdniveau. Det væsentlige er at *samme oplysninger*, indsamles *på samme vis* i både forsøgs- og kontrolklasse.

5. Overblik og gennemgang af evalueringsproces med forsøgs – og kontrolklasser

Når eksperimentteamet vælger et evalueringsdesign, kan teamet vælge en række metoder der styrker eksperimentets evaluering. I tabel 1 nedenfor, illustreres scenarier for en række evalueringsdesign.

Scenarie	Evalueringsstyrke	Evalueringsmetoder			
		forsøgshold	kontrolhold	Erfaringsgr undlag	Eftermåling
1	Meget stærkt	X	X	X	X
2	Stærkt	X	X		X
3	Mindre stærkt	X		X	X
4	Mindst stærkt	X			X
5	Svagt	X			

Som tommelfingerregel, vil flere evaluerings- og dataindsamlingsmetoder styrke evalueringsdesignet:

- Scenarie 1 er det *stærkeste* evalueringsdesign. Designet indeholder både erfaringsgrundlag, eftermåling, forsøgs- og kontrolhold.
- Scenarie 2 er et *stærkt* evalueringsdesign og indeholder forsøgs- og kontrolhold, eftermåling men intet erfaringsgrundlag.
- Scenarie 3 er et *mindre stærkt* evalueringsdesign og indeholder erfaringsgrundlag og eftermåling på forsøgshold, men ingen kontrolgruppe.
- Scenarie 4 kan betegnes som det *mindst stærke* evalueringsdesign. Designet indeholder kun eftermåling i forsøgsholdet og intet erfaringsgrundlag.
- Scenarie 5 kan betegnes som et "svagt" evalueringsdesign. Designet indeholder forsøgshold, men ingen målinger eller erfaringsgrundlag, hverken før eller efter aktioner.

Eksempel på scenarie 2 – det stærke design:

Uddannelseslaboratoriet har udarbejdet en fastholdelsesundersøgelse af elever som har deltaget i uddannelseseksperimenter på TEC og SOSU C.⁶

Undersøgelsen er udført på forskellig vis på de to institutioner:

Blandt data fra SOSU C, er fastholdelse sammenlignet i eksperimental- og kontrolklasser. For hver elev der indgår i undersøgelsen, er der tilsvarende minimum en elev fra en kontrolklasse hvor der ikke er eksperimenteret. Undersøgelsen er foretaget bagudrettet og sammenligner elevernes fastholdelse på uddannelsen. Havde data været indsamlet *inden aktionerne* kunne andre parametre have været inddraget, eksempelvis trivsel eller fagligt niveau.

Dette evalueringsdesign svarer til scenarie 2 – det stærke evalueringsdesign. Designet kunne styrkes, hvis der inden eksperimenterne var udarbejdet et erfaringsgrundlag eksempelvis på trivsel og fagligt niveau blandt forsøgs – og kontrolhold. Disse variable

⁶ <http://uddannelseslaboratoriet.dk/forskning-og-udvikling/analyser/>

ville styrke designet på to områder: dels ville det i endnu højere grad give en indikator på hvorvidt at forsøgs- og kontrolhold har haft samme forudsætninger inden aktionen. Dels kunne det i en eftermåling sammenholdes med tallene for fastholdelse. I undersøgelsen er hvert forsøgs- og kontrolhold sammenlignet på køn og alder. Derefter er der lavet en *eftermåling* af fastholdelse blandt elever i forsøgs- og kontrolklasse. I dette design indgår kun en variabel – fastholdelse – som eksperimenterne måles på. Eksperimenterne kan dog sagtens have været effektfulde på andre parametre, som ikke kan udledes af det kvantitative datamateriale.

Som illustreret i tabellen, er det vigtigt at planlægge evaluering og dataindsamling allerede inden eksperimentets aktioner udføres.

Eksempel på scenarie 4 – det mindst stærke design:

I Uddannelseslaboratoriets fastholdelsesundersøgelse⁷ indgår også forsøgshold fra TEC. I dette tilfælde er kontrolklasserne fundet bagudrettet. De fundende kontrolgrupper har ikke været parallelløb, men klasser fra optaget året før, hvilket giver usikkerhed om evalueringernes resultater. Derfor er fastholdelse blandt forsøgsholdene sammenlignet med den generelle fastholdelsesprocent på den respektive uddannelse. Dette er dog også behæftet med usikkerhed, da andelen af fastholdte elever kan være præget af historiske forskelle – eksempelvis er andelen af unge der gennemfører en erhvervsuddannelse faldet med 9 procentpoint de sidste 10 år⁸. Desuden indeholder tallene en tidsforskydning, da eleverne i forsøgs- og kontrolklasserne på undersøgelsestidspunktet endnu ikke har *gennemført*, og der derfor er en tidsperiode hvori flere elever endnu kan nå at frafalde uddannelsen.

