

Treatment Integrity: Implementation, Assessment, Evaluation, and Correlations with Outcome

Florian Weck · Christiane Bohn · Denise M. Ginzburg · Ulrich Stangier

Department of Clinical Psychology and Psychotherapy, Frankfurt University, Germany

Keywords

Treatment integrity · Adherence · Competence · Therapy research · Outcome

Summary

The verification of treatment integrity is an essential precondition for analysis of the effectiveness of psychotherapeutic research. Treatment integrity provides important information about to what extent a treatment was delivered as intended. 3 different aspects form the concept of treatment integrity: adherence, treatment differentiation, and competence. The current review gives recommendations regarding implementation, measurement, and evaluation of treatment integrity. Empirical studies and the correlations between treatment integrity and therapy outcome are also discussed. The methodological procedures, particularly measurement and evaluation, seem to be very heterogeneous, which makes it difficult to compare the results. However, for many studies, a direct or indirect correlation between treatment integrity and treatment outcome can be found. In further studies, the recommendations for implementation, measurement, and evaluation of treatment integrity should be taken into account to allow better comparability between studies.

Schlüsselwörter

Behandlungsintegrität · Adhärenz · Kompetenz · Psychotherapieforschung · Therapieerfolg

Zusammenfassung

Die Sicherung und Überprüfung der Behandlungsintegrität stellt eine essenzielle Voraussetzung für Wirksamkeitsanalysen in der Psychotherapieforschung dar. Die Behandlungsintegrität liefert wichtige Hinweise darüber, ob eine Behandlung so durchgeführt wurde, wie es intendiert war. Es lassen sich 3 unterschiedliche Aspekte der Behandlungsintegrität unterscheiden. Diese sind Adhärenz, Differenzierbarkeit und Kompetenz. Die vorliegende Arbeit gibt Empfehlungen zur Implementierung, Messung und Evaluation der Behandlungsintegrität. Zudem wird eine Übersicht über empirische Befunde gegeben, die einen Bezug zwischen Behandlungsintegrität und Behandlungserfolg herstellen. Es wird deutlich, dass das methodische Vorgehen, insbesondere bei Messung und Evaluation, heterogen und somit die Vergleichbarkeit der Ergebnisse erschwert ist. Dennoch findet sich für viele Studien ein direkter oder indirekter Zusammenhang zwischen Behandlungsintegrität und Behandlungserfolg. Bei nachfolgenden Studien sollten die vorliegenden Empfehlungen zur Implementierung, Messung und Evaluation der Behandlungsintegrität stärker berücksichtigt werden, um eine bessere Vergleichbarkeit der Studien zu ermöglichen.

Treatment Integrity

By treatment integrity (related concepts are treatment fidelity, treatment adherence, procedural reliability, therapist compliance, conformity to the manual) we mean the extent to which the interventions are carried out as intended [Yeaton and Sech-

rest, 1981]. Treatment integrity has 3 different aspects: adherence, treatment differentiation, and competence [Waltz et al., 1993]. Adherence here means the extent to which the therapist uses interventions and approaches as they are described in the treatment manual. Treatment differentiation, on the other hand, describes how the treatment differs from other forms of

treatment (if different treatments were administered). Treatment differentiation is therefore closely related to adherence, because the treatment can only be distinguished from other treatments if the therapist adheres to the treatment manual. Competence means that the therapist is able to carry out the interventions skillfully. This includes taking the treatment context into account appropriately [Waltz et al., 1993]. Relevant aspects of the context include the patient's characteristics (e.g., extent of impairment), specific problems (e.g., severe social anxiety), and the patient's living conditions (e.g., living with parents, being unemployed). Furthermore, the current phase of therapy and the progress of the patient to date should be taken into consideration. The timing of an intervention should be chosen with sensitivity. There is a close association between adherence and competence. A certain degree of adherence is a prerequisite for competent administration of the therapy. However, an adherent approach is not necessarily competent [Waltz et al., 1993]. There are also high empirical correlations between adherence and competence. Monitoring both aspects, however, appears to be reasonable and justified, because the predictive value of the 2 variables can be different, despite the high correlation between them [Barber et al., 2007].

