

Utvikling av en bedømmingsskala for mentaliseringsbasert terapi



Magnus Johan Engen

Innlevert som hovedoppgave ved Psykologisk institutt, Universitetet i Oslo

April 2009

Sammendrag

- Tittel:** Utvikling av en bedømmingsskala for mentaliseringsbasert terapi
- Forfatter:** Magnus Johan Engen
- Veiledere:** Asle Hoffart og Sigmund Karterud
- Bakgrunn:** Bedømmingsskalaer er nødvendige instrumenter i studier av behandlingseffekt. De anvendes for å gjøre mål på behandlingsintegritet; hvorvidt en behandling ble implementert som intendert. Mentaliseringsbasert terapi (MBT) er en manualisert terapiform som foreløpig ikke har utviklet bedømmingsskalaer skikket til å gjøre valide mål på behandlingsintegritet. Oppgaven beskriver et tidlig stadie av en utviklingsprosess der en bedømmingsskala for MBT (MBTEKS) utformes og testes. Hensikten er å skape en bedømmingsskala som kan supplere MBT behandlingsmanualen og inngå i fremtidige studier av behandlingseffekt.
- Metode:** Oppgaven beskriver metoden som ble benyttet til å utforme skalaens innhold. Dette ble gjort av en gruppe på seks medarbeidere. Skalaen ble så testet ved et interbedømmer (interrater) design der seks hele terapitimer ble skåret.
- Resultater:** Utviklingsprosessen ledet til begrepsdannelse og presisering av problemstillinger knyttet til utvikling av bedømmingsskalaer. De statistiske analysene viser at MBTEKS ved de innledende skåringene ikke lar seg tilfredsstillende reliabelt anvende. Resultatene viser dog til signifikant treningseffekt og tilfredsstillende reliabilitet for ulike ledd i skalaen målt ved intraklassekorrelasjoner.
- Konklusjon:** I det videre arbeidet må de avgrensede problemstillingene som dukket opp under utviklingen avklares. Treningseffekten og erfaringene fra utviklingen vil lede til utforming av et grundig treningsprogram, samt til utvikling av en manual for anvendelse av skalaen.

Forord

Jeg vil først og fremst takke Sigmund Karterud og Asle Hoffart for svært nyttige kommentarer og innspill underveis. Deres faglige tyngde ga meg ro og hjelp meg med å navigere gjennom et enormt og uoversiktlig forskningsfelt. Takk også til Dag-Erik Eilertsen for statistikkhjelp. Min far fortjener en stor takk for omfattende språkvaskarbeid. Til slutt vil jeg også takke kjæresten min, Kjersti Lohne, som har holdt ut med en fraværende og distré oppgavefokusert egoist under innspurten. Ikke bare holdt ut, men kommet med svært nyttige tips og rettelser.

April, 2009

*Art is not the application of a canon of beauty
but what the instinct and the brain can conceive
beyond any canon. When we love a woman
we don't start measuring her limbs.*

-Pablo Picasso

Innhold

Innledning.....	7
<i>Mentaliseringsbasert terapi.....</i>	8
<i>Mentaliseringsbegrepet, utviklingsteori og forståelse av lidelsen.....</i>	8
<i>Endringsmekanismer og endring i MBT.....</i>	9
<i>Terapeutisk holdning/rolle.....</i>	10
<i>Terapeutisk teknikk i MBT.....</i>	11
<i>Manualbasert psykoterapi.....</i>	11
<i>Psykoterapimanualen i forskning.....</i>	12
<i>Presisering av likhetsidealet.....</i>	13
<i>Om det å bedømme psykoterapi - behandlingsintegritet.....</i>	15
<i>Bedømmingsskalaer og empiri.....</i>	17
<i>Trening av bedømmere.....</i>	18
<i>Skillet mellom etterlevelse og kvalitet.....</i>	19
<i>Psykoterapiformens natur.....</i>	20
<i>Mål på behandlingsintegritet i forskningspraksis.....</i>	21
<i>Formål, målsetninger og problemstillinger.....</i>	21
Metoder.....	23
<i>Om måleinstrumentet - MBTEKS.....</i>	23
<i>Medarbeidere.....</i>	24
<i>Terapeuter.....</i>	24
<i>Pasienter.....</i>	24
<i>Utformingsstudien.....</i>	24
<i>Enighetsstudien.....</i>	25
<i>Plenumsdiskusjonen.....</i>	25
<i>Design på enighetsstudien.....</i>	26
Resultater.....	28
<i>Utformingsstudien.....</i>	28
<i>En terskel for skåring.....</i>	31
<i>Etterlevelse og kvalitet.....</i>	32
<i>Enighetsstudien.....</i>	34
<i>Plenumsdiskusjonen.....</i>	34

<i>De statistiske resultatene</i>	36
<i>Skalaen som helhet</i>	36
<i>Reliabilitet ved ICC</i>	38
<i>Variansanalyse - hva forklarer variasjon i skåringer</i>	39
<i>Bedømmeres enighet</i>	40
<i>Variasjon i enighet over terapitimer - treningseffekt</i>	41
<i>Sammenhengen mellom etterlevelse og kvalitet</i>	42
Diskusjon.....	43
<i>Viktige begrensninger ved studien</i>	43
<i>Skalaen som helhet og de spesifikke ledd</i>	44
<i>Terskelspørsmålet</i>	45
<i>Etterlevelse og kvalitet</i>	46
<i>Forskjeller i leddenes natur - aktivitets- og kvalitetsledd</i>	47
<i>Generelle psykoterapikriterier</i>	49
<i>Treningseffekt</i>	49
<i>Veien videre</i>	50
<i>Konklusjon</i>	52
Referanser.....	53
Appendiks A.....	59
Appendiks B.....	67

Innledning

Denne oppgaven beskriver arbeidet med å utvikle en bedømmingsskala for mentaliseringsbasert terapi (MBT); MBT etterlevels- og kvalitetsskala (MBTEKS). Slike bedømmingsskalaer (adherence and competence rating scales) benyttes for å vurdere i hvor stor grad terapeuter praktiserer psykoterapi i tråd med beskrivelser angitt i en behandlingsmanual. De siste 20-30 årene har både kunnskap om, og interesse for, behandlingseffekt av psykoterapi vært økende (Gullestad, 2001), og behandlingsmanualer med tilhørende bedømmingsskalaer anses som nødvendige for å besvare spørsmål om behandlingseffekt. Fordi bedømmingsskalaer har fått en slik viktig rolle i psykoterapiforskningen, er kvalitetssikring av bedømmingsskalaer helt avgjørende. Det skjer en gradvis utvikling av standarder og retningslinjer for hva en bedømmingsskala skal inneholde. De foreløpige bedømmingsskalaer som er laget for MBT (Bateman og Fonagy, 2004; Bateman og Fonagy, 2006) tilfredsstillende ikke nyere standarder for hva en bedømmingsskala skal inneholde, og det er heller ikke gjort undersøkelser av disse skalaenes reliabilitet og validitet. Utviklingen av MBTEKS er derfor et nødvendig bidrag til MBT som manualisert terapiform, og vil etter endt utvikling inngå i studier av MBT og behandlingseffekt (Anthony Bateman, personlig kommunikasjon, 06.03.09). Denne utviklingsstudien vil, ved grundig å beskrive utviklingen av MBTEKS, også gi et teoretisk bidrag til utvikling av bedømmingsskalaer generelt. Eksisterende litteratur på utvikling av bedømmingsskalaer (f. eks. Barber, Abrams & Liese, 2003; Carrol et al., 2000) beskriver i liten grad selve utformingen av innholdet i de respektive skalaer. Et fokus på sentrale problemstillinger knyttet til bedømmingsskalaens anvendbarhet, basert på bedømmernes erfaringer, vil derfor være sentralt i denne oppgaven.

Før selve utviklingsstudien presenteres, følger det her en litteraturgjennomgang av: (1) MBT, (2) behandlingsmanualer, og (3) bedømmingsskalaer.

Mentaliseringsbasert terapi - MBT

MBT er en psykodynamisk orientert langtidsbehandling for borderline personlighetsforstyrrelse (borderline pf) (DSM-IV-TR; American Psychiatric Association, 2000) utviklet av Bateman og Fonagy (2004). Bedømmingsskalaen MBTEKS er forankret i MBT som beskrevet i manualen «Mentalization-Based Treatment for Borderline Personality Disorder: A Practical Guide» av Bateman og Fonagy (2006). I det følgende presenteres relevante deler av teorigrunnlaget, og sentrale begreper innen MBT. Elementer av MBT-behandlingen som ikke er sentrale for utviklingen bedømmingsskalaen MBTEKS, er blant annet gruppeterapeutisk arbeid og retningslinjer for medikamentell behandling.

Mentaliseringsbegrepet, utviklingsteori og forståelse av lidelsen. Mentalisering er det mest sentrale begrepet i MBT. Bateman og Fonagy definerer mentalisering som prosessen der et «individ implisitt og eksplisitt tolker egne og andres handlinger som meningsfulle på grunnlag av intensjonelle mentale tilstander (f.eks. ønsker, behov, følelser, trosoppfatninger og begrunnelser)» (Bateman og Fonagy, 2004). Definisjonen understreker hvor omfattende mentaliseringsbegrepet er. Det beskriver et større konsept sammensatt av flere ulike elementer; det beskriver en prosess, samt en evne utviklet i en tilknytningsrelasjon. I MBT-teori finnes dessuten flere underbegreper for ulike aspekter relatert til mentalisering (f.eks. pretend modus, psykisk ekvivalens; Bateman og Fonagy, 2006). Bredden i begrepet gjør at det overlapper med flere eksisterende begreper i psykologien, og kan i noen tilfeller være vanskelig å skille tydelig fra disse. For eksempel har deler av begrepet i lange tider blitt forsket på under rubrikken «*Theory of Mind*»(ToM) (Premack og Woodruff, 1978). ToM-forskningen er relevant for mentaliseringsteori (Fonagy et al., 2002), men begrepet skiller seg fra mentaliseringsbegrepet ved at det utelukkende vektlegger det kognitive aspektet ved det å forstå andres perspektiver. Mentalisering favner noe bredere. I tillegg til å omfatte det kognitive ved å reflektere rundt ulike sinn, har mentaliseringsbegrepet sine røtter fra utviklingspsykologien, og knyttes til det affektive, selvet og selvopplevelser. Videre er en viktig forskjell fra ToM at mentalisering ofte foregår implisitt; førbevisst, automatisk og intuitivt (Allen, 2006).

Evnen til å mentalisere anses for å være et nødvendig steg på veien mot utviklingen av et sammenhengende (coherent) selv. Problemer i tilknytningsrelasjonen som hindrer utvikling av god mentaliseringsevne, er ifølge Fonagy, Gergely, Jurist og Target (2002) med på å forklare utvikling av borderline pf. I de tidlige leveår er barnet prisgitt omgivelsene, og inngår i en intersubjektiv

prosess med omsorgsgiver. Denne tilknytningsrelasjonen er avgjørende for utvikling av mentaliseringsevnen, ved at barnet under trygg tilknytning mottar markert speiling fra omsorgsgiver på sine selvopplevelser; det at mor kommuniserer empatisk innlevelse i barnets opplevelse (speiling), samtidig som hun kommuniserer at perspektivet er noe ulikt barnets eget (markering), gir barnet mulighet til å internalisere mors differensierte opplevelse (Fonagy et al. 2002). Gjennom disse prosesser i tilknytningsrelasjonen lærer barnet en selv/andre-differensiering som er nødvendig for den pågående utviklingsprosessen av et sammenhengende selv. Alvorlige forstyrrelser i disse prosessene i de tidlige leveår, enten ved at barnets affekter og opplevelser neglisjeres, eller ved at omsorgsgiver ikke makter å romme barnets affektive tilstand ved markert speiling, fører til forstyrrelser i utvikling av mentaliserende evner som igjen er skadelig for utviklingen av selvet. Bateman og Fonagy (2006) hevder denne forstyrrelsen i barnets selvopplevelse kan føre til at barnet tar til seg representasjoner av mor som ikke er differensiert fra dets egne, og som derved blir værende i barnet som «fremmed selv»-representasjoner. Det kan for eksempel være internaliserte representasjoner av den «fryktede» eller «fryktende» mor, og fenomener som proaktiv identifikasjon kan oppstå ved at individer forsøker å eksternalisere det «fremmede selvet», for å skape en følelse av indre sammenheng og kvitte seg med den truende «indre forfølgelsen».

Et viktig poeng ved MBT er at affektene får en sentral plass. Mentalisering og affektregulering er for Fonagy et al. (2002) uløselig tett forbundet. Et koherent selv bidrar til klare selvopplevelser, og mentaliserende evner er avgjørende for å forstå, og affektivt inntone seg, i en interpersonlig kontekst - da spesielt i tilknytningsrelasjoner. Slik forklares interpersonlige vansker hos pasienter med borderline pf i MBT.

Endringsmekanismer og endring i MBT. Det overordnede fokus i MBT er utviklingen av pasientens evne til å mentalisere. MBT er således prosessfokusert til forskjell fra enkelte andre dynamiske retninger som i større grad vektlegger innsikt (f.eks. Transference focused psychotherapy (TFP); Kernberg, Yeomans & Clarkin 2002). Prosessen innebærer at pasienten kontinuerlig, innen for et tålelig nivå, utfordres til å reflektere rundt sammenhenger mellom affekter, selvopplevelser og interpersonlige forhold. Dette oppnås av terapeuten ved kontinuerlig å ha fokus på å forstå pasientens indre tilstand, så vel som sin egen, mens prosessen i relasjonen pågår. Pasienten oppfordres av terapeuten til å forklare egne affektive tilstander og opplevelser, samt forsøke å knytte disse til en interpersonlig kontekst. Denne konteksten vil typisk være opplevelser i et parforhold. Terapeuten oppfordres også til å arbeide i overføringen med et «her og nå»-fokus. I et slikt arbeide er det vesentlig at terapeuten er oppmerksom på at siktemålet ikke er å

generere innsikt for pasienten, men snarere å demonstrere at terapeuten ikke kan vite hva pasienten tenker, samt tydelig demonstrere skillet mellom pasientens og terapeutens sinn. Prosessen i forholdet mellom pasient og terapeut kan beskrives som analog til den utviklingsteorien som MBT baserer seg på. I en MBT-terapeutisk relasjon, utøver terapeuten intervensjoner som i mange tilfeller kan tenkes å ha samme funksjon som markert speiling; terapeuten er en nysgjerrig tilknytningsperson som empatisk søker å forstå pasientens perspektiv, samtidig som vedkommende er utfordrende når pasienten ikke mentaliserer ved å demonstrere nysgjerrighet og en ikke-vitende holdning. Denne mentaliseringsutviklende terapeutiske prosess, er også ment å utvikle pasientens affektregulerende kapasitet og evner til sekundærrepresentasjoner av mentale tilstander. Den innsikt pasienten får, er ikke eksplisitte forklaringer og tolkninger av hvordan fortidige opplevelser og erfaringer henger sammen med nåtidige vansker. Innsikten vil være et resultat av prosess. En endring i MBT vil oppnåes ved at pasienten, både implisitt og eksplisitt, får forståelsen av at menneskesinnet er «ugjennomsiktig»; noe en ikke kan vite noe sikkert om og som derfor må mentaliseres. Denne innsikten gir et grunnlag for at pasienten kan danne relasjoner som i mindre grad preges av ustabilitet og affektive utbrudd (Bateman og Fonagy, 2006).

Bateman og Fonagy (2004) hevder videre at all behandling av borderline personlighetsforstyrrelse er virksom i den grad den fremmer pasientens evne til å mentalisere. Flere ulike behandlinger for denne pasientgruppen har det til felles at pasienten får en tilknytningsperson som gir erfaringer og opplevelser av at en annen har pasientens «sinn i sitt sinn». Denne opplevelsen hevder Allen (2006) er grunnen til at mange eksisterende terapiformer allerede benytter seg av teori og intervensjoner som er mentaliseringsfremmende.

Terapeutisk holdning/rolle. MBT fordrer en aktiv terapeutrolle. Det at terapeuten aktivt engasjerer seg i pasientens mentale liv, er avgjørende for alliansen og den samarbeidende relasjonen. Videre understrekes det også at terapeuten inntar en ikke-vitende, nysgjerrig holdning, og forholder seg håpefull og rolig i møtet med pasientens eventuelle affektive stormer. Dette blir også kalt «mentaliserende holdning» («mentalizing stance») (Bateman og Fonagy, 2006). Pasienten skal finne i terapeuten et forbilde som demonstrerer tydelig i sin væremåte en mentaliserende holdning i relasjonen. Terapeuten må derfor være oppmerksom på eventuelle motoverføringsreaksjoner og affekter, som kan være ødeleggende for posisjonen som den «mentaliserende andre».

Terapeutisk teknikk i MBT. Avgjørende for utvikling av MBTEKS, er fokuset på terapeutferdigheter knyttet til manualen, samt en forståelse av det underliggende målet for alle intervensjonsformer i MBT. Det underliggende målet er å hele tiden ha fokus på pasientens mentale tilstand, og hvordan pasienten forklarer, eller mentaliserer, denne. I den mentaliserende holdningen vektlegger Bateman og Fonagy (2004) det de kaller «mental nærhet» (mental closeness). Terapeuten beskriver og sjekker egen forståelse av pasientens tilstand, samtidig som han eller hun er nøye med å tilpasse forståelsen etter pasientens eventuelle korreksjoner. Dette arbeidet retter seg mot å knytte opplevelser opp mot indre representasjoner, med sikte på å styrke denne bindingen for å gi pasientens selvopplevelser og affekter mer velutviklede sekundære representasjoner. Dette arbeidet er ofte «her og nå»-fokuseret. Terapeuten oppfordres til kontinuerlig å opprettholde et slikt fokus for å arbeide med den mentaliserende prosessen direkte i terapirommet.

Senere i terapiforløpet, når pasientens evne til å knytte primære affektive opplevelser til symbolske representasjoner er bedret, vil MBT også ta i bruk mer klassiske psykoanalytiske intervensjonsformer som f.eks. tolkning av overføringen. Dette blir av Bateman og Fonagy (2006) kalt å «mentalisere overføringen», for å understreke forskjellen fra annen psykoanalytisk teoris arbeid med overføringen. I MBT blir tolkning av overføringen benyttet for å demonstrere skillet mellom selv og andre, og derigjennom styrke pasientens forståelse av hva som tilhører ens egne reaksjoner, og hvordan disse påvirkes av den andre. Slik beskriver manualen også et utviklingsforløp for det psykoterapeutiske arbeidet. Kontekstuelle variabler som i større eller mindre grad gjør bruken av intervensjonstyper aktuelle, er viktige hensyn av betydning for studien.

Manualbasert psykoterapi

En behandlingsmanual skal angi retningslinjer for hvordan man arbeider psykoterapeutisk med en gitt type pasienter. Luborsky og Barber (1993) har definert behandlingsmanualer som faglitteratur bestående av følgende tre elementer: (1) en presentasjon av hovedprinsippene som ligger bak teknikken i terapien, (2) konkrete eksempler på alle beskrevne terapiteknikker, og (3) skalaer og instrumenter for å bedømme terapeuters samsvar med terapimodellen. De fleste typer litteratur som omhandler psykoterapeutisk arbeid, vil i ulik grad inneholde ett eller to av disse elementene, men de vil ikke falle inn under manualkategorien med mindre de inneholder alle tre. Litteratur som passer med definisjonen ble først introdusert på 1960-tallet og inneholdt beskrivelser av behandlingsteknikker for adferdsterapi (Wolpe, 1969). Etter dette ble det lansert en rekke behandlingsmanualer fra ulike psykoterapitradisjoner, og ved midten av 1980-tallet fantes det et

utvalg av manualer innen psykodynamisk (Luborsky, 1984), kognitiv (Beck, Rush, Shaw & Emery, 1979) og interpersonlig terapi (Klerman, Weissman, Rousanville & Chevron, 1984).

Felles for alle psykoterapimanualer er at de tar sikte på, ved presist å beskrive behandlingen, å tjene som et viktig verktøy for opplæring av terapeuter i en spesifikk behandlingsmetode. Deres styrke som opplæringsverktøy er at de gir eksplisitte beskrivelser av terapiteknikk, og at de inneholder insutruer for å kontrollere egen prestasjon som mål på samsvar med manualen (Luborsky og DeRubeis, 1984)

Selv om de har klare fordeler i opplæringsøyemed, viser en studie av Addis og Krasnow (2000) at kun 6% av et utvalg på 891 kliniske psykologer rapporterer utstrakt bruk av psykoterapimanualer i praksis. Dette kan tyde på at manualene enda ikke har fått den aksept i det kliniske miljøet som de etterstreber. I tillegg til å være et verktøy for opplæring av terapeuter, er manualene også viktige i psykoterapiforskningen.

