

Måleegenskaper ved en skala for sosial akseptabilitet. En pilotstudie

Børge Strømgren¹, Jon A. Løkke² og Stian Orm³

¹OsloMet – storbyuniversitetet, ²Høgskolen i Østfold og ³Frambu

En vesentlig dimensjon ved anvendt atferdsanalyse er at tiltak skal være viktige for mottakeren og anerkjennes av relevante personer rundt; anvendt dimensjonen. En raffinering av dimensjonen førte til utvikling av begrepet sosial validitet; i hvilken grad et tiltak oppfattes som sosialt akseptabelt av omgivelsene. Formålet med studien er å gjøre tilgjengelig et norskspråklig instrument for vurdering av én dimensjon ved sosial validitet: akseptabilitet av tiltak. Vi har oversatt og revidert et engelskspråklig instrument sammen med andre profesjonelle terapeuter. Så testet vi ut den norske versjonen med andre-års vernepleierstudenter på to tidspunkter ved hjelp av to kliniske vignetter. Skalaens måleegenskaper ble vurdert gjennom analyser av strukturvaliditet (faktoranalyse) og reliabilitet. Instrumentet har god strukturvaliditet med én faktor og god reliabilitet. Skalaen er kort, enkel å administrere, og kan vise seg å ha nytteverdi i klinisk praksis og videre forskning. Videre forskning bør undersøke skalaens egenskaper med større og bredere sammensatt utvalg og i ulike kontekster.

Nøkkelord: Sosial validitet, sosial akseptabilitet, tiltaksvurdering, faktoranalyse, reliabilitet, strukturvaliditet

I 1968 ble *Journal of Applied Behavior Analysis* (JABA) lansert. I det første nummeret gjorde Baer et al. (1968) et forsøk på å beskrive innholdet i anvendt atferdsanalyse i form av sju dimensjoner. En av dimensjonene, den anvendte, handler om at anvendt atferdsanalyse i motsetning til eksperimentell atferdsanalyse tar sikte på å løse problemer som er viktig for mottakeren og samfunnet (Baer et al., 1968). Den anvendte dimensjonen har satt sitt preg på anvendt atferdsanalyse i form av at normative premisser og sosial validitet er ansett som viktig (Ferguson et al., 2019; Løkke & Salthe, 2012). Sosial validitet omhandler omgivelsenes bedøm-

ming av et tiltaks *mål*, teknikker og effekter. Sosial validitet dreier seg om tiltaket alt i alt oppfattes som sosialt akseptabelt (Wolf, 1978).

I en empirisk vitenskap er reliabel og valid innsamling av data viktig. Reliabilitet handler blant annet om instrumenters pålitelighet; dersom du måler noe flere ganger med samme instrument skal du få omtrent det samme resultatet. Validitet handler blant annet om instrumentets gyldighet, at instrumentet *måler det vi faktisk tror at det måler* – at de ulike utsagnene *på en skala* faller innenfor eller dekker det samme begrepet (Løkke & Løkke, 2014), i vårt tilfelle sosial validitet.

I anvendt atferdsanalyse er det minst tre mulige evalueringsområder for tiltak; (1) evaluering av effekt og måloppnåelse, (2) deltakerens subjektive opplevelse av tiltak og teknikker og (3) i hvilken grad tiltaket oppfattes som sosialt akseptabelt av folk i omgivelsene (Løkke & Salthe, 2012). Wolf

Børge Strømgren, ORCID: <https://orcid.org/0000-0001-7841-5693>

Det er ingen interessekonflikter. Alle forfatterne har godkjent siste versjon av manuskriptet.

Vi takker alle studentene som deltok i studien, samt Kristin Utgård og Anders Dechsling for bistand i forbindelse med oversettelse av skalaen for sosial validitet.

Setteredaktør har vært Torunn Lian.

Korrespondanse vedrørende manuskriptet kan rettes til Børge Strømgren; borge.stromgren@oslomet.no

(1978) skriver følgende om “social validity” eller “social importance”¹: «I wasn’t sure what «social importance» meant or, worse still, how to measure it. (...) Unfortunately, that sounded slightly subjective to me. And subjective criteria have not been very respectable in our field» (s. 203). Wolf tar opp to viktige poenger for atferdsanalyse; (1) sosial validitet er et samlebegrep uten klar avgrensning, og (2) det ligger i begrepet at vi er opptatt av folks holdninger og opplevelser, altså noe subjektivt.

Baer et al. (1987) påpeker i oppfølgingsartikkelen om de sju dimensjonene i anvendt atferdsanalyse, i likhet med Wolf (1978), at selvrapportering og måling av subjektive meninger og opplevelser ligger utenfor atferdsanalysens domene og forskningsmetodikk. Baer et al. mener at selvrapportering eller rapportering fra nærpersoner er problematisk og ikke er foretrukket som målemetode i anvendt atferdsanalyse. Vi oppfatter at metodekritikken er mer rettet mot kausale problemstillinger enn deskriptive. Vi har ikke direkte tilgang til folks oppfatninger, så vi må bruke indirekte mål i form av selvrapportering.

Hvis det er slik som Baer et al. (1987) påpeker, at sosial validitet er et *subjektivt anliggende som ikke kan måles*² i atferdsanalytisk forskningstradisjon, men likevel er en sentral dimensjon ved anvendt atferdsanalyse, står atferdsanalytikere overfor et problem. En løsning på problemet er å være opptatt av det første evalueringsområde skissert over, effekt og måloppnåelse, og ikke være videre opptatt av de to andre, deltakerens opplevelse og sosial validitet, som omhandler subjektive opplevelser og oppfatninger. Vi vil hevde at en slik tilnærming er problematisk og vil kunne føre til en praksis der «målet helliger

¹Wolf sin bruk av begreper «social importance» refererer her til samme begrep i Baer, Wolf og Risley (1968) sin artikkel om anvendt atferdsanalyse, at Applied Behavior Analysis skal løse problemer «which are socially important» (s. 92), og særlig dimensjonen «Applied».

²Denne argumentasjonen gjelder det subjektive elementet, noens mening om eller holdning til noe som selvrapporteres og ikke er basert på observerbar atferd. Skalaer brukt i atferdsanalytisk forskning som måler IQ eller adaptiv atferd er på den annen side basert på observert atferd, og vi refererer ikke i denne sammenheng til slike skalaer.

middelet». Anvendt atferdsanalyse har vært anklaget for en slik instrumentell tilnærming, der man ikke bryr seg om deltakerens subjektive opplevelse. Vektlegging av normative premisser og sosial validitet er blitt brukt som motargument til slike påstander (Løkke et al., 2019). Det er behov for systematiske metoder for å måle sosial validitet.

Baer et al. (1987) foreslår at mange ulike observatører kan observere tiltak og effekter, enten direkte eller ved videoopptak. En slik tilnærming er etter vår mening sjelden mulig i praksis, det vil være både tungvint og komme i konflikt med ønsket om og retten til privatliv. En annen løsning er, slik Wolf (1978) foreslår (9 år før han tilsynelatende ombestemte seg i 1987-artikkelen) å bruke intervjuer og spørreskjemaer (s. 209), og fordi bruk av intervjuer og spørreskjemaer for å måle sosial validitet er ...

«... not a rejection of our heritage. Our use of subjective measures does not relate to internal causal variables. Instead, it is an attempt to assess the dimensions of complex reinforcers in socially acceptable and practical ways. It is an evolutionary event that is occurring as a function of the contingencies of the applied research environment; contingencies that our founders would probably say they appreciate, if we had the nerve to ask them for such subjective feedback on our behavior» (s. 213).

Vi lar sitatet til Wolf sette punktum i diskusjonen, og mener at det er liten tvil om at atferdsanalytikere har behov for reliable, valide og anvendbare instrumenter for å måle folks subjektive oppfatning av et tiltak.

