

Hypotetiske Konstrukter er en Nødvendighet, og Derfor bør Atferdsanalytikere være mer Opptatt av Begrepsvaliditet

Gunn E. H. Løkke og Jon A. Løkke
Høgskolen i Østfold

Utgangspunktet for artikkelen er Svartdals «Hypotetiske konstrukter innenfor atferdsanalyse: Finnes de? Ja, men der lever de dessverre ikke i beste velgående». Vi er enige med Svartdal i at atferdsanalyse, uunngåelig, er preget av hypotetiske konstrukter, abstrakte begreper og samlebetegnelse. Siden atferdsanalysen inkluderer hypotetiske konstrukter, bør vi i konseptuell, eksperimentell og anvendt atferdsanalyse være mer opptatt av begrepsanalyse, begrepslutninger eller operasjonaliseringer som gjør begrepene eller konstruktene målbare. Begrepsvaliditet, eller mer eller mindre gyldige slutninger fra et teoretisk nivå med konstrukter til et empirisk nivå med målinger av konstruktene, burde være sentralt i atferdsanalysen. I anvendt atferdsanalyse bør operasjonaliseringen av både tiltak og effekt diskuteres i større grad.

Vi tar avstand fra en ekstrem form for operasjonisme – vi mener at de konstruktene vi bør være interessert i prinsipielt sett kan måles på ulikt vis, men målingene har alltid en andel feil inkludert. Det er viktig å minimere feilene gjennom stadig bedre begrepsanalyser og operasjonaliseringer. Stimulert av Svartdals hovedmoment om at hypotetiske konstrukter finnes i atferdsanalyse, og med utgangspunkt i Cook og Campbells validitetstypologi, utvider vi diskusjonen om hypotetiske konstrukter til å angå begrepsvaliditet i atferdsanalysen. Oppsummert er vår påstand at begrepsvaliditet er mer relevant for atferdsanalysen enn antatt, og begrepsvaliditeten bør få bedre levekår.

Nøkkelord: hypotetiske konstrukter, begrepsvaliditet, operasjonalisering, slutninger

Artikkelen tar utgangspunkt i Svartdals «Hypotetiske konstrukter innenfor atferdsanalyse: Finnes de? Ja, men der lever de dessverre ikke i beste velgående» (2014, dette nummer av NTA). Svartdal har lest Eilifsen, Vie og Arntzen (2011), og har latt seg inspirere. Vi kommenterer på det vi mener er det viktigste momentet hos Svartdal, som vi også er enige i, men utdyper og utvider argumentasjonen. Svartdals viktigste momenter er slik vi leser det: Det er uheldig at atferdsanalytikere avviser hypotetiske konstrukter. En avvisning medfører at atferdsanalyti-

kere avskjermer seg fra viktig informasjon innenfor eksempelvis kognitiv psykologi. Dessuten er det slik at «... *alle* forskere, også atferdspsykologer, formulerer hypotetiske konstrukter» (Svartdal, 2014, s. 120).

Hva er hypotetiske konstrukter i følge Svartdal (2014)? ««Hypotetiske konstrukter» er begreper eller modeller forskeren foreslår for å beskrive en klasse av fenomener som antas å høre sammen» (s. 122). Det dreier seg om organisering av observasjoner med utgangspunkt i en teori av en eller annen type. Svartdal nevner blant annet motivasjon, hukommelse, læring og operant som eksempler på konstrukter eller sammenslåing av en rekke fenomener med noe felles.

Korrespondanse angående denne artikkelen kan adresseres til Jon A. Løkke, Høgskolen i Østfold. E-post: jon.lokke@hiof.no.

Konstruktene har forskjellig innhold, men har til felles at «De er abstrakter og kan ikke direkte observeres» (s. 122). Svartdal skriver at man innenfor psykologi ikke gjør «... noe skarpt skille mellom konstrukturer av abstrakt form, og konstrukturer med en antatt eller påvist eksistens». Konstruktene er imidlertid forankret i observasjon, gjerne målt med indikatorer for eksempelvis «hukommelse», og akkurat det gjelder for alle psykologiske begreper; begrepene eller konstruktene er abstrakter som ikke kan observeres direkte.

Konklusjonen er at alle begreper fra atferdsanalysen er hypotetiske konstrukturer langs eksempelvis en dimensjon som dreier seg om det er mulig eller ikke mulig å påvise mekanismer av fysiologisk karakter som understøtter begrepet. Vi må leve med at vi må foreta slutninger fra abstrakte begreper til indikatorer, men vi bør bestrebe at slutningene blir best mulig. Vi er på linje med Svartdal når han skriver at deler Skinners produksjon er «... ideologisk også på områder der den burde være empirisk» (Svartdal, 2014, s. 129). Et eksempel på et slikt område er verbal atferd. Sundberg publiserte i 1991 en oversikt over 301 forskningstemaer med utgangspunkt i Skinners «Verbal Behavior». Selv om Skinner anså «Verbal Behavior» for sin viktigste bok, antyder Sundbergs omfattende liste at det er et overskudd av ubelyste problemstillinger, og flere begreper og påstander mangler empirisk og eksperimentell støtte.

Svartdal (2014) tar også opp, med henvisning til Eilifsen, Vie og Arntzen (2011), problemstillinger som angår kognitiv psykologis vekt på testing av konstrukturer, konstruktens forklaringskraft og anvendbarheten av resultatene fra kognitiv psykologi. Vi kommer ikke til å kommentere disse problemstillingene. Vi mener atferdsanalysens forhold til hypotetiske konstrukturer er av størst prinsipiell betydning.

Svartdal bruker "hypotetiske konstrukturer" og "konstrukturer" om hverandre. Dersom vi ikke legger noe mer mystisk i begrepet «hypotetisk» enn at noe er tenkt, påstått eller inne-

bærer forbehold eller betingelser, så kommer vi til å bruke «konstrukt» der det passer. Vårt hovedanliggende er å bringe momenter om begrepsvaliditet inn i diskusjonen. Siden «konstrukt» er mindre vanlig enn «begrep» på norsk, og den norske metodelitteraturen ofte bruker betegnelsen *begrepsvaliditet*, vil vi bruke både «konstrukt» og «begrep». Men, konstruktvaliditet er et godt alternativ for fremtiden. Før vi belyser begrepsvaliditetens, eller konstruktvaliditetens, plass i atferdsanalysen, gjengir vi noen momenter om konstrukturer fra generell vitenskapsteori og noen få eksempler på hva som må anses som vesentlige konstrukturdiskusjoner eller konseptuelle diskusjoner i atferdsanalysen.