Psykoterapimanualen i forskning. Psykoterapiforskningens øverste siktemål, er å finne svar på spørsmål om hvor vidt en gitt type psykoterapeutisk behandling har positiv effekt for pasientpopulasjonen som mottar behandlingen. Et ofte anvendt design på studier som søker svar på dette spørsmålet, kalles randomiserte kontrollerte forsøk (RCTer) (randomized controlled trials). For å kunne konkludere at psykoterapien er årsak til endring, må slike studier etterstrebe standarder for et eksperimentelt design. Et slikt syn på psykoterapiforskningen leder med nødvendighet til at psykoterapien må behandles som en uavhengig variabel i et eksperiment (Perepletchikova, Treat & Kazdin, 2007). Det er derfor viktig å i størst mulig grad vurdere og kontrollere for at behandlingsvariabelen (psykoterapien) er implementert som intendert, og behandlingsmanualen kan derfor sees på som psykoterapiens forsøk på å imøtekomme de krav som stilles i et eksperiment. Dette vil si at man i enhver klinisk studie etterstreber et ideal om å gi behandlingen likt for alle som mottar den (uniform), slik at den derved vil være repliserbar. Kriteriet om likhet i behandling betyr at man må beskrive behandlingen ved bruk av klare kriterier, samt beskrive metoder for å kontrollere og vurdere behandlingen (Caligor, 2005). Ved å imøtekomme disse kravene, kan man gjøre studier av behandlingseffekt på ulike pasientpopulasjoner, der resultatene lar seg sammenligne på en meningsfylt måte. Slike standarder er utfordrende for all behandling, men spesielt utfordrende er de i psykoterapiens verden. Det er ut ifra slike krav og standarder man må forstå psykoterapimanualenes plass i psykoterapiforskningen, ettersom psykoterapimanualens kanskje viktigste funksjon i forskningen er å sikre at behandlingen gis på en foreskrevet og repliserbar måte.

Denne måten å fremstille manualbasert psykoterapi på har av noen blitt kalt «teknologimodellen» (the technology model; Carroll et al., 2000). Med begreper lånt fra farmakologien angir modellen: (1) at den psykoteraeutiske behandlingen skal gis i en bestemt type dose (f.eks. antall timer per uke over tid), (2) at behandlingsmanualen skal inneholde beskrivelse av aktive ingredienser (bestemte typer intervensjoner som ifølge en modell er endringsfremmende for pasienten), (3) at det er eksplisitte betingelser for administrering (f.eks. gis behandlingen til pasienter med en bestemt diagnose), og (4) en vurdering/monitorering av hvorvidt behandlingen ble adekvat gjennomført (ulike skalaer for etterlevelse og kvalitet, samt en vurdering av andre forhold som anses å være relevante for vellykket behandling). Fra et teoretisk ståsted er dette idealer som det er vanskelig å være uenig i, da beskrivelsen inneholder prosedyrer som all annen klinisk forskning følger. Det kan dog argumenteres for at psykoterapi ikke er som «all annen klinisk behandling», og synet på psykoterapiens «annerledeshet» er grunnlaget for mye av kritikken som rettes mot paradigmet om manualbasert behandling.

Manualbasert psykoterapi regnes av mange for å være et viktig skritt på veien mot imøtekommelse av kravet om likhet i behandling, og derved bedre psykoterapistudiers eksperimentelle design (Luborsky og Barber, 1993; Perepletchikova et al., 2007). Mange hevder likevel at psykoterapi, fordi den er av en langt mer fleksibel art enn tradisjonell medisinsk behandling, ikke lar seg gjøre til gjenstand for vurdering etter like rigide kriterier. Dette argumentet benyttes gjerne mot psykoterapiforskning generelt, og manualbasert terapi spesielt. Synet på psykoterapi som en fleksibel og kompleks arbeidsform, er for mange uforenlig med ideen om manualer. Dermed befinner behandlingsmanualer seg i en kontroversiell posisjon, da de av mange er ansett for å være nødvendige for å imøtekomme krav om likhet i behandling (Perepletchikova et al., 2007; Perepletchikova og Kazdin, 2005), samtidig som de av mange anses for å være en tvangstrøye for psykoterapeuter ved at de er uforenlige med behov for fleksibilitet. Kritikken har særlig vært rettet mot: Manualers manglende appliserbarhet til det brede spekteret av pasientpopulasjoner og problemer (Abrahamson, 1999; Elliot 1998; Foxhall, 2000; Henry 1998; Norcross, 1999), et overdrevent fokus på foreskrevne behandlingsteknikker (Addis, Wade & Hatgis, 1999; Binder et al. 1993; Fonagy, 1999), restriksjoner de legger på terapeutens øvrige kliniske ekspertise (Addis, 1999; Elliot, 1998; Henry, 1998; Norcross, 1999) samt at manualer er ment å skulle kunne benyttes av av klinikere med svært ulik bakgrunn (Addis et al., 1999).

Mye av denne kritikken imøtegås i de samme artiklene som den fremmes. Det således ikke slik at debatten rundt behandlingsmanualer nødvendigvis er splittet i problemstillingen for eller mot behandlingsmanualen som fenomen. En konstruktiv debatt om behandlingsmanualer kan også anses

for å være nyttig i utviklingen av et teoretisk grunnlag for å bedre kvaliteten på manualene.

Debatten kan også bidra til å utvikle retningslinjer for hvor spesifikk en manual bør være, og hvor restriktivt/fleksibelt terapeutens rolle skal defineres.

Presisering av likhetsidealet. Det kan fremstå som et noe pussig paradoks at behandlingsmanualer, ved å ta høyde for nødvendighet av fleksibilitet (Kernberg et al. 2002; Bateman og Fonagy, 2004; Bateman og Fonagy 2006), også kan tenkes å undergrave egen etterlevelse av idealet om likhet i behandling. For eksempel er de tre eksisterende manualiserte dynamiske langtidspykoterapiene (definert av Caligor (2005) ved tilsiktet varighet på over ett år), Supportive-Expressive (SE) terapi (Luborsky, 1984) overføringsfokusert terapi (TFP) (Kernberg et al. 2002) og MBT (Bateman og Fonagy, 2004; Bateman og Fonagy, 2006), tydelige på nødvendigheten av fleksibilitet for psykoterapeuten i arbeidet med pasienten. Dette innebærer at de implisitt kommenterer på at de ikke har forestillinger om absolutt behandlingslikhet i ren terapiteknisk forstand. Behandlere oppfordres til å være fleksible i møtet med de tallrike kontekstuelle variablene som gjør seg gjeldende i terapirommet. Således omhandler spørsmålet om likhet i behandling, for psykoterapiens vedkommende, hvilket *nivå* det skal være likhet på.

For kognitive terapier (f. eks, Beck et al., 1979) vil mye av likheten kunne ligge på et nivå av spesifikk terapeutisk teknikk. Dette vil for eksempel kunne være terapeutens eksplisitte arbeid med hjemmelekser ved bruk av intervensjonsteknikker beskrevet i manualen. Men for flere intervensjoner, spesielt innen dynamisk terapi, vil likhet avgjøres av en teoretisk forståelse. I MBT vil dette kunne gjelde intervensjoner som er ment å stimulere mentalisering (Bateman og Fonagy 2006). Disse intervensjonene er ikke i samme grad forankret i eksplisitte prosedyrer som for eksempel hjemmelekser i kognitiv terapi. Beskrivelsene angitt i manualen er på et teoretisk nivå som tillater intervensjonene stor variasjon i form. Ved en vurdering av slike intervensjonsformer vil det derfor ligge en teoretisk forståelse til grunn for vurderingen av hvorvidt en intervensjon tjente manualens intenderte formål. Likheten ligger således ikke like klart i observerbare indikatorer, men krever at det observerte fortolkes ut fra kontekst, og vurderes opp mot forståelse av teori. Denne teorien (Fonagy et al. 2000; Bateman og Fonagy 2004; Bateman og Fonagy, 2006) omfatter: (1) en teoretisk forståelse av lidelsen (mentaliseringsvansker ved borderline pf i MBT), (2) en teoretisk modell for hva som utgjør endringsmekanismene, og (3) prinsipper for å arbeide med disse. Slik manualbeskrevet behandling, der likheten i stor grad knyttes til likhet i teoretisk forståelse, byr på spesielt store utfordringer ved mål av behandlingsintegritet (Perepletchikova et al. 2007).

Om det å bedømme psykoterapi - behandlingsintegritet

I Luborsky og Barbers (1993) definisjon av en behandlingsmanual heter det at den skal inneholde skalaer og instrumenter for å bedømme terapeuters samsvar med terapimodellen. Slik behandlingsmanualen kan anses for å være et middel for å sikre likhet i manipulasjon av den uavhengige variabelen (f.eks. MBT behandlingsintegritet) i et eksperiment, kan bedømmingsskalaen (f.eks. MBTEKS) anses for å være et verktøy for å kontrollere at den uavhengige variabelen er det den faktisk er ment å være. Behandlingsmanualen og bedømmingsskalaen er derfor to verktøy som skal sikre behandlingseksperimentets *indre validitet*. Ved psykoterapiforskning er den uavhengige variabelen redegjort for dersom man kan vise til tilfredsstillende behandlingsintegritet i terapeuten utførelse. Med behandlingsintegritet menes det i hvilken grad en behandling ble implementert som intendert (Perepletchikova og Kazdin, 2005). Dette innebærer tre aspekter i psykoterapi: (1) terapeuten *behandlingsetterlevelse*, i hvilken grad terapeuten benytter foreskrevne prosedyrer og ikke bruker prosedyrer som er frarådet; (2) terapeuten *kompetanse/kvalitet*, nivået av terapeuten ferdigheter og dømmekraft; (3) og *behandlingsdifferensiering*, hvorvidt behandlingsformer skiller seg fra hverandre på vesentlige punkter (Perepletchikova et al., 2007). Fidelity (renhet/trofasthet) er et overlappende begrep som benyttes i litteraturen for å angi i hvilken grad behandling er gjennomført etter planen (Forgatch, Patterson & DeGarmo, 2005). Begrepene brukes ofte om hverandre (Gresham, MacMillan, Beebe-Frankenberger & Bocian, 2000; Perepletchikova & Kazdin, 2005; Wilkinson, 2006). Oppgaven benytter begrepet behandlingsintegritet, slik det defineres av Perepletchikova et al. (2007), for å angi hvorvidt behandling er implementert som intendert.

Behandlingsintegritet kan vurderes på ulike måter. Det enkleste er å gjøre vurderinger ved å liste opp intervensjonene som konstituerer en gitt terapi, for så å krysse av for de intervensjonene som observeres i et observert psykoterapimateriale (Waltz, Koerner & Jacobson, 1993). Det vil da være mulig å plassere en gitt terapeuts arbeid i en terapitime innen en bestemt teoretisk tradisjon (f. eks. kognitiv terapi ved at det krysses av for intervensjoner rettet mot automatiske tanker). En mer avansert form er å sette mål på frekvens av relevante behandlingsteknikker (etterlevelse). Man har da i tillegg en liste over intervensjoner som inngår i terapien, men vurderer også i hvor stor utstrekning de ble benyttet. Dette gjøres som regel på en gitt skala som strekker seg fra lite til mye. Slik vurdering var den typiske form for samsvarsmål i de tidlige psykoterapimanualene, og blir fremdeles benyttet. Slike skalaer vurderer utelukkende frekvensen av relevante behandlingsteknikker, og er således egnet til å vurdere hvorvidt den observerte terapeuten praktiserte terapi med en bestemt type teoretisk forankring, samt at enkelte også har reliabelt kunnet

skille mellom ulike terapiformer (Luborsky og Barber, 1993). Svakheten ved denne type skalaer er at de ofte implisitt setter likhetstegn mellom etterlevelse av en behandlingsmanual og kvalitet i utførelse av terapi, ved at det antas at etterlevelsesmålet er tilstrekkelig for å si i hvilken grad behandlingsvariabelen ble manipulert på en vellykket måte. Waltz et al. (1993) kaller dette for et ubegrunnet konseptuelt sprang i psykoterapiforskningen, og hevder at synspunktet er en alvorlig feiloppfatning av begrepet om behandlingsintegritet i psykoterapiforskningen. Den teoretiske diskusjonen om nødvendige komponenter i en bedømmingsskala som er egnet til å gjøre valide mål på behandlingsintegritet, har dermed ført til at det nå er en forholdsvis bred enighet om at bedømmingsskalaer kun er mål på behandlingsintegritet i den grad de adresserer både etterlevelsen og kvaliteten av terapien (Perepletchikova og Kazdin, 2005; Perepletchikova et al., 2007; Barber og Crits-Cristoph, 1996; Barber et al., 2003).

En viktig styrke ved å ha kvalitet som en egen dimensjon, er at man åpner for muligheten til å trekke mer nyanserte slutninger fra de randomiserte kontrollerte forsøkene. Mer spesifikt åpner inndelingen for at man kan inndele behandlingsvariabelen i to for å studere hvorvidt det er teoritro etterlevelse, utførelse av høy kvalitet, eller en kombinasjon av disse variablene som forklarer eventuelle behandlingseffekter. Mulighetene for å gjøre en slik analyse påvirker også slutningsvaliditeten, da man mer nyansert kan beskrive hvordan eventuelle effektvariasjoner korrelerer med eventuelle spesifikke sammensetninger av behandlingskomponenter. Man kan ved en ideell analyse av en empirisk støttet behandling, eksempelvis vise til at det var kombinasjon av etterlevelse og kvalitet på (bestemte) kjerneområder av terapien som forklarte resultatene. En slik analyse har da et design med den styrken at det for eksempel i stor grad kan utelukke at effekten kun skyldtes terapeutvariabler knyttet til kvalitet på utførelse, men at denne for eksempel måtte kombineres med det spesifikke i manualen. Slik bidrar inndelingen til slutningsvaliditet ved at resultatene kan presenteres med større tyngde som en empirisk støtte til selve behandlingsformen. Barber og Crits-Christophs (1996) presentasjon av resultater fra en studie av SE Dynamic Psychotherapy er et eksempel på dette. Her viser de hvordan behandlingseffekten i deres studie var relatert spesifikt til kompetent (kvalitetsmessig høy) utførelse av tolkningsintervensjoner, og at dette funnet ikke ble endret ved å kontrollere for separate mål på allianse og generell terapeutisk kvalitet.

Bedømmingsskalaer og empiri. Den teoretiske diskusjonen som har ledet frem til synet på at en bedømmingsskala må skille etterlevelse og kvalitet, er også påvirket av empiri fra terapiforskningsfeltet. Ulike empiriske studier har undergravet tidligere teoretiske antagelser. En av disse antagelsene var oppfatningen om at kvalitetsmål var overflødig ved benyttelse av erfarne terapeuter. En slik antagelse inneholder premisset om at erfaring og utdanningsbakgrunn impliserer kvalitet. Dette mangler empirisk støtte, da ulike studier viser til manglende sammenheng mellom terapeuters erfaring og utfall av terapi (Berman & Norton, 1985; Durlak, 1979; Hattie, Sharpley, & Rogers, 1984). Dette er et av flere funn som støtter argumentet om nødvendigheten av at behandlingen er kontrollert for behandlingsintegritet. Slik kontroll forutsetter bruk av (tilgang til) reliable og valide mål på de egenskapene som definerer behandlingsintegritet. Videre viser flere studier til resultater der behandlingseffekt har blitt relatert til både etterlevelse og kvalitet. Resultatene fra disse studiene er foreløpig sprikende. Flere studier har vist til kvalitetsmål som positivt relatert til utfall av terapien (Barber & Crits-Cristoph, 1996; Luborsky, McLellan, Woody, O'Brien, & Auerbach, 1985; O'Malley et al., 1988;), og Havik (2001) har kommenterer på funn gjort i ulike studier ved å oppsummere at kvalitet, snarere enn blind etterlevelse, er det som synes å avgjøre om en intervensjon er vellykket. En studie av kognitiv terapi utdyper konklusjonene ytterligere ved å relatere symptomlette ved behandling av depresjon til kvalitet på terapeutens utførelse, og da spesielt til høye kvalitetsmål på terapeutens evne til å strukturere behandlingen (Shaw, et al. 1999). Andre studier ikke har ikke funnet disse sammenhengene mellom kvalitet og behandlingseffekt (Sandell, 1985). Empirien spriker ytterligere ved at man i én studie også fant at kvalitet predikerte *verre* utfall av terapien (Svartberg & Stiles, 1994).

Foreliggende evidens for effekten av etterlevelse er også varierende. Ulike studier finner sammenheng mellom etterlevelse og positivt utfall. Luborsky, McLellan, Woody, O'Brien og Auerbach (1985) fant at «renhet» i behandling, det vil si ratio av foreskrevne teknikker versus frarådede, predikerte positivt utfall. Det finnes dog en god del empiri som ikke gjør slike funn ved mål på etterlevelse, og en gjennomgang av empirien på området har blant annet ledet Wampold (2001) til å konkludere med at etterlevelse generelt ikke ser ut til å ha betydning for utfall av terapi.

Psykoterapiforskningen har enda til gode å finne klare, entydige forbindelser mellom mål på etterlevelse og kvalitet, og effekt av terapien. Man kan dog spørre seg om hva de ulike funnene av sammenheng mellom etterlevelse og behandlingseffekt betyr. Wampold (2001) etterlyser grundige metaanalyser for å få utredet spørsmålet nærmere. Det fremstår som noe uklart hvorfor han mener at dette spørsmålet er viktig å få avklart på et overordnet nivå. Når man ser på randomiserte kontrollerte forsøk som studier med eksperimentelt design, så er behandlingsintegritet den

uavhengige variabelen. Etterlevelse er et aspekt ved behandlingsintegritet, men både etterlevelse og behandlingsintegritet er begreper knyttet til en spesifikk form for psykoterapeutisk behandling. Dersom to ulike studier av to ulike terapiformer finner ulik sammenheng mellom etterlevelse og behandlingseffekt vil, vil dette ikke uten videre tilsi at etterlevelse ikke er av betydning. Man må alltid spørre hva det er etterlevelse av. Av den grunn vil metaanalyser, dersom de ikke tar for seg sammenligninger av samme terapiform, stå i fare for å trekke konklusjoner som ikke nødvendigvis er gyldige for bestemte terapiformer. Jeg vil derfor hevde at mål på etterlevelse er nødvendig for å kontrollere at en pasientpopulasjon faktisk mottok den behandlingen som var tilsiktet. Deretter vil resultatene av en slik behandlingseffektstudie kunne fortelle om terapien var virksom. Det vesentlige er at etterlevelse ikke er noe i seg selv - det må være etterlevelse av noe. Det er derfor problematisk å gjøre analyser av etterlevelse dersom man antar at etterlevelse er egen, og lik variabel, uavhengig av hvilken terapi som etterleves.

Et interessant poeng Wampold (2001) fremsetter er hvorvidt etterlevelse er et mål på noe ubetinget positivt. Dersom etterlevelse defineres utelukkende som et frekvensmål, vil ikke dette alltid være slik. Det er mulig å tenke seg høyfrekvent og upassende bruk av intervensjonsteknikker. I slike tilfeller skal et kvalitetsmål fange opp dette, og upassende, høyfrekvent intervensjonsbruk vil gi høy skåre på etterlevelse og lav skåre på kvalitet. Slike eksempler viser til hvordan etterlevelse og kvalitet i grunn er naive forsøk på å fange opp, og tallmessig uttrykke, hvorvidt behandlingen blir implementert som intendert (behandlingsintegritet).

Trening av bedømmere. Ulike studier viser til erfaringer med at bedømmingsskalaer kan være svært krevende å benytte. Selv om det er vist til akseptabel interrater-reliabilitet i ulike studier (f. eks. Barber & Crits-Cristoph 1996; Barber, et al., 2003; Carroll et al., 2000), så er det også slik at rimelig grad av reliabilitet mellom ulike bedømmere er oppnådd først etter omfattende trening. Nødvendigheten av opplæring og trening ved bruk av bedømmingsskalaer understrekes fra flere hold (f.eks. Carroll et al. 2000; Barber et al., 2003; Nordahl, Nysæter og Mikkelsen, 2006). Både Nordahl et al. (2006) og Carroll et al. (2000) anbefaler skåring av ti hele terapitimer med påfølgende diskusjon for å oppnå enighet ved bruk av henholdsvis CTACS og YACS. Det kommer dog ikke frem hva antallet timer baserer seg på.