Consideration of treatment integrity is essential to making an internally valid judgment as to whether the treatment is contributing to a significant improvement in symptoms [Vermilyea et al., 1984]. Only if treatment integrity is taken into account and found to be high, can the effectiveness or ineffectiveness of the interventions be attributed to the treatment at all. If, however, treatment integrity is not monitored or is low, and the treatment is ineffective, it remains unclear whether it is actually the type of treatment or its implementation that is responsible for these results. Even with effective treatment, no conclusion can be drawn about the effectiveness of the specific interventions if treatment integrity is not monitored or is markedly low.

Although the importance of treatment integrity in psychotherapy trials is undisputed, it is (too) often neglected in scientific practice. Thus Perepletchikova et al. [2007], after a review of 6 influential international professional journals (Archives of General Psychiatry, American Journal of Psychiatry, British Journal of Psychiatry, Journal of the American Academy of Child and Adolescent Psychiatry, Journal of Consulting and Clinical Psychology, and Journal of Clinical Psychiatry) from 2000 to 2004, came to the conclusion that treatment integrity was only addressed adequately in 3.5% of the randomized controlled trials. In a follow-up study, the authors of these treatment studies were asked about the reasons for their lack of attention to treatment integrity [Perepletchikova et al., 2009]. The primary obstacles proved to be lack of theoretical knowledge, lack of guidelines for implementation, time, and costs.

The aim of this study is to offer specific recommendations for implementation, measurement, and evaluation of treatment integrity. Empirical studies have also been assembled, to

study the relationships between treatment integrity and outcome. We hope to make clear here that the methodology has been heterogeneous, and in further research, it is imperative to use a uniform methodology for verifying treatment integrity.

Implementation of Treatment Integrity

There are several recommendations in the literature concerning how to achieve treatment integrity [Bellg et al., 2004; Gresham et al., 2000; Perepletchikova et al., 2009; Perepletchikova and Kazdin, 2005; Waltz et al., 1993]. First, there should be a clear *definition of treatment integrity*. Following the definition of treatment integrity given here in the introduction, adherence can be considered as a measure of the extent to which the desired interventions are done according to the treatment manual and undesired interventions are being avoided. Waltz et al. [1993] differentiate the therapist's conduct even further into 4 categories: therapist conduct that is specific and essential to the type of treatment; therapist conduct that is not specific to the type of treatment but is nevertheless essential; therapist conduct that is neither essential nor specific; and therapist conduct that is undesirable. Thus for each treatment, it must be established which types of therapist conduct should be classified under which category. Competence, on the other hand, should be defined as the ability and skill of the therapist in doing this work. This is not an overall assessment of the therapist's competence (e.g., knowledge or experience), but rather of how that competence is applied. That means that therapist conduct will only be considered competent if it can be viewed as competent from the standpoint of the therapy that is being used.

Furthermore, there should be an appropriate *description of treatment*, which establishes the content to be addressed, along with the exercises and techniques to be used. Specific descriptions contribute to unifying the approach and raising the level of treatment integrity [Ehrhardt et al., 1996]. It is important that the description be explicitly stated, e.g., in a treatment manual. There should also be a description of the number of therapy sessions (e.g., 16 sessions), their length (e.g., 50 min), frequency (e.g., weekly), and the qualifications of those providing care (e.g., qualified psychologists). This sets a minimum requirement to establish that the therapists are performing a comparable treatment. In addition, the various interventions (e.g., role play) should be described directly. Advice on the preparation of treatment manuals can be found in Carroll and Nuro [2002]. Carroll and Nuro suggest that consideration should be given to the context in which a treatment manual is used. They differentiate among the use of manuals in the testing phase (pilot trials), the controlled research phase (efficacy trials), and clinical practice (effectiveness trials). While during the pilot phase, for example, initial indications of techniques and goals are sufficient, in the research phase there should be more detailed instructions, e.g., about the training and supervi-

sion of the therapists. For clinical practice, there should also be information about how the manual can be applied to different clinical groups. Instructions should also be given regarding the 'indications' section of the manual.