Konstrukturer i Vitenskapsteorien og Konstruktproblemer i Atferdsanalysen

Konstrukturer i Vitenskapsteorien

Hypotetiske konstrukturer, konstrukturer og abstraksjoner er grovt sett synonymt med begreper, men konstrukturer er teoretiske *konstruksjoner* som har til hensikt å organisere de sansede og registrerte omgivelsene (Pedazur & Schmelkin, 1991). Konstrukturer er nettopp *konstruerte* med et vitenskapelig formål. Et konstrukt «...is not a visual image, nor is it external to the mind; it is analogous to a piece in a game which thought plays (Caws, 1959 s. 16). I og med at konstruktene, som angst, motivasjon, altruisme, forsterkning og frustrasjon, ikke kan sees direkte («...is not a visual image»), kan det være at «hypotetiske» konstrukturer innebærer en unødig fordobling av verden: Dersom (vitenskapelige) konstrukturer er abstraksjoner og uobserverte variabler; ja, da er de vel hypotetiske også i betydningen tenkte eller påståtte. Det holder nok å snakke om konstrukturer i mange tilfeller.

Caws (1957, s. 229) beskriver skiftene fra concept (begrep) til construct (konstrukt) til isolated construct (isolert konstrukt); vi kommer ikke til å omtale den formen for konstrukt). Begreper har røtter i alle

menneskers sansing og har nær tilknytning til objektet som betegnes. Abstraksjonsnivået er lavt. Konstrukter derimot kan sees på som begreper som har blitt sendt videre til en vitenskapelig disiplin for å fungere i teori. Konstrukter har et vitenskapelig formål og er knyttet til vitenskapelige krav, men hva kan kravene være?

Kravene til konstrukter (Caws, 1957) er et de bør koples til andre konstrukter (se dimensjonen «Conceptual systems» hos Baer, Wolf, & Risley, 1968 som innebærer noe liknende), de bør være stabile og ikke preget av motesvingninger. Enda viktigere er at konstruktene har det som kan kalles en epistemisk forbindelse og mulighet for å bli målt; konstruktet må i prinsippet kunne måles. Epistemisk forbindelse innebærer at det er noe i verden som vi kan observere og som kan være «partner» til konstruktet. Konstruktet må altså ha en empirisk karakter.

Når vi utvikler operasjoner, eller teknikker for å måle konstruktet, må vi ikke falle i fellen og tro at den utviklede og tilgjengelige teknikken er den eneste mulige: «Nor must it be assumed that, when such technique is developed, it must necessarily be single or of a single type. The difficulties that operationism has encountered over this question are well known» (Caws, 1957, s. 228). Operasjonisme på sitt ekstreme, med en enkelt operasjonisering av et begrep, er en form for essensialisme som er uakseptabel i vitenskap; vitenskapen er åpen for og stimulerer variasjon. Essensialister er opptatt av hva et fenomen «egentlig er», en gang for alle, men vi vet at selv sentrale begreper som positiv og negativ forsterkning (Michael, 1975) og variabilitet (Barba, 2012) er under diskusjon. Det er i prinsippet alltid en mulighet for utvikling av begreper, operasjoniseringer og måleteknikker. En slik begrepsutvikling er altså sammenfallende med fagutvikling, og er muligens mangelvare i atferdsanalysen (Marr, 2012).

Operasjoniseringer er også kalt empiriske definisjoner, epistemiske definisjoner, og korrespondanseregler (Pedazur & Schmelkin,

1991 s. 170). Forsøk på å diskutere seg frem til empiriske definisjoner i atferdsvitenskap eller psykososiale fag, enten definisjonene kalles epistemiske definisjoner, korrespondanseregler eller operasjonelle definisjoner, fører ofte til en kraftig oppvåkning der vi oppdager at vi ikke vet hva vi snakker om, eller at forståelsen er svak og diffus. Operasjonelle definisjoner spesifiserer hvordan et begrep eller konstrukt kan måles. Psykologiske og pedagogiske konstrukter rommer ofte en meningsfylde som operasjonelle definisjoner ikke greier å fylle. Operasjoniseringen kan i enkelte tilfeller dekke deler av et begrep, men ikke hele. Operasjoniseringen kan inneholde informasjon som er irrelevant for begrepet og utgjør en systematisk feilkilde i data. De to mulige problemene ved operasjoniseringer kalles konstrukt underrepresentasjon (vi måler for lite av begrepet) og konstrukt irrelevans (vi måler på siden av begrepet). Begge problemene utgjør systematiske feil i målinger (Kleven, 2002) fordi vår operasjonisering gjør at vi treffer feil ved *alle målinger*.

Problemene knyttet til begrepsvaliditet oppstår fordi vi opererer med begreper på både et abstrakt nivå; teorinivå, og et empirisk nivå eller målenivå. Konstruktens abstrakte karakter innebærer at vi må konkretisere og bestemme operasjoner eller prosedyrer for å måle begrepet – operasjonisering. I prinsippet foretar vi en *slutning* fra et abstrakt nivå til et konkret målenivå, og mulighetene for målinger kan vi tenke oss befinner seg i «rommet» mellom det abstrakte nivået og det empiriske nivået. Dette rommet, eller universet av mulige operasjoniseringer, skaper begrepsvaliditetsproblemet *nettopp* fordi det er flere muligheter (noen er ukjente for oss), og vi må velge. Konstrukt underrepresentasjon og konstrukt irrelevans er resultater av feilvalg. Essensialisten nekter å innse at det er flere muligheter i rommet. Atferdsanalytikeren bør ikke melde seg på essensialistens lag, og bør akseptere at det finnes hypotetiske konstrukter og begrepsvaliditetsproblematikk. Legg merke til at

metalismeproblemet (Moore, 2008) og forklaringsfiksjonsproblemet (Skinner, 1950) ikke først og fremst gjelder slutninger fra konstrukt til måling, men fra uavhengige variabler til avhengige variabler og effekt, og angår dermed ikke begrepsvaliditet, men indre validitet (Shadish, Cook, & Campbell, 2002).

Konstruktproblemer i Atferdsanalysen

Det finnes flere eksempler på konstruktproblemer og konstruktdiskusjoner i moderne atferdsanalyse. Et eksempel er redefineringsproblemet av etablerende operasjoner til motivasjonelle operasjoner med underbegreper (Laraway, Snyerski, Michael, & Poling, 2003) og en avgrensning til settinghendelser. Et annet og helt grunnleggende eksempel på konstruktdiskusjon er om det er grunn til å opprettholde skillet mellom positiv og negativ forsterkning (Baron, & Galizio, 2005; Michael, 1975; 2006).

Et tredje eksempel dreier seg om hvordan begrepet «variabilitet» skal forstås og måles. Barba (2012) har i en oversiktsartikkel kritisert variabilitetsbegrepet som særlig er kjent gjennom forskningen til Neuringer (se Barba, 2012). Neuringer hevder at variabilitet er en overordnet dimensjon ved operanter på linje med frekvens og styrke; variasjon i en sekvens med responser kan eksempelvis forsterkes. Kritikken fra Barba dreier seg både om selve begrepet variabilitet, om de operasjonaliseringene Neuringer legger til grunn for å måle variabilitet er gode nok, og om hovedkonstruktet «operant». Artikkelen til Barba ble fulgt opp av flere kommentarer (se Palmer, 2012 for en oversikt). Holth (2012a; 2012b) stiller kritiske spørsmål: Hvordan skal operanten avgrenses og responser telles når nye responser kan genereres og inngå i klassen?

