Psychotherapy Research
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tpsr20

The MBT Adherence and Competence Scale (MBT-ACS): Development, structure and reliability

Sigmund Karterud [a c], Geir Pedersen [a], Magnus Engen [b], Merete Selsbakk Johansen [a], Paul Niklas Johansson [a], Christian Schlüter [a], Øyvind Urnes [a], Theresa Wilberg [c] & Anthony W. Bateman [d]

[a] Department of Personality Psychiatry, Oslo University Hospital, Ullevaal, PB 4956 Nydalen, 0424, Oslo, Norway

[b] Josefinegate District Psychiatric Centre, Oslo University Hospital, Josefinegaten 30, Oslo, 0351, Norway

[c] University of Oslo, Institute of Clinical Medicine, PB 1039 Blindern, Oslo, 0315, Norway

[d] Barnet, Enfield and Haringey Mental Health NHS Trust, Halliwick Unit, St Ann's Hospital, London, UK

Version of record first published: 24 Aug 2012.

To cite this article: Sigmund Karterud, Geir Pedersen, Magnus Engen, Merete Selsbakk Johansen, Paul Niklas Johansson, Christian Schlüter, Øyvind Urnes, Theresa Wilberg & Anthony W. Bateman (2012): The MBT Adherence and Competence Scale (MBT-ACS): Development, structure and reliability, Psychotherapy Research, DOI:10.1080/10503307.2012.708795

To link to this article: http://dx.doi.org/10.1080/10503307.2012.708795

Psychotherapy Research,
2012, 1–13, iFirst article

The MBT Adherence and Competence Scale (MBT-ACS): Development, structure and reliability

SIGMUND KARTERUD[1,3]*, GEIR PEDERSEN[1], MAGNUS ENGEN[2], MERETE SELSBAKK JOHANSEN[1], PAUL NIKLAS JOHANSSON[1], CHRISTIAN SCHLÜTER[1], ØYVIND URNES[1], THERESA WILBERG[3], & ANTHONY W. BATEMAN[4]

[1]*Department of Personality Psychiatry, Oslo University Hospital, Ullevaal, PB 4956 Nydalen, 0424 Oslo, Norway;* [2]*Josefinegate District Psychiatric Centre, Oslo University Hospital, Josefinegaten 30, 0351 Oslo, Norway;* [3]*University of Oslo, Institute of Clinical Medicine, PB 1039 Blindern, 0315 Oslo, Norway &* [4]*Barnet, Enfield and Haringey Mental Health NHS Trust, Halliwick Unit, St Ann's Hospital, London, UK*

(Received 24 November 2011; revised 6 June 2012; accepted 28 June 2012)

Abstract
The properties of the 17-item Mentalization-Based Treatment Adherence and Competence Scale (MBT-ACS) were investigated in a reliability study in which 18 psychotherapy sessions, comprising two sessions by nine different therapists, were rated by seven different raters. The overall reliabilities for adherence and competence for seven raters were high, .84 and .88 respectively. The level of reliability declined by number of raters but was still acceptable for two raters (.60 and .68). The reliabilities for the various items differed. The MBT-ACS was found to be an appropriate rating measure for treatment fidelity and useful for the purposes of quality control and supervision. The reliability may be enhanced by redefining some items and reducing their numbers.

Keywords: mentalization; mentalization-based treatment; borderline personality disorder; psychotherapy research; reliability; Generalizability Theory

Introduction

Mentalization-based treatment (MBT) is an evidence-based treatment for borderline personality disorder (BPD). Two RCTs have demonstrated a superior effect (Bateman & Fonagy, 2001, 2009) and MBT has some recognition as a treatment to be implemented in the British National Mental Health Services (NICE, 2009). MBT is founded on the theories of mentalization, personality disorders and principles of psychodynamic treatment (Bateman & Fonagy, 2011; Bouchard & Lecours, 2008; Fonagy, Gergely, Jurist, & Target, 2002). There is a substantial link to the concept of metacognition, i.e., the capacity to reflect upon one's own and other's mind, in addition to mastery, which is the ability to use this knowledge to form adaptive problem-solving strategies (Lysaker, Gumley, & Dimaggio, 2011). It is a psychodynamic approach in the sense that the main instrument of change is believed to be the intersubjective transactions taking place between therapist and patient. Bateman and Fonagy (2006) suggest that all psychological therapies exert their effect through their impact on the patient's ability to mentalize. It is possible to modify different psychodynamic practices in the direction of MBT by doing more or less of mentalization-enhancing interventions. MBT cultivates this focus.

Treatment guidelines for MBT are described by Bateman and Fonagy (2004, 2006). The key element is the uncompromising focus on mentalizing within and without the therapy. Therapists focus on the patient's subjective sense of self. To do so they need to (a) identify and work with the patient's mentalizing capacities; (b) represent internal states in themselves and their patient; (c) focus on these internal states; and (d) sustain this in the face of constant challenges by the patient over a significant period of time. In order to achieve this level of focus, mentalizing techniques need to be (a) offered in the context of an attachment relationship; (b) consistently applied over time; and (c) used to reinforce the therapist's capacity to retain mental closeness

Correspondence concerning this article should be addressed to Sigmund Karterud, Oslo University Hospital, Department of Personality Psychiatry, PB 4956 Nydalen, 0424 Oslo, Norway. Email: sigmund.karterud@medisin.uio.no

ISSN 1050-3307 print/ISSN 1468-4381 online
http://dx.doi.org/10.1080/10503307.2012.708795

with the patient. Particular techniques are considered to enhance mentalizing more than others and the guidelines describe a number of levels of intervention—empathic and validating, exploratory, affect focus, and the patient–therapy relationship.

One shortcoming with the two RCTs on MBT is that they lacked formal adherence (and competence) ratings of the therapists. This limits a claim for the MBT technique being the causative agent of the superior effects that were demonstrated. This is the main reason why we developed the MBT Adherence and Competence Scale (MBT-ACS). We wanted to measure treatment integrity, i.e., the degree of consistency between the therapist's actual performance and the underlying theory, ideals, intentions and norms which are specified in the therapy manual (Perepletchikova, Treat, & Kazdin, 2007). In psychotherapy research, there is an increasing demand for documentation of treatment integrity. Earlier research is open to criticism in this regard. In a review of randomized psychotherapy studies, Perepletchikova and coworkers (2007) found that only 4% of studies satisfied their criteria for documentation of treatment integrity. Other reasons were that MBT-ACS might be beneficial for training and supervisory purposes and as a measure for monitoring the quality of treatment.

The MBT Adherence and Competence Scale, described in this study, was developed as part of a larger manual (Karterud & Bateman, 2010) that followed the recommendations of Luborsky and Barber (1993), containing: (1) a presentation of the main principles underlying the therapeutic techniques; (2) concrete examples of all techniques being described; and (3) scales and instruments which can assess the skills of the therapists for this particular treatment model. In this article, we limit ourselves to a description of the development, content and reliability of the scale.