Nødvendigheten av trening på selve *bruken* av bedømmingsskalaen understrekes av Jacobson (1998, ref. i Barber et al., 2003) som viser i en studie at erfarne eksperter i kognitiv terapi var langt fra å oppnå reliabel skåre ved benyttelse av Cognitive Therapy Scale (CTS) uten trening. Funnet tyder på at erfaring og teoretisk kunnskap ikke er godt nok til å bedømme terapier ved bruk av

bedømmingsskalaer, men at det er erfaring med selve bruken av bedømningsverktøyet som er avgjørende. Dette er ikke helt uproblematisk når det gjelder manualbasert terapi. Mange vil trekke validiteten til slike bedømmingsskalaer i tvil dersom ekspertise ikke har signifikant innvirkning på skåringer man får ved bruk av skjemaet. Barber et al. (2003) påpeker nødvendigheten av trening i drøftingen av egne funn i reliabilitetsstudien på CTACS. De hevder at dette ikke nødvendigvis bør ha fatale implikasjoner for bedømmingsskalaer som fenomen. Det kan dog argumenteres for at studier bør kunne vise til at eksperters skåringer er viser relativt høy enighet målt ved reliabilitet uten trening, selv om treningen er nødvendig for at bedømmere sammen utvikler presisjonsnivået og en felles forståelse. Funnet av at teoretisk kunnskap, uavhengig av trening, påvirker bruken av en skala i positiv retning, vil være et tegn på at skalaen er et valid mål på dens teoretiske forankring.

Særlig viktig ved utviklingen av et treningsprogram er en gjennomgang av det litteraturen forteller om problemstillinger knyttet til skillet mellom etterlevelse og kvalitet, og problemstillinger vedrørende terapiformer og deres natur; på hvilket nivå de beskriver sin metode (jfr. nivå av likhet i gjennomgangen av manualbasert terapi).

Skillet mellom etterlevelse og kvalitet. Skillet mellom kvalitet og etterlevelse er teoretisk utfyllende begrunnet (Waltz et al., 1993; Perepletchkova og Kazdin, 2005; Perepletchkova, Treat og Kazdin, 2007). Ved den praktiske anvendelsen av skalaer viser dog resultater fra flere studier til at bedømmere kan streve med å skille mellom de to begrepene. Her er det forskjeller i funn gjort ved benyttelse av ulike skalaer. Forskning på sammenhengen viser i enkelte studier til ikke-signifikante forhold (f.eks. Paivio, Holowaty og Hall, 2004), mens andre studier finner høye korrelasjoner mellom kvalitet og etterlevelse (f.eks. Barber et al., 2006; Shaw et al., 1999). McGlinchey og Dobson (2003) påpeker at det er et definisjonsmessig betinget forhold mellom de to begrepene: kvalitet fordrer etterlevelse, men etterlevelse fordrer ikke kvalitet. Denne forståelsen har ledet blant andre Barber et al. (2006) til å tolke egne funn av høye korrelasjoner mellom kvalitet og etterlevelse som et resultat av en feiloppfatning hos bedømmerne, der de har en tendens til å oppfatte en intervensjon til å være av høyere kvalitet jo hyppigere den ble benyttet.

En annen forklaring kan ligge i skalaenes faktiske ledd, og måten disse er formulert på. Som eksempel viser Waltz et al. (1993) til skalaen «the Collaborative Study Psychotherapy Rating Scale» (CSPRS). Denne er ment å måle etterlevelse, men gjennom måten leddene er operajonalisert på, måler den i enkelte tilfeller også kvalitet. En slik grundig tilnærming til leddenes faktiske innhold er avgjørende, sammen med utregninger av korrelasjonskoeffisienter mellom målene. Dersom forskjeller i mål på etterlevelse og kvalitet ikke er tydelig definert gjennom

operasjonaliseringen av skalaens ulike nivåer, vil problemet urettmessig kunne tilskrives manglende kompetanse hos bedømmere. Det blir derfor viktig under enhver utvikling av bedømmingsskalaer at det både må rettes fokus mot opplæring av bedømmere, samt gjøres kontinuerlige evalueringer av hvordan innholdet i bedømmingsskalaen er formulert i forhold til etterlevelse og kvalitet.

Psykoterapiformens natur. Perepletchikova et al. (2007) gjør ved en gjennomgang av ulike studiers mål på behandlingsintegritet en kategorisk inndeling av terapiformer basert på nivå av kompleksitet. Deres hypotese er at behandlingsintegritet i større grad gjøres rede for i behandlingsformer som er mindre komplekse i manualbeskrivelsene. De foretar en viktig hovedinndeling i følgende to kategorier: ferdighetstrenende tilnærminger (f.eks. kognitiv adferdsterapi) og ikke-ferdighetstrenende tilnærminger (f.eks. psykodynamisk terapi og andre prosessorienterte terapier). Denne inndelingen er viktig da MBTEKS tar sikte på å vurdere behandlingsintegritet av en ikke-ferdighetstrenende terapi. Dette byr i følge Perepletchikova et al. (2007) på spesielle utfordringer.

Ferdighetstrenende psykoterapiformer er i utgangspunktet enklere å vurdere. Dette skyldes terapiformenes eksplisitte og spesifikke fokus på veldefinerte strategier og teknikker som skal benyttes i terapiforløpet. En behandlingsmanual for ferdighetstrenende terapi vil således oftere presentere spesifikke problemer som skal møtes med spesifikke typer intervensjoner. Ikke-ferdighetstrenende terapier er derimot ofte mer generelle i beskrivelser av teknikk og strategi, der det overordnede fokus gjerne rettes mot underliggende prosesser. Dette fokuset på underliggende prosesser byr på åpenbare problemer når bedømmere blir utfordret til å bedømme terapeuters behandlingsintegritet. Fenomenene som vurderes er i mindre grad knyttet til eksplisitte, observerbare indikatorer, og bedømmeren blir derved i større grad utfordret til å benytte teoretisk kunnskap for å gi mening til det observerte; å fortolke det observerte. En mulig løsning på dette problemet vil kunne være å utarbeide mer eksplisitte og spesifikke beskrivelser i behandlingsmanualer for terapiformer som i hovedsak er prosessfokuserte (f.eks. MBT). Forskning på manualbasert psykodynamisk terapi viser dog til at et overdrevent fokus på behandlingsteknikker i psykodynamisk terapi, kan gå på bekostning av mål på terapeuters interpersonlige ferdigheter (Binder, 1993). Dette medfører at forskjellen i de ulike psykoterapiretningenes natur også må ivaretas når man utarbeider behandlingsmanualer og bedømmingsskalaer. Eventuelle problemer knyttet til reliabilitet ved skåringer av ikke-ferdighetstrenende terapi, må derfor antagelig løses på måter som ivaretar terapiformens kompleksitet og egenart. Oppsummert kan man si at en bedømmingsskala må tilpasses terapiformen, og ikke omvendt.

Mål på behandlingsintegritet i forskningspraksis. Til tross for mange de mange postulerte fordeler ved bruk av bedømmingsskalaer for å sikre behandlingsintegritet i randomiserte kontrollerte forsøk, er det en tendens ved flere studier at de i langt større grad redegjør for vurdering av den avhengige variabelen; behandlingseffekten. Denne tendensen beskrives av Peterson, Homer og Wonderlich (1982) som en «besynderlig dobbeltmoral» (s. 478), og synes fremdeles å være en aktuell problemstilling (Perepletchikova og Kazdin, 2007). I en analyse gjort av Perepletchikova og Kazdin (2007), ble et utvalg av randomiserte kontrollerte forsøk i ulike kjente tidsskrifter undersøkt. Av disse fant man at kun 3,5% av studiene tilfredsstilte deres definisjon for god standard ved redegjørelse av prosedyrer ment å sikre behandlingsintegritet. Slike funn aktualiserer spørsmålet om hvorfor grundig kontroll på behandlingsintegritet synes å være mangelvare.

En mulig årsak kan være kostnadene forbundet med å følge Perepletchikova og Kazdins (2007) prosedyrer. Ulike terapiformer har heller ikke utviklet bedømmingsskalaer eller utarbeidet gode treningsprogrammer for både terapeuter og bedømmere. Slikt arbeid er ressurskrevende, og kan tenkes å være en av årsakene til at man i noen studier velger å la disse variablene hvile på antagelser om at de er ivaretatt. Det som er særlig relevant for denne oppgaven, er den praktiske anvendelsen av bedømmingsskalaer. Idéen om å kunne kvalitetssikre terapi ved benyttelse av et skjema, er ikke uproblematisk når man beveger seg fra teori til praktiske anvendelsen. Utvikling av bedømmingsskalaer som er mer praktisk anvendelige, vil muligens kunne bidra til mer utbredt bruk av integritetsmål i forskning og klinikk.

Målsetninger og problemstillinger

Det overordnede hovedmål er å bidra til å utvikle en MBT bedømmingsskala som lar seg anvende for å gjøre *reliable* og *valide* målinger av *behandlingsintegritet*, slik det defineres av Perepletchikova et al. (2007).

Litteraturgjennomgangen viser at det er spesifikke problemstillinger knyttet til utvikling av bedømmingsskalaer generelt, og bedømmingsskalaer for psykodynamiske, ikke-ferdighetstrenende psykoterapitradisjoner spesielt. Når disse problemstillingene i litteraturen ofte nevnes på et overfladisk nivå, med henvisninger til statistiske resultater, vil denne studien av den innledende utviklingsprosessen gå nærmere inn på utformingen av de spesifikke ledd, og erfaringer fra arbeidet med dette. Hensikten med dette er å utdype, og grundigere tydeliggjøre problemstillingene på et nivå som er nærmere knyttet til den faktiske utformingen av innholdet i en bedømmingsskala. Med

dette arbeidet vil jeg søke å utvikle nyttige begreper som kan være klargjørende ved operasjonalisering av ledd, samt kategoribegreper for å skille mellom ledd av ulike typer.

Studien inndeles i to separate understudier. Utviklingsstudien viser til studien som helhet. Den inndeles i «utformingsstudien», som viser til prosessen med å ferdigstille selve produktet MBTEKS, og «enighetsstudien», som viser til prosessen der bedømmerne benytter den ferdig utformede skalaen til å bedømme terapitimer for å oppnå enighet.

Metoder

Om måleinstrumentet - MBTEKS

Skalaen er basert på Bateman og Fonagys behandlingsmanual «Mentalization-based treatment for the borderline patient - a practical guide» (Bateman og Fonagy, 2006). Den tok utgangspunkt i et arbeide av Bateman der det var utledet 15 punkter som skulle skåres i for etterlevelse (se Appendiks B). MBTEKS (se Appendiks A) består i sin nåværende form av 23 ledd (se tabell 1) som alle skal skåres for både kvalitet og etterlevelse. Videre er formatet endret med inspirasjon fra den norske utgaven av Cognitive Therapy Adherence and Competence Scale (CTACS) (Norrdahl et al. 2006). Skalaen inneholder 7-punkts Likert-skalaer for alle ledd, både under etterlevelse og kvalitet, med 4 referansepunkter for hver. Likert-skalaen har empirisk støtte i sammenligning med andre varianter, der syv synes å være det optimale nivåantallet (Finn, 1972). Referansepunktene under etterlevelse er like for alle ledd, der skårene 0, 2, 4 og 6 er blitt kalt henholdsvis «ingen», «noe», «betraktelig» og «fullstendig» etterlevelse. Kvalitetsreferansene er mer utdypende formulert (se Appendix A). Her er det beskrevet hva som tilsvarende kvalitet for skårene 0, 2, 4 og 6 med utfyllende tekst basert på innholdet i behandlingsmanualen (Bateman og Fonagy, 2006). Dette er gjort for alle MBT-kriterier med unntak av ledd 23 (totalvurderingen), som har kvalitetsreferansene «svært dårlig», «dårlig», «god» og «svært god» for henholdsvis skårene 0, 2, 4 og 6. De fire første leddene i skalaen (generelle psykoterapikriterier) er hentet direkte fra Norrdahl et al.s (2006) norske utgave av CTACS.

Tabell 1

Generelle psykoterapikriterier*

1. Varm/genuin/kongruent

2. Oppmerksomhet

3. Empati

4. Samarbeid

MBT-kriterier

5. "Ikke-vitende" holdning

6. Fremme utforskning og nysgjerrighet

7. Utfordrer uberettigede oppfatninger

8. Positiv mentalisering

9. Tilpasning til mentaliseringsnivå

10. Pretend modus

11. Psykisk ekvivalens

12. Optimalt spenningsnivå

13. Korte tilpassede intervensjoner

14. Bruk av metaforer, analogier og symboler

15. Affektfokus

16. Affekt og interpersonlige hendelser

17. Stimulere mentalisering

18. Innhold og prosess

19. Overføring

20. Motoverføring

21. Alliansebrudd

22. Sjekke egen forståelse og korrigere misforståelse

23. Helhetlig vurdering av MBT etterlevelse og kvalitet

*De generelle kriteriene er hentet fra Norrdahl et al. (2006)

Medarbeidere

Jeg har ledet arbeidet med å utvikle skalaen for vurdering av MBT etterlevelse og kvalitet. Dette ble gjort i samarbeid med fire psykiatere og en psykologspesialist ved Avdeling for Personlighetspsykiatri, Ullevål Universitetssykehus (UUS). Tre av psykiaterne har doktorgradsutdanning, og en av dem er også professor i psykiatri. Alle seks som var med i arbeidet med utviklingen av skalaen er fortrolige med innholdet i Bateman og Fonagys (2006) manual. De har deltatt på seminarer med Anthony Bateman og gått i MBT-veiledningsgrupper ved Avdeling for Personlighetspsykiatri, UUS. Tre av medarbeiderne har også erfaring med å lede veiledningsgrupper i MBT. Det er den samme gruppen som har deltatt i utformingsarbeidet som inngikk i skåringsarbeidet ved enighetsstudien. Sentral i utviklingen er også Anthony Bateman, som har fulgt prosessen og bidratt med veiledning og kommentarer underveis.

Terapeuter

I prosessen før vi påbegynte selve enighetsstudien, skåret vi og vurderte leddene basert på terapitimer der medlemmer av utviklingsgruppen selv var terapeuter. Ingen av disse skåringene er benyttet som datamateriale i enighetsstudien.

I enighetsstudien skåret vi seks terapitimer fordelt på fire terapeuter (to av terapeutene bidro med to timer hver). Terapeutene var en kvinnelig ergoterapeut med gruppeanalytisk videreutdanning, en psykiatrisk sykepleier med gruppeanalytisk videreutdanning, en assistentlege i psykiatrisk spesialisering og en psykolog. Alle terapeutene arbeider ved avdelinger og med prosjekter, der MBT praktiseres. Alle terapeutene deltar i veiledningsgrupper med fokus på MBT.

Pasienter

Det var 4 ulike pasienter som inngikk i enighetsstudien (de nevnte 2 terapeutene som bidro med 2 terapitimer hadde samme pasient i hver av timene). Tre av pasientene går i behandling ved Avdeling for Personlighetspsykiatri, UUS, mens én av pasientene deltar i et behandlingsprogram for spiseforstyrrelser ved Regional Avdeling for Spiseproblemer, UUS (RASP).

Utformingsstudien

Utviklingsarbeidet startet ved at vi først besluttet å endre Batemans opprinnelige 15-punkts skala (se Appendiks B) i retning av formatet til CTACS (Norrdahl et al. 2006), og arbeidet med å formulere referansepunkter ble påbegynt rett etter dette. Dette ble gjort ved at de opprinnelige 15

leddene i Batemans skala ble fordelt til alle medarbeiderne, som fikk i oppgave å formulere fire kvalitetsreferanser for hvert av leddene. Disse ble så levert til meg som innarbeidet dem i en samlet skala. Skalaen ble så testet over fire 2-timers møter der alle seks medarbeidere var samlet. Disse møtene var inndelt i fire deler. De første 15 minuttene gikk med til å gjennomgå og oppsummere nye endringer i skalaen, de påfølgende 45 minuttene gikk til å se en terapitime, så skåret alle det de hadde sett i terapitimen på 15 minutter, før vi avsluttet med en 45-minutters diskusjon rundt erfaringer med bruken av skalaen. I løpet av de fire møtene økte Batemans opprinnelige skala fra 15 til 19 ledd, pluss at det ble lagt til 4 ledd fra CTACS (Norrdahl et al. 2006). Resultater fra utformingsstudien ble etter hvert møte oversatt til engelsk og oversendt Anthony Bateman for kommentarer. Den endelige skalaen som inngikk i enighetsstudien ble godkjent av Bateman for å sikre skalaens teoretiske validitet. Eksempler på revisjoner og diskusjonsmomenter fra utformingsprosessen blir gjennomgått i resultatdelen.

Enighetsstudien

Enighetsstudien ble påbegynt straks etter arbeidet med utformingen av skalaen var fullført. Siktemålet med enighetsstudien er å undersøke hvorvidt man innledningsvis kan trekke noen foreløpige konklusjoner i forhold til skalaens anvendbarhet. Videre vil erfaringene fra enighetsstudien, samt de kvantitative resultatene, gi et grunnlag for videre revisjoner av skalaen. Jeg ønsker også å kunne benytte erfaringer og resultater til å på sikt utforme et treningsprogram for å bedre skalaens validitet.

Skåringer ble gjort ved møter der alle medarbeiderne deltok for å teste skalaen. Disse møtene har samme form som møtene fra utformingsfasen der de er inndelt i fire deler. Første del på 15 minutter går med til å oppsummere erfaringer fra tidligere møter, de påfølgende 45 minuttene går med til å se en hel terapitime, denne blir så skåret uten diskusjon i 15 minutter, før vi diskuterer de ulike skårene fra den aktuelle terapitimen i plenum.

Plenumsdiskusjonen. starter når alle bedømmere har skåret samtlige 23 ledd for både kvalitet og etterlevelse. Diskusjonen er lagt opp slik at alle kommer med innspill på lik linje. Først tar vi for oss skåringer på totalvurderingen (ledd 23) og begrunner disse. Ved totalvurderingen gir hver bedømmer en begrunnelse for hva det var som var avgjørende for det totale inntrykket. Dette leder gjerne diskusjonen videre til et annet ledd som var styrende for totalinntrykket. Store avvik i skåringer prioriteres også i diskusjonene. Dette må anses for å være et avgjørende arbeid i prosessen

med å oppnå større enighet rundt de enkelte leddenes betydning, og hvordan det observerte i terapitimen skal forstås i henhold til MBT.

I tillegg til store avvik, prioriteres også de tilfeller der det er uenighet om et ledd er anvendbart eller ikke. Denne problemstillingen har jeg valgt å kalle «terskelspørsmålet». Skalaen skiller mellom verdien «0» i etterlevelse (ingen etterlevelse på en intervensjonstype som var aktuell i terapitimen) og «ikke anvendbart» (IA) (ingen etterlevelse av en intervensjonstype som ikke var aktuell). Diskusjoner rundt terskelspørsmål dreier seg derfor om å komme til enighet om det som observeres i terapitimen overstiger en tenkt *terskelverdi* som gjør det aktuelt å sette en skåre på leddet. Diskusjoner om hvorvidt et ledd er anvendbart oppstår typisk på to måter: (1) terapeuten foretok en intervensjon som kunne minne om en beskrivelse fra et ledd i skalaen, eller (2) det forekom fenomener i terapitimen (f.eks. overføringsfenomener) som var tilstede i mindre grad, slik at bedømmerne falt på ulike vurderinger av hvorvidt dette skulle skåres som noe terapeuten overså (lav skåre), eller om det var «ikke anvendbart» (under terskelen) på skalaen.