Holth peker også på at forsterkning og ekstinksjon kan forklare variabilitet, og dermed trengs ingen andre forklaringer enn de som allerede er kjent. Marr (2012) bringer diskusjonen videre og mener at begrepet generaliserte operanter; som imitasjon, rela-

tering og regelfølgning, er dårlig analysert. I en, muligens for depressiv, kommentar skriver Marr: «Perhaps the greatest challenge in developing behavior analysis as a natural science is the lack of any clear notion about the units of analysis: What does it mean, if anything, to talk about units of behavior?» (s. 241). Eksempelene over viser at atferdsanalysene opererer med hypotetiske konstrukt slik Svartdal (2014) hevder, og analyser av konstrukt og mulige operasjonaliseringer er påkrevet. Analyser av begreper og slutninger fra teoretiske konstrukt til empirisk måling utgjør rett og slett et utviklingsområde for atferdsanalysen.

Omfanget av publiserte artikler og bøker om validitetsproblematikk i atferdsanalysen er lav, og søk i arkivene til Journal of Applied Behavior Analysis gir få treff på «construct» og «construct validity». Det er likevel mulig å skille mellom to ulike posisjoner når det gjelder behandling av begrepsvaliditet i atferdsanalytiske miljøer. Enkelte hevder at begrepsvaliditet ikke er relevant. Begrunnelsen er at man forholder seg til åpen og lett identifiserbar atferd hvor operasjonaliseringen «gir seg selv», jfr. «behavioral» dimensjonen i Baer, Wolf og Risleys (1968) klassiske artikkel om atferdsanalysens syv dimensjoner. Andre erkjenner at atferdsanalytikere beskjefter seg med hypotetisering og konstruerte begreper og derfor må forholde seg til begrepsvaliditetsdiskusjoner og slutninger (se Silva, 1993; Svartdal, 2014a; 2014b).

Validitet som Gyldige Slutninger

Alle målinger og bedømminger inneholder feil. I en ideell verden med full eksperimentell kontroll vil feilfrie målinger med fullstendige, utfyllende, og avgrensede operasjonaliseringer være mulig. Målingene og observasjonene vil være fri for tilfeldig fordelte feil (reliabilitetsproblematikk) og systematisk fordelte feil (validitetsproblematikk) i slik grad at man snakker om «sanne» skårer (Kleven, 2002). Vår tilgang til virke-

ligheten påvirkes derimot av flere faktorer, som tilkortkommenheten i vårt sansesystem, irrasjonalitet i valg og kulturelle faktorer. Det er med andre ord ikke et en-til-en forhold mellom det vi kan observere og det som anses som «sann kunnskap». Alle forsøk på å beskrive slutninger som valide er derfor tentative, eller omtrentlige. Omtrentligheten er bygget inn i et kritisk realistisk syn på at vi ikke kan vite helt hva som er sant (Cook & Campbell, 1979; Shadish, Cook, & Campbell, 2002). I en verden hvor direkte tilgang til sann kunnskap ikke er mulig, er kritisk vurdering av kunnskap sentralt. Målet er å komme så nær virkeligheten som mulig. Når vi gjennomfører målinger inneholder de observerte skårene, vårt datamateriale, en andel sanne skårer, usystematisk fordelte feil; reliabilitetsproblematikk, og systematisk fordelte feil; validitetsproblematikk. I en kritisk realistisk tilnærming til virkelighet og sannhet er validering sentralt – uansett psykologisk perspektiv.

Validitet har en sentral betydning i psykometri, og anses som det viktigste konseptet når det dreier seg om kvalitetsstandarder i vitenskap (Silva, 1993; se Svartdal, 2014b om behovet for statistikk og andre forskningsdesign enn $N=1$). Messick (1995) beskriver validering som empiriske vurderinger av betydningen og konsekvensene av målinger. Validering kombinerer vitenskapelige undersøkelser og rasjonelle argumenter eller logikk for å støtte eller avvise testskårer og effektmålinger. Fordi validitet handler om tolkninger av funn og slutninger basert på data, kan en test aldri valideres. Validitet kan aldri anslås helt presist, men kun estimeres. Valideringsprosessen er av betydning ikke bare for å gi mening til data: Meningsfylden i data, eller slutningene vi trekker på bakgrunn av data, har konsekvenser for eventuell behandling og videre forskning. Tolkninger og slutninger av data er alltid farget av personene som tolker og tiden og kulturen tolkningene gjøres i. Validering har med andre ord en sosial og politisk side (Messick, 1995).

Cook og Campbell (1979, se også Shadish, Cook, & Campbell, 2002) har utviklet en generell validitetstypologi eller system som rammeverk for validitetsvurderinger av eksperimenter og kausale problemstillinger. Validitet og invaliditet forstås som best mulige tilnærming til sannhet eller falsitet i påstander, inkludert påstander om årsaksforhold. Bakgrunnen for validitetstypologien er at det i vitenskapelige undersøkelser er nyttig å se etter spesifikke grunner til at vi kan trekke falske slutninger og konklusjoner. Trusler mot validitet er faktorer som kan føre til feilslutninger; det er slutninger som er resultat av systematiske feil ved undersøkelsen. Validitetssystemet er i prinsippet et resultat av en «oppsummering» av mulige feil i empiriske studier og logikk.

Validitetssystemet til Cook og Campbell tar for seg fire ulike validitetstyper og trusler mot validitet ved gjennomføring av undersøkelser; statistisk konklusjonsvaliditet, indre validitet, begrepsvaliditet og ytre validitet. Årsaken til at man tar utgangspunkt i fire typer validitet er rent praktisk. Typene korresponderer med fire grunnleggende spørsmål vi stilles ovenfor ved planlegging og gjennomføring av undersøkelser: (1) Eksisterer det en relasjon mellom variablene (X og Y)? (2) Gitt at det er en relasjon, er den sannsynlig kausal fra den ene variabelen (X) til den andre (Y), eller kunne samme effekt vært oppnådd i fravær av intervensjonen? (3) Gitt at relasjonen er sannsynlig kausal, og rimelig sikkert er fra en variabel til en annen, hvilke spesifikke årsaks- og effektbegreper og mekanismer er involvert i relasjonen? (4) Gitt at det er en sannsynlig kausal relasjon fra konstrukt A til konstrukt B, hvor generaliserbar er relasjonen over personer, settinger, tiltaksvarianter, effektmål og tider? Begrepsvaliditet er knyttet til spørsmål (3) om begreper og mekanismer involvert i relasjoner. Statistisk konklusjonsvaliditet er knyttet til statistiske prosedyrer og er derfor mindre aktuell for atferdsanalytikere, siden atferdsanalysens tradisjon for bruk av visuelle analyser gjør andre typer validering

mer aktuelt. Kazdin (2011) har foreslått at data fra innendeltaker-design valideres ved å se etter bestemte typer mønster og mønsterendringer (men, se Svartdal, 2014b).