På tross av stor motstand mot bruk av spørreskjemaer innenfor atferdsanalyse, tyder mye på at nettopp spørreskjemaer er vanlige metoder for å måle sosial validitet. I en litteraturgjennomgang av single-subjekt forskning med elever med emosjonelle og atferdsmessige problemer fant man at 16 av 22 studier hadde benyttet ulike former for spørreskjemaer til å måle sosial validitet

(Spear et al., 2013). En generell litteraturgjennomgang i JABA fra 1999 fant at kun 13% av artiklene som ble publisert i de første 31 årene av JABA inkluderte mål på sosial validitet (Carr et al., 1999). Verken Spear et al. eller Carr et al. har spesifisert om de studiene som har inkludert mål på sosial validitet har brukt valide og/eller reliable verktøy for dette.

Trenden ser ikke ut til å være endret. En oppfølgende litteraturgjennomgang fra 2019 fant at kun 12% av studier publisert i JABA fra 1999 til 2016 inkluderte mål på sosial validitet (Ferguson et al., 2019). En litteraturgjennomgang av studier publisert i Norsk Tidsskrift for Atferdsanalyse viste at kun 3% av studiene som er blitt publisert fra 2006 til 2019 inneholder målinger av sosial validitet (Løkke et al., 2020). Igjen, om målemetodene for sosial validitet som er brukt er reliable og valide rapporteres ikke, men det er grunn til å anta at det stort sett er snakk om hjemmelagede skalaer som ikke har gjennomgått evaluering av måleegenskaper. I en nylig publisert studie i JABA hvor et av forskningsspørsmålene eksplisitt handlet om sosial validitet, ble det ikke benyttet et evaluert måleinstrument og instrumentet er ikke åpent tilgjengelig (Mann & Karsten, 2020).

Sosial akseptabilitet av tiltak

Subjektiv oppfatning av et tiltak, eller dets sosiale validitet, er fremdeles definert slik som Wolf (1978) opprinnelig gjorde det: Sosial *akseptabilitet* av (1) mål for/behov for atferdsendring, (2) metoder og teknikker brukt, og (3) resultater og/eller utkommer. Sosial validitet oppfattes derfor som et multidimensjonalt konstrukt med de nevnte tre faktorer, eller underdimensjoner, i atferdsanalyse, generell undervisning, spesialpedagogikk og klinisk psykologi (Carter, 2010a; Common & Lane, 2017; Foster & Mash, 1999; Lane et al., 2009; Snodgrass et al., 2018). Videre påpeker Kazdin (1980) følgene om dimensjonen akseptabilitet; «Akseptabilitet refererer til vurderingene

av behandlingsprosedyrer som gjøres av ikke-profesjonelle, legfolk, klienter og andre mulige brukere av behandling» (s. 259, vår oversettelse). For formålet med denne studien ser vi ingen grunn til å endre definisjonene.

Det har vært utviklet måleinstrumenter for *akseptabilitet* innenfor anvendt atferdsanalyse siden 1980 (Carter, 2010c; Common & Lane, 2017; Miltenberger, 1990). Først ute var Kazdin (1980), som benyttet en eksperimentell analog metode for utprøving av en måleskala (ratingskala) (Miltenberger, 1990). Den analoge metoden går ut på at deltakere (ofte universitets- eller høskolestudenter på dette stadiet i utviklingen av skalaen) presenteres for en casebeskrivelse og ett eller flere foreslåtte tiltak mot problematferden beskrevet i casen, også kalt vignett. Deltakerne vurderer så tiltaksbeskrivelsen på den utviklede skalaen (Carter, 2010d; Miltenberger, 1990). Hvordan skalaen er utviklet, hvilke utsagn som inngår og hvorfor er ofte uklart beskrevet. Kazdin (1980) resonnerer at vurderinger av akseptabilitet sannsynligvis omfatter vurdering av om behandlingen er tilpasset problemet, er rettferdig, passende eller rimelig, inngripende eller innenfor konvensjonelle normer. Videre skriver han at 45 utsagn ble generert og at det det gjestod 16 som ut fra *face validity* så ut til å være mest relatert til behandlingsevaluering. Spørsmålene omhandler hvor akseptabelt man mener tiltaket er, hvor villig man er til å bruke det, hvor passende tiltaket er, osv. Denne måten å utvikle målingskaler på er ganske vanlig, noen ganger ut fra en klar definisjon eller et klart konsept, andre ganger er definisjon/konsept mindre avklart, og resultatene fra en psykometrisk evaluering av skalaen kan da avgrense og utvikle definisjonen eller konseptet videre (for eksempel Boateng et al., 2018).

Kazdin (1980) presenterte 144 studenter i psykologi (introduksjonskurs) med fire vignetter bestående av én beskrivelse av et barn med avvikende atferd, med fire påfølgende tiltaksbeskrivelser. Hver student vurderte hver behandling på en ratingskala,

Treatment Evaluation Inventory (TEI), som bestod av 16 utsagn med en Likert vurderingsskala fra 1-7. Resultatene, de fire ratingene fra hver student, ble faktoranalyser, hvorpå ett utsagn ble fjernet slik at skalaen bestod av én faktor eller konstrukt, som han betegnet *treatment acceptability*, med 15 utsagn. Med utgangspunkt i TEI har det senere blitt utviklet en kortversjon for foreldre til elever, TEI-SF med ni av de opprinnelige 15 utsagnene (Kelley et al., 1989), og en versjon for foreldre til barn som mottar behandling i klinikker (TARF) (Reimers & Wacker, 1988).

Witt og Martens (1983) undersøkte en skala, *Intervention Rating Profile* (IRP) der lærere og assistenter vurderte beskrivelser tiltak brukt i spesialundervisning. 180 lærerstudenter vurderte seks ulike intervensjoner. En faktoranalyse ga konstruert *generell akseptabilitet* fra 15 av utsagnene (IRP-15). Tarnowski og Simonian (Tarnowski & Simonian, 1992) baserte sin skala, *Abbreviated Acceptability Rating Profile* (AARP), på IRP-15 i barne- og ungdomspsykiatrien. Mødre til barn som var pasienter på poliklinikken vurderte beskrivelser av behandlinger som kunne tilbys deres barn, fem atferdsbehandlinger og én farmakologisk, med bruk av AARP og Likert-vurdering fra 1-6. En faktoranalyse av AARP-skalaen ga én faktor med åtte utsagn som ble betegnet *treatment acceptance*. Elliott og Treuting (1991) laget en utvidet versjon, BIRS (som inkluderte IRP-15), som et akseptabilitetsmål i skolen *i forkant av* behandling for å øke sjansen for implementering og vedlikehold blant lærere. Lærerne vurderte vignetter, eller beskrivelse av case med tiltak, og de fant tre faktorer som de betegnet *Akseptabilitet* (alle utsagn fra IRP-15), *Effektivitet* (endring, vedlikehold og generalisering), og *Tid til effekt* (hvor raskt atferdsendring vil vise seg).

Lignende framgangsmåter er også brukt i nyere tid i ulike kontekster, for eksempel i generell undervisning (Briesch et al., 2013; Harrison et al., 2016), innen implemen-

teringsforskning (Weiner et al., 2017) og psykiatri (Devilly & Borkovec, 2000) for å nevne noen. Det er sånn sett en lang tradisjon for å vurdere vignetter, eller tiltaksbeskrivelser, på en skala med ulike utsagn som rates på en Likertskala. Disse skalaene har så blitt evaluert med faktoranalyse for å undersøke om de ulike utsagnene, eller leddene, samler seg i en eller flere faktorer.