Design på enighetsstudien. Enighetsstudien baserer seg på et datamateriale fra seks skårede individualterapitimer på 45 minutter. Alle terapitimene er skåret av de samme seks bedømmerne i et interbedømmerdesign (interrater design). Alle skåringer gjøres av deltagerne hver for seg, med en påfølgende plenumsdiskusjon. Observasjonsenheten i studien er hele terapitimer. Disse er vurdert i henhold til et tidssamlingsdesign. Vårt tidssamlingsdesign innebærer at hele terapitimen blir vist før skåringen gjøres. Designet er også benyttet i lignende studier (Barber et al. 2003), mens Luborsky et al. (1985) benyttet et tidssamlingsdesign der observasjonsenheten var 15-minutters segmenter av terapitimer. Alternativet til et slikt design er hendelsessamlingsdesign (event sampling) der bedømmere bedømmer intervensjoner direkte etter observasjoner. Vi valgte et tidssamlingsdesign for hele terapitimer av hensyn til at vi ønsker på sikt å validere bedømmingsskalaen for det totale terapeutisk arbeidet som er gjort over en hel terapitime. Bedømmere begynner derfor ikke skåringen før en hel terapitime er observert.

Man kan tenke seg fem *hovedkilder* til variasjon i skårer ved bruk av bedømmingsskalaen: (1) leddet som skåres, (2) bedømmeren, (3) terapitimene, (4) terapeut, og (5) pasient. I de statistiske resultatene fra enighetsstudien, vil jeg se på skåringer ved MBTEKS i forhold til tre kilder til variasjon: ledd, bedømmer og terapitime. I det eksisterende datamaterialet fra enighetsstudien er det fire ulike terapeuter og fire ulike pasienter. Av hensyn til det lave antallet vil ikke disse inngå i designet. Terapeut og pasient vil derfor anses som de antatt viktigste skjulte kildene til variasjon i resultatene.

Plenumsskåring der bedømmerne hver for seg skårer terapimateriale med påfølgende diskusjon, er sterkt anbefalt for å øke interbedømmer reliabilitet (f.eks Nordahl et al., 2006; Barber et al., 2003). Ved slik trening er det konsensus i skåring opp mot det observerte man vil oppnå. Arbeidet kan være krevende, og det later til å bli mer krevende dersom terapiformen som bedømmes er av ikke-ferdighetstrenende art, som for eksempel MBT (Perepletchikova et al., 2007). Dette antas å ha sammenheng med hvor konkrete variablene som skal skåres er. I MBTEKS forsøker vi eksempelvis å skåre arbeid med motoverføringsfenomener. Plenumsdiskusjonen i enighetsstudien vil antas å være en viktig faktor som påvirker eventuell økt enighet studien kan vise til.

Resultater

Utformingsstudien

I prosessen som ledet frem til den vedlagte skalaen MBTEKS, dreiet mye av arbeidet seg om å bli enige om formuleringer av referansepunkter. Manualen skalaen er basert på (Bateman og Fonagy, 2006) inneholder ikke beskrivelser av kvalitetsnivåer av intervensjonene. Som de fleste andre psykoterapimanualer forholder den seg for det meste til beskrivelser av foreskrevne intervensjoner, eksempler på hvordan disse kan gjøres, det bakenforliggende rasjonalet, samt noen beskrivelser av teknikker som *ikke* bør anvendes i MBT. Vi hadde vi stort sett beskrivelser av de rette måtene å intervenere på i ulike kontekster, ofte kontrastert med beskrivelser av måter å intervenere på som i henhold til MBT er regnet for å være av lav kvalitet. Utfordringen for oss var derfor å operasjonalisere mellomnivåer på kvalitet på intervensjoner. Beskrivelser av intervensjoner som er frarådet i MBT er i manualen enten gitt implisitt, ved at det fremgår i beskrivelsen av prinsippene for en anbefalt intervensjon hvordan disse kan gjøres på feilaktig måte. Dessuten beskriver manualen noen steder også typiske intervensjoner som anses for å være direkte i strid med målet om å øke pasientens evne til å mentalisere. Batemans 15-punktsskala var til hjelp ved at den ga oss et innledende fokus, og fungerte som et utgangspunkt for vår videreutvikling.

Tidlig i utformingsprosessen, der vi i møter med plenumsdiskusjon vurderte hverandres referanseformuleringer, ble vi klar over nødvendigheten av å lage en regel for hvordan alminnelig god MBT-kvalitet skulle formuleres. Vi ble enige om at denne formuleringen i størst mulig grad skulle samsvare med en nøktern beskrivelse av den foreskrevne teknikken eller holdningen, og gjerne ligne på Batemans opprinnelige formulering. Denne regelen gjorde det lettere å justere formuleringer slik at de i størst mulig grad hang sammen på tvers av ledd. En skåre 6 ble da en mer utfyllende beskrivelse av fremragende prestasjon, gjerne knyttet til terapeutens kreativitet i arbeidet med intervensjonsformen, mens en skåre 2 gjerne beskriver terapeutaktivitet som minner om MBT, men som ikke når opp til et akseptabelt nivå; gjerne beskrevet som et overfladisk forsøk, eller på andre måter som en ufullstendig utført intervensjon.

For å illustrere problemstillinger knyttet til formulering av kvalitetsreferansepunkter, vises det til et konkret ledd i skalaen som anses som svært sentralt; leddet som har fått navnet «ikke-vitende holdning». Dette leddet beskriver et svært sentralt element ved MBT-behandling, og ved plenumsdiskusjoner ble det i økende grad tydelig hvordan dette leddet var med å påvirke terapeutens skåre på ulike andre ledd. (tabell 2)

Tabell 2

5. "Ikke-vitende" holdning

Etterlevelse	/ Kvalitet			
Etterlevelse:	0 Ingen	2 Noe	4 Betraktelig	6 Fullstendig
<i>Kvalitet:</i>				
0	Terapeuten var passiv og gjorde ingen forsøk på å finne ut om pasientens mentale tilstand, hvordan han/hun tolker andre personer og interaksjonen mellom dem. Intervensjonene var farget av at terapeuten ignorerte det at mentale tilstander er 'ugjennomsiktige' og vanskelig å si noe sikkert om.			
2	Terapeuten gjorde overfladiske forsøk på å finne ut om pasientens mentale tilstander, men terapeutens primære interesse virket å være hans egne underliggende motiver.			
4	Terapeuten tok en oppriktig "ikke-vitende" holdning og forsøkte å "finne ut". Unngikk påstander av typen "hva du egentlig mener er..."			
6	Terapeuten tok en oppriktig "ikke-vitende" holdning og forsøkte å "finne ut". Unngikk påstander av typen "hva du egentlig mener er..." Terapeuten utviste kreative måter å engasjere pasienten i en felles anstrengelse for å utforske både pasientens og andres sinn, og interaksjonen mellom dem.			

Leddets er et godt eksempel på hvordan punkt 4 er formulert i tråd med regelen om en nøktern beskrivelse av foreskreven holdning/teknikk, hvordan punkt 6 går videre med å beskrive terapeutens kreativitet og grundighet, og hvordan punkt 2 beskriver holdning/teknikk som er gjenkjennelig, men for overfladisk og lite grundig til å kunne oppnå en vurdering som «god MBT».

Under referansepunkt 2 i kvalitetsvurderingen kan det også nevnes at dette ble endret i utviklingsprosessen. Punkt 2 var først formulert mer utdypende:

Tabell 3

2	Terapeuten gjorde overfladiske forsøk på å finne ut om pasientens mentale tilstander, men terapeutens primære interesse virket å være hans egne underliggende motiver for å styre pasienten til en konklusjon om hva som foregår.
---	--

Ved skåringer i den innledende fasen kom det frem i plenumsdiskusjoner at bedømmerne ikke evnet å forankre vurderingene så spesifikt til at terapeuten forsøkte å «styre pasienten til en konklusjon om hva som foregår». Leddets spesifisitet på dette punktet begrenset også bedømmerne ved at det avgrenset en 2-skåre i så stor grad at det sjeldent ble gjort observasjoner og vurderinger som passet godt med punktet. Det ble derfor besluttet å gjøre en revisjon av punktet som var mer åpen for hvilke motiver terapeuten kunne tenkes å ha, slik at vurderinger av leddet i større grad kan forankres i det som observeres, snarere enn baseres på antagelser om hva terapeutens faktiske motiver er.

Denne revisjonen illustrerer et overordnet tema i utviklingen av skalaens ledd. Vi blir i utviklingsprosessen hele tiden tvunget til å avgjøre på hvilket nivå en operasjonalisering skal befinne seg, og denne vurderingen avhenger hele tiden av variabelenes natur. Noen variabler måler

konstrukter som lettere lar seg operasjonalisere med tilknytning til observerbare indikatorer, mens andre befinner seg på et nivå der bedømmernes tolkning av det observerte får større betydning. I eksempelet fra figur 2 ser man hvordan formuleringen, ikke bare med hensyn til spesifisitet i formulering, men også med hensyn til abstrakthet, fører til vanskeligheter for bedømmeren. Der indikatorene på intervensjonene blir vage, eller mindre direkte, blir plenumsdiskusjonene også vage i utgangspunktet. Figur 2 byr på det åpenbare problemet å få en felles forståelse av hvordan man kan observere en terapeuts egentlige motiv. Mer spesifikt går dette på en slags observert terapeutadferd, som man kan enes om er en indikator på en ikke-observerbar variabel (hva terapeutens egentlige motiv er). Her er det teoretiske fundamentet til behandlingsteorien vanskelig å overføre til den praktiske anvendelsen av skalaen.

Det ovennevnte problem er ikke uventet i utvikling av integritetsmål på MBT. Fordi behandlingsformen er av psykodynamisk art, plasserer den seg i Perepletchikova et al.s (2007) kategori «ikke-ferdighetstrenende» psykoterapi. Erfaringer oppsummert av Perepletchikova et al. (2007) viser at mål på behandlingsintegritet er noe vanskeligere å operasjonalisere opp mot observerbare indikatorer for psykoterapier i kategorien «ikke-ferdighetstrenende». På dette punkt må man derfor hele tiden veie teori mot praksis. Det vil si at vi noen ganger må fatte beslutninger som gjør at skalaen blir mindre ambisiøs i forhold til å fange opp de teoretiske ingrediensene i behandlingen fordi de simpelthen ikke lar seg reliabelt bedømme. Vår erfaring med utviklingen av MBTEKS tilsier at dette kan gjøres på to måter: (1) man kan begrense leddets beskrivelse, slik som i figur 2, eller man kan (2) operasjonalisere terskelbeskrivelser.

En terskel for skåring. Terskelbeskrivelser er noe vi gjennomgående drøftet i forhold til de mer abstrakte leddene i skalaen. Ledd nummer 19 (figur 3) har vært gjenstand for slik en diskusjon.

Tabell 4

19. Overføring

Etterlevelse	/ Kvalitet			
Etterlevelse:	0 Ingen	2 Noe	4 Betraktelig	6 Fullstendig

Kvalitet:

IA Ikke aktuelt.

0 Terapeuten tolket på en rigid måte overføring som enkle gjentakelser av relasjoner fra fortiden.

2 Terapeuten tolket overføringsreaksjoner uten å stille spørsmål om alternativ forståelse av pasientens reaksjoner i forhold til terapeuten.

4 Terapeuten påpekte overføringsreaksjoner, og oppfordret til refleksjon over alternative perspektiver på pasientens reaksjoner og oppfatninger av terapeuten.

6 Terapeuten påpekte overføringsreaksjoner, og oppfordret til refleksjon over alternative perspektiver på pasientens reaksjoner og oppfatninger av terapeuten. Terapeuten bidro videre til å skape refleksjon over terapeutenes forståelse av overføringsreaksjonene.

Figur 3 viser leddet i sin nåværende form. I diskusjoner rundt leddet, ble det også drøftet hvorvidt man skal vurdere terapeuten etter hvordan han *ikke* arbeider med overføringen. Dette ble forsøkt gjort i ledd 20 (figur 4).

Tabell 5

20. Motoverføring

Etterlevelse	/ Kvalitet				
Etterlevelse:	Ingen	0	2	4	6
		Noe	Betraktelig	Fullstendig	

Kvalitet:

IA Det forekom ingen klare motoverføringsresponser i den aktuelle timen.

0 Terapeuten registrerte ikke egne motoverføringsresponser overfor pasienten.

2 Terapeuten benyttet seg av sine motoverføringsresponser, men plasserte ansvaret for disse hos pasienten i et forsøk på å øke innsikt i pasients følelser.

4 Terapeuten utforsket åpent og adekvat visse motoverføringsresponser overfor pasienten og identifiserer disse i utgangspunktet som egne reaksjoner. Drøfter pasientens rolle i følelsene som ble vekket kun når dette er anerkjent.

6 Terapeuten utforsket åpent og adekvat visse motoverføringsresponser overfor pasienten og identifiserte disse i utgangspunktet som egne reaksjoner. Drøftet pasientens rolle i følelsene som ble vekket kun når dette var anerkjent. Han/hun gjorde dette på en kreativ måte og delte ansvaret for utviklingen i interaksjonen.

Forskjellen i disse leddene er *terskelen* for å kvalitetsvurdere leddet. Mens ledd nummer 19 nå kun skal vurderes de gangene terapeuten eksplisitt kommenterer på overføringsfenomener, krever ledd 20 at bedømmeren også vurderer hvorvidt terapeuten *burde* ha kommentert på fenomenet. Å vurdere hvorvidt man burde ha gjort en intervensjon, er i mindre grad knyttet til tydelige indikatorer

enn kun å vurdere de mer eksplisitte kommentarene på overføringen. Ledd 20 har dog den styrke at det i bredere forstand dekker MBT-intensjonen. Det er således liten tvil om at 0-skåren på ledd 20 faktisk er i tråd med MBT-teori. Spørsmålet må derfor besvares ut ifra de erfaringer som gjøres ved benyttelse av skalaen. Dersom bedømmere ikke klarer å komme til enighet i bruken av 0-skåren, kan løsningen bli en revisjon som endrer leddet mer i retning av ledd 19 (Figur 3).

Terskelspørsmålet har også blitt reist i diskusjoner rundt ledd 10 (pretend modus) og 11 (psykisk ekvivalens) (se Appendiks A). Disse leddene beskriver også fenomener og intervensjonsformer som ikke er ferdighetstrenende (Perepletchikova et al. 2007). Begge disse leddene er nå foreløpig definert slik at det er opp til bedømmeren å vurdere hvorvidt fenomenene overstiger en tenkt terskel som gjør leddet anvendbart. Dette betyr at leddene i sin nåværende form utfordrer den enkelte bedømmer til å avgjøre om fenomenene pretend modus og psykisk ekvivalens forekommer i så stor grad at terapeutens manglende intervensjoner fører til skåren «0» og ikke «IA».

Det jeg har valgt å kalle «terskelspørsmålet» er egentlig et resultat av formatet på skalaen, inspirert av CTACS (Barber et al. 2003; Nordahl et al. 2006). Det skyldes at skalaen åpner for at man enten kan skåre «ingen etterlevelse» (0) eller «ikke anvendbart» (IA). Denne muligheten finnes for eksempel ikke i the Yale Adherence and Competence Scale (YACS) (Carroll et al., 2000). Der er 0-nivået på etterlevelse operajonalisert slik at det tilsvarer både «IA» og skåren «0» i MBTEKS og CTACS. Således vil enhver terapeutisk intervensjon som ikke gjøres få skåren 0, uavhengig av terapikontekstuelle faktorer. Dette har klare fordeler og ulemper. Fordelen er at bedømmeren får mindre fortolkningsarbeid, og det kan tenkes at man får økt enighet i skåringer. Ulempen er at man går glipp av muligheten for bedømmeren til å gjøre et skille mellom det som er dårlig etterlevelse (0) og det som er ikke-etterlevelse (IA). Erfaringer fra studien tilsier at denne muligheten er av så stor verdi at den bør forbli en del av skalaen på tross av eventuell innflytelse på bedømmerenighet. Nødvendigheten av å kunne gjøre skillet mellom disse to typene av fravær kan også tenkes å være ulik for MBT enn den typen terapi som vurderes av YACS. En mulig forklaring kan ligge i at YACS er en skala som anvendes på en terapiform som i større grad er ferdighetstrenende. Disse har som hovedregel et langt større fokus på frekvens av foreskrevne, eksplisitte behandlingsteknikker enn manualiserte dynamiske psykoterapier.

Etterlevelse og kvalitet. Vanskelighetene knyttet til å skille mellom etterlevelse og kvalitet opplevde også bedømmerne i vår enighetsstudie, og dette ble ofte diskutert etter skåringer. Det som ble rapportert i denne forbindelse omhandlet en halo-effekt der bedømmere ofte gjorde vurderinger

ved først å sette en skåre som var forankret i kvalitetsreferansen for det spesifikke ledd, for så å gi en tilvarende vurdering av etterlevelse. Problemet kan ha oppstått av flere ulike grunner. Når bedømmere valgte først å forankre en vurdering i kvalitetsreferansen, kan skyldes at denne er mer utfyllende og presist rettet mot det som observeres i MBTEKS. Etterlevelsismålet innehar de samme generelle referansene for alle punkter, og gir videre assosiasjoner til et frekvensmål uten at det er helt overensstemmende med etterlevelsesebegrepet i MBTEKS. Etersom skalaen nå krever at bedømmeren tar høyde for kontekstuelle variabler ved målet av etterlevelse, blir den vanlige måten å måle etterlevelse på (ved å telle intervensjoner) ikke tilstrekkelig for å treffe den «sanne skåren». Altså er det her kanskje snakk om et mål på etterlevelse som for øyeblikket er for komplisert for studiens bedømmere. Etersom MBTEKS har samme format som CTACS, så er det verdt å merke seg at Barber et al. (2005) opplevde samme problem ved sin studie. De antydte at dette kunne være et spørsmål om manglende trening. Samtidig opplevde vi i vår studie at etterlevelsesebegrepet i mange tilfeller ikke er godt nok presisert for mange av leddene. Dette gjelder kanskje spesielt leddene som er hentet fra CTACS. Ledd nummer 1 (se tabell 1) der en skal vurdere terapeutens «varme» er et eksempel på dette. En vurdering av kvaliteten på terapeutens varme oppleves intuitivt som en mer overkommelig oppgave enn å forsøke å kvantifisere fenomenet «varme». Som påpekt av en bedømmer under utformingsprosessen, vil spørsmålene en må stille seg ved vurdering av leddet lyde omtrent som følger: «hvor ofte var terapeuten varm?» og «når terapeuten var varm, hvordan var kvaliteten på varmen?». Begge disse spørsmålene gir store utfordringer. Det skaper problemer for den enkelte bedømmer, og, etter vår erfaring, en diskusjon om felles forståelse av konseptualiseringen uklar og upresis. Dette leddet, som riktignok er hentet fra en kognitiv skala, er av ikke-ferdighetstrenende natur i følge Perepletchikova et al.s (2007) kategori. Leddet illustrerer godt hvordan målingsproblemer oppstår når leddets indikatorer er uklare og vanskelige å kvantifisere.

Et annet problem er at vi i utformingsstudien opplevde at det ved ulike kvalitetsformuleringer var vanskelig å skille disse fra etterlevelse. Ved en gjennomgang av skalaen, er det fortsatt ledd hvor kvalitetsreferansen gir klare assosiasjoner til etterlevelse. Tabell X viser til et eksempel fra ledd 6 (tabell 6).

Tabell 6

6	Terapeuten gjorde en fremragende jobb med å stille adekvate spørsmål, der egne og andres motiver ble utforsket og fulgt opp med kreative kommentarer for å øke pasientens nysgjerrighet (dette preget mesteparten av timen).
---	---

Tabellen viser hvordan leddet i sin nåværende form definerer høyeste skåre på kvalitet ved å referere til at intervensjonen *preget mesteparten av timen*. Det ble tidlig klart at dette var en formulering som må ut av skalaen, dersom det skal være mulig å også skille mellom høy kvalitet og lav etterlevelse for leddet.

Et spørsmål som ikke er drøftet i litteraturen, og som jeg her vil rette fokus mot, er hvorvidt man bør inndele noen ledd i separate ledd for etterlevelse og kvalitet, samtidig som man må innse at dette avhenger av leddets natur. Altså vil en kunne tenke seg at en bedømmer ikke klarer å skille mellom kvantitet og kvalitet på en terapeuts varme, og at dette i større grad bør gjøres i arbeid med ledd som rettes mot mer spesifikke intervensjonsteknikker. En lignende problemstilling blir drøftet av Hatcher og Barends (2006) i deres gjennomgang av alliansebegrepet. Her argumenterer de for at dette allianse skiller seg fra terapeutisk teknikk, ved at det er en mer gjennomgående kvalitet ved et terapeutisk forhold, og lett kan sammenblandes med ulike teknikker. Dette er overensstemmende med våre erfaringer ved skåringer på de ulike ledd. Foreløpig er dog trenden at man lager samme format for alle ledd i en skala. Videre er leddene sjeldent diskutert på dette nivået i litteraturen, slik at erfaringene med andre skalaer ofte kun oppsummeres ved kommentarer på statistiske resultater. Hatcher og Barends (2006) kaller sin artikkel «How a return to theory could help alliance research». Jeg vil hevde at forskning på bedømmingsskalaer også har mye å tjene på videre teoretisk avklaring på de nevnte problemstillinger.

