

Om kritiske vurderinger af psykologiske undersøgelser - validitet, reliabilitet og forstyrrende faktorer

I den første del af kapitlet har vi gennemgået, hvordan man kan udføre psykologisk forskning, og hvilke fordele og ulemper der er ved de forskellige metoder. Der er dog en række ting, som kan svække kvaliteten af psykologiske undersøgelser, og det skal vi kigge på i det følgende. Det er nemlig ofte nødvendigt at foretage en kritisk vurdering af værdien af en undersøgelse, så man ved, om man overhovedet kan bruge undersøgelsen til noget, dvs. om man kan bruge dens resultater i en analyse af eller diskussion i forhold til en psykologisk problemstilling. For at kunne foretage en kritisk vurdering er det godt at kende til begreberne validitet, reliabilitet og forstyrrende faktorer/variable (fejlkilder m.v.).

Validitet

En undersøgelses validitet betyder dens gyldighed, dvs. om den faktisk viser det, som den er designet til at vise. Validitet er ikke et spørgsmål om, hvorvidt en undersøgelse har eller ikke har validitet, men hvor høj en *grad* af validitet der er tale om. Man skelner mellem to former for validitet:

- **Intern validitet**, som handler om, hvorvidt en undersøgelse er designet på en måde, så den afdækker det, som forskerne gerne vil afdække. Der kan bl.a. være tale om, at designet sigter skævt. En spørgeskemaundersøgelse om unges alkoholforbrug har f.eks. en meget lav grad af validitet, hvis alle spørgsmålene kun handler om lektiepresset i gymnasiet. Så er det meget svært at konkludere noget om alkoholforbrug. Der kan også være tale om en lav grad af validitet i et eksperimentelt design, hvor det er nødvendigt, at deltagerne ikke må vide, hvad forsøget går ud på, men hvor eksperimentsituationen faktisk meget let bliver gennemskuet af deltagerne, så de handler ud fra dette.

PsykC

Grundbog til psykologi på C-niveau

© Frydenlund og forfatteren

1. udgave, 1. oplag, 2021

ISBN 978-87-7216-256-0

ISBN (e-bog) 978-87-7216-258-4

Fagredaktør: Flemming B. Olsen

Grafisk tilrettelæggelse: Morten Thunø Hansen

Korrektur: Peder Norup

Grafisk produktion: Pozkal, Polen

34.

Eksempel på undersøgelse, hvor der kan have været problemer med den interne validitet

Den amerikanske psykolog Stanley Milgram udførte i 1960'erne nogle berømte og berygtede forsøg om lydighed over for autoriteter, hvor han testede, om forsøgsdeltagere ville adlyde eksperimentlederen som en autoritetsperson og tildele en anden deltager elektriske stød, hvis denne deltager svarede forkert i en række huskeopgaver. Forsøget var manipuleret på den måde, at der ikke blev givet rigtige stød til den anden deltager (som var skuespiller), der befandt sig i et andet rum end den rigtige forsøgsdeltager. Den rigtige forsøgsdeltager var placeret foran en imponerende stødtildelingsmaskine og skulle for hvert forkert svar øge strømstyrken med 15 volt indtil en maksimal styrke på helbredstruende 450 volt – medmindre han nægtede at fortsætte. Arkivstudier har dog vist, at en af Milgrams assistenter analyserede sig frem til, at jo mere en deltager troede på, at den anden person ikke fik rigtige stød, desto mere tilbøjelig var vedkommende til at gå hele vejen til 450 volt. En stor portion af deltagerne i forsøgene har altså tilsyneladende gennemskuet, at der ikke blev tildelt rigtige stød – hvilket flere deltagere også gav udtryk for undervejs i forsøgene.

- **Ekstern validitet (også kaldet økologisk validitet)**, som handler om, hvorvidt man kan overføre resultater og konklusioner fra en undersøgelse til virkelighedens verden, dvs. virkeligheden uden for undersøgelsens 'verden'. Det er specielt i laboratorieeksperimenter med en meget høj grad af kontrol over variable, at den eksterne validitet er lav. Eksperimentsituationen (både omgivelserne og deltagernes opgaver) kan være så kunstig og virkelighedsfjern, at det kan være svært at konkludere noget om, hvordan mennesker uden for laboratoriet opfører sig.