Adherence and Competence

The concept of treatment integrity contains two components: (1) treatment adherence, i.e., the extent that the therapist uses prescribed techniques and avoids proscribed techniques, and (2) the therapist's competence, i.e., the level of skill and quality in his/her performance. The literature on assessment scales reveals an ongoing discussion about adherence versus competence. Adherence is usually easier to measure because it usually involves a quantitative judgment on a scale ranging from "no adherence" (absence of the intervention) or "some adherence/interventions" to "considerable" and "extensive" (frequency of the intervention in question). In contrast, competence is often judged as the level of consistency with short (qualitative) descriptions and, in psychotherapy, it will concern qualities of the discourse itself.

Several studies show that raters struggle to distinguish between adherence and competence (Perepletchikova et al., 2007). Frequent reasons are weak definitions in the manuals, comprehension problems on the part of the raters and lack of training. The ability to differentiate between adherence and competence seems to vary depending on the stage of the therapeutic process. Barber, Liese and Abrams (2003) found, for example, that raters had a tendency to interpret an intervention as being of a higher quality the more often it was used.

Such problems undermine the reliability of both adherence and competence measures, as well as the relationship between these variables at various stages. McGlinchey and Dobson (2003) have pointed out that there is a definition-contingent relationship between the two concepts: Competence presupposes adherence, but adherence does not necessarily presuppose competence. A moderate correlation, however, is to be expected between these two phenomena. A very low correlation or none at all is not a good sign. If this is the case, one should look closer at the definitions (validity). A very high correlation may also be a problem, because it may indicate that the two concepts are too similar and have not been clearly differentiated by the definitions. Most studies find moderate to high correlations. For example, Carroll et al. (2000) found that adherence and competence for different items of the "Yale Adherence and Competence Scale" (YACS-II) correlated somewhere between $r = .27$ and $r = .54$.

The architects of therapies would like adherence, competence and outcome to correlate in such a way that the more a therapist complies with the guidelines and the more he/she is able to practice the method in a qualitatively proficient manner, the better the outcome. Ideally, one would also expect competence to complement adherence: Given adequate adherence, the way in which the therapy is practiced should have a positive effect on the outcome. However, there is no consensus on this. Wampold (2001) states, for example, that there is no evidence for the claim of a positive relationship between adherence and outcome of psychotherapy. This touches on the issue of the significance of "unspecific" versus "specific" factors on treatment outcome. If it is generally the case that specific factors play a subordinate role in the outcome of psychotherapy, one should not expect that adherence to those techniques would play any significant role either. More recent research, however, seems to indicate that adherence plays a role in the treatment of "difficult" patients. Giessen-Bloo et al. (2006)

found a positive correlation between adherence and outcome in a long-term psychotherapy trial for borderline patients, and Høglend et al. (2006) showed that adherence played a role when differentiating between important patient characteristics. Furthermore, several other studies have found that quality measures are positively related to the outcome of psychotherapy (Barber, Crits-Christoph, & Luborsky, 1996; Luborsky, McLellan, Woody, O'Brien, & Auerbach, 1985; O'Malley, Bachman, & Johnston, 1988). For these reasons, we chose to design a rating scale containing both adherence and competence.

The Design of the MBT Adherence and Competence Scale

While working with the volume "Mentalization-based treatment for the borderline patient – a practical guide" (Bateman & Fonagy, 2006), Bateman developed an "MBT adherence scale" consisting of 15 items. In collaboration with Bateman, this scale was translated into Norwegian, tested and further developed by a research group at the Department of Personality Psychiatry, Oslo University Hospital (Engen, 2009). The developmental process and most important changes are described next.

International Consensus

The practice of psychotherapy is influenced by cultural factors (language, history, customs, norms, value systems, etc.). To avoid local/national idiosyncrasies, the developmental work with the manual (item selection, item definitions, quality descriptors, rating procedures, etc.) was performed in a dialog with the research group (including Anthony Bateman), and clinicians and researchers from Sweden (Stockholm) and Denmark (Copenhagen, Roskilde and Aarhus). This international team rated and discussed video recordings of therapy sessions from England, Norway, Sweden and Denmark and finally reached a consensus on the design and details of the manual and the scale.

Defining, Selecting, and Testing the Items

Since the original 15 items were chosen by the originators of MBT, the items have a high degree of content validity, cover a wide range of MBT interventions, and most can easily be identified in therapy sessions conducted in accordance with MBT guidelines. However, empirical studies revealed a need for substantial clarifications and redefinitions.

The original 15-item scale contained no items covering general factors of psychotherapy. Since MBT is a specialized form of psychodynamic psychotherapy, the research group felt that good MBT had to rest on a foundation of generally sound psychotherapeutic principles, and that it therefore was relevant to include some general factors known to be of importance for all types of psychotherapy. Four such items were selected from the Norwegian version of "Cognitive Therapy Adherence and Competence Scale" (CT-ACS) (Nordahl, Nysæter, & Mikkelsen, 2006): (1) Warm/genuine/congruent, (2) Attention, (3) Empathy and (4) Cooperation.

Furthermore, we redefined some items on transference, countertransference, understanding versus misunderstanding and validation of feelings.

Altogether we included 21 items which were explored in several video recordings before we conducted a more formal preliminary reliability test that included six raters who rated independently six different video recorded treatment sessions (with different therapists) from a MBT program (Engen, 2009). The ICC 2.1 values for the total scores were .58 for adherence and .62 for competence. However, the variations were large among the different items, and some items were hardly identified.

It turned out that the reliability of the four general factors were quite low (range .13–.59). Through discussion, we found that even though these items (which were identical to the items in the CT-ACS) were formulated in a non-specific cognitive language, they proved to communicate nuances that were somewhat in contradiction to MBT. One example of this is competence level 4 for the item of empathy: "The therapist exhibited good capacity for empathy. Seemed to understand patient's perspective (based on both subtle and obvious signs from patient)." Such a formulation suggests that the better the therapist understands the patient's perspective, the better is his/her competence. However, MBT emphasizes that the therapist should assume a not-knowing and inquisitive stance, i.e., try to explore, *together with the patient*, the patient's perspective, rather than investing effort in an understanding by him/herself. Thus, being "very empathetic" is not a main objective in MBT. The research group found that rating these items proved to be difficult from a MBT perspective, and the three items concerning attention, empathy and cooperation were eventually deleted. The item "warm/genuine/congruent" was retained.

Furthermore, we made some other adjustments and identified 17 items which are displayed in the Appendix and described by their quality rating of 4 (good enough). The complete worksheet can be downloaded from www.mentalisering.no.

A manual was written that described (1) the essence and general principles of MBT, and (2) the essence of each item with detailed indicators for quality ratings and examples of the intervention to facilitate adherence ratings (Karterud & Bateman, 2010). In order to emphasize that it is the therapist's activity that is being judged, the manual emphasized that qualifying statements for the items should be of the type "to what degree did therapist X do...?" with respect to adherence, and with respect to quality on the format: "the therapist's interventions were...", or "the therapist did...", etc. The items were defined using specific clinical examples based on observable behavior whenever possible. For example, "The therapist connects emotions and feelings to recent or immediate interpersonal events" (see appendix).