Konseptualisering av akseptabilitet

Akseptabilitet har vært konseptualisert ulikt, de som har utviklet ratingskalaene har hatt sine egne (Carter, 2010b). Carter omtaler noen enklere: (1) «Working model of treatment acceptability» *der faktorer som en behandlings akseptabilitet, effektivitet, integritet og anvendelse gjensidig påvirker hverandre* og (2) «Decision-making model of treatment acceptability» *der akseptabilitet er et tidlig ledd i en kjede av faktorer som kan påvirke blant annet overholdelse, effektivitet og vedlikehold*. Noen mer utvidede konseptualiseringer omtales også, slik som «Expansive view», «Epidemiological conceptualization», «Assessment model», «Maintenance model», og «Distributive model». For formålet med denne studien er konseptualiseringen «Decision-making model of treatment acceptability» mest relevant med tanke på at en skala for akseptabilitet skal kunne brukes i forkant av tiltak, underveis og i etterkant.

Så langt vi kjenner til finnes det ikke noen norskspråklig skala for å måle akseptabilitet. Vi vil derfor evaluere måleegenskapene for en skala for akseptabilitet som er oversatt fra Eilers og Hayes (2015) (Appendiks A og B). Skalaen fra Eilers og Hayes er kort og konsis, tilpasset anvendt atferdsanalytisk praksis, og muligens egnet for forskning. Eilers og Hayes benyttet skalaen for å måle akseptabiliteten til eksponering- og responsprevensjon kombinert med kognitiv defusjonering for fire barn med autisme. Skalaen ble distribuert til tre foreldre og tre lærere, det er uklart om det ble gjort før eller etter behandlingen. Det ble ikke rapportert psykometriske analyser. Eilers og Hayes har ikke oppgitt noen kilder for de

utsagnene de har brukt.

Formålet med denne studien er å undersøke den oversatte skalaen med tanke på strukturvaliditet og reliabilitet.

Metode

Deltakere og prosedyre

Totalt ble 82 andre-års vernepleierstudenter på emnet «Innføring i psykologi» rekruttert. Deltakerne fikk distribuert skalaen for sosial validitet sammen med to kliniske vignetter (se Appendiks C). Første tidspunkt for administrering (T1) var i starten av emnet og i pausen til andreforfatters forelesning. Andre tidspunkt for administrering (T2) var midt i emnet, i pausen til tredjeforfatters forelesning. Totalt 76 studenter besvarte spørreundersøkelsen ved både T1 og T2.

Deltakerne ble informert om at det var helt frivillig å delta, at svaret deres ville være helt anonymt, og at de kunne trekke seg når som helst. For å kunne kombinere deltakernes svar på T1 og T2 ble de bedt om å lage et kodenavn og skrive det på besvarelsen. Forfatterne oppfordret til å lage et kodenavn som var lett å huske, eksempelvis en kombinasjon av yndlingsrett, yndlingssang og de fire første sifrene i fødselsdatoen sin.

Design

Deltakerne fikk utdelt skalaen for akseptabilitet og halvparten fikk utdelt vignett 1 først, mens den andre halvparten fikk utdelt vignett 2 først for å utjevne eventuell overføringseffekter fra den ene vignetten til den andre. Deltakerne ble bedt om å lese vignetten og deretter ta stilling til utsagnene på skalaen og vurdere hvor godt de passet til vignetten. Videre ble studentene ble gitt en kort definisjon på sosial validitet, om skalaen «måler om omgivelsene vil finne et tiltak akseptabelt», og stilt spørsmålet «Ser det ut som at skalaen måler sosial validitet?»

Materiell

Skala for akseptabilitet. Skalaen for akseptabilitet som ble benyttet er oversatt

og revidert fra Eilers og Hayes (2015). Skalaen ble oversatt av andre- og tredjeforfatter. Oversettelsen ble deretter kritisk gjennomgått av førsteforfatter og to andre atferdsanalytikere med gode kunnskaper i både norsk og engelsk. Skalaen ble lettere revidert, der spørsmål som i originalversjonen inneholdt ordet *child* ble endret til «deltaker» for å favne en bredere aldersgruppe, og et spørsmål om hvorvidt det er akseptabelt å bruke tiltaket overfor barn som ikke kan velge selv ble tatt ut (utsagn 8 i Appendiks B).

Den oversatte skalaen for vurdering av akseptabilitet består av 11 av 12 opprinnelige utsagn. Skalaen ble oversatt med den originale Likert-skalaen med fem svaralternativer fra svært uenig (1) til svært enig (5) (se Appendiks A og B). Til vår kjennskap er skalaens måleegenskaper ikke vurdert tidligere.

Vignetter. Vi utviklet to kliniske vignetter (små illustrasjoner av case med forslag til behandling) basert på reelle kasus som forfatterne har kjennskap til (se Appendiks). Den ene vignetten omhandler et tiltak rettet mot restriktiv og repeterende atferd (vignett 1, RRB) og ble laget for å illustrere et tiltak med antatt lav sosial akseptabilitet. Den andre vignetten (vignett 2, ANG) illustrerer et angstbehandlingstiltak og var ment å illustrere et tiltak med antatt høy sosial akseptabilitet. Begge vignettene ble på forhånd pilottestet med seks deltakere med ulik bakgrunn innen psykologi og pedagogikk. Pilottestingene viste at vignettene fungerte etter intensjonen; vignett 1 ble skåret betydelig lavere på sosial akseptabilitet enn vignett 2.

Statistiske analyser

Dataene fra T1 og T2 ble sortert slik at skårer fra begge vignetter ble analysert samlet, altså stablet opp på hverandre før analyse, men også hver for seg i separate analyser. For T1 inkluderte det strukturvaliditet og reliabilitet, for T2 kun reliabilitet. All dataanalyse er gjort med SPSS 26. Utsagnet «stablet opp på hverandre» betyr at både en

persons vurdering av vignett 1 og vignett 2 analyseres samlet. En slik fremgangsmåte er vanlig praksis når man for eksempel vurderer egenskaper til skalaer som måler sosiale ferdigheter og problemer, én lærer vurderer flere elever (eks. syv til 11) i sin klasse, og alle inngår i samlet datasett for analyse av psykometriske egenskaper (se for eksempel Arnesen et al., 2018; Tobia et al., 2013).

Manglende data. Data for T1 bestod av 11 variabler, alle hadde manglende data. Av 164 cases, eller svar, hadde 131 (80%) ikke manglende data på noen utsagn, 10 (6.1%) manglet svar på ett utsagn, og 23 (14%) manglet data på alle utsagn. Data for T2 bestod av 11 variabler, alle hadde manglende data. Av 164 cases, eller svar, hadde 84 (51.2%) ikke manglende data på noen utsagn, 29 (17.7%) manglet svar på ett utsagn, én (0.6%) manglet data på to utsagn, og 50 (30.5%) manglet data på alle utsagn.

Analysen ble gjennomført med valgene «missing listwise» (kun T1- eller T2-data) eller «missing pairwise» (T1- og T2-data), slik at kun caser med ingen manglende data ble med i analysene, i SPSS kalt «valid cases», heretter kun kalt cases. For T1-data betyr det at antall cases var 131 når alle 11 variabler var med i analyse, og antall cases var 133 når åtte variabler var med i analysen. For T2-data var det kun åtte variabler med i alle analyser, og da var det 84 caser med i analysen.

Eksplorative faktoranalyse (EFA) for T1-data. Da vi ønsket å finne en eller flere latente underliggende faktorer ble faktoranalyse valgt framfor prinsippal komponentanalyse (Field, 2018; Watkins, 2018). Vi sjekket faktorabiliteten til T1 data gjennom Bartlett's test for «sphericity» og brukte Kaiser-Maier-Olkin (KMO) sitt kriterium for tilstrekkelig «sample» (minimum $> .50$ og $\geq .90$ er marvelous) (Watkins, 2018). Faktoranalysen ble gjennomført med *Principal Axis Factoring* og *Promax* rotasjon. Resultatene ble tolket basert på *factor matrix*, og koeffisientene måtte være statistisk signifikante for å være meningsfulle. Signifikans ble estimert med formelen for 1 % signifikansnivå (α) og 5 %

signifikansnivå (β , henholdsvis $\geq .45$ og $\geq .36$) (Watkins, 2018), faktorkorrelasjoner under $.36$ ble derfor ikke med i utskrift. Endelig bør kommunalitetens verdier (delt varians) for hvert utsagn overstige $.5$ (Field, 2018).

