

Til
Undervisningsministeriet

Dokumenttype
Bilag

Dato
Maj 2014

BILAG 2

METODE OG

FORSKNINGSDESIGN

BILAG 2 FORSKNINGSDESIGN

INDHOLD

| | | |
|-----------|-------------------------------|----------|
| 1. | Design- og metodebilag | 1 |
| 1.1 | Forskningsdesign | 1 |
| 1.2 | Analysemetoder | 2 |
| 1.3 | Sammenfatning | 4 |

1. Design- og metodebilag

I forskningskortlægningen er studierne blevet vurderet på baggrund af deres forskningsmæssige kvalitet og forsknings- og emnemæssige relevans i relation til indeværende reviewspørgsmål. Dette bilag giver et overblik over de forskningsdesign og analysemetoder, som anvendes i de inkluderede studier. Bilaget har dermed karakter af en læsevejledning til de metodiske begreber, som bliver brugt i abstracts og syntese.

1.1 Forskningsdesign

De inkluderede studier i forskningskortlægningen er baseret på en række forskellige forskningsdesign. I nedenstående tabel forklares de forskellige forskningsdesign, som er anvendt i studierne.

Tabel 1: Oversigt over forskningsdesign

Systematisk review

Et systematisk review er et review af den samlede forskning inden for et givent område. Systematiske reviews er baseret på en systematisk strategi for søgningen efter studier (fx hvilke databaser der søges i, hvilke keywords der søges efter, hvilke tidsskrifter der er blevet gennemgået ved manuelle søgninger etc.) og indeholder kriterier for at inkludere og ekskludere studier i reviewet samt metoder til at vurdere og sammenfatte resultaterne af de inkluderede studier.

Randomiseret, kontrolleret forsøg (RCT)

Randomiserede, kontrollerede forsøg indebærer, at forsøgspersonerne inddeles i mindst én indsats- og kontrolgruppe. Indsatsgruppen modtager den indsats, som studiet undersøger effekten af. Kontrolgruppen modtager ingen indsats og fortsætter typisk med den gængse praksis på området, fx et bestemt undervisningsforløb. Kontrolgruppen fungerer således som et sammenligningsgrundlag, som resultaterne i indsatsgruppen kan holdes op imod. Herved kan effekten af en indsats vurderes.

For at der er tale om et RCT, er det afgørende, at fordelingen mellem indsats- og kontrolgruppen er tilfældig. Den tilfældige tildeling til grupper sikrer mod selektionsbias og medfører, at indsats- og kontrolgruppen har de samme karakteristika inden for statistisk tilfældige grænser. Målet er dermed at sikre, at påvirkningen, som indsatsgruppen modtager, er den eneste forskel mellem de to grupper, således at eventuelle effekter for indsatsgruppen kan tilskrives indsatsen.

Kvasi-eksperiment

Kvasi-eksperimenter indebærer ligesom RCT-studier typisk en indsats- og kontrolgruppe (i nogle tilfælde kaldt en sammenligningsgruppe). Indsatsgruppen modtager en indsats, mens kontrolgruppen ikke gør. Alternativt kan et kvasi-eksperiment bestå af flere indsatsgrupper og ingen kontrolgruppe, hvilket muliggør vurdering af relative effekter.

Til forskel fra RCT-studier er tildelingen til grupper ikke tilfældig, hvilket eksempelvis kan skyldes, at tilfældig udvælgelse ikke er mulig, eller at man har interesse i en særlig målgruppe. Tildelingen til grupper kan foregå på forskellige måder, eksempelvis via matchning hvor grupperne er sammensat, så de ligner hinanden på variable, som antages at være centrale i forhold til det specifikke forskningsspørgsmål, som undersøges.

Longitudinelt studie

I longitudinelle studier følges de samme individer over tid, og de modtager som udgangspunkt ingen indsats. Typisk indhentes måleresultater for deltagerne på flere tidspunkter i løbet af undersøgelsens forløb.