Enighetsstudien

Plenumsdiskusjonen. I plenumsdiskusjonene ble uklarheter i leddene ofte svært tydelig. Til tross for stor enighet blant flere, skjedde det at noen hadde forstått leddet, eller deler av et ledd, helt forskjellig. For å belyse dette, vil jeg her presentere transkriberte utdrag fra en diskusjon der det kommer frem ulike forståelser av ledd fem («ikke-vitende» holdning) og ledd seks (fremme utforskning og nysgjerrighet).

Utdraget begynner cirka 20 minutter ut i plenumsdiskusjonen;

2: *Jeg syntes at det var et gjennomgående problem med hvordan hun ordla seg i intervensjonene sine. Altså at det ble, altså at hun i sin måte å intervenere på konsekvent skåret litt lavt hos meg på ikkevitende holdning. Altså bar det veldig preg av «ja, dette er vanskelig for deg, fordi sånn og sånn».*

3: *Ja, på ikkevitende... hva hadde du på ikke-vitende holdning, da?*

- 2: *To én.*
- 1: *Det har jeg også.*
- 2: *Og jeg syntes at det preget, altså, selv når hun hadde rett fokus at det kom litt skjevt ut.*
- 3: *Mhm...*
- 2: *Så sånne som, eh, som dette her med intervensjonen på at pasienten lett er negativ og at dette kan være vanskelig å avlære, og «spesielt når man er to også». Altså hun kom lett med forklaringer på hvorfor ting var som de var og ble litt sånn belærende for pasienten med å forklare...*
- 3: *Altså hun hadde noen sånne ikke-, eller mere fremme utforskning og nysgjerrighet da...?*

Her trekker bedømmer 2 frem intervensjonsstilen til terapeuten som negativt for skåren på leddet «ikke-vitende» holdning. Bedømmeren opplyser om den lave skåren på leddet, og bedømmer 1 sier seg enig og oppgir at sin skåre på leddet er identisk. Bedømmer 3 avslutter så ved å antyde at terapeuten antagelig har kommet bedre ut på ledd 6 (fremme utforskning og nysgjerrighet), og svaret kommer i det følgende:

- 1: *Jeg har null jeg der.*
- 4: *Null null har jeg.*
- 3: *Ja, det synes jeg er alt for strengt. Det med å fremme utforskning og nysgjerrighet, der har jeg faktisk tre tre, fordi du har flere eksempler her... altså det starter opp med referanser til denne konflikten på jobben, hvor terapeuten spør «hvordan ble det for deg?»*
- 1: *Den har jeg tatt på ikke-vitende holdning, jeg.*
- 3: *Ja, ja, det kan også være ikke-vitende holdning, eller er det å fremme utforskning og nysgjerrighet, altså å gjøre forsøk på å stille spørsmål for å finne ut om egne og andres motiver... eh... ja... motiver er kanskje...*
- 6: *Ja, så hun utforsket jo, hun følte jo at hun hadde mer ansvar for konflikten enn de andre. «I den situasjonen du fortalte om, hva var det som...» altså, hun var jo utforskende...*
- 3: *Ja, i forhold til mannen: «hvordan opplevde du det?». Altså: «hvordan var det for deg?», «hvordan opplevde du det?», eh, «hvorfor gjorde du det, tro?», altså en del,*

det var en del slik som jeg vil si er å fremme utforskning... slik at en nuller der synes jeg er for strengt.

4: *Jeg tror ikke jeg kunne høre noe i forhold til mannen, jeg, altså...*

2: *Nei...*

Det transkriberte utdraget viser her til hvordan det innledningsvis i en treningsfase er store uklarheter knyttet til de ulike leddene. Denne diskusjonen fremhever kanskje også en svakhet i skalaen, da leddene fem og seks kan være overlappende i for stor grad. Det viktige for treningseffekten er her at de bedømmerne får etter hvert et større antall klare holdepunkter å knytte til de ulike leddene. Det er vår erfaring fra diskusjonene at læringskurven er brattest når det som skåres er tvetydig, eller vanskelig å få til å passe med et referansepunkt. Det som kommer frem og blir avklart i slike situasjoner vil på sikt inngå i en skriftlig manual for MBTEKS.

De statistiske resultatene

Skalaen som helhet. Tabell 7 inneholder de deskreptive dataene fra enighetsstudien for alle leddene i skalaen. N refererer her til antall ganger leddet er skåret av en rater, og derved er det maksimale antallet ganger et ledd kan ha blitt skåret seks ganger seks (6 bedømmere og 6 hele terapitimer) 36. Det mest påfallene i denne tabellen er ledd 14 som kun har blitt benyttet av én rater én gang, samt ledd 21 som kun er benyttet to ganger. Videre viser også skalaen relativt lave gjennomsnittsverdier. Til tross for at behandlere i studien hadde som intensjon å praktisere MBT, var påfallende mange av skårene under nivået av «god» MBT (skåren 4 for både etterlevelse og kvalitet).

Tabell 7

Ledd #	N	Etterlevelse			Kvalitet		
		Min/Maks	M	SD	Min/Maks	M	SD
1	36	2/6	3.86	1.02	1/6	3.81	1.06
2	36	0/5	3.03	1.25	1/5	2.94	1.12
3	36	1/6	3.22	1.05	2/6	3.14	1.05
4	36	0/5	3.06	1.15	1/5	3.08	1.05
5	36	1/5	2.81	1.35	0/5	2.69	1.33
6	36	0/6	2.75	1.50	0/6	2.47	1.50
7	31	0/5	1.71	1.72	0/6	1.74	1.79
8	19	1/5	2.37	1.16	0/4	2.21	1.03
9	36	0/5	2.47	1.32	0/6	2.36	1.25
10	29	0/4	.97	1.45	0/4	.86	1.30
11	25	0/4	1.12	1.27	0/5	1.28	1.59
12	36	0/5	2.33	1.59	0/6	2.17	1.58
13	36	1/6	3.47	1.23	1/6	3.28	1.29
14	1	1/1			1/4		
15	36	1/5	2.92	1.23	1/5	2.58	1.08
16	36	0/6	2.86	1.59	0/6	2.72	1.49
17	36	0/6	2.61	1.36	0/6	2.44	1.36
18	36	0/5	2.69	1.19	0/5	2.42	1.20
19	17	0/4	2.00	1.50	0/5	1.71	1.69
20	9	0/2	.22	.67	0/0	.00	.00
21	2	0/1	.50	.71	0/3	1.50	2.12
22	36	0/5	1.86	1.59	0/6	1.83	1.63
23	36	0/5	2.50	1.34	0/6	2.25	1.38

Reliabilitet ved ICC. Som et mål på enighet mellom bedømmerne, ble det utregnet intraklassekorrelasjoner ICC (2, n; Shrout og Fleiss, 1979), der bedømmer blir ansett for å være tilfeldig effekt og n er antallet bedømmere. I vår enighetsstudie var antallet bedømmere 6 (ICC 2, 6), men samtidig vil resultater presenteres med et estimat på hvor reliabelt én enkelt bedømmer vil antas å være på hvert enkelt ledd (ICC 2, 1). Fordi det antas at bedømmere vil ha læringseffekt av å bruke skalaen over tid, og fordi vårt utvalg er lite, er ikke disse intraklassekorrelasjonene ment å benyttes for å felle noen endelig dom over skalaen. Korrelasjonene vil inngå som en nyttig del av vurderingsgrunnlaget når revisjoner av skalaen foretas.

Tabell 8

Ledd #	Etterlevelse		Kvalitet	
	ICC (2, 6)	ICC (2, 1)	ICC (2, 6)	ICC (2, 1)
1	.47	.13	.68	.26
2	.55	.17	.52	.16
3	.53	.16	.81	.41
4	.75	.34	.90	.59
5	.81	.41	.75	.34
6	.85	.49	.90	.60
7	.99	.96	.97	.83
9	.87	.53	.85	.48
12	.88	.56	.88	.54
13	.84	.47	.87	.52
15	.68	.26	.85	.49
16	.85	.48	.88	.55
17	.79	.39	.86	.51
18	.76	.35	.70	.28
22	.69	.27	.87	.53
23	.83	.45	.87	.53

ICC (2, 1) og ICC (2, 6) for skalaen som helhet var henholdsvis .58 og .87 for etterlevelse og .62 og .88 for kvalitet. Det ble ikke utregnet ICC for leddene 8, 10, 11, 14, 19, 20 og 21 på grunn av et for stort antall manglende verdier (IA skåringer).

Testen viser (tabell 8) tilfredsstillende reliabilitet $ICC > .5$ for alle de MBT-spesifikke leddene ved ICC som et gjennomsnitt av studiens seks bedømmere ($ICC 2, 6$). Selv om foreliggende design har den styrken i reliabilitetsøyemed som seks bedømmere tilsier, vil antagelig senere anvendelse av skalaen begrenses til en eller to bedømmere. Derfor er resultatene presentert i kolonnene ICC (2, 1) av størst betydning for vurdering av skalaens anvendbarhet.

ICC (2, 1)-kolonnenene viser at ledd 1-3 har svært lave korrelasjoner ($r < .20$). Disse leddene er alle hentet fra CTACS (Nordahl et al., 2006), og det blir derfor naturlig å gjøre en ny vurdering av hvorvidt disse leddene skal inngå i fremtidige versjoner av MBTEKS.

De øvrige leddene viser store sprik. For etterlevelse varierer de MBT-spesifikke leddene i ICC 2, 1 fra .27 til .96, mens for kvalitet er variasjonen fra .28 til .83. Ledd 7 hadde høyeste ICC for både etterlevelse og kvalitet (henholdsvis .96 og .83). For etterlevelse er leddene 7, 9 og 12 de eneste leddene med $ICC(2, 1) > .5$. For kvalitet er har leddene 6, 7, 12, 13, 16, 17 og 23 $ICC(2, 1) > .5$. Analysen viser altså at vurdering av kvalitet gjøres noe mer reliabelt enn vurdering av etterlevelse i enighetsstudien. Det er ikke påfallende likheter mellom leddene med $ICC(2, 1) > .05$ som kan forklare deres relativt høye ICC, men ledd 7 er, ved å beskrive en spesifikk intervensjonstype, av den typen ledd som erfaringsmessig gir bedre reliabilitet (Perepletchikova et al. 2007).

Variansanalyse - hva forklarer variasjonen i skåringene? Som beskrevet kan designet for enighetsstudien slik at det vil kunne si noe om effekten av bedømmer, ledd og terapitime på skåringene.

Tabell 9

Tre-veis variansanalyse av etterlevelse og kvalitet avhengig av bedømmer, terapitime og ledd.

	Etterlevelse	Kvalitet
	Kvadratsum	Kvadratsum
Terapitime	324.48	409.66
Ledd #	324.01	300.75
Bedømmer	66.74	57.08
Terapitime*Ledd	134.09	102.18
Terapitime*Bedømmer	208.99	226.49
Terapitime*Ledd*Bedømmer	376.65	294.67
Total	1546.16	1513.04

Utifra tabell 9 kan man lese styrken av de ulike komponentene på variasjonen i skåringene. I designet skjuler det seg som sagt to antatt viktige kilder til variasjon: terapeut og pasient.

En styrke ved disse tallene er den forholdsvis lave variansen i skåringer som kan forklares ved bedømmer alene. Dersom testen virkelig fanger opp variasjon i observasjonsenheten, med bedømmere som reliable måleinstrumenter, skal i teorien variasjonen være i større grad forklart av den spesifikke terapitimen og det spesifikke ledd, enn av bedømmeren som gjør vurderingen. Interaksjonseffektene for terapitime*bedømmer er noe høy. Dette skyldes antagelig at skåringene er fra den innledende fasen. Senere variansanalyser, gjort etter omfattende trening, bør sammenlignes.

Bedømmeres enighet. For å se på forskjeller i enighet mellom bedømmere, ble det utregnet korrelasjoner for etterlevelseshet- og kvalitetsskåringene til de seks bedømmerne som inngikk i studien. Her er N=6.

Tabell 10

Bedømmer	Etterlevelse		Kvalitet	
	Korrelasjon med gruppen*	ICC (2, 6) uten bedømmer	Korrelasjon med gruppen*	ICC (2, 6) uten bedømmer
1	.89	.84	.93	.87
2	.89	.84	.90	.87
3	-.07	.95	.04	.95
4	.87	.85	.88	.88
5	.78	.86	.85	.88
6	.92	.85	.90	.88

*Korrelasjonen mellom den enkelte bedømmer og de øvrige bedømmeres samlede skårer.

Av tallene i tabell 10 fremgår det en påfallende svakhet ved enigheten mellom bedømmerne. Bedømmer tre er et tydelig avvik ved vurderinger av både etterlevelse og kvalitet. Ved etterlevelse er bedømmerne enige med korrelasjoner på rundt 0,8, mens rater tre er på -,07. Kolonnen ICC (2, 6) uten bedømmer forteller hva den samlede ICCen ville vært uten bedømmeren. Dette vil si at skalaen som helhet øker fra henholdsvis .87 og .88 til .95 for både etterlevelse og kvalitet uten bedømmer 3.

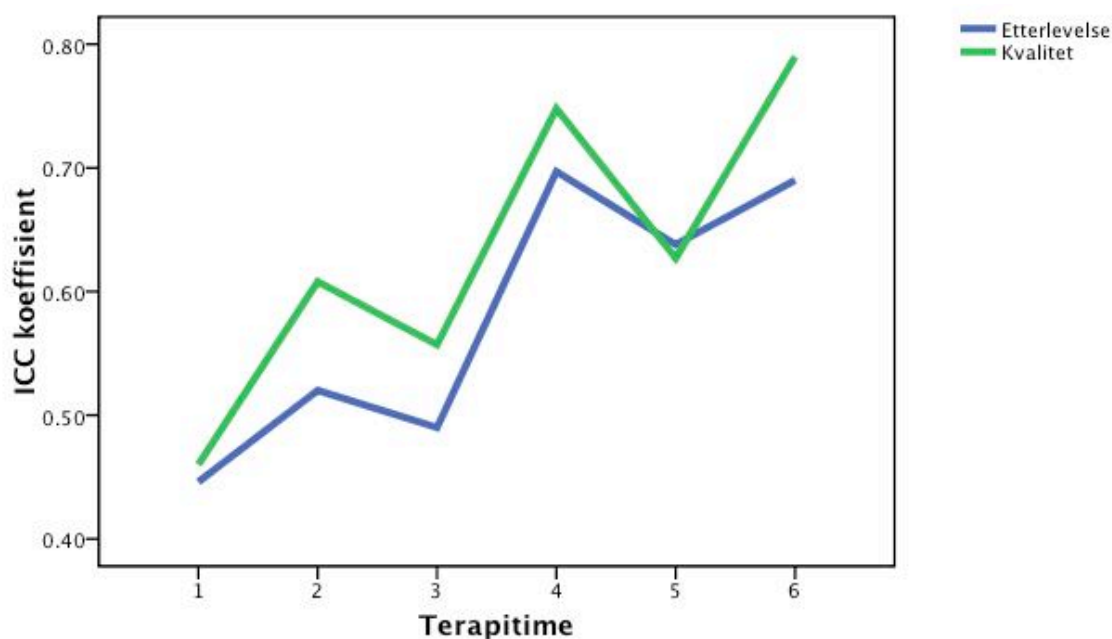
Variasjon i enighet over terapitimer - treningseffekt. Tabell 11 viser utviklingen av enighet over terapitimer i enighetsstudien. Terapitimene er listet opp i kronologisk rekkefølge, slik at man kan se utviklingen av enighet ved å bevege seg nedover radene.

Tabell 11

Terpitime	Etterlevelse		Kvalitet	
	ICC (2, 6)	ICC (2, 1)	ICC (2, 6)	ICC (2, 1)
1	.83	.45	.84	.46
2	.87	.52	.90	.61
3	.85	.49	.88	.56
4	.93	.70	.95	.75
5	.91	.64	.91	.63
6	.93	.69	.96	.79

Figur 1 fremstiller utviklingen av enighet over tid (terapitimer i kronologisk rekkefølge). En utregning av korrelasjonen mellom terapitimennummer og mål på enighet (ICC 2, 1) viser til en signifikant utvikling av enighet for både etterlevelse ($r = .88$, $p < .05$) og kvalitet ($r = .83$, $p < .05$) på et $p < .05$ nivå. Dette resultatet støtter antagelsen om at enighet utvikles ved skåringer og plenumsdiskusjoner.

Figur 1



Sammenhengen mellom etterlevelse og kvalitet. Forholdet mellom etterlevelse og kvalitet for de enkelte ledd og for skalaen samlet, ble beregnet ved å benytte «Pearson korrelasjonskoeffisienter». Den samlede korrelasjonen mellom kvalitet og etterlevelse for alle ledd var i enighetsstudien var på .89 ($r = .89, p < .01$). Tabell 12 inneholder en oversikt over korrelasjonskoeffisientene for alle ledd. Alle korrelasjonene er her signifikante ($p < .05$). Korrelasjoner for leddene 14, 20 og 21 ble ikke tatt med grunnet få skåringer. Oversikten viser at korrelasjonen mellom etterlevelse og kvalitet for samtlige ledd er forholdsvis høy ($r > .70$). Kun fem ledd (1, 4, 8, 15 og 19) hadde korrelasjoner lavere enn .80 ($r > .80$), men dette må også anses for å være relativt høyt. Det er ingen påfallende likhetstrekk mellom disse leddene som skiller dem fra ledd med høyere korrelasjonskoeffisienter. Ledd 10 har høyeste korrelasjon mellom kvalitet og etterlevelse ($r = .98$). Denne høye koeffisienten skyldes forholdsvis hyppige 0-skårer på etterlevelse som normalt også gir 0 i kvalitet. Den relativt høye forekomsten av 0-skårer på ledd 10 kan gjenspeiles også i gjennomsnittet for dette leddet i tabell 7 ($M = .97$).

Tabell 12

Korrelasjonskoeffisienter for forholdet mellom etterlevelse og kvalitet			
Ledd #	<i>r</i>	Ledd #	<i>r</i>
1	.77	13	.86
2	.82	14	-
3	.81	15	.75
4	.73	16	.90
5	.86	17	.92
6	.90	18	.89
7	.89	19	.79
8	.76	20	-
9	.92	21	-
10	.98	22	.86
11	.89	23	.89
12	.96		

Diskusjon

Studiens formål var å beskrive prosessen med å utvikle en bedømmingsskala for MBT. Dette innebar en redegjørelse for generering av ledd, beskrivelse av operasjonalisering på ulike etterlevelsens- og kvalitetsnivåer, og problemstillinger forbundet med dette. I tillegg skulle studien vurdere den innledende enighetstesten der man forsøkte å opparbeide en felles forståelse av hvordan de enkelte ledd beskriver fenomener og intervensjoner observert i terapitimer.

Resultatene fra utformings- og enighetsstudien viser at skalaen fremdeles står foran uløste problemstillinger. Dette gjelder følgende: (1) den store variasjonen i reliabilitet ved skåringer på de ulike ledd, (2) at noen ledd svært sjeldent ble benyttet, (3) en avklaring på terskelspørsmålet ved skåringer av ledd som har skåringsalternativet «ikke anvendbart», (4) en redegjørelse for, og mulige løsninger på, den høye korrelasjonen mellom etterlevelse og kvalitet, og (5) en avklaring vedrørende leddene for generelle psykoterapikriterier som er hentet fra CTACS (Nordahl et al., 2006; Barber et al., 2003).

Erfaringene og resultatene viser at det er potensiale for videre utvikling og revisjoner. Enighetsstudien viser en signifikant forhøyet enighet over de første seks skåringene. Leddene med lavest reliabilitet var hentet fra CTACS (Nordahl et al., 2006; Barber et al., 2003), og Barber et al. (2003) fant også lave reliabilitetskoeffisienter for to av disse fire leddene (empati og oppmerksomhet). Utvikling av nye og mer presist operasjonaliserte ledd under generelle psykoterapikriterier, vil derfor kunne gjøres i det videre arbeidet.