The 17 items were considered sufficient to cover most of the variations of MBT when used with a wide range of patients, in different treatment contexts, and at varied therapeutic stages. They were considered to be a combination of essential *and* unique items, and essential but not unique items (Waltz, Addis, Koerner, & Jacobson, 1993). For example, the item "exploration, curiosity and a not-knowing stance" is essential, but not unique, while the item on "psychic equivalence" is essential and unique. In practice, the difference between essential and unique proved to be vague. Many psychotherapies attempt to promote exploration and curiosity, challenge the patients, focus on affects, link affects to interpersonal events, etc. The manual states that *the unique aspects of MBT lies less in each item than in the overall "package" of items, i.e., the total design.* While many therapies may advocate interventions that "promote exploration and curiosity" on the part of the patient, the unique feature of MBT is the consistent emphasis on an exploration of one's own *and* others' motives. This is not something that takes place sporadically, by chance or on certain occasions. It is a dominating characteristic in terms of frequency, scope and quality of the therapeutic dialog as a whole.

Quantifying the Scale

The 17 items are rated in accordance with a 1–7 Likert scale for adherence. Adherence primarily involves frequency and extensiveness. Frequency is simply the number of times the therapist carries out an intervention, and extensiveness is the time and attention that the therapist gives to the intervention. The range is from "not at all" (score 1) to "extensively" (score 7). For two of the items ("engagement, interest and warmth" as well as "adaptation to mentalizing capacity") a frequency assessment was deemed to be irrelevant.

All items are also scored on a 0–7 Likert scale for competence, in which "0" signifies "Not applicable (the intervention was not observed)," "1" is very poor and "7" is excellent. In the event of no occurrence, one should assume that there would be no need for any competence rating. However, things are more complicated. The rater may observe unequivocal signs of a phenomenon that the manual instructs the therapist to address; e.g., clear signs of pseudomentalizing discourse. If the therapist does not address this, the adherence rating is 1 (no occurrence). However, the quality with respect to this item will then be low since the phenomenon is ignored, and this should be noted with a low competence rating (e.g., 2) (cf., The Penn-adherence scale for supportive-expressive therapy; Barber, Crits-Christoph, & Luborsky, 1996). The anchoring points in the 1–7 Likert scales were adopted from the "Yale Adherence and Competence Scale" (YACS-II) (Carroll et al., 2000).

On the scale worksheet, the notation 4 contains a brief description of what is deemed adequate (good enough) competence (see Appendix). The rater's starting point should be at "4 = adequate." The basic assumption is that the therapist is average ("good enough"). One should therefore be aware of deviations in a positive or a negative direction from this starting point. The manuals contain examples to guide raters when they are determining the degree of deviation from a "good enough" practice.

After each item has been assessed, the rater decides on an *overall score* for the specific therapy session, for both adherence and competence. A global assessment is made, not on the basis of an arithmetic average of the 17 items, but on the basis of an overall clinical judgement, with particular emphasis on the following items: (2) Exploration, curiosity and a not-knowing stance; (6) Stimulating mentalization through the process; (10) Affect focus; and (11) Affect and interpersonal events. These four items are, from a clinical point of view, considered somewhat more important than the others.

An overall score of 4 is defined as an adequate performance in terms of both adherence and competence. For adherence, this means that most individual items have received a minimum score of 4, indicating that the general impression of the rater is that a broad range of MBT-type interventions has clearly been demonstrated. A score of 4 indicates that the therapist has adequate knowledge about MBT and that he/she is able to implement the recommended interventions in practice to a reasonable degree. Similarly, the competence of the performance should also have been demonstrated

sufficiently. Most individual items should have been scored at least 4, and the general impression should be that the therapist masters the technique to a reasonable degree. This means that there is a relatively good "flow" during the session; that the sequences fit together; that word choice and timing are adequate; that the therapist tries out new MBT strategies if/when some interventions fail; that the entire session is guided by an attempt to engage the patient in a mentalizing dialogue, and that there are no extended sequences featuring other types of techniques (supportive therapy, problem solving, guidance, etc.). A low competence rating means that the therapist did other things than those which are prescribed for MBT, that he/she delivered MBT interventions in an inflexible or clumsy way or that he/she failed to follow up interventions adequately.

Rating Procedures

The manual contains guidelines on a worksheet notation system and what counts as a specific intervention, and determines how many interventions qualify for the different adherence ratings, etc.

Competence of Raters for the Reliability Test

There is not necessarily a strong association between being an expert clinician (therapist) and a good rater (Barber et al., 2003). Both skills are important. Clinical experts need training on *using* the assessment scale. On the other hand, it is hard to become a good rater without extensive clinical experience with this type of treatment.

The seven raters for this reliability study participated in the research group that developed the scale from its beginning. During a 1-year period, they rated around 15 sessions. One was a psychology student who wrote his MA thesis on the first 21-item version of the scale. The others were clinicians and researchers who had been trained in MBT by attending seminars with Anthony Bateman and who were involved in the implementation of the MBT program at the Department of Personality Psychiatry at Oslo University Hospital. They were experienced psychotherapists and worked as clinicians in the MBT program. Four were psychiatrists and two were clinical psychologists. Three of them had a PhD research degree.

During the preliminary reliability test of the 21-item version of the scale, the six raters met and discussed their ratings after each of the six sessions. As can be seen from Figure 1, there was a significant learning effect or improvement in agreement, from the first to the last session.

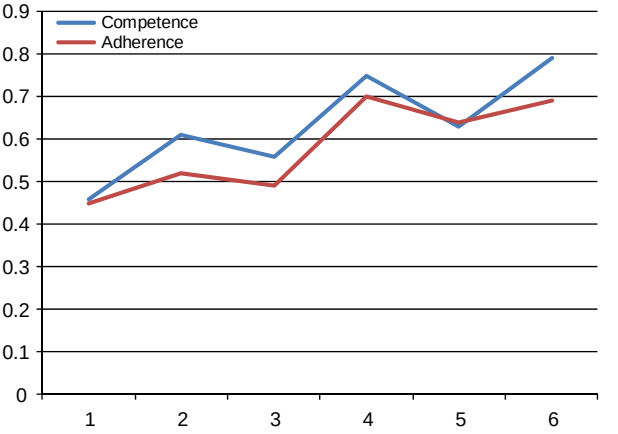

Figure 1. Increase in ICC 2.1 reliabilities between six raters during the training phase.