Reliabilitet for T1-data. Etter EFA beregnet vi indre konsistens for totalskalaen med *Cronbach's alpha* (α) (Tabell 3). Vi beregnet også *Corrected Item-Total Correlations* (CI-CT) (bør være $> .3$), og *Cronbach's alpha if Item Deleted* (α if ID) (kan være lik men ikke større enn α for totalskalaen) (Chalmers, 2018; Field, 2018; Watkins, 2018) for å anslå reliabilitet for totalskalaen. Vi beregnet også gjennomsnitt og standardavvik for hvert utsagn og for hele skalaen.

Test-retest reliabilitet. Vi målte test-retest mellom T1-data og T2-data, og beregnet både *Pearson's r* og også *Spearman's rho* (r_s) da skalaen er ordinal.

Validitet. Vi undersøkte face validity, altså om utsagnene ser ut til å handle om sosial validitet slik deltakerne forstår begrepet. Videre testet vi om også skalaen kunne skille mellom høy og lav grad av sosial validitet (de to vignettene RBB og ANG) med en t-test for forskjellen mellom T1 og T2. Vi brukte rådata og manglende data ble sortert ut ved «listwise deletion».

Resultat

Initial Exploratory Factor Analysis for T1-data

Resultatene fra initial EFA vises i Tabell 1, kolonner Faktorladninger, øvre panel. *Bartlett's test of sphericity* indikerte at korrelasjonsmatrisen ikke var tilfeldig, $\chi^2(55) = 1304.14$, $p < .001$, og den var sånn sett egnet for EFA. *Kaiser-Meyer-Olkin* verdi på KMO = $.94$, verifiserte *sampling adequacy* for analysen (*marvelous*), utvalget på 131 cases var altså tilstrekkelig med god margin.

For initial EFA med alle 11 variabler var alle korrelasjoner untatt én i *faktorkorrelasjon* over $.45$ som kan anses som adekvat siden de var over signifikansestimaten for 1 % nivå.

Tabell 1. Faktorielladninger og reliabilitet ved initial EFA (øvre panel) og ved ny EFA med beholdte utsagn (nedre panel).

	Faktorielladninger			Reliabilitet	
	Faktor-korrelasjon	Delt varians	Unik varians	CI-CT	α if Item Deleted
Initial EFA, uthevede utsagn ble ikke beholdt (Valid N = 131), total α = .94					
1. Jeg mener det er viktig å endre atferden som tiltaket er rettet mot.	.58	.33	.66	.56	.94
2. Tiltaket er en akseptabel måte å endre deltakerens atferd	.90	.81	.19	.88	.92
3. Jeg er villig til å bruke tiltaket for å endre deltakerens atferd	.89	.79	.21	.86	.92
4. Jeg mener at det er akseptabelt å benytte tiltaket, selv om deltakeren ikke kan samtykke	.45	.51	.49	.43	.94
5. Jeg liker fremgangsmåten som brukes i tiltaket	.86	.79	.21	.86	.93
6. Jeg mener at tiltaket er effektivt for å endre deltakerens atferd	.90	.87	.13	.84	.93
7. Jeg tror tiltaket sannsynligvis fører til varige forbedringer	.87	.80	.20	.83	.93
8. Jeg tror at deltakeren opplevde ubehag under gjennomføringen av tiltaket	-	.25	.75	.12	.95
9. Jeg mener at tiltaket var fordelaktig for deltakeren	.90	.81	.19	.86	.93
10. Jeg mener at fremgangsmåten som ble brukt i tiltaket er velbegrunnede	.82	.70	.30	.81	.93
11. Samlet sett synes jeg tiltaket er positivt	.92	.85	.15	.90	.92
Ny EFA med beholdte utsagn (Valid N = 133), total α = .97					
2. Tiltaket er en akseptabel måte å endre deltakerens atferd	.89	.79	.21	.87	.96
3. Jeg er villig til å bruke tiltaket for å endre deltakerens atferd	.88	.78	.22	.86	.96
5. Jeg liker fremgangsmåten som brukes i tiltaket	.88	.77	.23	.86	.96
6. Jeg mener at tiltaket er effektivt for å endre deltakerens atferd	.90	.81	.19	.88	.96
7. Jeg tror tiltaket sannsynligvis fører til varige forbedringer	.87	.76	.24	.86	.96
9. Jeg mener at tiltaket var fordelaktig for deltakeren	.90	.81	.19	.88	.96
10. Jeg mener at fremgangsmåten som ble brukt i tiltaket er velbegrunnede	.80	.64	.36	.78	.97
11. Samlet sett synes jeg tiltaket er positivt	.92	.86	.14	.91	.96

Note: Delt varians er her det samme som Kommunalitetsverdi.

CI-CT = Corrected Item-Total Correlations, hert enkelt utsagns korrelasjon med totalakala

Kommunalitetsverdier for de enkelte utsagn var over akseptabelt nivå på $> .5$ for ni av 11. En faktor forklarte 62 % av variansen. To variabler, utsagnene 4 og 8, ladet på en 2. faktor, og forklarte 6.3 % av variansen. En faktor med kun to utsagn kan ikke beholdes.

For utsagnene 1 og 8 var delt varians (kommunalitet) under 0.5, noe som tilsier at utsagnene bør tas ut (markert med utheving i Tabell 1, øvre panel). Utsagn 8 hadde i tillegg ingen signifikant faktorkorrelasjon på faktor 1, kun på faktor 2. Utsagnene 1 og 8 ble derfor ikke beholdt.

Reliabilitet for T1-data etter initial EFA.

Reliabilitet etter initial EFA kan sees i Tabell 1, kolonner Reliabilitet, øvre panel. For hele skalaen (11 utsagn) var indre konsistens $\alpha = .94$, og utvalget var 131 cases. Ved inspeksjon av α if ID ser en at utsagn 8 vil gi en $\alpha \geq .95$ dersom det slettes. I tillegg ser en av CI-CT at utsagn 4 og 8 korrelerer dårlig med totalskalaen ($< .5$), derfor ble heller ikke utsagn 4 beholdt videre.

Ny EFA med T1-data

Vi utførte så en ny EFA med de åtte beholdte utsagn (Tabell 1, kolonner Faktorkorrelasjoner, nedre panel). *Bartlett's test of sphericity* indikerte at korrelasjonsmatrisen ikke var tilfeldig, $\chi^2(28) = 1190.04$, $p < .001$, og den var sånn sett egnet for EFA. *Kaiser-Meyer-Olkin* verdi på $KMO = .95$, verifiserte *sampling adequacy* for analysen (*marvelous*), utvalget på 133 var altså tilstrekkelig med god margin.

Separate analyser. Vignettene 1 (66 cases) og 2 (67 cases) viste *sphericity* på henholdsvis $\chi^2(28) = 401.30$, $p < .001$ og $\chi^2(28) = 327.44$, $p < .001$, og KMO var på henholdsvis $.91$ og $.89$.

En samlet analyse viste at én faktor ble foreslått som forklarte 77.77 % av variansen. Faktorkorrelasjoner var alle over $.77$ som kan anses som adekvat siden de var over signifikansestimaterne for 1 % nivå, $\geq .45$. Kommunalitetsverdier (delt varians) for de enkelte utsagn var alle over akseptabelt nivå

på $> .5$.

Separate analyser. Vignettene 1 (66 cases) og 2 (67 cases) viste at én faktor forklarte henholdsvis 65.81 og 57.33 % av variansen. Faktorkorrelasjoner var alle over $.78$ for vignett 1 ($.79$ -.85) og over $.57$ ($.57$ -.90) for vignett 2. Kommunalitetsverdier (delt varians) for de enkelte utsagn var alle over akseptabelt nivå på $> .5$ for vignett 1 ($.64$ -.73) og for vignett 2 ($.57$ -.81) med unntak for utsagnene 6 [.49] 7 [.40] og 10 [.33]).