Deltagerudvælgelse

For at sikre en høj grad af validitet i en undersøgelse skal man sørge for, at man har udvalgt en passende deltagergruppe (et såkaldt *sample*). I nogle undersøgelser (typisk de kvantitative) er man interesseret i at lære noget om store grupper eller hele befolkninger, og i andre undersøgelser (typisk de kvalitative) er man blot interesseret i at lære noget om en lille gruppe.

En meget passende deltagergruppe er en gruppe, som udgør et repræsentativt udsnit af den større gruppe (dvs. den *population*), som man gerne vil vide noget om. Når et udsnit er repræsentativt, så

afspejler det den større gruppe, dvs. det indeholder den samme sammensætning af forskellige folk. Det kan være svært at lave et udsnit, der repræsenterer alle, for i princippet skal udsnittet afspejle både kønsfordelingen, aldersfordelingen, den kulturelle sammensætning, den geografiske spredning m.v. i populationen. Dertil kommer, at man ikke nødvendigvis bare kan få fat i de deltagere, som man gerne vil have. Hvis deltagerne skal svare på f.eks. en annonce om at medvirke i et eksperiment, så er det muligvis nogle bestemte typer af folk, som vælger at besvare annoncen. Mange psykologiske undersøgelser er uheldigvis præget af, at man har udvalgt en meget ensartet og snæver undersøgelsesgruppe (f.eks. universitetsstuderende), som sjældent afspejler den bredere population, som man gerne vil lære noget om.

Reliabilitet

En undersøgelses reliabilitet betyder dens pålidelighed, dvs. om undersøgelsen kan gentages (eventuelt af andre forskere) og give de samme resultater. Reliabilitet måles ligesom validitet i grader. Umiddelbart kunne man tro, at undersøgelser selvfølgelig giver de samme resultater, hvis de gentages, men der er en række forstyrrende faktorer (ofte kaldet fejlkilder), som kan påvirke en undersøgelses pålidelighed. Dette vil blive uddybet nedenfor, hvor vi vil kigge på forskellige typer af forstyrrende faktorer/variable (også kaldet bias).

Forstyrrende faktorer (fejlkilder m.v.)

- **Forsøgseffekt** (eng. *observer effect*): Fordi deltagerne i en undersøgelse ved, at de bliver undersøgt, så kan der være en tendens til, at de præsterer bedre eller på en anden måde, end de ville gøre, hvis de ikke blev observeret. Forsøgseffekten kan være særlig udtalt i eksperimenter og åbne observationer.
- **Selvrepræsentationseffekt**: De fleste mennesker vil gerne fremstå som bl.a. ordentlige, høflige, pålidelige og generelt gode mennesker over for andre. Dette ønske om at præsentere sig selv på en god og socialt acceptabel måde kan også påvirke deltageres adfærd i undersøgelser. Selvrepræsentationseffekten kan finde sted i både observationer (åbne), eksperimenter, spørgeskemaer og interviews. I specielt spørgeskemaer og interviews kan effekten medføre, at deltagerne lyver, fordi de er bange for, at det sande svar vil stille dem i et dårligt lys.

Kontroleffekt (kaldes også fordringseffekt – eng. *demand characteristics*): Vi mennesker er meget nysgerrige, og vi vil meget gerne vide, hvad det er, der sker med os. Når folk deltager i en undersøgelse, så ved de, at forskerne har en hypotese, og denne intention med undersøgelsen kan deltagerne ikke lade være med at prøve at gætte. Det er underordnet, om deltagerne rent faktisk gætter rigtigt, for pointen med kontroleffekten er, at de retter deres adfærd ind, så den passer med deres gæt. På den måde kan det give unaturlige resultater – specielt hvis deltagerne har fået penge for at deltage og derfor gerne vil være gode deltagere, der hjælper forskerne. For at modvirke denne effekt bruger man ofte en hvid løgn om, hvad undersøgelsen handler om (såkaldt *enkeltblinding*), så man forhåbentlig lidt bedre kan styre deltagernes gætteri. Kontroleffekten kan være særlig udtalt i eksperimenter, men kan også finde sted i åbne observationer og spørgeskemaer.