Reliability Issue

There is no standard procedure for reliability testing of psychotherapy adherence and competence scales. It is customary to use different versions of intra-class correlations (ICC), most often ICC 2.1 (Shrout & Fleiss, 1979). ICC 2.1 implies that each subject is measured by each rater and that raters are considered representative of a larger population of similar raters. Reliability is calculated based on the total design (average measurement) and from a single measurement. However, several previous studies have included the same therapists and the same patients more than once in the sessions being studied, although papers are unclear about these important methodological matters (Barber, Mercer, Krakauer, & Calvo, 1996; Hilsenroth, Bonge, Blagys, Ackerman, & Blais, 2005). The study design may thus violate the random requirement of ICC 2.1. There are often unknown component variations due to the same therapists and the same patients which are confounded with rater variance. Barber et al. (2003) take account of such component variations, but they do not report the results other than rater variation. The design of the present study makes it possible to disentangle the component variations and estimate the reliability for a decreasing number of raters.

Reliability refers to the scores obtained from a measurement method. Thus, it is less appropriate to speak of reliable raters, subjects, or measurement methods. It is the level of score stability and the dependability of these that is the main focus in a reliability analysis. With respect to the stability of measurement scores, there are two aspects of reliability. One concerns the stability of rank ordering of the scores, the other concerns the stability of absolute scores. Which of these estimates is used depends on the intention. If the intention is to use the measurement method to rank order subjects with

respect to some properties, reliability estimates for relative decisions are important. If the intended use of the scores is to categorize subjects based on some clearly defined scores, reliability estimates for absolute decisions are necessary. The intended use of the MBT-ACS concerns decisions of whether subjects are below or above some specific level of adherence or competence. Therefore, the most relevant reliability estimate is *absolute decisions*.

Material

Eighteen videotaped sessions of individual psychotherapy, performed by nine different therapists (two sessions by each) were used for this study. The therapists were clinicians in the MBT program of Department of Personality Psychiatry. Most of them were experienced clinicians with basic psychotherapy training in group analysis. Mean age was 50 years (range 32–60). By profession, there was one psychiatrist, one psychiatric resident, one clinical psychologist, one social worker, one occupational therapist, one physiotherapist and three psychiatric nurses. They had been trained in individual MBT locally when the department changed its policy from a previous group-oriented day hospital program in 2008. When the reliability test was performed, some had been doing MBT for a year, while some had only done around 3 months. The therapists were asked to deliver two videotaped recordings each, preferably one which they considered to be of high quality and one they considered to be of rather low quality, in order to enhance the variation of the phenomena under study.

Most of the patients had a diagnosis of borderline personality disorder. They were predominantly females, aged 20–30 years, with a baseline level of global functioning of GAF = 45. They were offered weekly sessions of individual MBT during the first year and less frequent sessions the second year, while also attending a weekly MBT group for a maximum of 3 years. The patients in this study had been in the program for various length of time, e.g., 2–15 months.

Study Design and Methods

In this study design, in which the observed score is compounded by three or more sources of variance, intra-class correlation is not an appropriate method to estimate the level of reliability. When the measurement design contains multiple sources of variance, Generalizability Theory (G-theory) is more meaningful. Within the design of G-theory, several variance components can be disentangled in just one analysis (Shavelson, Webb, & Rowley, 1989).

In the current research design, two therapy sessions from each of nine therapists were videotaped. This makes 18 unique therapy sessions, and all seven raters rate all 18 sessions. In the framework of Generalizability Theory (Shavelson & Webb, 1991), this is a two facet partially nested "(s:t) × r" design, where sessions (s) are nested within therapists (t) and raters (r) are crossed over sessions within therapists. The design is partially nested because the effect of session (s) is both nested (within t) and crossed (over r). With respect to generalizations beyond this study, therapists, sessions and raters are considered as randomly selected from the whole "universe" of admissible therapists, sessions and raters. The object of measurement is therapist behavior, and the measurement design is balanced since all therapists are rated by the same number of raters. Furthermore, this study has two so-called differentiation variance components, which are individual variance between therapists (t) and systematic variance between sessions for each therapist (st). This makes three sources of instrumentation variance that directly affect the reliability of the observed scores. These are (1) the rater effect (r) indicating the consistency of how much 'behavior' the raters see, averaged over therapists and sessions, (2) the interaction between raters and therapists (tr), indicating the raters different rank ordering of the therapists, and lastly (3) the unique rater-therapist-session interaction plus other unknown error variance (rst, e). See Figure 2. Within this design, sessions (s) cannot be separated from therapist (t) and neither can the session-rater interaction (sr) be separated from the rater-session-therapist interaction.

Applying classical intra-class correlation to such a design would give comparable estimates of the degree in which observers rank order their object of measurement, and how much they agree upon a random object's absolute score. However, several ICCs have to be done to differentiate the different variance components, and the interpretation of the whole picture will be more challenging.

Based on the sample data, the relative impacts of different sources of variation are estimated by a G-study (Shavelson et al., 1989), from which

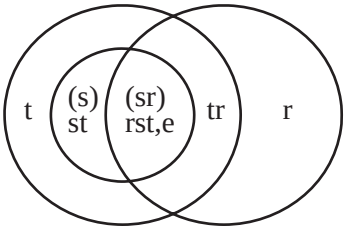

Figure 2. Venn diagram of the variance components in the (s:t) × r design.

generalizability coefficients are computed. Conceptually, a generalizability coefficient is equivalent to reliability coefficients such as intra-class correlations (ICC). Reliability coefficients based on a sound research design will often be rather expensive and unrealistic for treatment studies (here, seven raters rating 18 sessions from nine therapists), so an important pragmatic question is *how few raters are necessary to get a score with acceptable reliability.* Based on the obtained G-study components, the generalizability framework offers a subsequent study called D-study, or optimization study. By the D-study it is possible to estimate the reliability of scores based on, for example, four, two, or only one rater. The current G- and D-studies have been processed through the EduG program (Swiss Society for Research in Education Working Group, 2006; Cardinet, Johnson & Pini, 2009).

Results

Table I displays the reliability coefficients for relative decisions for all seven raters as well as estimated coefficients for two raters (D-study). The overall reliability for adherence and competence by seven raters were .84 and .88 respectively, and the overall estimates for two raters were .60 and .68.

Table I reveals a large reliability variation among the different items. Highest reliabilities were found for interventions that concerned the relation to the therapist (item 14) and the concurrent group therapy (item 17). The reliabilities here were in the range .70–.96. Lowest reliabilities were found for interventions that concerned pretend mode (item 8) and stop and rewind (item 12). The reliabilities here were in the range .07–.54.

There was a general trend for the reliability of competence to be somewhat higher than of adherence.

The correlations (Pearson's *r*) between adherence and competence were moderate to high, ranging from .50 (pretend mode) to .96 (exploring, not-knowing stance and stimulating mentalization). The correlation between overall adherence and competence was .93.

Table II displays a rank ordering of the reliabilities for the case of two raters. The following nine items had a particular low reliability: Challenge, psychic equivalence, praise, countertransference, regulating emotional arousal, validating understanding, validating feelings, stop and rewind and pretend mode.