Det må presiseres at vår enighetsstudie inneholder skåringer gjort uten trening av bedømmerne. De innledende resultatene fra enighetsstudien er derfor ikke sammenlignbare med f.eks. Barber et al. (2003) sine funn, da deres studie av CTACS ble først gjort etter at det var etablert prosedyrer for opplæring av bedømmere, samt utformet skriftlige veiledninger som supplement til bedømmingsskalaen. Det videre arbeidet med MBTEKS vil derfor, i tillegg til revisjoner av skalaens konkrete innhold, også innebære utarbeidelse av treningsprosedyrer og en skåringsmanual som supplement til skalaen.

Viktige begrensninger ved studien

Enighetsstudiens observasjonsenhet var terapitimer. Da datamaterialet kun baserer seg på skåringer av seks hele terapitimer, må dette tas i betraktning ved slutninger tuftet på de statistiske

resultatene. Videre var flere av terapeutene kollegaer av bedømmerne. Dette kan ha gitt haloeffekter ved skåringer.

Begrensningene i studiens design er ikke av avgjørende betydning for konklusjonene oppgaven forsøker å trekke. Oppgavens fokus er utviklingsprosessen og erfaringene høstet fra denne. Idéer til revisjoner av skalaen vil derfor ikke baseres alene på kvantitative resultater, men vil gjøres på grunnlag av erfaringene fra utviklingsprosessen der de kvantitative resultatene inngår som supplerende informasjon. En studie for å teste skalaens psykometriske egenskaper mer inngående, vil gjøres etter at ytterligere revisjoner av skalaen er foretatt, samt et treningsprogram for bedømmere er utviklet.

Skalaen som helhet og de spesifikke ledd

Skalaen som helhet viser i enighetsstudien lovende reliabilitetsmål for både etterlevelse og kvalitet. Manglende treningsprogram, svake reliabilitetsmål på ledd hentet fra CTACS, samt den påviste treningseffekten taler til fordel for MBTEKS i den videre utviklingen. Variansanalysen viser at skalaen fanger opp variasjon i observasjonene, og at variasjonen i større grad forklares av de spesifikke ledd, terapitimen som bedømmes og ulike interaksjonseffekter, enn av hvem som benytter skalaen (bedømmer).

De generelle psykoterapikriteriene vil ikke inngå i fremtidige utgaver av MBTEKS. De ekskluderes av to grunner: (1) Enighetsstudien viser til lave mål på reliabilitet ved utregninger av ICC sammenlignet med øvrige ledd, og (2) erfaringer fra plenumsdiskusjonene tilsier at begrepene og referansene som benyttes i leddene ikke er forenlige med MBT.

De lave reliabilitetsestimatene kan skyldes at plenumsdiskusjonene tenderte i retning av et fokus på å opparbeide enighet rundt de MBT-spesifikke leddene til fordel for de hentet fra CTACS. Barber et al.s (2003) funn av lave reliabilitetskoeffisienter for to av disse fire leddene taler dog til fordel for vår beslutning om å utarbeide nye ledd for generelle psykoterapikriterier. Viktigst er kanskje funnet av at leddenes formuleringer ved ulike anledninger viste seg ikke å være kompatible med MBT-beskrivelser angitt i manualen (Bateman og Fonagy, 2006).

Ledd 14 ble kun benyttet av én bedømmer én gang i løpet av studien. Det foreligger dermed ikke noen statistiske resultater på hvorvidt leddets psykometriske egenskaper er tilfredsstillende, men erfaringene tilsier at leddet svært sjeldent har vært oppe til diskusjon fordi det som oftest har vært tydelig «ikke anvendbart». Det er flere mulige forklaringer på hvorfor dette leddet ikke ble anvendt hyppigere i enighetsstudien. Leddet omhandler «Bruk av metaforer, analogier og symboler». Utvalget vårt bestod av terapeuter, med ulik opplæringsbakgrunn, som alle hadde en

intensjon om å praktisere MBT. Bruk av metaforer, analogier og symboler er stort sett frarådet ved MBT-intervensjoner i henhold til manualbeskrivelser. Således er leddets anvendbarhet på MBT-terapier muligens begrenset, da dette i stor grad etterleves av terapeutene. Derimot skal mål på behandlingsintegritet også inneholde beskrivelser av teknikker og intervensjoner som er frarådet. Dersom mål på etterlevelse også inneholder beskrivelser av frarådede intervensjonstyper og teknikker, er et mål på etterlevelse ansett for å være tilstrekkelig for å vurdere skalaens evne til å skille mellom ulike terapiformer (Waltz et al. 1993). Dette kan bety at leddet kan tenkes å komme til sin rett ved studier der skalaen testes på et heterogent utvalg av terapier bestående av terapiformer som i større grad tillater bruk av metaforer, analogier og symboler. Om leddet skal bestå ved senere studier avhenger altså av hvorvidt leddet egner seg til å skille MBT fra andre terapiformer. En annen mulig løsning er å omformulere leddet slik at det i større grad også er anvendbart på MBT.

Ledd 21 er kun benyttet to ganger i studien. Leddet beskriver foreskrevne MBT-metoder for å arbeide med «alliansebrudd». Manglende antall skåringer gjør også dette leddets psykometriske egenskaper uklare i de kvantitative analysene. Det lave antallet skåringer henger sammen med at leddet beskriver et helt spesifikt fenomen. Om leddet skal bestå avhenger av hvor viktig det MBT-spesifikke arbeidet med alliansebrudd vurderes å være, samt i hvor liten grad dette arbeidet fanges opp av resterende MBT-ledd. Arbeid med alliansebrudd anses som sentralt i MBT ved behandling av borderline pf. Derfor vil leddet antagelig bestå i sin nåværende form inntil en større studie avdekker eventuelle svakheter med hensyn til nivåformuleringer og/eller svakheter relatert til reliabilitetsmål.

Terskelspørsmålet

Et viktig moment som ble fremtredende under utviklingen og i plenumsdiskusjonene, var de ganger det oppstod uenighet om hvorvidt et ledd skulle skåres «IA» eller få skåren «0» på etterlevelse. Drøftningen av ulike syn på dette ble presentert som resultater av utviklingsstudien da vi har valgt å forsøke oss på dette skillet. Andre skalaer av lignende format (f.eks. YACS; Carroll et al., 2000) har omgått dette problemet ved at skalaen ikke skiller mellom etterlevelse «0» og skåren «ikke anvendbar». Slike skalaer forholder seg kun til det observerte ved at de ikke bedømmer i hvilken grad fravær av en intervensjon var upassende (0 i etterlevelse) eller «ikke anvendbar». Selv om dette forenkler arbeidet til bedømmeren, kan det antas å være en svakhet dersom en terapitime viser at en terapeut gjentatte anledninger forsømmer seg ved å ikke ta tak i fenomener som utspiller på en måte som angitt i manualen. Det empiriske grunnlaget fra enighetsstudien er ikke tilstrekkelig

for å kunne konkludere i forhold til hvorvidt skillet mellom en nullskåre i etterlevelse og «ikke anvendbart» er for ambisiøst. Erfaringene fra enighetsstudien tilsier at bedømmere ved flere anledninger opplevde skillet som meningsfullt. Dette oppstod typisk i terapier preget av fenomener som dominerte timen (f.eks. pretend modus), uten at terapeuten viste til intervensjoner som arbeidet med dette. Slike erfaringer fra utviklingsstudien og plenumsdiskusjonene medfører at den foreløpige konklusjonen taler til skillet fordel. Senere og mer omfattende reliabilitetsstudier vil så kunne besvare om skillet er forsvarlig rent empirisk.

Etterlevelse og kvalitet

Resultatene fra enighetsstudien viste høy korrelasjon mellom etterlevelse og kvalitet. Dette funnet er i tråd med Barber et al.s (2003) erfaringer med CTACS. Ettersom CTACS er av samme format som MBTEKS, vil denne sammenligningen være naturlig, og våre funn kan indikere at det er egenskaper ved selve utformingen av skalaen som forklarer den høye korrelasjonen.

Barber et al. (2003) presenterer sitt funn av den høye korrelasjonen som et problem. De argumenterer for at det i teorien bør være større variasjon i kvalitet på intervensjonene enn det korrelasjonskoeffisienten tilsier, og antar at deres funn henger sammen med mangelfull opplæring av bedømmere i forkant av studien. Våre erfaringer fra enighetsstudien har ledet til følgende fem mulige forklaringer på den høye korrelasjonen: (1) mangelfull trening og/eller avklaring/definisjon av forskjellen på etterlevelse og kvalitet for de enkelte ledd, (2) en mulig naturlig sammenheng mellom terapeuters høye etterlevelse og kvaliteten på denne, (3) egenskaper ved skalaens og det komplekse syn på etterlevelse, (4) kvalitetsoperasjonaliseringer som glir over i etterlevelseshbegrepet, og (5) grunnleggende forskjeller i leddenes natur (kvalitetsvariabler og aktivitetsvariabler).

Enighetsstudien ble påbegynt rett etter skalaen var ferdig utviklet. Et treningsprogram og skriftlig veileder/manual for benyttelse av skalaen var derfor ikke-eksisterende. Dette gjør mangelfull trening til en mulig forklaring på vårt funn av høy korrelasjon mellom etterlevelse og kvalitet.

Det er uklart om MBT som beskrevet i manualen gir opphav til en naturlig sammenheng mellom etterlevelse og kvalitet, eller om det er en naturlig sammenheng blant terapeutene i deres behandlingsteoretiske kunnskap og kvaliteten på deres utførelse. Barber og Crits-Cristoph (1996) fant i sin studie av Penn Adherence/Competence Scale for Supportive-Expressive (SE) Dynamic Psychotherapy (PACS-SE; Barber, 1988) en langt lavere korrelasjon mellom kvalitet og etterlevelse. Dette funnet taler imot en forklaring som retter seg mot terapeutpopulasjonen som

helhet, ettersom det ikke synes å være noen generell tendens til at kvalitet og etterlevelse naturlig hører sammen som egenskaper ved terapeuter. Konstruktene har vist seg å være forskjellige ved andre studier, slik at årsakene til våre funn sannsynligvis må rettes mot skalaens spesifikke egenskaper, eller egenskaper ved MBT som terapiform. Ettersom Barber et al.s (2003) funn ved studien av CTACS er i tråd med våre resultater, antyder dette at forklaringen kan ligge i skalaens spesifikke utforming.

Skalaens komplekse konseptualisering av etterlevelse, med blant annet skillet mellom «IA» og «0» (ikke rent frekvensmål på intervensjoner), ble rapportert som et problem for bedømmerne under enighetsstudien. Det kan tenkes at et mer enhetlig fokus på observerte intervensjoner, og kvantifiseringen av disse (f.eks. YACS; Carroll et al., 2000), leder til større skiller mellom etterlevelse og kvalitet. Utfordringen med en slik definisjon av etterlevelse, er at den bryter med vår oppfatning at frekvens og etterlevelse ikke er det samme. Etterlevelse som mål på hvorvidt terapeuten etterlever beskrivelser angitt i manualen, fører til en mer kompleks operasjonalisering. Derfor vil vi anse et rent frekvensmål som, muligens mer reliabelt, men svekket i validitetsøyemed ettersom vi ser på etterlevelse som et mål både på hva terapeuten gjør, og hva terapeuten burde gjøre i henhold til behandlingsmanualen. En forståelse av etterlevelse utelukkende som frekvens vil ikke gi valide mål på det vi forstår som etterlevelse av MBT.

The Collaborative Study Psychotherapy Rating Scale (CSPRS) ble kritisert av Waltz et al. (1993) fordi den inneholdt etterlevelsledd som i ulik grad også gjorde mål på kvalitet. Ved benyttelse av MBTEKS oppdaget vi lignende problemer underveis. Under enighetsstudien ble det tydelig at enkelte av leddene inneholdt kvalitetsreferanser som overlappet med begrepet om etterlevelse. Disse erfaringene vil lede til flere reformuleringer i skalaen. Hvor store problemer dette ga bedømmerne utover i studien er vanskelig å anslå. Ettersom disse problemene ble tidlig tematisert, og det var enighet om å forstå referansen annerledes enn det som stod skrevet, er det mulig at effekten av disse referanseformuleringene ikke forklarer så mye av den høye korrelasjonen, men endringer vil gjøres slik at referansepunkter i tydeligere skiller etterlevelse og kvalitet.

Forskjeller i leddenes natur - aktivitets- og kvalitetsledd

Det er tidligere påpekt hvordan bedømmere i enighetsstudien opplevde vanskeligheter med å skille mellom etterlevelse og kvalitet på leddet som omhandler terapeutens varme (ledd 1; «varm/genuin/kongruent»). Lignende problemer oppstod for flere andre ledd. Jeg vil belyse dette problemet, og komme med et forslag til løsning, ved å opprette to hovedkategorier for ledd som inngår i bedømmingsskalaer: aktivitetsledd og kvalitetsledd.

Skillet mellom aktivitets- og kvalitetsledd understreker det som antas å være grunnleggende forskjeller i leddene som variabler. Enkelte ledd i bedømmingsskalaer beskriver spesifikke terapeutaktiviteter (f.eks. ledd 9 «utfordrer uberettigede oppfatninger»). Slike ledd er det i stor grad mulig å kvantifisere. Dette kan gjøres ved å utelukkende telle terapeutens intervensjoner rettet mot pasientens uberettigede oppfatninger, eller ved å telle terapeutens intervensjoner rettet mot uberettigede oppfatninger, og sette dette opp mot antall ganger i terapien pasienten kom med utsagn preget av uberettigede oppfatninger. Kvaliteten vil så kunne bedømmes ved å gjøre en vurdering av hvor grundig arbeid terapeuten gjorde ved de tilfellene uberettigede oppfatninger ble konfrontert. Min erfaring tilsier at denne vurderingen blir straks mer problematisk når bedømmere skal vurdere ledd som i mindre grad beskriver en terapeutaktivitet. Ledd 1, «varm/genuin/kongruent», er et eksempel på et slikt ledd som er nevnt tidligere. Ved at leddet i sin natur er rettet mot det som antas å være en mer gjennomgående kvalitet ved terapeuten i arbeidet, gjør dette skillet mellom etterlevelse og kvalitet langt mer problematisk. Dersom det senere viser seg at dette skillet ikke lar seg gjøre, selv ved grundig opplæring av bedømmere vil, jeg derfor foreslå at bedømmingsskalaer bør inndele ledd i kategoriene aktivitets- og kvalitetsledd. En slik inndeling vil føre til at kvalitetsledd utelukkende vurderes for kvalitet, mens aktivitetsleddene forblir inndelt i dimensjonene etterlevelse og kvalitet. Denne inndelingen i aktivitets- og kvalitetsledd bør gjøres etter en presis definisjon de to kategorier av ledd. Jeg foreslår at disse definisjonene tar utgangspunkt i begrepet etterlevelse. Dersom et ledd ikke beskriver en aktivitet som har operasjonaliserte, observerbare og kvantifiserbare indikatorer, men i vesentlig grad beskriver en gjennomgående kvalitet ved terapeutens arbeid, vil leddet kunne kategoriseres som et kvalitetsledd. Slike ledd vil da vurderes utelukkende for kvalitet.

Ledd 1 er hentet fra CTACS (Nordahl et al., 2006; Barber et al., 2003). Barber et al. (2003) går i sin drøftning av CTACS ikke inn på mulige forskjeller i ledd på dette nivået. Deres regjøreelse for høy korrelasjon mellom etterlevelse og kvalitet retter seg mot mangelfull trening, samt at de spekulerer i hvorvidt teknikker beskrevet for kognitiv terapi er så spesifikke og detaljerte, at de vanskeligjør et skille. Vår erfaring tilsier at en drøftning av de enkelte ledds spesifikke egenskaper er svært nødvendig som et tillegg til statistiske analyser. Statistiske analyser alene vil ikke kunne avdekke slike problemer. En diskusjon på leddnivå, og en avklaring om hvordan ledd kan plasseres i spesifikke kategorier, vil jeg lansere som et forslag til hvordan man kan bedre bedømmingsskalaers anvendbarhet.

Generelle psykoterapikriterier

De foreliggende leddene for generelle psykoterapikriterier vil erstattes av nye ledd i fremtidige versjoner av MBTEKS. Dette gjøres i hovedsak på grunn av leddenes manglende kompatibilitet med MBT-manualen som ligger til grunn for MBTEKS (Bateman og Fonagy, 2006). Endringene er drøftet med Anthony Bateman (personlig kommunikasjon, 09.03.09), som er enig i denne vurderingen. Han har i tillegg foreslått at forslag til nye kriterier kan hentes fra skalaer som måler allianse. Videre ble det foreslått en endring i operasjonaliseringen av empati som i større grad er overensstemmende med MBT ved at det beskriver terapeutens engasjement for å *forsøke* å forstå pasientens perspektiv, i stedet for det kognitive leddets fokus på terapeutens *faktiske* forståelse av pasientens perspektiv.

Arbeidet med å utvikle nye kriterier for generelle psykoterapikriterier er ikke påbegynt. Det er foreløpig uklart hvilke ledd denne delen av skalaen skal inneholde, hvorvidt delen vil inneholde såkalte kvalitetsledd, og hvor mange ledd som må til for å fange opp en god samlet beskrivelse av generelle terapeutferdigheter. Dette er viktige valg i forhold til skalaens validitet. Dersom skalaen som helhet skal nå sitt mål om å fange opp i hvilken grad man observerer god MBT-terapi, må videre diskusjoner strekke seg lenger enn reliabilitetsspørsmålet, og mer i retning av leddenes innholdsvaliditet.

Treningseffekt

Enighetsstudien viser et signifikant forhold mellom trening og mål på enighet utregnet ved ICC. Dette tyder på at MBTEKS, for å kunne la seg sammenligne med studier av andre skalaer, først må utarbeide et omfattende treningsprogram og en manual/veileder for benyttelse av skalaen. Det må understrekes at studien kun baserer seg på seks skåringer, men tendensen til utvikling av enighet er tydelig og i tråd med anbefalinger basert på erfaringer gjort i andre studier av bedømmingsskalaer (Barber et al., 2003; Barber og Crits-Cristoph, 1996; Barber et al., 1996; Carroll et al., 2000). Disse skalaene har gjennomført interbedømmerstudier først etter å ha etablert treningsprogrammer, utviklet manualer til skåringsskjemaene og gitt bedømmerne som inngår i studien grundig trening. MBTEKS i enighetsstudien viser til treningseffekt allerede ved de seks første skåringene. Dette resultatet viser at det vil være nødvendig å utvikle et strukturert treningsopplegg med hensikt å bidra til utvikling av enighet. Skalaens anvendbarhet kan deretter avgjøres ved reliabilitetsmål gjort etter omfattende opplæring av bedømmerne.

Erfaringer fra utviklingsstudien vil danne grunnlaget for utviklingen av et treningsprogram spesifikt for MBTEKS, samt en manual for anvendelse av skalaen. Dette arbeidet er i startfasen, og

foreløpig vurderes manualen tilhørende Yale Adherence and Competence Scale YACS (YACS II; ikke publisert) som en mulig inspirasjonskilde til en manual for MBTEKS. Et treningsprogram vil ta utgangspunkt behandlingsmanualen for MBT (Bateman og Fonagy, 2006), en manual tilpasset MBTEKS, samt resultater av videre studier på enighetsutvikling. Enighetsstudien viser foreløpig den innledende utviklingen i enighet. Ved senere studier bør denne utviklingen av enighet studeres over flere terapitimer parallelt med at bedømmerne mottar opplæring. En slik studie vil kunne se på effekten av enighetsutvikling under treningsfasen. Målet vil være å se hvor mange skåringer og hvor mye opplæring som skal til for at enighetsutviklingen når et optimalt nivå. Dette optimale nivået kan slutes fra statistiske analyser av hvor mange terapitimer som skal bedømmes før *utviklingen* i enighet ikke lenger er signifikant økende.