The *training of therapists* is a key task in the assurance of treatment integrity. Direct training (e.g., by a qualified trainer) is preferable to indirect training (e.g., by reading a treatment manual) or to indirect training supplemented by direct training, since it has been shown empirically that direct training is superior to indirect in ensuring treatment integrity [Sholomskas et al., 2005; Sterling-Turner et al., 2002]. Direct training should provide opportunities for practice. This can be done by reciprocal role playing, videos of pilot treatment programs, model conduct by the trainer, and the trainer's feedback about the conduct of the therapist. While theoretical instructions should be primary at the beginning of the training program, more specific exercises should be gradually introduced, with the therapist as an active participant. The scope of training should be appropriate both to the treatment manual and to the disorder being treated. To ensure that a therapist is proceeding in an adherent and competent way, a standard role play should be performed, in which the therapist demonstrates the desired strategies. Therapy situations that are particularly relevant should be selected (e.g., conveying the rationale for exposure therapy to agoraphobic patients). The therapist's approach should then be judged according to certain criteria (e.g., use of guided discovery rather than 'persuasion' of the patient), with respect to adherence and competence. Only when the therapist meets the desired requirements can it be concluded that the training was successful from the standpoint of treatment integrity. Ongoing training sessions are advisable, so that the training content and exercises can be reviewed.

To supplement in addition to sessions, there should also be regular *supervision of therapists*, including audiotape or video recordings. Otherwise, even well-trained therapists may vary a great deal in how they implement the manual in the course of treatment, and may drift away from it with respect to treatment integrity. The supervisor should pay special attention to particularly difficult therapeutic challenges, for example using role play to test the adequacy of a therapeutic approach.

Measurement of Treatment Integrity

Treatment integrity can be monitored directly or indirectly. An indirect measurement is made, for example, if the therapist himself provides information on his adherence to the therapeutic method. This approach is problematic, however, because the therapist might tend to overestimate his own adherence to the manual. Thus empirical studies have found only slight agreement between self-assessment (54% adherence) and external assessment (4% adherence) of treatment integrity [Wickstrom et al., 1998]. Accordingly, indirect moni-

toring of treatment integrity should always be supplemented with direct monitoring by an independent rater.

Treatment integrity could be directly assessed, for example, using transcripts or audio/video recordings of the session. Empirical studies show, however, that estimates of treatment that use different methods, such as assessments by supervisors or assessments by independent raters using video tapes, are largely uncorrelated [Chevron and Rounsaville, 1983]. Materials with high information density (e.g., video recordings) are therefore preferable to those with lower information density (e.g., session transcripts) [Waltz et al., 1993].

The assessment of treatment integrity should be done by means of standardized rating scales or questionnaires. For the monitoring of competence in particular, a multi-level response format should be chosen that adequately portrays differences between therapists. For monitoring of both adherence and competence, all important aspects should be covered by several items. Expert judgments can be sought to ensure that all aspects of treatment were given sufficient attention [Weck et al., 2010a].

Two German-language works provide a good overview of methods of monitoring treatment integrity [Pohl et al., 2000; Staats et al., 2003]. The current paper therefore concentrates on new developments in the German-speaking countries.

The German version of the Cognitive Therapy Scale (CTS) [Weck et al., 2010a] is currently widely used in a research context for external assessment procedures. The CTS was developed by Young and Beck [1980] to assess therapeutic competence in cognitive therapy. The German version of the CTS contains 14 items, which are rated on a 7-point Likert Scale, ranging from poor (0) to excellent (6). Descriptions are also given as an anchor for every second point on the scale. The CTS includes a subscale for *general therapeutic competences* (dealing with problems/questions/objections, clarity of communication, interpersonal effectiveness, resource activation, analysis of homework, use of and feedback from summaries), and a subscale for *competences that pertain to the structure of the therapy session* (agenda, time aspects, guided discovery, focus on key cognitions and behavior, rationale, selection of appropriate policies, appropriate implementation of techniques, and assigning of homework). Additional items of the CTS are also assessed on the 7-point scale, to evaluate the difficulty of the treatment and the competence of the therapist overall. There was good to very good rater agreement and internal consistency for the subscales of Structuring Competences ($ICC_{(2,2)} = 0.93$; $\alpha = 0.88$) and General Therapeutic Competences ($ICC_{(2,2)} = 0.85$; $\alpha = 0.71$), as well as the Total Scale ($ICC_{(2,2)} = 0.90$; $\alpha = 0.86$). In addition to the CTS, there is a similar method designed to evaluate treatment competences for psychoeducational treatments [Weck et al., 2011a].