Der findes flere forskellige typer longitudinelle studier, herunder kohortestudier og casekontrolstudier. I kohortestudier følges individer, som er kendetegnet ved at have oplevet den samme begivenhed inden for en given tidsperiode. Et typisk eksempel på kohortestudier er at følge børn, der er født i samme år. I casekontrolstudier udvælger man forsøgspersonerne på baggrund af, om de er kendetegnet ved den tilstand, man ønsker at undersøge (casegruppen) eller ikke (kontrolgruppen).

Før- og eftermålinger

Design med før- og eftermålinger indebærer, at en gruppe forsøgspersoner testes på et outcome af interesse, før og efter de gives en indsats. Formålet er at undersøge, om der er sket en forandring på de områder, som indsatsen søger at påvirke. Der kan eksempelvis være tale om en indsats, der sigter mod at forbedre eleveres matematikkompetencer. Elevernes matematikkompetencer måles før og efter indsatsen for at belyse, om indsatsen har en virkning.

Eftermålinger

Design med eftermålinger indebærer, at en gruppe testes på et outcome af interesse, efter at de har modtaget en indsats. Eksempelvis kan en gruppe eleveres syn på undervisningsmiljøet måles, efter at en ny klasseledelsesstrategi er indført. Udfordringen med forskningsdesignet er, at forsøgspersonernes niveau på målet af interesse før indsatsen er ukendt, hvilket besværliggør udlægning af en eventuel virkning.

Tværsnitsstudie

I tværsnitsstudier indsamles data på tværs af en gruppe individer eller organisationer på et tidspunkt. I et tværsnitsstudie vil data have form af kvantitative eller kvantificerbare data om mere end to variable, idet formålet er at belyse variation. Tværsnitsstudier kan afdække samvariation mellem variable, men kan ikke fastslå retningen af en sammenhæng, idet der kun indgår data fra ét tidspunkt.

Casestudie

Casestudier er kvalitative studier af en eller få enheder, fx skoler, som sigter mod dybdegående viden om de undersøgte cases. Dataindsamlingen i casestudier kan foregå på mange forskellige måder, eksempelvis via interviews, fokusgruppeinterviews, observation og tekstanalyse. Casestudier er velegnede til at undersøge komplekse sammenhænge, men kan kun i begrænset omfang belyse effekten af en indsats.

Kilder: Rieper, Olaf & Hanne Foss Hansen (2007): Metoddebatten om evidens. København: AKF Forlaget.

1.2 Analysemetoder

Studierne i forskningskortlægningen anvender mange forskellige analysemetoder til at belyse forskellige indsatser, praksisser og redskabers betydning for eleveres matematikkompetencer (*mathematical literacy*)¹. I abstracts anvendes en række samlebetegnelser for analysemetoderne, som uddybes nedenfor.

¹ *Mathematical literacy* er en samlet betegnelse for eleveres matematikkompetencer. *Mathematical literacy* omfatter således de basale kompetencer, der er nødvendige for at kunne opstille og løse matematiske problemer i dagligdagen, bearbejde og fortolke data og/eller at vende matematik til at beskrive eller forudsige en udvikling eller en begivenhed. Samtidig handler *mathematical literacy* også om at ville anvende de erhvervede matematiske færdigheder og kundskaber. Derfor indgår elevernes tiltro til egne matematiske evner (ofte beskrevet som *self-efficacy*) også i begrebet om *mathematical literacy*.

Tabel 2: Oversigt over analysemetoder**Variansanalyse**

Variansanalyse bruges som en fællesbetegnelse, der dækker over en række statistiske analysemetoder, der analyserer forskelle i flere gruppers gennemsnit. De forskellige typer variansanalyse er af varierende statistisk styrke.

Den mest simple form for variansanalyse er en statistisk test af, om flere gruppers gennemsnit er forskelligt (ANOVA). Der er flere forskellige varianter af denne type variansanalyse. Eksempelvis anvendes envejs-ANOVA, når der er én uafhængig variabel, mens tovejs-ANOVA anvendes, når der er mere end en uafhængig variabel. Multivariat variansanalyse (MANOVA) benyttes, når der er mere end en afhængig variabel.