Enigheten mellom bedømmere i enighetsstudien viser et interessant funn ved at en bedømmer (bedømmer 3) er markant mindre enig med øvrige bedømmere i studien. Dette kan også spille en viktig rolle ved utvikling av treningsprosedyrer. Dersom flere bedømmere trener på utvikling av enighet sammen, bør en analyse av enighet for hver enkelt bedømmer i forhold til de øvrige, gjøres som en «screening» i forkant av eventuelle studier av behandlingseffekt. Hvilke konsekvenser en slik screening vil få for den enkelte bedømmer, er ikke helt entydig. En mulighet er å operere med «cut-off»-verdier på korrelasjonskoeffisienter mot slutten av en treningsfase der bedømmere under denne ekskluderes som bedømmere i videre studier. Det er dog ikke undersøkt om denne manglende enigheten kan bedres ved å intervensere på ulike måter. Det er for eksempel ikke undersøkt effekten av å gjøre bedømmeren oppmerksom på problemet.

Veien videre

De ulike erfaringene som er drøftet kaster lys over veien videre i utviklingsarbeidet med MBTEKS. Forestående oppgaver som er drøftet består i å utvikle og definere et anbefalt treningsprogram, utvikle en manual til veiledning for bedømmere, forkaste svake ledd og erstatte enkelte av disse (f. eks. generelle psykoterapikriterier) med nye, undersøke om en inndeling i kategoriene aktivitets- og kvalitetsvariabler er meningsfull og empirisk holdbar, reformulere kvalitetsreferanser som overlapper med etterlevelsesebegrepet, tydeligere definere og operasjonalisere terskler for skåring for de ledd der dette er aktuelt, og avklare hva som skal være tilfredsstillende mål på enighet samt stillingstagen til bedømmere som ikke tilfredsstillende definerte kriterier for tilfredsstillende enighet. Mer omfattende vurderinger vil være om den grunnleggende operasjonaliseringsmetoden er hensiktsmessig. Hoffart (1997) har for eksempel tatt til orde for å vurdere kvalitet på utførelse av terapi utifra en skjematilnærming. Denne forståelsen kan

inkorporeres ved utvikling av en manual til MBTEKS. En slik skjematilnærming vil i større grad se på helhetlig vurdering terapeutens arbeid og hvordan dette henger sammen med den teoretiske intensjonen, til fordel for et mer konkret fokus på kvalitet knyttet til enkeltstående intervensjoner.

Ved en fremtidig reliabilitetsstudie må det også gjøres viktige vurderinger i forhold til valg av design. En slik studie må samle videobånd med terapitimer fra ulike terapitradisjoner slik at MBTEKS kan vurderes i forhold til hvor vidt den kan skille MBT fra andre former for terapi. Videre vil det være naturlig i en slik studie å operere med to trente bedømmere da dette vil føre til et bedre sammenligningsgrunnlag med andre studier gjort på bedømmingsskalaer (f. eks. Barber et al., 2003; Barber og Crits-Cristoph, 1996; Barber et al., 1996; Carroll et al., 2000).

Et annet viktig spørsmål vil være om det også bør settes mål på treffsikkerhet i tillegg til interbedømmer reliabilitet. Wu, Whiteside og Neighbours (2007) argumenterer for at gullstandarden for validering av en bedømmingsskala er ved å gjøre separate mål på treffsikkerhet (grad av samsvar med ekspert) og interbedømmer reliabilitet (enighet; grad av samsvar mellom bedømmere). Slike studier krever dog at man først definerer en autoritet som blir gitt status som ekspertbedømmer (fortrinnsvis en som er godt trent i behandlingsformen og har omfattende kjennskap til bedømmingsskalaen). Ekspertbedømmeren skårer så terapitimer uavhengig av gruppen som skal trenes opp i enighet. Disse skåringene vil så fungere som en fasit bedømmerne måles mot for treffsikkerhet. En slik studie vil være krevende. Den forutsetter at skalaen først er ferdig utviklet og at man klarer å definere en ekspert på bruk av skalaen. Det er foreløpig uklart hva som skal ligge til grunn som kriterier for vurdering av ekspertkompetanse. Ettersom Jacobson (1998 ref. i Barber et al., 2003) fant lav reliabilitet mellom eksperter på kognitiv terapi ved benyttelse av CTS, kan man ikke uten videre anta sammenheng mellom kompetanse på bedømming ut ifra kompetanse på behandlingsteori. Dette spørsmålet bør derfor avklares nærmere før en utnevner en ekspertrolle til en MBTEKS-bedømmer.

Under utviklingen meldte det seg også et behov for begrepsdannelse. Terskelbegrepet, og en mulig inndeling i aktivitets- og kvalitetsvariabler må undersøkes nøye i den videre utviklingen. Dersom fremtidige erfaringer tyder på at disse begrepene forteller noe viktig om bedømmingsskalaer, bør det lede til funn som kan føre til at det utarbeides retningslinjer for hvordan en skala skal løse terskelspørsmål, samt danne klare definisjoner for kategoriene aktivitets- og kvalitetsvariabler, slik at ledd lar seg enklere plassere innenfor rette kategori.

Konklusjon

Erfaringene som er gjort i utviklingsstudien viser til mangler i retningslinjer for hvordan en bedømmingsskala skal utformes. Derfor anbefales det at forskning på bedømmingsskalaer i større grad studerer selve utformingsprosessen og problemer knyttet til denne, i tillegg til å gjøre statistiske analyser. Et slikt arbeid vil i større grad sette et fokus på validitet, og ikke anta at reliabilitetsmål alene er tilstrekkelig for å validere en skala. Et overordnet fokus ved utvikling av en bedømmingsskala bør være en beskrivelse av erfaringer gjort ved den praktiske anvendelsen av en skala. Det er i samspillet mellom skalaens faktiske innhold og bedømmernes praktiske anvendelse at de viktige spørsmålene dukket opp i Utviklingsstudien. En skala må altså tilpasses både det teoretiske fundamentet i behandlingsmanualen og bedømmerne som skal benytte skalaen. Erfaringer fra utviklingsstudien tilsier at et slikt arbeid vil være svært nyttig for å gjøre skalaer mer praktisk anvendelige.

Referanser

- Abrahamson, D. J. (1999). Outcomes, guidelines, and manuals: On leading horses to water. *Clinical Psychology: Science and Practice*, 6, 467-471.
- Addis, M. E., & Krasnow, A. D. (2000). A national survey of practicing psychologists' attitudes toward psychotherapy treatment manuals. *Journal of Consulting and Clinical Psychology*, 68(2), 331-339.
- Addis, M. E., Wade, W. A., & Hatgis, C. (1999). Barriers to dissemination of evidence-based practices: Addressing practitioners' concerns about manual-based psychotherapies. *Clinical Psychology: Science and Practice*, 6, 430-441.
- Allen, J. G., Fonagy, P., & Bateman, A. (2008). *Mentalizing in clinical practice*. Washington, DC: American Psychiatric Publishing.
- American Psychiatric Association (APA). (2000). *Diagnostic and Statistical Manual of Mental Disorders-IV-TR*. Washington, DC: American Psychiatric Association.
- Barber, J. P., & Crits-Cristoph, P. (1996). Development of a Therapist Adherence/Competence rating scale for supportive-expressive dynamic psychotherapy: A preliminary report. *Psychotherapy Research*, 6(2), 81-94.
- Barber, J. P., Liese, B. S., & Abrams, M. J. (2003). Development of the Cognitive Therapy Adherence and Competence scale. *Psychotherapy Research* 13(2), 205-221
- Barber, J. B., Mercer, D., Krakauer, I., & Calvo, N. (1996). Development of an adherence/competence rating scale for individual drug counseling. *Drug and Alcohol Dependence* 43, 125-132.
- Bateman, A., & Fonagy, P. (2006). *Mentalization-based treatment for borderline personality disorder: A practical guide*. New York: Oxford University Press.

- Bateman, A. W., & Fonagy, P. (2004). *Psychotherapy for borderline personality disorder: Mentalization based treatment*. Oxford: Oxford University Press.
- Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). *Cognitive therapy of depression*. New York: Guilford.
- Berman, J. S., & Norton, N. C. (1985). Does professional training make a therapist more effective? *Psychological Bulletin*, 89, 401-407.
- Binder, J. L., Bongar, B., Messer, S., Strupp, H. H., Lee, S. S., & Peake, T. H. (1993). Recommendations for improving psychotherapy training based on experiences with manual-guided training and research: Epilogue. *Psychotherapy*, 30, 599-600.
- Caligor, E. (2005). Treatment manuals for long-term psychodynamic psychotherapy and psychoanalysis. *Clinical Neuroscience Research*, 4, 387-398.
- Carroll, K. M., et al. (2000). A general system for evaluating therapist adherence and competence in psychotherapy research in the addictions. *Drug and alcohol dependence*, 57, 225-238.
- Durlak, J. A. (1979). Obsolescence or lifelong education: A choice for the professional. *American Psychologist*, 27, 486-498.
- Elliott, R. (1998). Editor's introduction: A guide to the empirically supported treatments controversy. *Psychotherapy Research*, 8(2), 115-125.
- Finn, R.H., (1972). Effects of some variation in rating scale characteristics on the means and reliabilities of ratings. *Educational and Psychological Measurement*, 32, 255-265.
- Fonagy, P. (1999). Achieving evidence-based psychotherapy practice: A psychodynamic perspective on the general acceptance of treatment manuals. *Clinical Psychology: Science and Practice*, 6, 442-444.
- Fonagy, P., Gergely, G., Jurist, E., & Target, M. (2002). *Affect regulation, mentalization and the development of the self*. New York: Other Press.

- Forgatch, M. S., Patterson, G. R., & DeGarmo, D. S. (2005). Evaluating fidelity: predictive validity for a measure of competent adherence to the Oregon model of parent management training. *Behavior Therapy*, 36(1), 3-13.
- Foxhall, K. (2000). Research for the real world. *APA Monitor on Psychology*, 31, 28-36.
- Gresham, F. M., MacMillan, D. L., Beebe-Frankenberger, M. E., & Bocian, K. M. (2000). Treatment integrity in learning disabilities intervention research: do we really know how treatments are implemented? *Learning Disabilities Research & Practice*, 15(4), 198-205.
- Gullestad, S. E. (2001). Hva er evidensbasert psykoterapi? *Tidsskrift for norsk psykologforening*, 38(10), 942-951.
- Hatcher, R. L., & Barends, A. W. (2006). How a return to theory could help alliance research. *Psychotherapy: Theory, Research, Practice, Training*, 43(3), 292-299.
- Hattie, J. A., Sharpley, C. F., & Rogers, H. J. (1984). Comparative effectiveness of professional and paraprofessional helpers. *Psychological Bulletin*, 95, 534-541.
- Havik, O. E. (2000). Behandlingsveiledere og terapimanualer- bedre eller verre enn sitt rykte? In A. Holte, M.H. Rønnestad & G.H. Nielsen, (Red.), *Psykoterapi og psykoterapiveiledning. Teori, empiri og praksis*. Oslo: Gyldendal Akademisk.
- Henry, W. P. (1998). Science, politics, and the politics of science: The use and misuse of empirically validated treatment research. *Psychotherapy Research*, 8(2), 126-140.
- Hoffart, A. (1997). A schema model for examining the integrity of psychotherapy: A theoretical contribution. *Psychotherapy Research*, 7(2), 127-143.
- Kernberg O. F., Clarkin J. F., & Yeomans F. E. (2002). *A primer of transference focused psychotherapy for the borderline patient*. New York: Jason Aronson.

- Klerman, G. L., Weissman, M. M., Rounsaville, B. J., & Chevron, E. (1984). *Interpersonal psychotherapy of depression*. New York: Basic Books.
- Luborsky, L. (1984). *Principles of psychoanalytic psychotherapy: A manual for supportive-expressive treatment*. New York: Basic Books.
- Luborsky L., & Barber, J. P. (1993). Benefits of adherence to psychotherapy manuals and where to get them. I N. C. Miller, L. Luborsky, J. P. Barber & J. P Doherty (Eds.), *Psychodynamic treatment research. A handbook for clinical practice* (ss. 211-226). New York: Basic Books.
- Luborsky, L., & DeRubeis, R. J. (1984). The use of psychotherapy treatment manuals: a small revolution in psychotherapy research style. *Clinical Psychology Review*, 4(1), 5-14.
- Luborsky, L., McLellan, A. X, Woody, G. E., O'Brien, C. P., & Auerbach, A. (1985). Therapist success and its determinants. *Archives of General Psychiatry*, 42, 602-611.
- McGlinchey, J. B., & Dobson, K. S. (2003). Treatment integrity concerns in cognitive therapy for depression, *Journal of Cognitive Psychotherapy: An International Quarterly*, 17, 299-318.
- Norcross, J. C. (1999). Collegially validated limitations of empirically validated treatments. *Clinical Psychology: Science and Practice*, 6, 472-476.
- Nordahl, H. M., Nysæter, T. E., & Mikkelsen, B. (2006). Cognitive Therapy Adherence and Competence Scale. *Tidsskrift for Kognitiv Terapi*. Nr 3. s 14-29.
- O'Malley, S. S., Foley, S. H., Rounsaville, B. J., Watkins, J. T, Sotsky, S. M., Imber, S. D, & Elkin, I. (1988). Therapist competence and patient outcome in interpersonal psychotherapy of depression. *Journal of Consulting and Clinical Psychology*, 56, 496-501.
- Paivio, S. C., Holowaty, K. A. M., & Hall, I. (2004). The influence of therapist adherence and competence on client reprocessing of child abuse memories. *Psychotherapy: Theory, Research, Practice, Training*, 41, 56-68.

- Perepletchikova, F., & Kazdin, A. E. (2005). Treatment integrity and therapeutic change: Issues and research recommendations. *Clinical Psychology: Science and Practice, 12*(4), 365-383.
- Perepletchikova, F., Treat, T. A., & Kazdin, A.E. (2007). Treatment integrity in psychotherapy research: analysis of the studies and examination of the associated factors. *Journal of Consulting and Clinical Psychology, 75*(6), 829-841.
- Peterson, L., Homer, A., & Wonderlich, S. (1982). The integrity of independent variables in analysis. *Journal of Applied Behavior Analysis, 15*, 477-492.
- Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral Brain Science, 1*, 515-526.
- Sandell, R. (1985). Influence of supervision, therapist's competence, and patient's ego level on the effects of time-limited psychotherapy. *Psychotherapy and Psychosomatics, 44*, 103-109.
- Shaw, B. F., Elkin, I. E., Yamaguchi, J., Olmsted, M., Vallis, T. M., Dobson, K. S., et al. (1999). Therapist competence ratings in relation to clinical outcome in cognitive therapy of depression. *Journal of Consulting and Clinical Psychology, 67*, 837-846.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass Correlations: Uses in Assessing Rater Reliability. *Psychological Bulletin, 86*(2), 420-428.
- Svartberg, M., & Stiles, T. C. (1994). Therapeutic alliance, therapist competence, and client change in short-term anxiety-provoking psychotherapy. *Psychotherapy Research, 4*, 20-33.
- Waltz, J., Addis, M. E., Koerner, K., & Jacobson, N. S. (1993). Testing the integrity of a psychotherapy protocol: assessment of adherence and competence. *Journal of Consulting and Clinical Psychology, 61*(4), 620-630.
- Wampold, B. E. (2001). *The great psychotherapy debate: Models, methods and findings*. Mahwah: Lawrence Erlbaum Associates.

Wilkinson, L. A. (2006). Monitoring treatment integrity: an alternative to the "Consult and Hope" strategy in school-based behavioural consultation. *School Psychology International*, 27(4), 426-438.

Wolpe, J. (1969). *The practice of behavior therapy*. New York: Pergamon Press.

Wu, S. M., Whiteside, U., & Neighbors, C. (2007). Differences in inter-rater reliability and accuracy for a treatment adherence scale. *Cognitive Behaviour Therapy*, 36(4), 230-239.

Appendiks A

MBT Etterlevels- og Kvalitetsskala

Bedømmer ID _____

Timenummer _____

Timekode _____

Pasient ID _____

Dato _____

Terapeut ID _____

Introduksjon

MBT Etterlevels- og Kvalitetsskala er et verktøy som skal hjelpe veiledere og forskere til å oppnå et reliabelt og valid mål på hvor godt en terapeut utfører behandling i henhold til prinsippene for mentaliseringsbasert terapi. Under følger 23 ledd som alle skal skåres på en skala fra 0-6 der 0 er laveste og 6 er høyeste oppnåelige skåre. *Etterlevelse* refererer til hyppigheten/frekvensen av MBT-relevante behandlingsteknikker og *kvalitet* refererer til hvor godt utført disse er når de benyttes (f.eks. timing).

Nederst på hvert ledd er det listet opp fire referanser som forklarer hva som legges i kvalitetsbegrepet for det aktuelle ledd. **Hele skalaen fra 0-6 skal benyttes i vurderingen av hvert ledd slik at et av oddetallene 1, 3 og 5 skal oppføres dersom en vurdering faller mellom to referansepunkter i etterlevels- eller kvalitetsvurderingen.** For noen ledd er det også oppført "IA (ikke aktuelt)" under kvalitetsreferansen. Denne skåringen benyttes utelukkende når leddet ikke lar seg vurdere i henhold til kvalitet. Dette kan forekomme f.eks. dersom fenomenet leddet beskriver ikke forekommer, eller at intervensjonsformen av andre grunner er irrelevant i den aktuelle terapitimen.

Siste ledd (23) er en helhetlig vurdering av den (MBT) etterlevelse og kvalitet terapeuten utviste i den aktuelle terapitimen. Denne vurderingen gjøres skjønnsmessig på grunnlag av det totale inntrykket, og ikke som et gjennomsnitt av de øvrige 22 punktene.

Generelle psykoterapikriterier**1. Varm/genuin/kongruent**

_____	/	_____
Etterlevelse	/	Kvalitet

<i>Etterlevelse:</i>	0	2	4	6
	Ingen	Noe	Betraktelig	Fullstendig

Kvalitet:

- 0 Terapeuten fremstod kald, uengasjert, omsorgsløs.
- 2 Terapeuten fremstod litt reservert, distansert, inkongruent.
- 4 Terapeuten fremstod rimelig varm og genuin.
- 6 Terapeuten fremstod optimalt varm, genuin, omsorgsfull og kongruent.

2. Oppmerksomhet

_____	/	_____
Etterlevelse	/	Kvalitet

<i>Etterlevelse:</i>	0	2	4	6
	Ingen	Noe	Betraktelig	Fullstendig

Kvalitet:

- 0 Terapeuten ignorerte viktige åpenbare og subtile (f.eks. verbale og ikke verbale) signal.
- 2 Terapeuten var oppmerksom på noen tydelige signal, men overså subtile signal.
- 4 Terapeuten var oppmerksom på tydelige signal og noe oppmerksom på subtile signal.
- 6 Terapeuten var særdeles oppmerksom på viktige åpenbare og subtile signal.

3. Empati

_____/ _____
Etterlevelse / Kvalitet

Etterlevelse: 0 2 4 6
 Ingen Noe Betraktelig Fullstendig

Kvalitet:

- 0 Terapeuten viste dårlige evner til empati; virket ikke å forstå pasienten.
- 2 Terapeuten viste begrenset evner til empati; kunne reflektere og omskrive hva pasienten eksplisitt sa, men overså viktige subtile tema.
- 4 Terapeuten viste gode evner til empati. Virket generelt å skjønne pasientens perspektiv (fra både pasientens subtile og åpenbare signal).
- 6 Terapeuten viste særdeles gode evner til empati og forståelse; delte forståelse med pasient.

4. Samarbeid

_____/ _____
Etterlevelse / Kvalitet

Etterlevelse: 0 2 4 6
 Ingen Noe Betraktelig Fullstendig

Kvalitet:

- 0 Terapeuten monopoliserte timen eller overlot alt ansvar til pasienten.
- 2 Terapeuten forsøkte samarbeid men tok for lite eller for mye ansvar for å definere eller løse pasientens problemer.
- 4 Terapeuten var noe samarbeidende; delte ansvar med pasienten.
- 6 Terapeuten var svært samarbeidende; delte ansvar for å definere pasientens problem og mulige løsninger; fungerte som et "team".