For å studere et konstrukt må avgjørelser tas om hvilke observerbare variabler man skal bruke som indikator på, eller operasjonalisering, av konstruktet. Resultater av empiriske undersøkelser er direkte knyttet til det operasjonaliserte begrepet. Eksempelvis er en IQ-skåre resultatet på en intelligens-test som er én av flere mulige operasjonaliseringer av kognitiv fungering. Begrepsvaliditet defineres som den grad man har lyktes å operasjonalisere et begrep i samsvar med den meningen som legges i begrepet. Fordi man i psykologisk forskningssammenheng måler prosesser og konstrukturer som ikke alltid er direkte tilgjengelig og observerbare er det umulig å sikre perfekt begrepsvaliditet (Kleven, 2002). For å underbygge antakelser om høy grad av begrepsvaliditet bør en operasjonell definisjon korrelere høyt med andre operasjonelle definisjoner og samtidig korrelere lavt med variabler som måler andre konstrukturer (Campbell & Fiske, 1959; Shadish, Cook & Campbell, 2002). Campbell og Fiske (1959) betegner korrelasjonene som konvergerende validitet og diskriminerende validitet.

Begrepsvalidering i forskning er aktuelt for uavhengig (X) og avhengig variabel (Y) hver for seg, pluss den kausale relasjonen forstått som mekanisme. Faktorer som utgjør trusler mot begrepsvaliditet varierer på tvers av undersøkelser. I tillegg til selve operasjonaliseringen kan evalueringsforståelse, hypotesegjetting, eksperimentatorforventning, og generell oppmerksomhet som gis til en tiltaksgruppe føre til begrepsinvaliditet eller at vi måler andre faktorer enn vi hadde tenkt å måle. Svak testreliabilitet reduserer begrepsvaliditeten fordi varierende resultater og mange tilfeldige feil kan være et tegn på manglende eller skjev dekning av det relevante begrepet i operasjonaliseringen (for inngående beskrivelse av validitetstypologien, les Shadish, Cook, & Campbell, 2002).

Atferdsanalyse og Variabler med Relevans for Begrepsvaliditet

Iboende eller Intrinsisk Validitet

I det atferdsanalytiske miljøet har interessen for begrepsvaliditet vært liten av flere grunner. Ett hovedskille mellom atferdsanalysen og andre psykologiske retninger er av metodologisk karakter. Det sentrale trekket ved atferdsanalytisk metodologi er vektlegging av direkte observasjon av atferd og et relativt lite omfang av abstrakt teoriutvikling og metodeutvikling sammenliknet med andre psykologiske retninger. Målinger av observerbare responser, motoriske eller verbale, er ofte ikke antatt å måle noe underliggende begrep – kun den operasjonaliserte responsen (Baer, Wolf, & Risley, 1987).

Dersom vi kobler direkte måling av atferd til validitetsteori er iboende validitet eller intrinsisk validitet et aktuelt begrep fordi enkelte målinger kan oppfattes å ha iboende validitet eller gyldighet (Ebel, 1961). Måling av ferdigheter som dansing, lesing eller pianospilling er eksempler. Måling hvor flere eksperter kan bli enige om hvordan mestrings kan operasjonalisere er ikke ment å måle mer enn akkurat de spesifiserte ferdighetene. Jo mer konkrete responsenheter vi kan observere, jo bedre er det. En av de grunnleggende forutsetningene for mange atferdsbedømminger er altså at atferdsutvalget eller operasjonaliseringen ikke måler noe utover seg selv; intrinsisk validitet (Goldefried & Sprafkin, 1974, i Silva, 1993).

Mellomliggende variabler og Latente variabler

En årsak til liten vekt på begrepsvaliditet i atferdsanalyse skyldes en mulig forveksling av konstrukturer med underliggende intrapsykiske enheter som eksempelvis «ego» og «selvtillit». Et kjennetegn ved radikalbehaviorismen er dens radikale standpunkt til private hendelser: Atferdsanalysen har som standpunkt at private hendelser, som tenking og føling, ikke skiller seg vesentlig og prinsipielt fra observerbar atferd (Skinner,

1953). Private hendelser kan forstås i lys av en persons forsterkningshistorie, og kan analyseres på samme måte som observerbar atferd, men uten mulighet for enighet mellom flere observatører. Der andre psykologiske perspektiver beskriver at offentlig atferd er forårsaket av tenkning, hevder eksempelvis Skinner (1953) at «...de to hendelsene enkelt kan attribueres til de samme variablene» (s. 279, vår oversettelse). Atferdsanalysen har som utgangspunkt at underliggende intrapsykiske enheter, som «sinnnet», ikke eksisterer og heller ikke er nødvendige for å forklare menneskelig atferd og utvikling. Moore (2008) er klar i sin tale:

Presumably, the hypothetical construct is not useful because it affords some unique logico-theoretical insight into another dimension. That kind of insight does not exist for anybody, including those who appeal to hypothetical constructs. There is no such dimension, so there can be no such insight (s.280).

MacCorquodale og Meehl (1948) beskrev tidlig at diskusjonene om teoretiske konstrukt i atferdsanalysen har bakgrunn i at det psykologiske samfunnet bruker to typer teoretiske konstrukt. Den ene typen er «mellomliggende variabler», og den andre er «latente» variabler». Mellomliggende variabler er teoretiske konstrukt som er mulige å redusere ned til observerbare fenomen – et eksempel kan være «leseflyt». De inneholder ingen antakelser om tilstedeværelse av uobserverbare prosesser. Muligheten til å få direkte tilgang til det teoretiske begrepet, gitt at reduksjonen og operasjonaliseringen til observerbare fenomener er god nok, har vi omtalt som operasjonisme. Mellomliggende variabler har ingen underliggende meningsfylde utover det som kan observeres (Moore, 1995), j.fr. intrinsisk validitet.

Latente variabler er på den annen side teoretiske begreper som inneholder en meningsfylde som ikke kan måles, i hvert fall ikke med dagens tilgjengelige teknologi. Vi kan altså ikke måle latente begreper

direkte, men må bruke observerbar atferd som indikator på det begrepet som måles; for eksempel selvfølelse; IQ og aggresjon (Bollen, 2003). Enhver operasjonalisering og reduksjon til observerbare fenomener vil alltid mangle en del av meningsfylde av begrepet som skal undersøkes. Ergo innehar slike begreper ingen iboende validitet og operasjonaliseringene gir mulighet for over- eller underrepresentasjon av begrepet.