Reliabilitet for T1-data etter ny EFA.

Reliabilitet etter ny EFA kan sees i Tabell 1, kolonner Reliabilitet, nedre panel. For hele skalaen (8 utsagn) var indre konsistens $\alpha = .97$, og utvalget var 133. Siden α if ID ikke gir en større α enn for totalskalaen kan totalskalaen anses som reliabel. Videre korrelerer alle utsagn bra med totalskalaene (CI-CT) og har verdier langt over $.3$. Gjennomsnitt for ny totalskala var $M = 25.42$, $SD = 8.46$.

Separate analyser. Separate analyser for vignettene 1 (66 cases) og 2 (67 cases) viste indre konsistens på henholdsvis $.94$ og $.91$. α if ID ga ikke større α enn for totalskalaen for noen av vignettene, henholdsvis $.93$ for vignett 1 og $.89$ -.91 for vignett 2. Endelig korrelerer alle utsagn bra med totalskalaene (CI-CT) for både vignett 1 ($.76$ -.82) og 2 ($.54$ -.85).

Reliabilitet for T2-data

Reliabilitetesten for T2-data hadde 84 valide caser med de åtte variablene som ble beholdt fra EFA for T1-data. Indre konsistens for hele skalaen var $\alpha = .96$. α if ID varierte fra $.95$ til $.96$ (utsagn 6), og alle utsagn korrelererte bra med hovedskalaen (CI-CT), verdiene var fra $.74$ til $.93$. Gjennomsnitt for totalskalaen var $M = 26.82$, $SD = 8.20$.

Test-retest reliabilitet

Vi undersøkte også test-retest reliabilitet mello T1 og T2 med bruk av sumskårer for de åtte utsagnene som ble beholdt med «pairwise deletion», totalt 95 caser. *Pearsons* $r = .80$, *BCa* [.69, .88], $p < .001$ og *Spearman's* $r_s = .79$, *BCa*

[.67, .87], $p < .001$.

Validitet

Ved T1 ble alle studentene stilt spørsmålet «Ser det ut som at skalaen måler sosial validitet?». Sekstisju studenter svarte på spørsmålet, og av de 67 svarte over tre fjerdedeler (76%) «ja» og under en fjerdedel (24%) «nei». Et stort flertall mente at skalaen måler sosial akseptabilitet. Dette er en vanlig fremgangsmåte for å vurdere *face validity*, altså en vurdering av om utsagnene synes å handle om et tiltaks sosiale validitet.

Vi fant også en forskjell i gjennomsnitt og standardavvik mellom T1-dataene for vignett 1 (RRB), $N = 69$, $M = 19.19$, $SD = 6.56$, $SE = .80$ og vignett 2 (ANG), $N = 72$, $M = 31.06$, $SD = 5.01$, $SE = .59$. En t-test med «listwise deletion» fra rådata viste at forskjellen var signifikant, $t(139) = 12.11$, $p < .001$, $d = 2.04$, 95 % CI [1.63, 2.45].

Forskjell i gjennomsnitt og standardavvik mellom T2-dataene var for vignett 1 (RRB), $N = 57$, $M = 20.14$, $SD = 6.23$, $SE = .83$ og vignett 2 (ANG), $N = 57$, $M = 31.82$, $SD = 4.94$, $SE = .66$. En t-test med «listwise deletion» fra rådata viste at forskjellen var signifikant, $t(112) = 11.09$, $p < .001$, $d = 2.08$, 95 % CI [1.62, 2.53].

Analysen antyder at skalaen oppførte seg slik vi forventet fra teorien; vignetten som ble skrevet for å illustrere lav sosial validitet fikk en lavere totalskår enn vignetten som ble skrevet for å illustrere høy sosial validitet.

Diskusjon

Formålet med studien var å undersøke en norsk versjon av en skala for vurdering av tiltaks akseptabilitet med tanke på strukturvaliditet og reliabilitet. Resultatene viser at de 8 utsagnene som inngår i skalaen tilhører en underliggende faktor, og at denne faktoren har god reliabilitet. Hovedinntrykket holdt seg når de to vignettene ble analysert hver for seg i separate analyser, og selv om resultatene er usikre med så få caser fant vi enkelte lavere faktorkorrelasjoner for vignetten som beskrev et tiltak med antatt høy akseptabilitet.

Skalaen klarte også å skille de to vignettene, vignetten med antatt høy akseptabilitet hadde høyere sumskåre enn vignetten med antatt lav akseptabilitet.

Resultatene ligner på funn med lignende skalaer. Kazdin (1980) rapporterer én faktor med faktorladninger på .61 til .94 for sine 15 beholdte utsagn på TEI, og de forklarte 51.4 % av variansen. Tarnowski og Simonian (1992) rapporterer én faktor for AARP med faktorladninger på .89 til .96 som forklarte henholdsvis 84.9 og 90.3 prosent av variansen.

Tre utsagn ble tatt ut av den opprinnelige skalaen på 11 utsagn basert på for lave faktorkorrelasjoner, lav kommunalitet og også lave reliabilitetsmål. Slik oppnådde vi en skala bedre psykometriske egenskaper. De tre utsagnene som ble tatt ut var; (1) «jeg mener det er viktig å endre atferden som tiltaket er rettet mot», (4) «jeg tror at deltakeren opplevde ubehag under gjennomføringen av tiltaket», og (8) «jeg mener at det er akseptabelt å benytte tiltaket, selv om deltakeren ikke kan samtykke».

En spekulasjon rundt utsagn 1 er at det er mer normativt ladet enn de andre utsagnene i skalaen, og dermed er mer rettet mot de normative premisene og vurderingene i forkant av et tiltak, heller enn estetikk og normative forhold ved selve tiltaket. Utsagn 4 ble også tatt ut, det omhandler samtykke og samtykkekompetanse. I norsk sammenheng kan det tenkes at utsagnet hører mer til det juridiske feltet enn det anvendte, da samtykke og samtykkekompetanse i Norge er regulert i lovverket. Utsagn 8 hadde ingen samvariasjon med de andre utsagnene i skalaen, vi vet ikke hvorfor. Det kan tenkes at både tiltak med lav og høy grad av sosial validitet kan innebære et vist ubehag for deltakeren. Atferdsendrende tiltak innebærer ofte anstrengelser fra deltakerens side, eksempelvis utprøving av nye aktiviteter eller handlinger og eksponering. Vi understreker at de tre utsagnene som er tatt ut av skalaen omhandler viktige forhold og bør inngå i beslutninger om tiltak, men ikke nødvendigvis i en skala som omhandler

akseptabilitet av selve tiltaket.

Det må understrekes at resultatene er usikre og foreløpige og må tolkes med forsiktighet. Det er flere grunner til dette, pilotstudien har noen svakheter som vil vi kommentere, (1) validitet - utsagnene som skalaen består av og kontekst, (2) bruk av vignetter og faktoranalyse, (3) design og utvalgets størrelse og sammensetning.

Validitetsspørsmålet, om utsagnene i skalaen måler akseptabilitet eller ikke, er viktig. Vi har som nevnt i innledningen ikke utviklet skalaen selv, vi har oversatt en som er brukt av Eilers og Hayes (2015). De oppgir ikke hvordan skalaen er utviklet eller om utsagnene er hentet fra en annen skala. Ved første øyekast ser man at de minner om Kazdins (1980) resonnering om at akseptabilitet sannsynligvis omfatter vurdering av om behandlingen er tilpasset problemet, er rettferdig, passende eller rimelig, inngripende eller innenfor konvensjonelle normer. Ved nærmere gjennomgang av de skalaene som er nevnt i innledningen ser man at utsagnene er ord for ord like noen av de som er omtalt, eller de har likt meningsinnhold.