Analyse af samvarians (ANCOVA) er en mere avanceret analysemetode, som belyser, om gennemsnittene for en afhængig variabel er ens på tværs af forskellige niveauer af en kategorisk, uafhængig variabel. Multivariat analyse af samvarians (MANCOVA) anvendes, når der er mere end én afhængig variabel.

Regressionsanalyse

Regressionsanalyse er en bred betegnelse, som dækker over forskellige typer statistiske analyser, der undersøger sammenhængen mellem en afhængig variabel og flere uafhængige variable. Analysemetoden estimerer effekten af en given uafhængig variabel på den afhængige variabel, mens der kontrolleres for betydningen af øvrige uafhængige variable (kontrolvariable) i analysen.

Der findes mange forskellige typer regressionsanalyse, som kan anvendes, alt efter hvilke typer variable der indgår i modellen og deres statistiske sammenhæng. Hvis den afhængige variabel eksempelvis antages at være intervalskaleret, kan multivariat lineær regression anvendes, mens logistisk regression anvendes, hvis den afhængige variabel er binær. Ikke-parametriske statistiske regressionsmodeller anvendes, når de sædvanlige antagelser om populationens fordeling (normalitetsantagelsen) ikke er opfyldt.

En anden analysemodel inden for denne gruppe er structural equation modelling (SEM), hvor statistiske data og kvalitative kausalantagelser bruges til at teste og estimere kausalsammenhænge.

Hierarkiske modeller

Hierarkiske modeller er en samlebetegnelse for forskellige statistiske analysemodeller, som er kendetegnet ved variation på flere niveauer. Eksempelvis både et skoleniveau og et elevniveau. Centralt i forhold til hierarkiske modeller er, at forskelle mellem populationen på forskellige niveauer er relateret til netop de forskellige niveauer.

Der er flere forskellige typer multilevel regressionsanalyse inden for de hierarkiske modeller, herunder både lineære og ikke-lineære modeller. Brugen af fixed effects giver mulighed for at fjerne uobserverede effekter, eksempelvis effekten på eleverne af at være i en bestemt klasse på en skole. Random effects kan bruges, når man antager, at den uobserverede effekt ikke er korreleret med nogle af de uafhængige variable.

Kvasi-eksperimentelle analysetilgange

Kvasi-eksperimentelle analysetilgange bruges som en samlebetegnelse for analysemetoder, hvor det ikke er muligt at skabe en faktisk kontrolgruppe. Kvasi-eksperimentelle analysemetoder dækker blandt andet over difference-in-difference og regression discontinuity modellering.

Difference-in-difference bruges til at estimere effekten af en indsats ved at udregne forskellen mellem før- og eftermåling for både indsats- og kontrolgruppe, og herefter trækkes ændringen for kontrolgruppen fra ændringen for indsatsgruppen, hvilket giver et estimat for effekten af indsatsen. Regression discontinuity udnytter tærskelværdier til at estimere effekt ved at sammenligne observationer, der ligger tæt på hver side af tærskelværdien.

Kilder: Agresti, Alan & Barbara Finlay (2009): *Statistical methods in the social sciences* (4. udgave). Pearson Prentice Hall. Wooldridge, Jeffrey M. (2009): *Introductory econometrics – a modern approach* (4. udgave). South-Western Cengage Learning.

1.3 Outcomemål

I de inkluderede studier sondres der – når dette er relevant – mellem effekter målt med standardiserede måleredskaber og effekter målt med måleredskaber, som er udviklet af forskerne bag studiet (her benævnt *forskergenererede måleredskaber*). Dette skyldes, at flere af de inkluderede systematiske reviews viser en forskel i effekt, afhængig af om der anvendes standardiserede eller forskergenererede måleredskaber til at adressere indsatsens effekt på elevernes matematiske kompetencer (bla. Bottge et al., 2004; Clarke et al., 2004; Jitendra et al., 2009). Tendensen er,

at forskergenererede måleredskaber i højere grad end standardiserede måleredskaber kan dokumentere effekter. Dette er et vigtigt opmærksomhedspunkt, da en betydelig andel af de inkluderede studier udelukkende adresserer indsatsens effekt ved hjælp af forskergenererede måleredskaber². Boksen nedenfor beskriver forskellen mellem de to typer måleredskaber.