As mentioned in the introduction, the overall rating was not calculated as a mean of the 17 items. The raters should consider the items 2, 6, 10 and 11 to carry a stronger weight. Table I shows that these items had a high reliability, e.g., 7R for competence in the range .84–.86.

Figure 3 displays the mean ratings of all items, for adherence as well as competence, for all sessions. The overall adherence was judged as being close to "good enough," while the overall quality was judged as being between "acceptable" and "good enough." Figure 3 reveals that the nine items with the lowest reliability were rated quite seldom for adherence. Actually, the two "worst" items, pretend mode and stop and rewind, were rated with an occurrence of 1–2 interventions in each session. When the occurrence is that low, disagreement with respect

Table I. Reliability coefficients (generalizability coefficients) for MBT-ACS

	Adherence 7R	Adherence 2R	Competence 7R	Competence 2R
1. Engagement			.81	.54
2. Exploring	.84	.60	.86	.63
3. Challenging	.79	.52	.65	.35
4. Adjustment			.78	.51
5. Regulating arousal	.61	.31	.61	.31
6. Stimulating mentalization	.72	.42	.86	.63
7. Acknowledging positive mentalizing	.71	.41	.55	.26
8. Pretend mode	.22	.07	.54	.25
9. Psychic equivalence	.72	.43	.64	.33
10. Focus on affects	.79	.51	.84	.61
11. Focus on interpersonal affects	.82	.57	.85	.61
12. Stop and rewind	.27	.10	.49	.21
13. Validating feelings	.50	.23	.57	.28
14. Relation to therapist	.91	.74	.88	.67
15. Counter-transference	.67	.37	.50	.22
16. Validating understanding	.61	.31	.73	.43
17. Integrating group experiences	.96	.88	.90	.71
Overall	.84	.60	.88	.68

Note. Adherence/competence 7R: Results for seven raters. Adherence/competence 2R: Results estimated (D-study) for two raters. Generalizability coefficients are for relative decisions.

Table II. Rank order of the items according to the reliability of two raters

Item	Adherence 2R	Competence 2R
Integrating group experiences	.88	.71
Relation to therapist	.74	.67
Exploration and not knowing stance	.60	.63
Focus on interpersonal affects	.57	.61
Engagement and warmth		.54
Adjustment to level of mentalization		.51
Challenging unwarranted beliefs	.52	.35
Focus on affects	.51	.61
Dealing with psychic equivalence	.43	.33
Stimulating mentalization	.42	.63
Acknowledging positive mentalization	.41	.26
Use of countertransference	.37	.22
Regulating arousal	.31	.31
Validating own understanding	.23	.43
Validating patient's feelings	.23	.28
Stop and rewind	.10	.21
Dealing with pretend mode	.07	.25

to the presence of interventions will have large consequences for the reliability.

The raters are a major source of variation in this study. Some raters deviated more from the mean than others. The effects on the reliabilities varied. Excluding the least reliable rater resulted in the overall adherence 7R increasing from .84 to .85, and the overall competence 7R increasing from .88 to .89, a rather marginal effect. The effect on the item level was greater. Excluding the least reliable rater on pretend mode increased the adherence from .22 to .42, while adherence for stop and rewind increased from .27 to .38.

The therapists also varied with respect to their overall adherence and overall competence, on the item level, as well as from session to session. The mean A/C over two sessions for the least competent (in MBT) therapist was 2.4, while the mean A/C for the most competent therapist was 5.2.

Table III summarizes the major sources of variation in adherence 7R for the two items with highest and the two with lowest reliability. From a reliability point of view, it is favorable that the residual variance is low, that the raters' ranking variation is low, and that there is some variation among therapists from session to session. The item on interventions that aim to integrate experiences from the concurrent group therapy was the one with highest reliability ($r = .96$). Table III specifies the reasons: The residual variance was very low (16%), which means that the specified variables accounted for the major part of the variance. There was a complete agreement among the raters on the ranking order of the therapists. There was a large variation in therapist behavior from session to session (78%), e.g., in some sessions the intervention was absent while in other sessions the intervention was frequent. There was a low variation (6%) in how much of the behavior (the specific intervention) the raters observed, and there

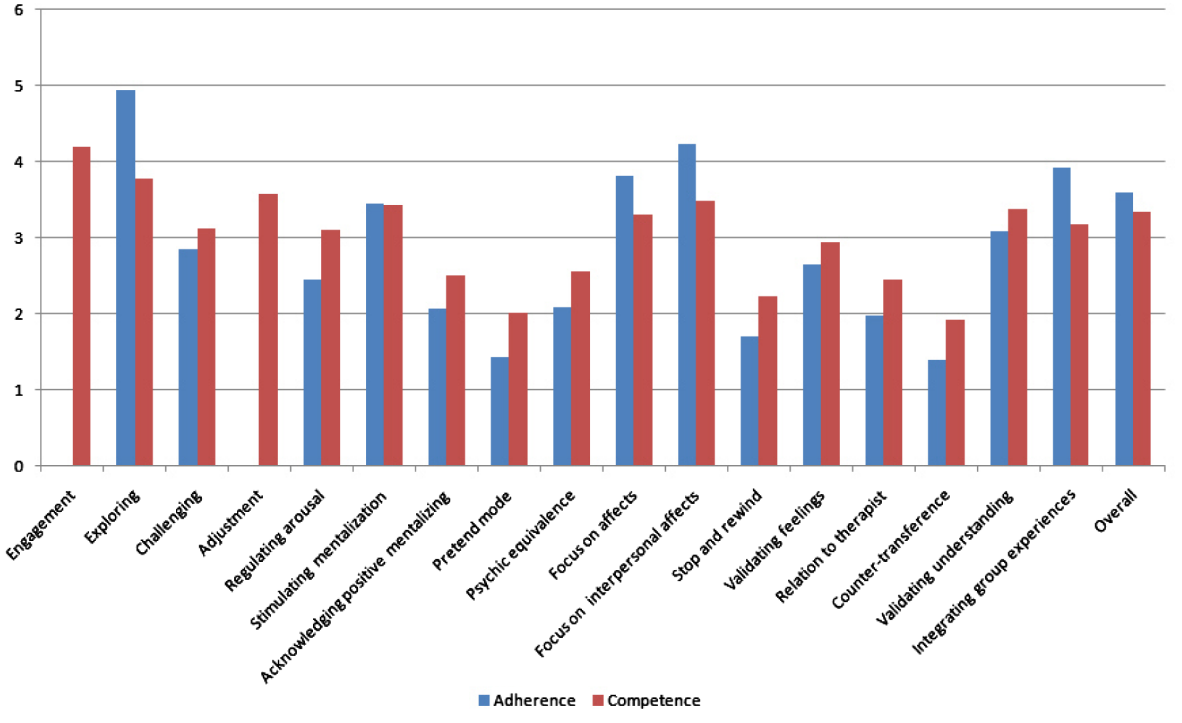

Figure 3. Intervention profile (mean) for all (nine) therapists by all (seven) raters.