Atferdsanalytisk teori hevdes, på bakgrunn av sine direkte observasjoner, å ha fordeler med henblikk på operasjonalisering av teoretiske begreper. Begrepene som benyttes hevdes å være empirisk avledet og innebærer ikke måling av latente variabler. De fenomenene man er opptatt av å forklare og undersøke har sin rot i observasjoner av stimulusendringer og atferdsendringer hos mennesker og dyr. Derfor oppfattes behovet for omfattende operasjonalisering som mindre og sannsynligheten for feilslutninger derfor redusert (Baer et al., 1987).

Noen konstrukt i atferdsanalyse kan hevdes å være mellomliggende variabler i tråd med Nunnally (1978) som beskriver at begreper om årsakssammenhenger kun er hypoteser om hvilke atferder som vil samvariere på en pålitelig måte. Standpunktet har likhetstrekk med behavioristisk vitenskapsteori og epistemologi; menneskets atferd beskrives som ikke-determinert av indre eller ytre faktorer, men som probabilistisk i sin eksistens. Man advarer i atferdsanalytisk litteratur mot forvirring mellom konstrukt og virkelighet eller foreslåtte enheter, som i et essensialistisk perspektiv skulle være underliggende for observerbare responser (Silva, 1993).

Atferdsanalytikere vil hevde at en stor ulempe ved bruk av samlebetegnelser, som psykologiske begreper ofte er, er at slike samlebetegnelser eller ”merkelapper” kan føre til en konklusjon om at ”merkelappen” er en *ting* inne i en person og mulig årsak til personens handlinger (Grant & Evans, 1994). Når denne slutningen foretas har merkelappen blitt det Skinner (1950) omtalte som en

”forklaringsfiksjon”. Observert atferd har med andre ord endret seg til å bli en indre forklaring på den samme atferden, men legg merke til at slutninger om effekter dreier seg om indre validitet. Konstrukturer befinner seg i atferdsanalytisk tenkning slik sett ikke ”bakenfor” atferd, men mellom den som handler og personen som forsøker å forstå handlingene. Med andre ord kan man si at konstrukturer ikke henger sammen med personen som handler, men «ligger i øynene» hos den som ser eller mer presist forekommer som verbal atferd (Cronbach, 1971; i Messick, 1995). Konstrukturer er i sin natur epistemologiske, ikke ontologiske. Her vil atferdsanalytikerene hevde at psykometrisk litteratur innimellom har en tendens til ikke å skille ordentlig mellom konstrukturer og indre egenskaper med ontologiske kvaliteter som er antatt å styre ytre atferd (Silva, 1993).

Så langt kan det sies å være et skille i psykologien mellom måling av hypotetiske konstrukturer og konstrukturer som er basert på klasser av observerbare hendelser, hvor atferdsanalytiske begreper hevdes å befinne seg i den andre gruppen. Derfor har forfattere av atferdsanalytisk metodelitteratur tradisjonelt hatt liten tro på indirekte målinger av atferd. Kritikken mot indirekte målinger ligger i at indirekte målinger i stor grad er unøyaktige og i verste fall ikke på noen måte måler det de er tenkt å måle. De er altså antatt å være invalide. Eksempelvis er det i følge Baer og kolleger (1987) slik at en personlighetstest, eller en test på depresjonsnivå, ikke nødvendigvis måler personlighet eller depresjon, men kun evne til å snakke om eller skrive om begrepene personlighet og depresjon; altså ren verbalatferd.

Fordelen med en atferdstilnærming er, i følge Baer og kolleger (1987), at konstrukturer som angst, oppmerksomhet, spontanitet, lagring og gjenhenting ofte representerer atferdsmessige realiteter som ennå ikke har blitt oppfattet som det. En atferdsmessig bedømming «tvinger» fram de realitetene som skuler seg bak begrepet. De fortsetter sin redegjørelse for atferdsfokuset med å peke

på at gjennom historien har mange skydd atferdsanalysen som teori, men benyttet atferdsbaserte teknikker i behandling. Baer og kolleger beskriver det slik:

The past 20 years have shown us again and again that our audiences respond very negatively to our systematic explanations of our programs and their underlying assumptions, yet very positively to the total spectacle of our programs—their procedures and their results—as long as they are left “unexplained” by us” (Baer, Wolf, & Risley, 1987, s. 316).

En mer gjennomtenkt språkpraksis vil, i følge mange atferdsanalytikere, redusere bruken av unødige mentalistiske eller latente forklaringer av atferd og gjøre validering mer oversiktlig. Eksempelvis brukes konstruktet personlighet som startpunkt for store deler av et menneskes atferdsrepertoar, på tross av at de fleste er enige om at atferd startes med utgangspunkt i ytre stimuli. Atferdsanalytikere er altså opptatt av observerbar atferd og ønsker minst mulig grad å benytte hypotetiske konstrukturer. Er det dermed mulig å si at begrepsvaliditet rent faktisk er irrelevant i atferdsanalytisk sammenheng, eller fins det en motstand mot å se egne operasjonaliseringer som avledning av konstrukturer?

Det er mulig å spørre om de teoretiske begrepene i atferdsanalysen alltid beskjeftiger seg med er observerbare hendelser, eller om noen faktisk er konstrukturer. Et eksempel er begrepet forsterkning. Forsterkning, i betydningen forsterkningsprosedyre, er mulig å redusere til observerbar atferd. Når vi derimot snakker om begrepet forsterkning som i forsterkningsprosess, får vi en utfordring. Forsterkning i form av økt sannsynlighet for forekomst av liknende atferd, ved liknende diskriminative stimuli på et senere tidspunkt, skaper et tidsmessig gap. Det aktuelle spørsmålet som gjenstår er hvordan vi kan beskrive det som skjer mellom forekomst av forsterkere og senere forekomst av atferd. Her må vi ty til hypotetiske begreper som kanskje ikke har intrinsisk meningsfylde.

Om Begrepsvaliditet er Uinteressant for Atferdsanalytikere

I diskusjoner av atferd, indirekte og direkte målinger og operasjonalisering av konstrukter kommer man fort inn på en diskusjon om hva et begrep egentlig er; hvorvidt et begrep er summen av sine empiriske operasjonaliseringer eller noe mer. De fleste som er opptatt av begrepsvaliditet vil si at et konstrukt ikke er mulig å redusere ned til sine empiriske referenter, fordi det er definert av et nettverk av relasjoner mellom indikatorer (Messick, 1975). De fleste konsepter og begreper som benyttes inneholder, i følge Hogan og Nicholson (1988), en betydelig ekstra dimensjon som ikke kan operasjonaliseres fullstendig.