Eksempel på demonstration af kontroleffekten

Martin Orne og hans kolleger demonstrerede i 1960'erne, hvor stærk kontroleffekten kan være. De gav bl.a. forsøgsdeltagere en bunke med 2000 ark papir og satte dem til selv at skrive 224 plusstykker pr. ark papir, som de efterfølgende skulle regne. Når de var færdige med et ark, skulle de blot fortsætte på det næste ark og blive ved, indtil de blev bedt om at stoppe. Forskerne havde taget deltagernes armbåndsure, så deltagerne ikke var klar over, hvor længe de havde været i gang. De fleste deltagere fortsatte med den meningsløse opgave med plusstykkerne i over fem timer (!) uden at brokke sig. De fortsatte altså i unaturligt lang tid, blot fordi de vidste, at de deltog i et psykologisk eksperiment og sandsynligvis tænkte, at det handlede om udholdenhed eller noget i den stil.

Forsøgsledereffekt (eng. *researcher bias*): Denne effekt kaldes også Rosenthal-effekten efter den amerikanske psykolog Robert Rosenthal, der har forsket meget i den. Effekten handler om, at forskeres hypoteser om deres undersøgelser (dvs. deres forventninger til, hvad de finder) kan komme til at påvirke deltagerne i undersøgelsen på en sådan måde, at forskeren får de resultater, som han eller hun gerne vil have. Det er ofte på ubevidste og meget subtile måder, at forskere får hjulpet deltagere hen i den ønskede retning – det er sjældent for at fuske med resultaterne. For at modvirke denne

effekt bruger man i specielt eksperimenter såkaldt *dobbeltblinding*, hvor hverken deltagerne eller den forsker, som har med deltagerne at gøre, ved, hvad hypotesen er, eller om deltagerne er i enten kontrol- eller eksperimentgruppen. Forsøgsledereffekten kan være særlig udtalt i eksperimenter, men den kan også finde sted i åbne observationer, spørgeskemaer og interviews.

Eksempel på undersøgelse, hvor forsøgsledereffekten lader til at have spillet en rolle

I 1950'erne udførte den tyrkisk-amerikanske psykolog Muzafer Sherif og hans kolleger en række klassiske forsøg, hvor 11-årige amerikanske drenge fra den hvide middelklasse blev opdelt i to grupper på en sommerlejr i et stort naturområde (bl.a. i Robbers Cave State Park i Oklahoma). De to grupper skulle konkurrere mod hinanden i forskellige dyster, og det viste sig, at de to grupper hurtigt udviklede en meget stærk negativ holdning til hinanden, og det kom ligefrem til voldelige sammenstød såsom slåskampe og ødelæggelse af hinandens hytter. Arkivstudier af bl.a. forskerens noter fra undersøgelseerne har dog vist, at forskerne i høj grad opmuntrede drengene til at være fjendtlige over for hinanden. Et af Sherif og kollegers forsøg (i 1953) endte også med, at drengene i samlet flok gjorde oprør mod forskerne, fordi de ikke ønskede at være uvenner, sådan som forskerne havde forsøgt at præge dem til. Forskerne havde bl.a. ødelagt den ene gruppes telt i håbet om, at gruppen ville give den anden gruppe skylden.

Om brud på etiske retningslinjer i psykologisk forskning

Når man som forsker udfører psykologiske undersøgelser, så er der en række etiske retningslinjer, som man skal overholde. Disse retningslinjer handler i bund og grund om, at man skal respektere sine undersøgelsesdeltagere og opføre sig ordentligt over for dem. Følgende retningslinjer er alment accepteret:

- a) Deltagere må ikke lide hverken fysisk eller psykisk overlast.
- b) Deltagere skal have givet informeret samtykke, dvs. de skal have sagt ja til at deltage baseret på, at de ved, hvordan undersøgelsen kommer til at foregå. Dette punkt bliver man nogle gange nødt til at omgå, hvis deltagerne ikke må vide alt om undersøgelsens hypotese og design.