For eksempel er første utsagn i TEI (Kazdin, 1980) formulert slik «How acceptable do you find this treatment to be for this child's problem behavior?», og svaralternativene er fra «Not acceptable at all» til «Moderately acceptable» og «Very acceptable». Det er to blanke avkrysningsmuligheter mellom hvert svaralternativ, slik at det blir en 7-delt skala. Utsagn 2 hos Eilers og Hayes er formulert slik: «The treatment is an acceptable way to change the child's problem behavior», med svaralternativer som «Strongly agree», «Agree», «Neutral», «Disagree» og «Strongly disagree», en fem-delt skala. Witt og Elliott sin IRP-15 (som gjengitt i Carter, 2010c) har denne formuleringen i utsagn 1: «This would be an acceptable intervention for the child's problem behavior», med svaralternativ «Strongly disagree», «Disagree», «Slightly disagree», «Slightly agree», «Agree» og «Strongly agree».

Selv om Kazdins skala er formulert som et spørsmål og Eilers og Hayes sin er formulert som er påstand, mener vi at meningsinnholdet er det samme. Og Witt og Elliott sin formulering er enda likere Eilers og Hayes. De 12 utsagnene som Eilers og Hayes har brukt gjenfinnes i TEI, IRP-15 og AARP (Tarnowski & Simonian, 1992). Vi mener derfor at skalaen har god innholdsvaliditet, den samsvarer godt med andre validerte og anvendte skalaer.

En annen validitetsvurdering er kontekst, de ulike skalaene er formulert for ulike kontekster. TEI var utviklet med tanke på profesjonelles akseptabilitet av behandling av atferdsvansker hos barn i en klinikksetting, mens TARF er en omformulering av TEI ment for foreldres akseptabilitet av behandling av atferdsvansker hos barn i en klinikksetting. IRP ble utviklet med tanke på bruk av tiltak overfor problematferd hos elever, altså en skolesetting. Disse kontekstuelle tilpasningene endrer etter vår mening ikke meningsinnholdet i utsagnene, begreper som «akseptabel måte», «villig til å bruke», «fordelaktig for» går igjen uavhengig av kontekst. Kontekstendringer er begreper som «child» og «student» som angir alder og sted for behandling.

Vi har ment å lage den foreliggende skalaen «kontekstnøytral» ved å bruke «deltakeren» gjennomgående. Allikevel kan utsagnene i den foreliggende skalaen lett tilpasses ulike kontekster, «deltaker» kan byttes ut med «pasient», «elev», «ungdom», osv. Om det er nødvendig eller nyttig kan bare besvares med framtidige studier. I denne sammenhengen er betegnelsen «deltaker» i skalaen å forstå den som skal motta behandling.

Antallet deltakere som vurderte skalaen har også vært lavt og relativt homogent (andre-års vernepleierstudenter), *selv om det har vært tilstrekkelig for faktoranalyse samlet, er det få i hver vignett. Dersom man skal gå videre fra en pilotstudie må man ha en bredere tilnærming med flere caser og flere tiltaksbeskrivelser per case.*

Antallet deltakere må betraktelig opp, og man må ha et mer heterogent tilfang av deltakere.

De samme deltakerne fylte ut skalaen to ganger, en gang for hver vignett, og begge vignetter ble analysert samlet (stabled) og hver for seg. En kritikk her kan være at de samme personene inngår to ganger i samme datasett. Dette mener vi ikke er relevant, det er en vanlig fremgangsmåte i for eksempel analyse av skalaer som måler sosiale ferdigheter eller vansker, der for eksempel én lærer vurderer flere av sine elever og alle elevvurderingene fra samme lærer inngår i analysen, slik som omtalt i metoden.

Vi var opptatt av at begge vignetter skulle analyseres samlet, fordi faktorstrukturen og reliabilitet bør være uavhengig vignett som er vurdert, slik som for eksempel med skalaer for sosial kompetanse og vansker. Allikevel er hver vignett evaluert separat i tillegg, og det gir ikke andre resultater en samlet evaluering. I tillegg ble vignettene presentert i motsatt rekkefølge for halvparten av de som vurderte. Allikevel er det en usikkerhet med hensyn på om man ville fått et annet resultat om ulike deltakere hadde vurdert kun en vignett hver, noe som er en mer robust framgangsmåte og som bør gjøres fremover.

Med de begrensningene som ligger i designet som er brukt, er foreliggende skala valid for de to vignettene og med de respondentene vi brukte, *andre-års vernepleierstudenter*. Vi mener det har vært tilstrekkelig for en pilotstudie, å undersøke om det er en skala det kan forskes videre på før en større studie iverksettes.

En bedre design, flere caser med flere forslåtte tiltak til hver case, flere og mer heterogene deltakere, vil kunne bidra til en mer valid skala. Den vil da kunne anvendes som et mål på akseptabilitet innenfor en konseptualisering som for eksempel «Decision-making model of treatment acceptability» (Reimers et al., 1987) gjengitt i (Carter, 2010b). I modellen er akseptabilitet er et tidlig ledd i en kjede av faktorer som kan påvirke blant annet overholdelse,

effektivitet og vedlikehold. Her inngår ikke kun mottaker, familie og andre i vurderingen, men også de som skal gjennomføre tiltaket, miljøarbeider, teraperuter, eller lærere. Kjeden ser slik ut når alt går greit: Tiltaksbeskrivelse → forståelse av tiltaket → høy aksept → god overholdelse → god effekt → lite negativ påvirkning av familie eller skole → høy grad av vedlikehold. Med lav aksept vil mann kunne få en kjede preget av dårlig overholdelse som gir dårlig effekt og behov for avslutning, justering eller skifte av behandling. Skalaen for akseptabilitet, TARF (Reimers & Wacker, 1988; Reimers et al., 1987) benyttes i denne konseptualiseringen i forkant som en prediktor for videre implementering. En slik anvendelse av en skala for akseptabilitet anser vi er nyttig. Vi tror også det er nyttig å bruke slike skalaer underveis og ved tiltakets opphør, da med justert grammatikk i utsagnene.

Oppsummering

Begrepet sosial validitet har vært en del av anvendt atferdsanalyse siden starten, og er ifølge Wolf (1978) hvordan anvendt atferdsanalyse «finder sitt hjerte» med subjektive målinger eller vurderinger av foreslåtte tiltak; om målene er sosialt signifikante, om prosedyrene er sosialt akseptable eller passende og om effektene eller utkommene er sosialt viktige. Noen tok opp hansken og utformet ulike måleskalaer som har vært brukt i varierende til liten grad (Ferguson et al., 2019). I Norsk Tidsskrift for Atferdsanalyse har slike mål vært brukt i enda mindre grad (Løkke et al., 2020). Den foreliggende studien har evaluert psykometriske egenskaper ved en skala for måling av sosial akseptabilitet som har god reliabilitet og stabilitet over tid. Skalaen synes også valid; den måler sosial akseptabilitet og skiller på grader av sosial akseptabilitet. Den er kort, enkel å bruke og lett tilgjengelig. Noen usikkerheter gjenstår gitt antall vignetter, design og utvalgets størrelse og homogenitet. Videre forskning er derfor anbefalt. I mellomtiden kan vi anbefale at skalaen brukes i vurdering av planlagte atferdsanalytiske og andre tiltak.