Selv begrepet «atferd», slik det brukes i atferdsanalytisk praksis kan anses å være et begrep med en andel vanskelig definerbare referenter. Moore (2008) understreker den relasjonelle dimensjonen ved atferd som nærmest essensiell for atferdsanalysen:

...the sense of behavior that is relevant to radical behaviorism is that it is an interaction between organism and environment that has particular properties as a result of certain functional relations that obtain between the features of the behavior and features of the environment. The interaction may have developed phylogenetically or ontogenetically, and represents a central characteristic of the organism as it progresses through its life cycle (s. 68).

Atferd, slik den er interessant for atferdsanalysen, er altså ikke i form av enkeltresponser, men respondering i relasjon med betingelser i omgivelsene (Lee, 1988). Atferd kan slik anses å være et teoretisk, og hypotetisk konstrukt. Iversen (2013) gir et godt bilde av problemet med uobserverbare relasjoner som brukes i atferdsanalytisk litteratur når han diskuterer «operant responsklasse». Skinner beskrev i sin redegjørelse for seleksjon av atferd at en organismes responser selekteres tilfeldig fra en klasse med responser som antas å ha samme funksjon i omgivel-

sene (Skinner, 1935). Iversen (2013) stiller spørsmål ved hva Skinner mente da han skrev at responser «... selekteres tilfeldig fra hele klassen» (vår oversettelse).

Det sentrale problemet med Skinners beskrivelse er at dersom det fins en responsklasse, så må responsene befinne seg et sted når de selekteres. Seleksjonsprosessen innebærer naturlig at det er en samling å velge fra, på samme måte som vi kan velge fra en samling klinkekuler i et glass. Hvor befinner responser i en klasse seg før de selekteres? Hvor befinner de responsene som ikke selekteres seg? Iversen hevder at spørsmål om hvor responsklasser befinner seg mellom hver gang en respons selekteres muligens ikke er tilgjengelig for empirisk undersøkelse. Han spør om atferdsanalysen faktisk har nytte av begreper som referer til responser som ennå ikke har forekommet eller til uobserverbare seleksjonsprosesser. Vi kan dermed stille spørsmål ved begrepsvaliditeten i begrepet «operant responsklasse». At forskere *post hoc* definerer responsklasser i tråd med kriterier for hvilke responser som forsterkes i et bestemt eksperiment er en sak. Beskrivelser av teoretiske funksjonelle responsklasser som er utilgjengelig for empiriske studier medfører derimot spekulasjoner og kan lett være en inngang til forklaringsfiksjoner uten at det er hensikten (Fryling, 2013).

Som i det tidligere eksemplet, vil vi teknisk sett operere med et teoretisk begrep når vi går fra å snakke om forsterker til å snakke om forsterkning eller når vi beskriver funksjonelle responsklasser (Messick, 1981). Svartdal har tidligere beskrevet at atferdsanalysen på mange måter forholder seg som en "bortkommen sønn" som ikke vil hjem i forhold til andre psykologiske retninger (1998). Han hevder at atferdsanalytiske kjernebegreper i like stor grad som andre psykologiske begreper må kunne oppfattes som konstruerte teoretiske begreper.

Ulike Hvorfor-spørsmål i Psykologien

Holth (2013) gir en beskrivelse av ytterligere en mulig årsak til at ulike psykologiske

grener ikke ser ut til å kunne enes om bruk av begreper og forklaringer. I forlengelse av Holths beskrivelse finner vi en årsak til at mange behaviorister ikke anser begrepsvaliditet som viktig i konseptuelle diskusjoner. Holth hevder at grunnlaget for uenighet mellom ulike psykologiske retninger følger av at de ulike psykologiske retningene stiller ulike «Hvorfor»-spørsmål om det folk gjør, og at svarene er ulike med hensyn til abstraksjonsnivå.

Hvis vi spør hvorfor noe skjer, eksempelvis hvorfor et vindu knuses, vil en beskrivelse av forutgående hendelser aksepteres av de fleste; vinduet ble truffet av en stein. Dersom vi går videre og spør hvorfor vindusglass knuses er det mulig å henvise til at glass er et skjørt materiale. En slik forklaring gjør det mulig å predikere at vinduer kan knuses når bestemte forhold ligger til rette for det. Kognitiv psykologi er i stor grad opptatt av denne typen forklaringer. Årsaker til at en person gjør som hun eller han gjør finnes i psykiske strukturer. De psykiske strukturene brukes som forklaring på at personen gjør som hun eller han gjør. Forklaringen antas å gjøre prediksjon av senere atferd mulig, men gir ikke mulighet til å påvirke atferden fordi årsakene ligger i strukturen og er gitt. En tredje type «Hvorfor»-spørsmål er de som handler om mekanismene som ligger bak det folk gjør. I vinduseksemplet dreier det seg om hvilke mekanismer i vindusglasset som gjør at det knuser lett; molekylstrukturen i glass. Nevrologien, nevropsykologien og nevrovitenskapen har denne type spørsmål som sine kjernes spørsmål.

Svarene på «Hvorfor»-spørsmål som dreier seg om psykiske strukturer og indre mekanismer er ikke svarene på det «Hvorfor»-spørsmålet atferdsanalytikere stiller. Når atferdsanalytikere spør hvorfor en person gjør som hun/han gjør, er vi i hovedsak interessert i atferdsforklaringer som er knyttet til manipulerbare hendelser i omgivelsene. Slike forklaringer gir mulighet til påvirkning ved å endre omgivelsene. De «Hvorfor»-spørsmålene som stilles av andre psykologiske grener

gir mulighet til beskrivelse og prediksjon, men ikke til kontroll og påvirkning.

Perspektivet om at atferd kun kan forklares fullstendig med de omgivelseshendelsene den er en funksjon av er sannsynligvis en grunn til at begrepsvaliditet ikke har vært mye omtalt i atferdsanalytisk metodelitteratur. Når forklaringer og årsaker til atferd antas å ligge i observerbare omgivelser, er det lett å videre anta at begrepene som vi bruker om lovmessigheter i læring og utvikling er observerbare per se.

Atferdsbaserte vurderinger og det man kan kalle ”tradisjonell psykologisk måling” kan kontrasteres med hensyn til flere faktorer. En sentral forskjell er atferdsanalysens vekt på direkte observasjon av atferd i naturlige omgivelser. Psykometrien fokuserer på data som er akkumulert i unaturlige settinger via indirekte målinger av latente variabler. Selv om kontrastene mellom tradisjonene sier noe om idealet innen atferdsbasert forskning, forteller forskjellene lite om hvorvidt dataakkumuleringen faktisk skiller seg fra tradisjonell måling og bedømming (Achenbach, 1982). I praksis er forskjellen ofte mindre enn det som kan forventes med tanke på grunnleggende ulikheter. Ingen tilnærming er monolittisk, og atferdsanalytisk måling varierer på samme måte som tradisjonelle målingstilnærminger. Også atferdsbasert behandling kan baseres på indirekte data som intervjuer, skåringsskjemaer og sjekklister.