Boks 1-1: Brug af måleredskaber i studierne

Den centrale forskel mellem forskergenererede og standardiserede måleredskaber er, at førstnævnte er specifikke, mens sidstnævnte er generelle. Dette har betydning for, *hvilken form for effekt* de to slags redskaber måler:

- **Forskergenererede måleredskaber giver typisk viden om indsatsens nære effekt.** De er ofte udviklet med henblik på at måle effekten af den konkrete indsats og er derfor typisk tilpasset indsatsens konkrete elementer og indhold. Det vil sige, at forskergenererede måleredskaber måler, om indsatsen har en effekt på elevernes matematiske kompetencer, som er meget lig de matematiske problemer og opgaver, de har øvet sig på i løbet af indsatsen.
- **Standardiserede måleredskaber giver viden om indsatsens transfereffekt.** I denne sammenhæng er transfereffekt den effekt, som den viden og de kompetencer, eleven har tilegnet sig i kraft af indsatsen, har på elevens generelle kompetencer på fx matematisk problemløsning (*problem solving*). Standardiserede måleredskaber adresserer elevernes evne til at anvende deres tilegnede matematiske kompetencer til at forstå og løse matematiske problemer, som på den ene eller anden måde er forskellig fra de problemstillinger, som blev anvendt som undervisningsmateriale i indsatsen.

I bilag 1 fremgår det for de enkelte studier hvor vidt outcome er målt gennem et forskergenereret måleredskab eller et standardiserede måleredskab. Derudover er det i rapportens tabeller over studier i de fem forskellige temaer præciseret hvorvidt outcomemål er forskergenererede eller standardiserede måleredskaber.

1.4 Sammenfatning

Ovenstående giver en kort indføring i de forskningsdesign og analysemetoder, som er anvendt på tværs af de 41 studier i forskningskortlægningen. Det er centralt at have for øje, at kombinationen af forskningsdesign og analysemetode har betydning for styrken af et studies resultater.

Eksempelvis kan en simpel variansanalyse af data indsamlet via et randomiseret, kontrolleret forsøg medføre robuste resultater, idet den tilfældige udvælgelse gør, at indsats- og kontrolgruppe har samme karakteristika (også uobserverede faktorer). Hvis kontrol- og indsatsgruppe derimod ikke er tilfældigt udvalgt – som i et kvasi-eksperimentelt design – så vil en variansanalyse ikke give helt lige så robuste resultater. På den anden side kan en regressionsanalyse være en måde at udbedre svagheder i undersøgelser, hvor indsats- og kontrolgruppe ikke er tilfældigt udvalgt, idet analysen kan kontrollere for relevante baggrundsfaktorer, eksempelvis køn, alder og socioøkonomisk baggrund.

Størrelsen af et studies stikprøve har ligeledes betydning i sammenhæng med forskningsdesign og analysemetode. Eksempelvis kræver et randomiseret, kontrolleret forsøg en langt mindre stikprøve end et tværnsnitsstudie. Selv for randomiserede, kontrollerede forsøg er der dog en nedre grænse for størrelsen af stikprøven, hvis resultaterne skal være robuste. Analysemetoden er også relevant i den sammenhæng, idet brug af eksempelvis hierarkiske modeller medfører et behov for en større stikprøve.

² Forskergenererede måleredskaber anvendes ofte i mangel af standardiserede måleredskaber.

Forskningsdesign, analysemetode og stikprøvestørrelse bør altså ikke ses som uafhængige faktorer i vurderingen af et studies kvalitet, men skal sammenholdes for at danne et retvisende billede af styrken af et studies resultater.