- c) Deltagere skal debriefes efter undersøgelsen, dvs. at de skal have at vide, hvordan det gik. Hvis man har været nødt til at føre deltagere bag lyset, så de ikke gættede, hvad undersøgelsen gik ud på, så er det ekstra vigtigt, at de efterfølgende får at vide, hvad det egentlige formål med undersøgelsen var.
- d) Deltagerne skal være sikret anonymitet, dvs. det kun er forskerne, der må kende deres identitet.
- e) Deltagere har til enhver tid lov til at afbryde deres deltagelse i en undersøgelse – og de skal være informeret om dette. Hvis de benytter sig af denne ret, så skal deres data elimineres.

I en kritisk vurdering af en undersøgelse kan man kommentere på, om forskerne har brudt et eller flere af ovenstående punkter. Det er vigtigt at bemærke, at et brud på de etiske retningslinjer ikke umiddelbart påvirker en undersøgelses validitet eller reliabilitet negativt, dvs. dens videnskabelige værdi. Det betyder 'blot', at forskerne ikke har handlet etisk forsvarligt. Mange psykologiske studier gennem tiden har været etisk uforsvarlige, og det betyder blot, at man ikke vil tillade, at de bliver udført i dag.

Tjekliste til kritisk vurdering af psykologiske undersøgelser

- Er der problemer med **den interne validitet**? Er forsøgsopstillingen f.eks. troværdig for deltagerne? Er spørgsmålene i en spørgeskemaundersøgelse passende til emnet?
- Er der problemer med **den eksterne validitet**? Dette er særligt relevant for laboratorieeksperimenter: Ligner eksperimentsituationen en virkelig situation, så man kan overføre viden fra eksperimentet til verden udenfor?
- Vedrørende **deltagerudvælgelsen**: Svarer den udvalgte deltagergruppe (sample) til den målgruppe (population), som man gerne vil vide noget om? Er deltagergruppen repræsentativ, dvs. er alle køn, aldersgrupper, kulturer m.v. repræsenteret? (Bemærk, at forskere godt kan have en bestemt målgruppe i sigte – f.eks. piger i alderen 16-20 år – og så behøver alle køn og aldersgrupper selvfølgelig ikke at være repræsenteret).
- Kan det tænkes, at deltagerne i undersøgelsen handlede på en speciel og unaturlig måde, blot fordi de var med i en psykologisk undersøgelse (**deltagereffekt**)?
- Kan deltagerne på nogen måde have påvirket resultaterne, hvis de var klar over – eller troede, de var klar over – hvad undersøgelsen handlede om (**kontroleffekt**)?

- Kan forskeren på nogen måde have påvirket resultaterne, hvis han eller hun har haft en forestilling (hypotese) om, hvad undersøgelsen ville vise (**forsøgsleder-effekt**)?
- Overholder undersøgelsen **de etiske retningslinjer** for psykologisk forskning? (Bemærk, at dette ikke umiddelbart har en indflydelse på undersøgelsens videnskabelige værdi).

OPGAVE:

Arbejd i grupper: Lav en virkelig, virkelig interessant og fængende video, som på maksimum 2 minutter fortæller, hvordan man foretager en kritisk vurdering af en psykologisk undersøgelse.

Brug din viden!

Læs nedenstående beskrivelse af, hvordan det klassiske eksperiment The Stanford Prison Experiment forløb. Udpeg de problematiske aspekter og forstyrrende faktorer ved undersøgelsen. Husk at bruge fagbegreberne fra dette kapitel.

Stanford-fængelseksperimentet

I 1971 udførte de amerikanske psykologer Philip G. Zimbardo, Craig Haney og Curtis Banks et eksperiment i kælderens under Stanford University, USA. Formålet var at undersøge, hvad der skabte den brutalitet hos fangevogtere i de amerikanske fængsler, som man havde set mange eksempler på. Zimbardo havde en hypotese om, at det ikke skyldtes en specielt aggressiv personlighed hos fangevogterne, men at selve fængselssituationen, hvor de havde magt over fangerne, fik dem til at udføre ondsksfulde handlinger.