Table III. Sources of variation for items with high versus low reliability on adherence 7R: Percentages of total variation

Item	T: Between therapist variation	R: Variation in how much raters observe	S:T: Therapist variation across sessions	TR: Variation in raters ranking of therapists	RS:T: Residual (including error) variance
Integrate group	0	5.8	78.4	0	15.8
Relation to therapist	44.6	6	14.6	0	34.8
Pretend mode	2.9	5.6	0	24.4	67.1
Stop and rewind	3.3	25.1	1.8	7.6	62.3

was no systematic variance between therapists regarding the intervention.

Interventions that concerned the patient–therapist relationship also had a very high 7R reliability ($r = .91$). Here, the pattern is a bit different. The residual variance was somewhat higher (35%). There was a complete agreement on the ranking order of the therapists. The variance of the therapists from session to session was smaller (15%), indicating that this kind of intervention was used more evenly. The variance in how much of the behavior the raters observed, was low (6%), while the variation among therapists in how much they used the intervention was high (45%), indicating that some therapists hardly used the intervention, while other therapists used it frequently.

Interventions aimed at pretend mode had a low 7R reliability ($r = .22$). Table III shows that the residual variance for this item was very high (67%). Furthermore, there was a high disagreement on the ranking order of the therapists (24% of the variance). There was no session variation, e.g., the intervention was observed seldom in both sessions. There was a small rater variation (6%), and little systematic variation among therapists (3%) in how much they used the intervention. When the residual variance is this high, it suggests that there is something wrong with the validity of the item. It does not seem sufficiently operationalized so that therapists know when and how to apply it, and so that raters can recognize when therapists actually perform it. However, the raters apparently did have some opinions on this topic, since they disagreed so much (24% variance) on the ranking order of the therapists.

Interventions of "stop and rewind" also had a low 7R reliability ($r = .27$) and the residual variance was very high (62%). The other major source of variance was rater variation (25%). There was considerable disagreement among raters on how much of this intervention they observed, although they did not deviate much in their ranking order (8% variance). Strong disagreement on the presence of a phenomenon may have large consequences for reliability when the therapist variation (between therapists and between sessions) is low.

Inspecting all the items on the variance components, we found a connection between low reliability and residual variance. The eight items with lowest 7R reliability had a mean residual variance of 58%, compared to 37% for the seven items with the highest 7R reliability.

Concerning the 7R reliabilities of the overall adherence and competence ratings, the residual variation was quite low (21–24%), the variance in therapist behavior was quite high (32–47%), the variance in rater observation quite low (6–12%), the variation in ranking of therapists was moderate (19–24%) and the therapist variation between sessions rather low (6–11%). Concerning competence we saw the same trends as we have discussed for adherence. However, some results require emphasis. The intervention validating feelings showed a high 7R competence residual variance at 78%, suggesting weak references for the raters as to how to rate this item. The same holds true (high residual variance) for competence in regulating the emotional arousal, dealing with psychic equivalence, and stop and rewind. There was a considerable variance from item to item in how much the therapists varied in their competence. For some items, it was close to zero so the therapists did not vary at all. For example, all therapists handled dealing with psychic equivalence with low competence. For other items, they varied considerably, e.g., the items considered most important for the overall rating of competence: Not-knowing stance, stimulating mentalization, focus on feelings and dealing with feelings in interpersonal encounters. The therapist variation in competence on these items was in the range 35–40%.

Discussion

The results demonstrate that the MBT adherence and competence scale is a reliable instrument for rating overall adherence to and competence in mentalization-based treatment. The reliabilities among seven raters were very high (.84/.88). They declined gradually with fewer raters, but they were still acceptable for two raters (.60/.68). The instrument can thus be used for research purposes where the question of overall treatment fidelity needs to be documented. The level of reliability is comparable to that of The Cognitive Therapy Adherence and Competence Scale (Barber et al., 2003).

If overall treatment fidelity was the only task for the scale, either the adherence or the competence rating could be deleted, since they correlated by $r = .93$. However, rating both adherence and competence gave additional information at a profile (item) level. It is important to identify low adherence to specific items, since this may be a problem in its own right (e.g., low adherence to "transference and the relation to the therapist").

The reliabilities at item level varied considerably. This is a common finding among rating scales (Barber et al., 2003). Some items had a satisfactory reliability, while others had a low to very low reliability. The study confirmed the soundness of giving more weight to items number 2, 6, 10 and 11 when performing an overall rating. These items had high reliability.

The study also confirmed the usefulness of the methods (G-study and D-study) for disentangling variation components and calculating effects of decreasing number of raters and effects of specific raters with deviant reliability. Deviant raters had a moderate to large effect upon the items with lowest reliability, but not upon the overall ratings.

The G-study yielded data which are useful for modification of the scale. Final modifications will depend upon the intended use of the scale and theoretical considerations. For research which aims to answer whether treatment fidelity is good enough or not, 17 items are more than enough and the number of items could be reduced, e.g., by deleting the least reliable items. However, for the purpose of treatment process research, it is necessary to retain items with low reliability if the content is regarded as important, and address the sources of residual variance. Also for the purposes of training and supervision, a more detailed intervention profile, e.g., retaining items if possible, is useful. And lastly, the different items have a different status within the theory of mentalization. Weighing all these arguments against each other and considering the data from the G-study, we suggest the following:

The item "validate feelings" may be deleted and the intention behind it can be incorporated in the general item on "focus on feelings" as a competence marker.

The item "monitoring own understanding" (e.g., "have I got you correctly by . . .") may be deleted and integrated in the item "exploration, curiosity and not-knowing stance."

However, the item "stop and rewind" is more central to the technique of MBT. It has a parallel to "chain analysis" in dialectical behavioral therapy (DBT) and "conversation analysis" in interpersonal psychotherapy (IPT). The intended purpose is to calm down an emotionally driven narrative and consider the details in intersubjective encounters more carefully, e.g., "can we stop here, and go back to . . ." It refers both to external interpersonal events, and to events in the here and now enacted by the patient–therapist couple. The data suggest that most therapists in this study seldom performed such interventions. Reasons may be inadequate training, or that the therapists did not see any benefit from such interventions. In addition, the raters disagreed on what counted as an intervention belonging to this item. The remedy here seems to be to sharpen the definition of the item for the raters, and to instruct the therapists to perform these kinds of interventions more frequently.