Referanser

- Arnesen, A., Braeken, J., Ogden, T., & Melby-Lervåg, M. (2018). Assessing children's social functioning and reading proficiency: A systematic review of the quality of educational assessment instruments used in Norwegian elementary schools. *Scandinavian Journal of Educational Research*, 1–26. <https://doi.org/10.1080/00313831.2017.1420685>
- 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. <https://doi.org/10.1901/jaba.1968.1-91>
- Baer, D. M., Wolf, M. M., & Risley, T. R. (1987). Some still-current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis*, 20(4), 313–327. <https://doi.org/10.1901/jaba.1987.20-313>
- Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best Practices for Developing and Validating Scales for Health, Social, and Behavioral Research: A Primer. *Frontiers in Public Health*, 6(149). <https://doi.org/10.3389/fpubh.2018.00149>
- Briesch, A. M., Chafouleas, S. M., Neugebauer, S. R., & Riley-Tillman, T. C. (2013, 2/1). Assessing influences on intervention implementation: Revision of the Usage Rating Profile-Intervention. *Journal of School Psychology*, 51(1), 81–96. <https://doi.org/http://dx.doi.org/10.1016/j.jsp.2012.08.006>
- Carr, J. E., Austin, J. L., Britton, L. N., Kellum, K. K., & Bailey, J. S. (1999). An assessment of social validity trends in applied behavior analysis. *Behavioral Interventions*, 14(4), 223–231. [https://doi.org/10.1002/\(sici\)1099-078x\(199910/12\)14:4<223::Aid-bin37>3.0.Co;2-y](https://doi.org/10.1002/(sici)1099-078x(199910/12)14:4<223::Aid-bin37>3.0.Co;2-y)
- Carter, S. L. (2010a). Chapter 1 - Background of Social Validity. In S. L. Carter (Ed.), *The Social Validity Manual* (pp. 1–20). Academic Press. <https://doi.org/https://doi.org/10.1016/B978-0-12-374897-3.00001-5>
- Carter, S. L. (2010b). Chapter 2 - Conceptualizations of Social Validity. In S. L. Carter (Ed.), *The Social Validity Manual* (pp. 21–36). Academic Press. <https://doi.org/https://doi.org/10.1016/B978-0-12-374897-3.00002-7>
- Carter, S. L. (2010c). Chapter 3 - Instruments for Evaluating Social Validity. In S. L. Carter (Ed.), *The Social Validity Manual* (pp. 37–67). Academic Press. <https://doi.org/https://doi.org/10.1016/B978-0-12-374897-3.00003-9>
- Carter, S. L. (2010d). Chapter 4 - Research on Social Validity. In S. L. Carter (Ed.), *The Social Validity Manual* (pp. 69–106). Academic Press. <https://doi.org/https://doi.org/10.1016/B978-0-12-374897-3.00004-0>
- Chalmers, R. P. (2018). On misconceptions and the limited usefulness of ordinal alpha. *Educational and Psychological Measurement*, 78, 1056–1071. <https://doi.org/10.1177/0013164417727036>
- Common, E. A., & Lane, K. L. (2017). 4 - Social Validity Assessment. In J. K. Luiselli (Ed.), *Applied Behavior Analysis Advanced Guidebook* (pp. 73–92). Academic Press. <https://doi.org/https://doi.org/10.1016/B978-0-12-811122-2.00004-8>
- Deville, G. J., & Borkovec, T. D. (2000). Psychometric properties of the credibility/expectancy questionnaire. *Journal of Behavior Therapy and Experimental Psychiatry*, 31(2), 73–86. [https://doi.org/https://doi.org/10.1016/S0005-7916\(00\)00012-4](https://doi.org/https://doi.org/10.1016/S0005-7916(00)00012-4)
- Eilers, H. J., & Hayes, S. C. (2015). Exposure and response prevention therapy with cognitive defusion exercises to reduce repetitive and restrictive behaviors displayed by children with autism spectrum disorder. *Research in Autism Spectrum Disorders*, 19, 18–31. <https://doi.org/https://doi.org/10.1016/j.rasd.2014.12.014>
- Elliott, S. N., & Treuting, M. V. B. (1991).

- The behavior intervention rating scale: Development and validation of a pretreatment acceptability and effectiveness measure. *Journal of School Psychology*, 29(1), 43–51. [https://doi.org/https://doi.org/10.1016/0022-4405\(91\)90014-I](https://doi.org/https://doi.org/10.1016/0022-4405(91)90014-I)
- Ferguson, J. L., Cihon, J. H., Leaf, J. B., Van Meter, S. M., McEachin, J., & Leaf, R. (2019). Assessment of social validity trends in the journal of applied behavior analysis. *European Journal of Behavior Analysis*, 20(1), 146–157. <https://doi.org/10.1080/15021149.2018.1534771>
- Ferguson, J. L., Cihon, J. H., Leaf, J. B., Van Meter, S. M., McEachin, J., & Leaf, R. (2019). Assessment of social validity trends in the journal of applied behavior analysis. *European Journal of Behavior Analysis*, 20(1), 146–157. <https://doi.org/10.1080/15021149.2018.1534771>
- Field, A. (2018). *Discovering statistics using IBM SPSS Statistics* (5th ed.). SAGE.
- Foster, S. L., & Mash, E. J. (1999). Assessing Social Validity in Clinical Treatment Research. Issues and Procedures. *Journal of Consulting and Clinical Psychology*, 67(3), 308–319. <https://doi.org/10.1037/0022-006X.67.3.308>
- Harrison, J. R., State, T. M., Evans, S. W., & Schamberg, T. (2016). Construct and Predictive Validity of Social Acceptability: Scores From High School Teacher Ratings on the School Intervention Rating Form. *Journal of Positive Behavior Interventions*, 18(2), 111–123. <https://doi.org/10.1177/1098300715596135>
- Kazdin, A. E. (1980). Acceptability of alternative treatments for deviant child behavior. *Journal of Applied Behavior Analysis*, 13(2), 259–273. <https://doi.org/10.1901/jaba.1980.13-259>
- Kelley, M. L., Heffer, R. W., Gresham, F. M., & Elliott, S. N. (1989). Development of a modified treatment evaluation inventory. *Journal of Psychopathology and Behavioral Assessment*, 11(3), 235–247. <https://doi.org/10.1007/BF00960495>
- Lane, K. L., Kalberg, J. R., Bruhn, A. L., Driscoll, S. A., Wehby, J. H., & Elliott, S. N. (2009). Assessing Social Validity of School-wide Positive Behavior Support Plans: Evidence for the Reliability and Structure of the Primary Intervention Rating Scale. *School Psychology Review*, 38(1), 135–144.
- Løkke, G., & Løkke, J. A. (2014). Hypotetiske konstrukt er en nødvendighet, og derfor bør atferdsanalytikere være mer opptatt av begrepsvaliditet. *Norsk Tidsskrift for Atferdsanalyse*, 41(2), 157–170. <https://nta.atferd.no/journalissue.aspx?IdDocument=507>
- Løkke, J. A., & Salthe, G. (2012). Sjekkliste for målrettet tiltaksarbeid: fra normative og deskriptive premisser til tiltak og evaluering. *Norsk Tidsskrift for Atferdsanalyse*, 39(1), 17–32. <https://nta.atferd.no/journalissue.aspx?IdDocument=308>
- Løkke, J. A., Orm, S., & Strømgren, B. (2020). Kort rapport: Måling av sosial validitet i studier publisert i Norsk Tidsskrift for Atferdsanalyse. *Norsk Tidsskrift for Atferdsanalyse*, 47(1), 1–4. <https://nta.atferd.no/journalissue.aspx?IdDocument=819>
- Mann, C. C., & Karsten, A. M. (2020). Efficacy and social validity of procedures for improving conversational skills of college students with autism. *Journal of Applied Behavior Analysis*, 53(1), 402–421. <https://doi.org/10.1002/jaba.600>
- Miltenberger, R. G. (1990). Assessment of Treatment Acceptability: A Review of the Literature. *Topics in Early Childhood Special Education*, 10(3), 24–38. <https://doi.org/10.1177/027112149001000304>
- Reimers, T. M., & Wacker, D. P. (1988). Parents' Ratings of the Acceptability of Behavioral Treatment Recommendations Made in an Outpatient Clinic: A Preliminary Analysis of the Influence of Treatment Effectiveness. *Behavioral Disorders*, 14(1), 7–15. www.jstor.org/stable/23886135
- Reimers, T. M., Wacker, D. P., & Koeppl, G. (1987). Acceptability of Behavioral Interventions: A Review of the Literature.