MBT-kriterier

5. "Ikke-vitende" holdning

_____/ _____
Etterlevelse / Kvalitet

Etterlevelse: 0 2 4 6
 Ingen Noe Betraktelig Fullstendig

Kvalitet:

- 0 Terapeuten var passiv og gjorde ingen forsøk på å finne ut om pasientens mentale tilstand, hvordan han/hun tolker andre personer og interaksjonen mellom dem. Intervensjonene var farget av at terapeuten ignorerte det at mentale tilstander er 'ugjennomsiktige' og vanskelig å si noe sikkert om.
- 2 Terapeuten gjorde overfladiske forsøk på å finne ut om pasientens mentale tilstander, men terapeuten primære interesse virket å være hans egne underliggende motiver.
- 4 Terapeuten tok en oppriktig "ikke-vitende" holdning og forsøkte å "finne ut". Unngikk påstander av typen "hva du egentlig mener er..."
- 6 Terapeuten tok en oppriktig "ikke-vitende" holdning og forsøkte å "finne ut". Unngikk påstander av typen "hva du egentlig mener er..." Terapeuten utviste kreative måter å engasjere pasienten i en felles anstrengelse for å utforske både pasientens og andres sinn, og interaksjonen mellom dem.

6. Fremme utforskning og nysgjerrighet

_____/_____
Etterlevelse / Kvalitet

Etterlevelse: 0 2 4 6
 Ingen Noe Betraktelig Fullstendig

Kvalitet:

- 0 Terapeuten stilte ikke adekvate spørsmål egnet til å fremme utforskning og nysgjerrighet rundt egne og andres motiver.
- 2 Terapeuten gjorde forsøk på å stille spørsmål for å fremme utforskning og nysgjerrighet rundt egne og andres motiver, men dette ble gjort ufullstendig og overfladisk.
- 4 Terapeuten stilte adekvate (f.eks. åpne) spørsmål for å fremme utforskning og nysgjerrighet om egne og andres motiver.
- 6 Terapeuten gjorde en fremragende jobb med å stille adekvate spørsmål, der egne og andres motiver ble utforsket og fulgt opp med kreative kommentarer for å øke pasientens nysgjerrighet (*dette preget mesteparten av timen*).

7. Utfordrer uberettigede oppfatninger

_____/_____
Etterlevelse / Kvalitet

Etterlevelse: 0 2 4 6
 Ingen Noe Betraktelig Fullstendig

Kvalitet:

- IA Det forekom ingen sekvenser med uberettigede oppfatninger om selv og andre.
- 0 Terapeuten ignorerte uttalelser som demonstrerte uberettigede oppfatninger om selv og andre.
- 2 Terapeuten konfronterte og utfordret uberettigede oppfatninger om selv og andre, men ikke ledsaget av utforskning.
- 4 Terapeuten konfronterte og utfordret uberettigede oppfatninger om en selv og andre på en passende måte.
- 6 Terapeuten konfronterte og utfordret uberettigede oppfatninger om en selv og andre på en utmerket måte. Utfordringene ble alltid ledsaget av en utforskning av pasientens underliggende følelser og antagelser.

8. Positiv mentalisering

_____/_____
Etterlevelse / Kvalitet

Etterlevelse: 0 2 4 6
 Ingen Noe Betraktelig Fullstendig

Kvalitet:

- IA Det forekom ingen sekvenser der pasienten demonstrerte positiv mentalisering.
- 0 Terapeuten ignorerte pasientens positive mentalisering.
- 2 Terapeuten identifiserte positiv mentalisering på en overfladisk måte.
- 4 Terapeuten identifiserte og utforsket positiv mentalisering, dette ble ledsaget av skjønnsom bruk av ros.
- 6 Terapeuten brukte ros på en skjønnsom måte, identifiserte og utforsket positiv mentalisering. Terapeuten oppdaget når pasientens mentalisering har gitt "positive resultater" og undersøkte hvordan dette kan ha hjulpet pasienten ikke bare til å forstå hvordan han/hun følte det, men også til å anerkjenne den andre personens følelser.

9. Tilpasning til mentaliseringsnivå

_____/_____
Etterlevelse / Kvalitet

Etterlevelse: 0 2 4 6
 Ingen Noe Betraktelig Fullstendig

Kvalitet:

- 0 Terapeuten tilpasset ikke intervensjonene etter pasientens mentaliseringsnivå.
- 2 Terapeuten tok i noen grad hensyn til mentaliseringsnivå, men intervensjonene var for mye/lite utfordrende.
- 4 Terapeutens intervensjoner var godt tilpasset pasientens mentaliseringsnivå.
- 6 Terapeutens hadde hele tiden øye for pasientens mentaliseringsnivå og tilpasset intervensjonene på en utmerket måte.

10. Pretend modus

_____/_____
Etterlevelse / Kvalitet

Etterlevelse: 0 2 4 6
 Ingen Noe Betraktelig Fullstendig

Kvalitet:

- IA Ikke aktuelt, pretend modus forekom så å si ikke
- 0 Det forekom sekvenser preget av pretend modus, men terapeuten overså dem og intervensjonene ikke.
- 2 Terapeuten identifiserte og omtalte pretend modus, men hadde ingen klare intervensjoner for å forsøke å gjenopprette mentaliseringsevnen.
- 4 Terapeuten identifiserte pretend modus og intervensjonene for å forsøke å gjenopprette mentaliseringsevnen.
- 6 Terapeuten identifiserte pretend modus og utviste stor og hensiktsmessig kreativitet i forsøk på å gjenopprette mentaliseringsevnen

11. Psykisk ekvivalens

_____/_____
Etterlevelse / Kvalitet

Etterlevelse: 0 2 4 6
 Ingen Noe Betraktelig Fullstendig

Kvalitet:

- IA Det var ingen vesentlige hendelser preget av psykisk ekvivalens
- 0 Terapeuten ignorerte sekvenser med psykisk ekvivalens hos pasienten.
- 2 Terapeuten identifiserte og navnga aspekter av psykisk ekvivalens, men viste ingen klare intervensjoner for å gjenopprette evnen til mentalisering.
- 4 Terapeuten identifiserte psykisk ekvivalens og intervensjonene for å gjenopprette mentaliseringsevnen.
- 6 Terapeuten identifiserte psykisk ekvivalens og viste stor kreativitet i sine forsøk på å gjenopprette mentaliseringsevnen.

12. Optimalt spenningsnivå

_____/_____
Etterlevelse / Kvalitet

Etterlevelse: 0 2 4 6
 Ingen Noe Betraktelig Fullstendig

Kvalitet:

- 0 Terapeuten var passiv og så ikke ut til å bry seg om å regulere spenningsnivået.
- 2 Terapeuten var oppmerksom på spenningsnivået, men intervensjonene ikke.
- 4 Terapeuten forsøkte å holde spenningsnivået på et optimalt nivå.
- 6 Terapeuten var hele tiden aktiv med henhold til å regulere spenningsnivået, og gjorde dette på en utmerket måte.

13. Korte tilpassede intervensjoner

_____/_____
Etterlevelse / Kvalitet

Etterlevelse: 0 2 4 6
 Ingen Noe Betraktelig Fullstendig

Kvalitet:

- 0 Terapeutens intervensjoner var enten for kompliserte eller for lange, og bommet (fullstendig) på pasientens aktuelle mentaliseringsnivå.
- 2 Terapeutens intervensjoner var enten for kompliserte eller for lange, og er lite tilpasset pasientens aktuelle mentaliseringsnivå.
- 4 Terapeutens intervensjoner var korte, og rimelig tilpasset pasientens aktuelle mentaliseringsnivå.
- 6 Terapeutens intervensjoner var korte, og svært godt tilpasset pasientens aktuelle mentaliseringsnivå.

14. Bruk av metaforer, analogier og symboler

_____/_____
Etterlevelse / Kvalitet

Etterlevelse: 0 2 4 6
 Ingen Noe Betraktelig Fullstendig

Kvalitet:

- IA Det forekom ingen bruk av metaforer, analogier eller symboler
- 0 Terapeuten gjorde en rekke intervensjoner der innholdet gjennomgående var preget av komplisert bruk av metaforer, analogier og symboler.
- 2 Terapeuten gjorde noen intervensjoner som var preget av komplisert bruk av metaforer, analogier og symboler.
- 4 Terapeutens intervensjoner inneholdt sjeldent bruk av metaforer, analogier og symboler, men da det forekom var det i hovedsak tydelig og stort sett lettfattelig.
- 6 Terapeutens intervensjoner inneholdt svært sjeldent bruk av metaforer, analogier og symboler, men da det ble gjort var det friskt og med stor kreativitet.

15. Affektfokus

_____/_____
Etterlevelse / Kvalitet

Etterlevelse: 0 2 4 6
 Ingen Noe Betraktelig Fullstendig

Kvalitet:

- 0 Terapeuten ignorerte alle signaler om eller referanser til affekt, både i forbindelse med hendelser utenfor timen og i timen her og nå.
- 2 Terapeuten var oppmerksom på mulige affektive reaksjoner, men disse ble bare overflatisk referert til eller navngitt av terapeuten.
- 4 Terapeutens intervensjoner var hovedsakelig fokusert på affekt snarere enn innholdet i pasientens narrativ. Oppmerksomheten var rettet mot følelser slik de kom til uttrykk i øyeblikket, og særlig følelser mellom pasient og terapeut i timen.
- 6 Terapeuten forsøkte på en utmerket måte å opprettholde et affektfokus og stimulere til pasientens utforskning av mange aspekter ved egne affekter, eller blandede affekter, både mht hendelser utenfor timen og slik de utspilte seg i øyeblikket i forhold til terapeuten.

16. Affekt og interpersonlige hendelser

_____/_____
Etterlevelse / Kvalitet

Etterlevelse: 0 2 4 6
 Ingen Noe Betraktelig Fullstendig

Kvalitet:

- 0 Terapeuten ignorerte enhver sammenheng mellom affekter og nylige interpersonlige hendelser.
- 2 Terapeuten refererte kun overflatisk til mulige sammenhenger mellom aktuelle affekter og nylige interpersonlige hendelser.
- 4 Terapeuten koplet aktuelle affekter til nylige interpersonlige hendelser ved aktivt å spørre etter pasientens affekter i forbindelse med slike hendelser.
- 6 Terapeuten forsøkte på en utmerket måte å stimulere til utforskning av sammenhengen mellom affekter og nylige interpersonlige hendelser, med særlig fokus på pasientens forestillinger og fortolkninger av egne og andres reaksjoner.

17. Stimulere mentalisering

_____/_____
Etterlevelse / Kvalitet

Etterlevelse: 0 2 4 6
 Ingen Noe Betraktelig Fullstendig

Kvalitet:

- 0 Intervensjonene var i liten grad rettet mot opplevelser av selv og andre og la ikke vekt på å utforske dette.
- 2 Intervensjonene var av og til rettet mot opplevelser av selv og andre, men det ble lagt liten vekt på å utforske dette.
- 4 Intervensjonene var siktet mot å stimulere mentalisering av opplevelser av selv og andre, og mot å opprettholde prosessen
- 6 Intervensjonene var i utpreget grad, og på kreativt vis, siktet mot å stimulere mentalisering av opplevelser av selv og andre, og mot å opprettholde prosessen

18. Innhold og prosess

_____/_____
Etterlevelse / Kvalitet

Etterlevelse: 0 2 4 6
 Ingen Noe Betraktelig Fullstendig

Kvalitet:

- 0 Terapeuten hadde et konsekvent fokus på innhold uten noen forsøk på å utforske mentale prosesser, eller benyttet innhold utelukkende som middel for å generere innsikt for pasienten.
- 2 Intervensjonene var overveiende fokusert på innhold, mens fokus på å utvikle prosessen kom klart i annen rekke.
- 4 Terapeuten benyttet seg av innhold på en passende måte med det formål å utforske mentale prosesser.
- 6 Terapeuten benyttet innholdet i pasientens fortellinger på en utmerket og kreativ måte for å utforske mentale prosesser.

19. Overføring

_____/_____
Etterlevelse / Kvalitet

Etterlevelse: 0 2 4 6
 Ingen Noe Betraktelig Fullstendig

Kvalitet:

IA Ikke aktuelt.

0 Terapeuten tolket på en rigid måte overføring som enkle gjentakelser av relasjoner fra fortiden, eller ignorerte åpenbare overføringsfenomener.

2 Terapeuten tolket overføringsfenomener uten å stille spørsmål om alternativ forståelse av pasientens reaksjoner i forhold til terapeuten.

4 Terapeuten påpekte overføringsfenomener, og oppfordret til refleksjon over alternative perspektiver på pasientens reaksjoner og oppfatninger av terapeuten.

6 Terapeuten påpekte overføringsfenomener, og oppfordret til refleksjon over alternative perspektiver på pasientens reaksjoner og oppfatninger av terapeuten. Terapeuten bidro videre til å skape refleksjon over terapeutenes forståelse av overføringsfenomenene.

20. Motoverføring

_____/_____
Etterlevelse / Kvalitet

Etterlevelse: 0 2 4 6
 Ingen Noe Betraktelig Fullstendig

Kvalitet:

IA Det forekom ingen klare motoverføringsresponsers i den aktuelle timen.

0 Terapeuten registrerte ikke egne motoverføringsresponsers overfor pasienten.

2 Terapeuten benyttet seg av sine motoverføringsresponsers, men plasserte ansvaret for disse hos pasienten i et forsøk på å øke innsikt i pasients følelser.

4 Terapeuten utforsket åpent og adekvat visse motoverføringsresponsers overfor pasienten og identifiserer disse i utgangspunktet som egne reaksjoner. Drøfter pasientens rolle i følelsene som ble vekket kun når dette er anerkjent.

6 Terapeuten utforsket åpent og adekvat visse motoverføringsresponsers overfor pasienten og identifiserte disse i utgangspunktet som egne reaksjoner. Drøftet pasientens rolle i følelsene som ble vekket kun når dette var anerkjent. Han/hun gjorde dette på en kreativ måte og delte ansvaret for utviklingen i interaksjonen.

21. Alliansebrudd

_____/_____
Etterlevelse / Kvalitet

Etterlevelse: 0 2 4 6
 Ingen Noe Betraktelig Fullstendig

Kvalitet:

IA Ikke aktuelt. Det forekom ikke noe alliansebrudd.

0 Det forekom klart alliansebrudd, men terapeuten valgte å overse det eller reagerte med å bebreide eller straffe pasienten.

2 Det forekom alliansebrudd som ble omtalt, men ganske overflatisk.

4 Alliansebrudd ble drøftet og ledsaget av en klargjøring av pasientens og terapeutenes respektive roller.

6 Alliansebrudd ble drøftet nøye og ledsaget av en inngående klargjøring av pasientens og terapeutenes respektive roller.

22. Sjekke egen forståelse og korrigere misforståelse

_____/_____
Etterlevelse / Kvalitet

Etterlevelse: 0 2 4 6
 Ingen Noe Betraktelig Fullstendig

Kvalitet:

- 0 Terapeuten gjorde ingen forsøk på å sjekke ut om egen forståelse av pasientens sinnstilstand korresponderte med pasientens egen forståelse.
- 2 Terapeuten gjorde bare overflatiske forsøk på å sjekke ut sin egen forståelse av pasientens sinnstilstand.
- 4 Terapeuten sjekket ut sin egen forståelse av pasientens sinnstilstand, lot sin egen forståelse påvirkes av dette og innrømmet åpent egne misforståelser dersom dette forekom.
- 6 Terapeuten sjekket ut med stor sensitivitet sin egen forståelse av pasientens sinnstilstand, demonstrerte tydelig at slik utveksling har reell betydning for forståelsen av pasientens sinn og håndterte eventuelle egne misforståelser på en måte som forsterket den pågående mentaliseringsprosessen.

23. Helhetlig vurdering av MBT etterlevelse og kvalitet

_____/_____
Etterlevelse / Kvalitet

Etterlevelse: 0 2 4 6
 Ingen Noe Betraktelig Fullstendig

Kvalitet: 0 2 4 6
 Svært dårlig Dårlig God Svært god

Oppsummering

#	ETTERLEVELSE	KVALITET	#	ETTERLEVELSE	KVALITET
1	0-1-2-3-4-5-6	0-1-2-3-4-5-6	13	0-1-2-3-4-5-6	0-1-2-3-4-5-6
2	0-1-2-3-4-5-6	0-1-2-3-4-5-6	14	0-1-2-3-4-5-6	0-1-2-3-4-5-6
3	0-1-2-3-4-5-6	0-1-2-3-4-5-6	15	0-1-2-3-4-5-6	0-1-2-3-4-5-6
4	0-1-2-3-4-5-6	0-1-2-3-4-5-6	16	0-1-2-3-4-5-6	0-1-2-3-4-5-6
5	0-1-2-3-4-5-6	0-1-2-3-4-5-6	17	0-1-2-3-4-5-6	0-1-2-3-4-5-6
6	0-1-2-3-4-5-6	0-1-2-3-4-5-6	18	0-1-2-3-4-5-6	0-1-2-3-4-5-6
7	0-1-2-3-4-5-6	0-1-2-3-4-5-6	19	0-1-2-3-4-5-6	0-1-2-3-4-5-6
8	0-1-2-3-4-5-6	0-1-2-3-4-5-6	20	0-1-2-3-4-5-6	0-1-2-3-4-5-6
9	0-1-2-3-4-5-6	0-1-2-3-4-5-6	21	0-1-2-3-4-5-6	0-1-2-3-4-5-6
10	0-1-2-3-4-5-6	0-1-2-3-4-5-6	22	0-1-2-3-4-5-6	0-1-2-3-4-5-6
11	0-1-2-3-4-5-6	0-1-2-3-4-5-6	23	0-1-2-3-4-5-6	0-1-2-3-4-5-6
12	0-1-2-3-4-5-6	0-1-2-3-4-5-6			

Appendiks B

MBT adherence skala – kortversjon (Bateman 2007, oversatt av Karterud 2008)

Kortversjonen kan brukes for å fremme refleksjon etter hver time. Formålet er å stimulere terapeutene til å mentalisere om sine roller i timen og sjekke hvor godt de holder seg til modellen.

Gi en skåring av i hvilken grad terapeutens intervensjoner reflekterer optimale MBT prinsipper, i lys av arbeidet med denne pasienten i denne timen. Bruk følgende skala

1. Anvender ikke MBT prinsipper
2. Forsøker å anvende MBT prinsipper men avviker fra modellen
3. Bruker MBT prinsipper adekvat
4. Svært god anvendelse av MBT prinsipper

1. Lar seg ikke måle

Terapeutisk holdning og intervensjoner:

1. Terapeuten inntar en oppriktig ”ikke-vitende” holdning og forsøker å ”finne ut”. Unngår påstander av typen ”hva du egentlig mener er ...”
2. Terapeuten stiller adekvate spørsmål for å fremme utforskning og fremmer nysjerrighet om egne og andres motiver.
3. Passende konfrontasjon og utfordring av uberettigede oppfatninger om en selv og andre
4. Intervensjonene er tidsmessig (”timing”) avpasset pasientens mentaliseringskapasitet til enhver tid.
5. Identifiserer pretend mode hos pasienten og intervensjoner for å gjenopprette mentaliseringsevnen
6. Identifiserer psykisk ekvivalens hos pasienten og intervensjoner for å gjenopprette mentaliseringsevnen
7. Terapeuten tar en aktiv rolle med hensyn til å holde spenningsnivået på et optimalt nivå (ikke for høyt slik at pasienten mister sin evne til å mentalisere; ikke for lavt slik at timen blir meningsløs).
8. Intervensjonene er enkle og korte, innenfor pasientens aktuelle evne til å mentalisere. Terapeuten avstår fra å bruke metaforer, analogier og symboler
9. Intervensjonene er i hovedsak affekt-fokuserte, og mindre grad rettet mot adferd. Spesielt rettet mot aktuelle affekter slik de skjer her og nå.
10. Kobler affekter til umiddelbare eller nylige interpersonlige hendelser

11. Intervensjonene har som formål å stimulere mentalisering av opplevelser av selv og andre i en pågående prosess, og er mindre rettet mot innhold og tolkning av innhold for å fremme innsikt.
12. Identifiserer, påskjønner og utforsker positiv mentalisering
13. Drøfter/forhandler om alliansebrudd ved å klargjøre pasientens og terapeutens respektive roller
14. Overføring forklares ikke som enkle gjentakelser av fortiden
15. Overføring brukes for å demonstrere alternative perspektiver mellom selv og andre

Hvor godt arbeidet terapeuten etter modellen? Eller: Hvor mye MBT kompetanse totalt sett viste terapeuten i timen?

1

2

3

4

Svært lite

Svært mye

Hva er det terapeuten strever med i sin aktuelle terapeutiske stil med hensyn til å kunne anvende modellen?