The item "dealing with pretend mode" is central to the theory of mentalization and should be retained for that reason. However, the residual variance was very high for this item, indicating (1) that the therapists had difficulties with identifying pretend mode, (2) that the therapists had difficulties with knowing what to do with it, and (3) that the raters had difficulties with identifying interventions aimed to modify pretend mode. Renaming the item, for example, to "dealing with pseudomentalization" might clarify the concept clinically. However, the "pretend mode" label seems a better fit for capturing the intersubjective discourse of this item, e.g., the tendency for the therapist to collude with the patient discourse and join a rather aloof conversational style. Part of the problem might be that "dealing with pretend mode" cannot be reduced to a question of particular interventions. This means that one should delete the adherence rating, but retain the competence rating. This implies that therapists usually handle this issue more indirectly, by using a series of individually tailored interventions, e.g., by shifting topics to areas of higher emotional arousal and greater vitality. It may be difficult for raters to detect that the intention of an indirect intervention which resulted in a change in the topic was to "deal with pretend mode." However, for the session as a whole it may be possible to rate whether the therapist seems to accept a pretend mode discourse (low competence rating) or if he/she seems to be aware of the phenomenon and persistently intervenes. Nevertheless the question of pseudomentalization/pretend mode seems to be a topic for more detailed clinical studies, e.g., do different observers agree on the very phenomenon as displayed by the patients? If there is a low agreement as to what counts as pretend mode, the reliability on interventions that aim at the phenomenon will of course be dubious.

"Dealing with psychic equivalence" is also central to the theory of mentalization. There is a moderate agreement on identifying interventions aimed at psychic equivalence. However, the competence

reliability is lower (.33). The manual should be more specific with respect to what counts as a high versus low competence for this item.

The item "regulating arousal" also had low reliability. The item refers to "emotional arousal," e.g., to feelings and vitality, "not too high so that the patient loses his or her ability to mentalize; not too low so that the session becomes meaningless emotionally." However, vitality and arousal are already covered by "dealing with pretend mode" and the topic of emotion regulation can be defined as part of "focus on affects" and "dealing with psychic equivalence." This item could be deleted if the arousal aspect was more explicitly integrated in the above mentioned items. This would leave a 14-item scale.

The MBT-ACS ratings in this study provided important feedback to the clinical unit in question, although care should be taken in generalizing the findings to the unit as a whole as only two-thirds of the therapists were rated and since they were the least formally qualified therapists. However, the overall mean for adherence as well as for competence was below the level of 4, which is defined as "good enough." The average profile revealed where the problems resided. They concerned identifying and dealing with pretend mode, the use of the stop and rewind technique, and a low adherence on the items that dealt with the patient–therapist relationship. Lastly, the individual profiles revealed large differences among therapists, suggesting that some should be more adequately trained, while others were doing well.

There are several limitations to this study. MBT is a flexible approach that emphasizes that therapeutic strategy should be modified according to the mentalizing capabilities and the contextual state of the patient. For example, a more supportive and containing stance might be appropriate during the initial phase, while "mentalizing the transference" becomes more important later on during treatment. Likewise, the "stop and rewind" technique is more appropriate when patients are highly aroused. An ideal patient sample would cover a wide range of therapeutic phases and situations. Most of the borderline patients in this sample were in their "middle phase" and none displayed any acute suicidal crisis. A larger situational variance might have provided more interventions of the type that were observed with low frequencies in this study, e.g., psychic equivalence, giving these items "a better chance" to achieve a higher reliability.

Concerning therapists, recruiting therapists from the same MBT program also limits the variance. We tried to increase the variance by suggesting to the therapists that they provided one session which they regarded as "good" and one which they regarded as "poor." Unfortunately, we do not know if they followed this suggestion, and the raters were not aware of which sessions were regarded as "good" or "poor." The main implication of the limited therapist sample is that the correlation between adherence and competence become artificially high if we consider the appropriate universe for generalization to be psychotherapy in general. We know from previous ratings of non-MBT psychotherapy sessions that non-MBT therapists might display high adherence on items such as "affect focus" and "transference and the relation to the therapist." However, *the way of doing this* is different from an MBT approach. This gives a low competence rating and the correlation between adherence and competence declines, although by how much remains an empirical question. A subsequent study should try to increase the variance by increasing the number of sessions to be rated, including patients from all phases of treatment, and also including other kind of therapeutic strategies for the same type of patients, e.g., DBT. By this, the adherence-competence issue could be explored more in depth and the discriminant validity of the MBR-ACS could be tested.

The results reported in this study should be considered initial estimates of the reliability of the MBT Adherence and Competence Scale. Further studies are needed for more robust conclusions. Such studies should also incorporate the question of a minimum level of rater competence. Our own preliminary studies indicate that comparison with a gold standard is the best way, e.g., rating verbatim text (that has been given expert ratings) while watching the video recordings.

In conclusion, the 17-item version of the MBT-ACS was found to be a useful instrument for measuring overall treatment fidelity of MBT. It also yields useful data for evaluating and providing feedback to therapists by identifying strong and weak aspects of their therapeutic style. It might also be a valuable instrument for quality control. However, the reliability of the detailed intervention profile should be enhanced. One option might be condensing the scale to a 14-item version.

References

Barber, J.P., Crits-Christoph, P., & Luborsky, L. (1996). Effects of therapist adherence and competence on patient outcome in brief dynamic therapy. *Journal of Consulting & Clinical Psychology*, *64*, 619–622.

Barber, J.P., 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.

Barber, J.P., Liese, B.S., & Abrams, M.J. (2003). Development of the Cognitive Therapy Adherence and Competence scale. *Psychotherapy Research*, *13*, 205–221.

Bateman, A., & Fonagy, P. (1999). The effectiveness of partial hospitalisation in the treatment of borderline personality

disorder: a randomised controlled trial. *American Journal of Psychiatry*, *156*, 1563–1569.

Bateman, A.W., & Fonagy, P. (2001). Treatment of borderline personality disorder with psychoanalytically oriented partial hospitalization: an 18-month follow-up. *American Journal of Psychiatry*, *158*, 36–42.

Bateman, A., & Fonagy, P. (2004). *Psychotherapy for borderline personality disorder: Mentalization-based treatment*. Oxford: Oxford University Press.

Bateman, A.W., & Fonagy, P. (2006). *Mentalization-based treatment for borderline personality disorder: A practical guide*. Oxford: Oxford University Press.

Bateman, A., & Fonagy, P. (2009). Randomized controlled trial of out-patient mentalization based treatment versus structured clinical management for borderline personality disorder. *American Journal of Psychiatry*, *1666*, 1355–1364.

Bateman, A.W., & Fonagy, P. (2011). *Handbook of mentalizing in mental health practice*. Arlington, VA: American Psychiatric Publishing.

Bion, W.R. (1970). *Attention and interpretation*. London: Tavistock [reprinted London: Karnac Books 1984]. Reprinted in Seven Servants (1977e).

Bouchard, M.-A., & Lecours, S. (2008). Contemporary approaches to mentalization in the light of Freud's *Project*. In F.N. Busch (Ed.), *Mentalization: Clinical considerations, research findings, and clinical implications*. New York: Analytic Press.

Cardinet, J., Johnson, S., & Pini, G. (2009). *Applying generalizability theory using EduG*. New York: Routledge.

Carroll, K.M., Nich, C., Sifry, R.L., Nuro, K.F., Frankforter, T.L., Ball, S.A., Fenton, L., & Rounsaville, B.J. (2000). A general system for evaluating therapist adherence and competence in psychotherapy research in the addictions. *Drug and Alcohol Dependence*, *57*, 225–238.