Forskerne omdannede kælderen under universitetet til en fængselslignende korridor med celler, isolationskammer og vagtstue. 75 mandlige universitetsstuderende meldte sig til eksperimentet på baggrund af en annonce i den lokale avis, og ud af disse valgte forskerne 24 deltagere, som var socialt veltilpassede, som ikke havde psykiske lidelser, og som ikke tidligere havde begået kriminalitet. Ved tilfældig lodtrækning fik 10 deltagere rollen som fangevogtere, 11 fik rollen som fanger, og tre var reserver. Den ene af forskerne, Zimbardo, tog rollen som fængselsinspektør (eng. *prison superintendent*). Forsøget var fastsat til at skulle vare i 14 dage.

På dagen i sommerferien, hvor forsøget skulle begynde, blev deltagerne, som skulle være fanger, hentet af politibiler på deres hjemmeadresse og bragt til det nyoprettede universitetsfængsel, hvor de fik fangedragter på og fik numre i stedet for navne.

Inden fangernes ankomst holdt Zimbardo et møde med fangevogterne, hvor han briefede dem om, at de ikke måtte bruge fysisk vold mod fangerne, og at deres opgave var at holde ro og orden i fængslet. Han fortalte dem dog også, at de gerne måtte indgyde fangerne frygt og en følelse af håbløshed og manglende kontrol.

Allerede efter få dage begyndte der at opstå gnidninger mellem fangevogterne og fangerne, fordi fangerne protesterede mod tilstandene i fængslet. Fangevogterne begyndte at køre på lederen af fangernes 'oprør', fange nr. 8612, og da dette blev for meget for ham, fik han et møde med Zimbardo, hvor han bad om at komme ud af eksperimentet. Zimbardo opretholdt dog rollen som fængselsinspektør og foreslog fangen, at han kunne blive stikker og levere informationer til Zimbardo mod så at blive skånet af fangevogterne. Fange nr. 8612 blev så forvirret over dette, at han vendte tilbage til de andre fanger og fortalte dem, at man ikke kunne blive løsladt fra fængslet. Eftersom han havde en lederrolle blandt fangerne, medvirkede dette til at intensivere håbløsheden hos fangerne. Fange nr. 8612 gled herefter ind i en mental tilstand, der var så ustabil, at forskerne valgte at løslade ham. Han har dog flere år senere udtalt, at han blot spillede skuespil for at blive løsladt, så han kunne fortsætte sin eksamenslæsning.

Fangevogternes nedværdigende behandling af fangerne eskalerede dog efter løsladelsen af fange nr. 8612, og fangerne blev bl.a. sat til at skrubbe toiletter samt lave meningsløse fysiske øvelser på alle tidspunkter af døgnet.

En af de fangevogtere, som viste sig at være den mest opfindsomme i sit tyranni af fangerne, har senere i en dokumentarfilm om eksperimentet udtalt, at han havde den opfattelse, at forsøget handlede om at vise, hvor modbydelige fængsler er. Dette ville han gerne hjælpe med at demonstrere, og han modellerede sin rolle som fangevogter efter en led fangevogter, som han havde set i en film.

Den tidligere indsatte Carlo Prescott, som op til fængselseksperimentet fungerede som konsulent for forskerne, har også udtalt, at de ondskabsfulde ting, som fangevogterne fandt på at udsætte fangerne for, lader til at være inspireret af de historier fra San Quentin-fængslet, som han fortalte forskerne om. Forskerne har derfor sandsynligvis forsynet fangevogterne med inspiration.

Forskerne afbrød (efter pres fra en kollega) forsøget efter kun seks dage, fordi situationen i fængslet var blevet for brutal. På dette tidspunkt var fem af fangerne blevet løsladt før tid på grund af følelsesmæssige sammenbrud, depression eller akutte angstanfald.

(Kilde: Dokumentarfilmen *The Stanford Prison Experiment* (BBC, 2002) samt Carlo Prescott (2005). "The lie of the Stanford Prison Experiment." *The Stanford Daily*, vol. 227, nr. 50)