En diskusjon mellom Friman, Wilson og Hayes (1998) og Lamal (1998) om hva som kan være gangbare operasjonaliseringer av atferd er et eksempel på hvordan ulike perspektiver på begreper i atferdsanalysen arter seg. Friman og kolleger (1998) beskriver at det er viktig å inkludere klientens egne uttalelser om emosjoner i forskning. Lamal (1998) svarer i en kommentar at følelser og tanker ikke er atferd, men hypotetiske konstrukt som ikke kan observeres eller verifiseres av andre, og derfor ikke er relevant for atferdsanalyse. Han henviser til Skinners uttalelser om at det ofte er liten korrespondanse mellom private hendelser og verbale

responser. Lamal går så langt som å hevde at: "This criticism does not deny the existence and importance of private events. Rather, it questions the values of referring to undefined internal processes that can be observed by others indirectly, if at all" (s. 705). Han står steilt på at verdien av folks egne rapporter om eller målinger av indre atferd har liten verdi, og at atferdsanalysen kun bør omfatte seg med det som hevdes å være rene observerbare responser. På samme måte hevder Baum (2011) at private hendelser er irrelevante for atferdsanalytikere, og at begrepene «tenking» og «følning» ikke bør inngå i omtale av private hendelser, fordi ordene leder legfolk inn i forestillingen om «sinnnet». Han sammenlikner verbal atferd med fuglesang slik verbal atferd får sin funksjon først når andre hører den. Når noen snakker om tanker og følelser, trenger vi derfor ikke å se de indre atferdene som årsak til snakkingen. Vi må heller lete etter tidligere og samtidige hendelser i omgivelsene som årsaker til atferd. Hendelsene kan være usynlige når atferden forekommer, men har vært offentlige tidligere i forløpet. Baum hevder at når verbale utsagn ses i en videre tidssammenheng, enn det ett enkelt øyeblikk gjør mulig, blir skjulte hendelser overflødige i forklaringen av atferd.

Palmer (2011) svarer Baum med at hans perspektiv på betydningen av private hendelser kan være nyttig i eksperimentelle analyser av atferd, men ikke i tolkning av atferd på utsiden av laboratoriet. Han understreker rollen private hendelser har i en fullstendig teori om atferd.

Samtidig med utviklingen mot forskning på mer kompleks uobserverbar atferd, har nevrovitenskapen gjort funn som i stor grad overensstemmer med grunnleggende funn og teoretiske begreper innen atferdsanalysen. Operant betinging, ekstinksjon, habituering og metning er fenomener som nevrovitenskapen forholder seg til med den største selvfølge (Carlson, 2009). Når forskningens jobb endres fra slutninger om observerbar atferd til slutninger om forholdet mellom atferd og hjernefunksjon, er begrepsvali-

ditet åpenbart sentralt. Nyere forskning på forholdet mellom hjernefunksjon og atferd omhandler eksempelvis nevrobiologisk bakgrunn for ADHD (som for øvrig er en samlebetegnelse for en rekke ulike observerbare atferder). Deler av forskningen har en atferdsanalytisk basis, og begrepsvalidering er en naturlig del av denne forskningsprosessen (Sagvolden, Johansen, Aase & Russell, 2005).

Oppsummering

Relevansen av atferdsanalytisk teori, filosofi, vitenskapsteori og metodologi for praktisk atferdsanalyse er ikke alltid umiddelbart åpenbar (Fryling, 2013). Relevansen er sannsynligvis minst åpenbar for personer som i liten grad blir eksponert for det filosofiske og teoretiske grunnlaget for atferdsanalyse. I enkelte tilfeller kan opplæring i atferdsanalyse bære mer preg av avansert operasjonalisering, måling og grafisk framstilling av data enn teoretisk og konseptuell utgreiing av rasjonalen for atferdsanalytisk tilnærming til praksis. Det er en fare for at utdanninger og miljøer hvor tjenesteyting står i sentrum i liten grad kommer i kontakt med de konseptuelle sidene av atferdsvitenskapen.

Atferdsanalysen involverer få hypotetiske konstrukturer sammenliknet med andre psykologiske tradisjoner, og vekten på operasjonaliserbare begreper er stor. Konseptualisering og teoriutvikling innebærer likevel mer enn operasjonelle definisjoner og målingssystemer. De fem arbeidsområdene i atferdsanalysen henger tett sammen, og god praksis og tjenesteyting er avhengig av en grundig forståelse av konseptuelt og teoretisk tilfang (Løkke, Arntzen, & Løkke, 2012). Erkjenningen av at mange av våre atferdsanalytiske begreper er samlebetegnelser, abstraksjoner og hypotetiske konstrukturer er viktig for videre utvikling av den konseptuelle atferdsanalysen.

Deler av den atferdsanalytiske litteraturen kan leses som at begrepsvaliditet ikke har vært eller er relevant siden studier av enkle atferdsformer ikke innebærer trusler mot begreps-

validitet på samme måte som begreper med høyere grad av abstraksjon. Motstanden mot å erkjenne bruk av teoretiske konstrukturer, uansett om atferdsformene er komplekse eller ikke, har sannsynligvis sammenheng med atferdsanalysens motstand mot latente variabler og mentalistiske forklaringer. Dersom det, i metodologisk sammenheng, er mulig å skille mellom hypotetiske konstrukturer, abstraksjoner, og begrepsvaliditet på den ene siden, og reifisering av atferdsprosesser med mentalistiske forklaringer som resultat på den andre siden, vil kanskje færre atferdsanalytikere ha motstand mot teoretiske konstrukturer og begrepsvalidering.