Engen, M.J. (2009). *Utvikling av en bedømningsskala for mentaliseringsbasert terapi*. Hovedoppgave i psykologi-Universitetet i Oslo. http://urn.nb.no/URN:NBN:no-26206.

Fonagy, P., & Bateman, A. (2006). Mechanisms of change in Mentalization-Based Therapy of borderline personality disorder. *Journal of Clinical Psychology*, *62*, 411–430.

Fonagy, P., Gergely, G., Jurist, E.L., & Target, M. (2002). *Affect regulation, mentalization and the development of the self*. New York: Other Press.

Giesen-Bloo, J., Van Dyck, R., Spinhoven, P., van Tilburg, W., Dirksen, C., van Asselt, T., et al. (2006). Outpatient psychotherapy for borderline personality disorder. Randomized trial of schema-focused therapy vs transference-focused psychotherapy. *Archives of General Psychiatry*, *63*, 649–658.

Hilsenroth, M.I., Bonge, D.R., Blagys, M.D., Ackerman, S.J., & Blais, M.A. (2005). Measuring psychodynamic-interpersonal and cognitive-behavioral techniques: Development of the comparative psychotherapy process scale. *Psychotherapy: Theory, Research, Practice, Training*, *3*, 340–356.

Høglend, P., Amlo, A., Marble, A., Bøgwlad, K.P., Sørbye, Ø., Sjaastad, M.C., & Heyerdahl, O. (2006). Analysis of the patient-therapist relationship in dynamic psychotherapy: An experimental study of transference interpretations. *American Journal of Psychiatry*, *163*, 1739–1746.

Karterud, S., & Bateman, A. (2010). Manual for mentaliseringsbasert terapi (MBT) og MBT vurderingsskala: versjon individualterapi. Gyldendal Akademisk.

Luborsky, L., McLellan, A.T., Woody, G.E., O'Brien, C.P., & Auerbach, A. (1985). Therapist success and its determinants. *Archives of General Psychiatry*, *42*, 602–611.

Luborsky, L., & Barber, J.P. (1993). Benefits of adherence to psychotherapy manuals – and where to get them. In N. Miller, L. Luborsky, J.P. Barber, & J. Docherty (Eds.), *Psychodynamic treatment research: A handbook for clinical practice* (pp. 211–226). New York: Basic Books.

Lysaker, P.H., Gumley, A., & Dimaggio, G. (2011). Metacognitive disturbances in persons with severe mental illness: Theory, correlates with psychopathology and models of psychotherapy. *Psychology and Psychotherapy: Theory, Research and Practice*, *84*, 1–8.

McGlinchey, J., & Dobson, K.S. (2003). Treatment fidelity assessment in cognitive behavioral therapy. *Journal of Cognitive Psychotherapy: An International Quarterly*, *17*, 299–318.

NICE. (2009). *Borderline Personality Disorder: treatment and management* (*NICE guideline*). National Institute for Health and Clinical Excellence. www.nice.org.uk.

Nordahl, H.M., Nysæter, T.E., & Mikkelsen, B. (2006). Cognitive Therapy Adherence and Competence Scale. *Tidskrift for Kognitiv Terapi*, *3*, s14–29.

O'Malley, P.M., Bachman, J.G., & Johnston, L.D. (1988). Period, age, and cohort effects on substance use among young Americans: A decade of change, 1976–1986. *American Journal of Public Health*, *78*, 1315–1321.

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*, 829–841.

Shavelson, R.J., Webb, N.M., & Rowley, G.L. (1989). Generalizability theory. *American Psychologist*, *44*, 922–932.

Shavelson, R.J., & Webb, N.M. (1991). *Generalizability theory. A primer*. Newbury Park, CA: Sage.

Shrout, P.E., & Fleiss, J.L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *2*, 420–428.

Swiss Society for Research in Education Working Group. (2006). EDUG user guide. Euchatel, Switzerland: IRDP

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*, 620–630.

Wampold, B.E. (2001). *The great psychotherapy debate: Models, methods, and findings*. Mahwah, NJ: Erlbaum.

Appendix: The 17 items of the MBT adherence and competence scale and the "good enough" quality level

Item name	Good enough quality level (4)
1. Engagement, interest and warmth	The therapist appears genuinely warm and interested. The rater gets the impression that the therapist cares. Several concrete comments communicate this positive attitude
2. Exploration, curiosity and a not-knowing stance	The therapist poses appropriate questions designed to promote exploration of the patient's and others mental states, motives and affects and communicate a genuine interest in finding out more about them
3. Challenging unwarranted beliefs	The therapist confronts and challenges unwarranted opinions about oneself or others in an appropriate manner
4. Adaptation to mentalizing capacity	The therapist seems to have adapted to the patient's mentalizing level and the interventions are for the most part short, concise and unpretentious
5. Regulation of arousal	The therapist plays an active role in terms of maintaining emotional arousal at an optimal level (not too high so that the patient loses his or her ability to mentalize; not too low so that the session becomes meaningless emotionally)
6. Stimulating mentalization through the process	The aim of the interventions clearly seems to be to stimulate the mentalizing of experiences of self and others in an ongoing process and is less concerned about content and interpretation of content in order to promote insight
7. Acknowledging positive mentalizing	The therapist identifies and explores good mentalization and this is accompanied by approving words or judicious praise
8. Pretend mode	The therapist identifies pretend mode and intervenes to improve mentalizing capacity
9. Psychic equivalence	The therapist identifies psychic equivalence functioning and intervenes to improve mentalizing capacity
10. Affect focus	The interventions focus primarily on affects, more than on behavior. The attention is directed at affects as they are expressed in the here and now, and particularly in terms of the relationship between patient and therapist
11. Affect and interpersonal events	The therapist connects emotions and feelings to recent or immediate interpersonal events
12. Stop and rewind	The therapist identifies at least one incident in which the patient reacts in a maladaptive way to an interpersonal event, then tries to slow down the pace and find out about the incident step-by-step
13. Validation of emotional reactions	The therapist expresses a normative view on the warranted nature of the patient's emotional reaction(s) after these are sufficiently investigated and understood
14. Transference and the relation to the therapist	The therapist comments on and attempts to explore – together with the patient – how the patient relates to the therapist during the session and stimulates reflections on alternative perspectives whenever appropriate
15. Use of countertransference	The therapist actively utilizes his/her own feelings and thoughts about the relationship to the patient and attempts by this to stimulate an exploration of the relationship between them
16. Monitoring own understanding and correcting misunderstanding	The therapist checks out his/her understanding of the patient's state of mind and to what extent this corresponds with the patient's understanding. Then he/she lets his/her own understanding be influenced by the patient's understanding and openly admits to any misunderstanding whenever they occur
17. Integrating experiences from concurrent group therapy	The therapist stimulates exploration of the patient's experiences from the group therapy sessions and helps to integrate the material so that the treatment as a whole is coherent