- School Psychology Review*, 16(2), 212–227. <https://doi.org/10.1080/02796015.1987.12085286>
- Snodgrass, M. R., Chung, M. Y., Meadan, H., & Halle, J. W. (2018, 2018/03/01/). Social validity in single-case research: A systematic literature review of prevalence and application. *Research in Developmental Disabilities*, 74, 160–173. <https://doi.org/https://doi.org/10.1016/j.ridd.2018.01.007>
- Spear, C. F., Strickland-Cohen, M. K., Romer, N., & Albin, R. W. (2013). An Examination of Social Validity Within Single-Case Research With Students With Emotional and Behavioral Disorders. *Remedial and Special Education*, 34(6), 357–370. <https://doi.org/10.1177/0741932513490809>
- Tarnowski, K. J., & Simonian, S. J. (1992). Assessing treatment acceptance: The abbreviated acceptability rating profile. *Journal of Behavior Therapy and Experimental Psychiatry*, 23(2), 101–106. [https://doi.org/https://doi.org/10.1016/0005-7916\(92\)90007-6](https://doi.org/https://doi.org/10.1016/0005-7916(92)90007-6)
- Tobia, V., Gabriele, M. A., & Marzocchi, G. M. (2013). The Italian Version of the Strengths and Difficulties Questionnaire (SDQ)—Teacher: Psychometric Properties. *Journal of Psychoeducational Assessment*, 31(5), 493–505. <https://doi.org/10.1177/0734282912473456>
- Watkins, M. W. (2018). Exploratory factor analysis: A guide to best practice. *Journal of Black Psychology*, 44, 219–246. <https://doi.org/10.1177/0095798418771807>
- Weiner, B. J., Lewis, C. C., Stanick, C., Powell, B. J., Dorsey, C. N., Clary, A. S., Boynton, M. H., & Halko, H. (2017, 2017/08/29). Psychometric assessment of three newly developed implementation outcome measures. *Implementation Science*, 12(1), 108. <https://doi.org/10.1186/s13012-017-0635-3>
- Witt, J. C., & Martens, B. K. (1983). Assessing the acceptability of behavioral interventions used in classrooms. *Psychology in the Schools*, 20(4), 510–517. [https://doi.org/10.1002/1520-6807\(198310\)20:4<510::Aid-pits2310200420>3.0.Co;2-1](https://doi.org/10.1002/1520-6807(198310)20:4<510::Aid-pits2310200420>3.0.Co;2-1)
- Wolf, M. M. (1978). Social validity: the case for subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis*, 11(2), 203–214. <https://doi.org/10.1901/jaba.1978.11-203>

Psychometric properties of a Norwegian social validity scale. A pilot study

Børge Strømgren¹, Jon A. Løkke², and Stian Orm³

¹OsloMet – Oslo Metropolitan University, ²Østfold University College, and ³Frambu

A principal dimension of applied behavior analysis is that interventions must be important for the recipient and acknowledged by relevant stakeholders; the applied dimension. A refinement of this dimension led to the development of the term social validity; the extent to which an intervention is perceived as socially acceptable. The purpose of the current study is to present, and make available, a Norwegian instrument for assessing one dimension of social validity: the acceptability of an intervention. We have translated and revised the Norwegian version with professional social educators. Next, 2nd year social educator students were presented with two clinical vignettes and assessed those using the scale at two different points in time. Some psychometric properties were evaluated thru analyses of structure validity (factor analysis) and reliability. The scale shows good structure validity with one factor and good reliability. The scale is short, easy to administer, and may be of utility value in clinical practice and further research. Future research should evaluate the scale properties with a larger and broader sample of respondents and in different contexts.

Key words: Social validity, social acceptability, intervention assessment, factor analysis, reliability, structure validity

Appendiks A

Bedømming av et tiltaks anseelse (akseptabilitet)

Oversatt og revidert fra Eilers & Hayes (*Research in Autism Spectrum Disorders*, 2015, s. 25)

av Jon Arne Løkke, Stian Orm & Børge Strømgren, 2020

	Svært enig	Enig	Nøytral	Uenig	Svært uenig
Tiltaket er en akseptabel måte å endre deltakerens atferd					
Jeg er villig til å bruke tiltaket for å endre deltakerens atferd					
Jeg liker fremgangsmåten som brukes i tiltaket					
Jeg mener at tiltaket vil være effektivt for å endre deltakerens atferd					
Jeg tror tiltaket sannsynligvis fører til varige forbedringer					
Jeg mener at tiltaket vil være fordelaktig for deltakeren					
Jeg mener at fremgangsmåtene som blir brukt i tiltaket er velbegrunnede					
Samlet sett synes jeg tiltaket er positivt					

Appendiks B

Original skala fra Eilers & Hayes (*Research in Autism Spectrum Disorders, 2015, s. 25*)

1.	I believe it is important to change the problem behaviors targeted with this treatment
2.	The treatment is an acceptable way to change the child's problem behavior
3.	I am willing to use this treatment to change the child's problem behavior
4.	I believe it is acceptable to use the treatment, even though the child is unable to consent to the treatment
5.	I like the procedures used in this treatment
6.	I believe the treatment is an effective way to reduce the child's problem behavior
7.	I believe the treatment is likely to result in permanent improvement
8.	I believe it is acceptable to use this treatment with children who cannot choose treatments for themselves
9.	I believe the child experienced discomfort during the treatment
10.	I believe the child benefited from this treatment
11.	I believe the procedures used in this treatment are reasonable
12.	Overall, I have a positive reaction to this treatment

Utelatte utsagn er markert med fet skrift, utsagn 8 ble utelatt før skalaen ble testet.

Appendiks C

Vignett 1 (RRB) : Redusere repeterende atferd

Problem: Kristian (13 år) har stor interesse for og bruker mye tid på å se på YouTube videoer av en amerikansk komiker, og blir sterkt påvirket av ting som sies av komikeren. Kristian tar det som sies bokstavelig, så når komikeren sier at juice er farlig, fører det til at Kristian ikke drikker juice på lang tid.

Tiltak: For å redusere Kristians tid på YouTube og interesse for den amerikanske komikeren, så skal personalet og foreldrene ikke snakke om eller kommentere komikeren. Og dersom Kristian tar det opp eller snakker om det, skal det ignoreres, og personalet skal ta opp et annet samtaletema. Tiltaket er detaljert beskrevet i en tiltaksbeskrivelse.

Effekt: Kristian bruker like mye tid på YouTube og komikeren, men han snakker ikke om komikeren med hverken personalet eller foreldrene. I starten maste Kristian mye for å få andre til å respondere på hans snakking om komikeren, men det har gradvis gitt seg.

Begrunnelse: Ekstinksjon (ignorering) benyttes ofte som tiltak når man ønsker å redusere forekomsten av en atferd hos barn, unge og voksne som har atferd som er utfordrende for omgivelsene.

Vignett 2 (ANG): Trening på å sitte alene med døren igjen

Problem: Kristian (13 år) er redd for å være alene i et rom med døren igjen. Det fører til at Kristian er avhengig av at en voksen er i nærheten av ham, hele tiden. Det går utover søsken fordi foreldrene må bruke mye tid med Kristian.

Tiltak: Kristian skal øve på å sitte alene med døren igjen. Treningen foregår ved gradvis eksponering styrt av tilbakemeldinger fra Kristian om hvor mye ubehag han føler ved å sitte alene med døren igjen. Når Kristian ikke føler ubehag trappes tiden oppover.

Effekt: Kristian tør stadig å sitte lengre tid alene med døren igjen. Det tok litt tid før Kristian ga tilbakemelding om at han kunne gå fra det første trinnet i eksponeringen til det andre, men deretter gikk det jevnt fremover.

Begrunnelse: Gradvis eksponering er en evidensbasert teknikk for å redusere angst og engstelse hos barn, unge og voksne som har angst, fobier eller er redd for ting.