Referanser

- Achenbach, T. A. (1982). Assessment and taxonomy of children's behavior disorders. I B.B. Lahey & A.E. Kazdin (eds.) *Advances in Clinical Child Psychology, vol 5*. New York: Plenum Press.
- Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis, 1*, 91–97.
- Baer, D. M., Wolf, M. M., & Risley, T. R. (1987). Some still-current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis, 20*, 313–327.
- Barba, L. S. (2012). Operant variability: A conceptual analysis. *The Behavior Analyst, 35*, 213–227.
- Baron, A., & Galizio, M. (2005). Positive and negative reinforcement: Should the distinction be preserved? *The Behavior Analyst, 28*, 85–98.
- Baum, W. M. (2011). Behaviorism, private events, and the molar view of behavior. *The Behavior Analyst, 34*, 185–200.
- Bollen, K. A. (2003). Latent variables in psychology and the social sciences. *Annual Review of Psychology, 53*, 605–634.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait–multimethod matrix. *Psychological Bulletin, 56*, 81–105.
- Caws, P. (1957). A reappraisal of the conceptual scheme of science. *Philosophy of Science, 24*, 221–234.
- Caws, P. (1959). Definition and measurement in physics. I C. W. Churchman & P. Ratoosh (Red.), *Measurement: Definitions and theories* (ss.3–17). New York: Wiley.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analytical issues for field settings*. Chicago: Rand McNally.
- Carlson, N. R. (2009). *Physiology of behaviour (10. utg)*. New York: Pearson Education.
- Ebel, R. (1961). Must all tests be valid? *American Psychologist, 16*, 640–647.
- Eilifsen, C., Vie, A. & Arntzen, E. (2011). Eksperimentelle studier av hukommelse innen kognitiv psykologi og atferdsanalyse. *Norsk Tidsskrift for Atferdsanalyse, 38*, 115–135.
- Friman, P. C., Hayes, S. C., & Wilson, K. G. (1998). Why behavior analysts should study emotion: the example of anxiety. *Journal of Applied Behavior Analysis, 31*, 137–156.
- Fryling, M. J. (2013). Theory, philosophy, and the practice of applied behavior analysis. *European Journal of Applied Behavior Analysis, 14*, 45–54.
- Grant, L., & Evans, A. (1994). *Principles of behavior analysis*. New York: Harper Collins College Publishers.
- Hogan, R., & Nicholson, R. A. (1988). The meaning of personality test scores. *American Psychologist, 43*, 621–626.
- Holth, P. (2012a). The creative porpoise. *European Journal of Behavior Analysis, 13*, 87–89.
- Holth, P. (2012b). Variability as an operant? *The Behavior Analyst, 35*, 243–248.
- Holth, P. (2013). Different sciences as answers to different why questions. *European Journal of Behavior Analysis, 14*, 165–170.
- Iversen, I. (2013). Response classes: What are they, where do they come from, and where do they go? *European Journal of Applied Behavior Analysis, 14*, 57–68.

- Kazdin, A. E. (2011). *Single-case research designs. Methods for clinical and applied settings* (2. utg). New York: Oxford University Press.
- Kleven, T. A. (2002). Begrepsoperasjonalisering. I T. Lund (red.) *Innføring i forskningsmetodologi*. (s.141–183). Oslo: Unipub.
- Lamal, P. A. (1998). Advancing backwards. *Journal of Applied Behavior Analysis*, *31*, 705–706.
- Laraway, S., Snyerski, S., Michael, J., & Poling, A. (2003). Motivating operations and terms to describe them: some further refinements. *Journal of Applied Behavior Analysis*, *36*, 407–414.
- Lee, V. (1988). *Beyond behaviorism*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Løkke, J. A., Arntzen, E., & Løkke, G. E. H. (2012). Inndeling av atferdsanalysen i subsystemer eller arbeidsområder. *Norsk Tidsskrift for Atferdsanalyse*, *39*, 63–71.
- MacCorquodale, K. & Meehl, P. E. (1948). On a distinction between hypothetical constructs and intervening variables. *Psychological Review*, *55*, 95–107.
- Marr, M. J. (2012). Operant variability: Some random thoughts. *The Behavior Analyst*, *35*, 237–241.
- Messick, S. (1975). The standard problem: Meanings and values in measurement and evaluation. *American Psychologist*, *30*, 955–966.
- Messick, S. (1981). Evidence and ethics in the evaluation of tests. *Educational Researcher*, *10*, 9–20.
- Messick, S. (1995). Standards of validity and the validity of standards in performance assessment. *Educational Measurement: Issues and Practice*, 5–8.
- Michael, J. (1975). Positive and negative reinforcement, a distinction that is no longer necessary; or a better way to talk about bad things. *Behaviorism*, *3*, 149–155.
- Michael, J. (2006). Comment on Baron and Galizio (2005). *The Behavior Analyst*, *29*, 117–119.
- Moore, J. (1995). Radical behaviorism and the subjective-objective distinction. *The Behavior Analyst*, *18*, 33–49.
- Moore, J. (2008). *Conceptual foundations of radical behaviorism*. Cornwall-on-Hudson, NY: Sloan.
- Nunnally, J. C. (1978). *Psychometric Theory*. New York: McGraw-Hill.
- Palmer, D. C. (2011). Consideration of private events is required in a comprehensive science of behavior. *The Behavior Analyst*, *34*, 201–207.
- Palmer, D. C. (2012). Is variability an operant? Introductory remarks. *The Behavior Analyst*, *35*, 209–211.
- Pedhazur, J. E. & Schmelkin, L. P. (1991). *Mesurement, design, and analysis: An integrated approach*. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.
- Sagvolden, T., Johansen, E. B., Aase, H., & Russell, V. A. (2005). A dynamic developmental theory of attention – deficit /hyperactivity disorder (ADHD) predominately hyperactive/ impulsive and combined subtypes. *Behavioral Brain Sciences*, *28*, 397–419.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin Company.
- Silva, F. (1993). *Psychometric foundations and behavioral assessment*. Newbury Park, USA: Sage Publications.
- Skinner, B. F. (1935). The generic nature of the concepts stimulus and response. *Journal of General Psychology*, *12*, 40–65.
- Skinner, B. F. (1950). Are theories of learning necessary? *Psychological Review*, *57*, 193–216.
- Skinner, B. F. (1953). *Science and human behavior*. New York: Macmillan.
- Sundberg, M. L. (1991). 301 research topics from Skinner's book *Verbal Behavior*. *The Analysis of Verbal Behavior*, *9*, 81–96.
- Svartdal, F. (1998). Med Farsethås på telttur. Et svar til Farsethås sin leder i nr. 1 1998.

- Diskriminanten*, 25, 31–34.
- Svartdal, F. (2014). Hypotetiske konstrukter innefor atferdsanalyse: Finnes de? Ja, men der lever de dessverre ikke i beste velgående. *Norsk Tidsskrift for Atferdsanalyse*, 41, 119–131.
- Svartdal, F. (2014b). Reforming behavior analysis: The third way. *European Journal of Behavior Analysis*, 15, 99–102.
- Williams, B. A. (1986). On the role of theory in behavior analysis. *Behaviorism*, 14, 111–124.

Hypothetical Constructs are a necessity, and Behavior Analysts should be more concerned with Construct Validity

Gunn E. H. Løkke og Jon A. Løkke
Høgskolen i Østfold

This article is a comment on Svartdal (2014) and a theoretical account and discussion of the position of construct validity in behavior analysis. Based on Cook and Campbell's validity typology, we discuss the view on construct validity found in behavior analytic literature and practice. We describe possible traits of behavior analysis that reduce the focus on construct validity, and why behavior analytic terms, contrary to common assumptions, can be coined as non-observable constructs. Our claim is that construct validity is far more relevant to behavior analysis than assumed, due to frequent use of constructs. As technology makes it possible to study increasingly more complex behavior, discussions of construct validity should become more numerous.

Keywords: hypothetical constructs, construct validity, operationalization, inferences