⁷ <http://uddannelseslaboratoriet.dk/forskning-og-udvikling/analyser/>

⁸ <http://www.danmarksstatistik.dk/da/Statistik/NytHtm.l.aspx?cid=19298>

Derfor er dette evalueringsdesign det *mindst stærke*, men designet kan stadig angive en retning for arbejdet med uddannelseseksperimenter. Som det også er tilfældet med de stærke evalueringsdesign, kan fastholdelse aldrig stå alene som parameter for, hvorvidt et eksperiment kan bidrage til ønsket forandring på institutionen.

Det er vigtigt at notere sig, at et eksperiment altid producerer viden og erfaringer der kan anvendes af andre eksperimentteams, uanset om eksperimentet har det forventede udkom.

I tabellen nedenfor illustreres eksempler på data, hvem der indsamler dem, hvordan og hvornår.

Hvem indsamler hvad hvornår?					
Data	Hvem indsamler	Hvordan	Hvornår indsamles data		
			Inden aktioner	Under aktioner	Efter aktioner
Generelt erfaringsgrundlag	Ekspertimentteam	Afdækning af viden på området, eksempelvis via litteratursøgning, interviews med mere	X		
Data til brug ved evaluering					
Trivsel	Ekspertimentteam	Fx elektroniske spørgeskemaer,	X		X
		Deltagerobservation,	X	X	X
		Interview	(X)		X
Fagligt niveau		Eksempelvis teste i forsøgs- og kontrolklasser	X		X
Fravær	Ekspertimentteam	Indsamles af undervisere i forsøgs- og kontrolklasse	X	(X) ⁹	X
Frafald¹⁰, køn, alder og karakter	Koordinator eller kvalitetsmedarbejder	Ekspertimentteam indleverer holdnumre på forsøgs- og kontrolhold til koordinator eller kvalitetsmedarbejder, som indhenter oplysninger på hver elev via Elevadministrationen.	X – obs. kun køn og alder		X- obs. kun frafald og karakterer
Anden information til perspektivering af resultater					
Pædagogisk evaluering og/eller trivselsundersøgelser	Leder eller koordinator	Generelle trivselsundersøgelser samt Undervisningsevalueringer, som umiddelbart gennemføres ved et uddannelsesforløbs afslutning, kan bruges til			X

⁹ Afhænger af aktionens varighed

¹⁰ Det er væsentligt at skelne mellem "afbrudt uddannelse med omvalg" og "uden omvalg"

at understøtte eksperimentets fund og resultater. Praksis for disse evalueringer er dog forskellige fra institution til institution og indsamling af disse data		
---	--	--

Overstående oversigt er ikke en udtømmende liste over mulige data. Oplysningerne kan suppleres med andre datakilder, eksempelvis uddybende oplysninger om elever og studerende. Det væsentlige er at *samme oplysninger*, indsamles *på samme vis* i både forsøgs- og kontrolklasse. Det er ligeledes vigtigt at tage højde for hvornår oplysningerne er indsamlet. Data der anvendes som før-målinger skal altid være indsamlet inden datoen for at eksperimentets aktioner har været igangsat. Hvis ikke, er der en risiko for at målingerne være påvirket af eksperimentets aktion(er).

Hvornår fortages evalueringen?

I tabellen nedenfor illustreres centrale faser i eksperimenthjulet i forhold til eksperimentets evaluering, og hvem der har ansvar for hvert enkelt element. For en uddybning af evalueringselementerne i hver enkelt fase se *Håndbog for evaluering og forandringsteori for uddannelseseksperimenter*.

Fase i eksperimenthjulet	Centralt for evaluering	Hvem udfører
Præfase	Udpeg forsøg- og kontrolgruppe	Eksperimentteam, eventuelt i samarbejde med det strategiske niveau
	Indrporter eksperiment- og kontrolhold til koordinator eller kvalitetsmedarbejder	Eksperimentteam
Designfase	Planlæg evaluering	Eksperimentteam
	Udform evalueringsdesign	Eksperimentteam
	Lav før-måling	Eksperimentteam
Aktionsfase	Udfyld aktionslog, herunder dato(er) for aktioner	Eksperimentteam
Evaluering – og analysefase	Lav efter-måling	Eksperimentteam

	Analyser datamateriale fra før – og eftermålinger samt aktionslog	Eksperimentteam
Konceptualiseringsfase	Indsend analyse, bilag og eksperimentkoncept til koordinator eller kvalitetsmedarbejder	Eksperimentteam
	Vurdering af hvorvidt eksperimentet kan bidrage til ny forbedret praksis på institutionen	Koordinator eller kvalitetsmedarbejder
Implementeringsfase	Implementér værdiskabende eksperimenter	Det strategiske niveau

På uddannelseslaboratoriets hjemmeside:

<http://uddannelseslaboratoriet.dk/metoder/>

- findes materialer og værktøjer til evaluering og implementering af uddannelseseksperimenter. Her kan du også finde uddybende materiale og vejledning til den eksperimenterende metode i praksis målrettet undervisere, eksperimentteam og kvalitetskonsulenter.