The Verhaltenstherapie-Checkliste (VTKC, Behavioral Therapy Checklist) [Linden et al., 2007a; Linden and Langhoff, 2010] is an instrument that was developed specifically for use in clinical practice and for training of behavioral thera-

pists. The VTKC is based on the multi-level model of psychotherapeutic competences [Linden et al., 2007b] and makes it easier (a) to operationalize what is considered to be good therapeutic practice in cognitive behavioral therapy, (b) to measure how good a specific therapy is on this scale, and (c) to be used in training and continuing education as a training instrument to optimize therapist conduct [Linden and Langhoff, 2010]. It allows assessment of how often various base competences in behavioral therapy were applied (quantity) and how well they were applied (quality). It is particularly noteworthy that the VTKC exists in different forms and can include different sources of assessment: a detailed version (86 items) for supervision and therapist training (VTKC-S), a version for patients (VTKC-P), a shortened version for research purposes (71 items), as a self-assessment for therapists (VTKC-T), and an external assessment for external raters (VTKC-R). Overall, 12 major categories can be differentiated (discussion of homework results, microanalysis, macroanalysis, development of a disorder model, problem solving, modification of cognitions, self-management, assigning homework, ensuring a working therapeutic alliance, structuring of hours and processes, global assessment of behavioral therapy competence skills, as well as heuristics). The VTKC allows an economical approach: Only if 1 of the major categories is rated with a 'yes' are the items associated with it then rated on a 7-point Likert Scale, ranging from '(1) does not apply at all' to '(7) absolutely true.' As part of an extensive evaluation, 4 independent raters found the VTKC to have good reliability (ICC = 0.72) and moderate to very good internal consistencies for the 11 evaluation categories ($\alpha = 0.51$ to 0.90) [Linden et al., 2007a].

As already stated, supervision is a key element in achieving treatment integrity. The Stundenbogen zur Supervisionsbeurteilung (SSB, Post-Session Report on Supervision Assessment) [Zabocke et al., 2009] provides a method for evaluation of supervision and thus for verification and assurance of treatment integrity. The SSB is based on Grawe's general psychotherapy model [Grawe, 1998] and integrates the components *therapeutic relationship*, *problem solving*, and *motivational clarification*. The method includes a version for the supervisees as well as one for the supervisor and consists of 12 items that are rated on a 7-point Likert Scale (1–7). Both the version for the supervisees ($\alpha = 0.87$) and that for the supervisors ($\alpha = 0.83$) showed satisfactory to good internal consistencies. Both scales showed their concurrent validity by significant correlations with 3 items that covered overall satisfaction with the supervisory session [Zarbock et al., 2009].

All 3 methods presented can be used to assure and verify treatment integrity. While the CTS is an external assessment process, which requires independent raters and a previous intensive training program, the VTKC and the SSB permit assessment by the therapist/supervisor. Thus it can be stated that there are extensively validated methods in the German-speaking countries for the implementation and verification of treatment integrity.

Evaluation of Treatment Integrity

To evaluate treatment integrity, it must first be determined what sort of observations will be the basis for assessment of the treatment. Despite the high survey cost, as many therapy sessions as possible should be taken into account. It is advantageous here to divide the treatment into several phases (e.g., psychoeducation, confrontation, relapse prevention) and to evaluate treatment integrity at each stage if possible [Barber et al., 2007]. Earlier studies indicated that the use of segments of a session (segments of 3–5 min) for the assessment of treatment integrity was less reliable and less valid than using the entire session [Bachrach et al., 1981; Mintz and Luborsky, 1971]. A recent study demonstrated, however, that session segments do provide a reliable and valid basis for the assessment of treatment integrity, if the middle third of the session (20 min) was observed [Weck et al., 2010b]. In particular, the evaluation of competence should be undertaken by experienced therapists, as recent empirical findings show that assessments made by novices (students without clinical experience) are not sufficiently reliable [Weck et al., 2011b]. In principle, a higher number of raters leads to increased reliability of their assessments [Wirtz and Caspar, 2002]. Good inter-rater reliability (≥ 0.75 [Portney and Watkins, 2009]) is the prerequisite for further analyses, which could refer, for example, to the relationship between treatment integrity and treatment outcome. Sufficient reliability should be ensured by several hours of rater training, in which the scales to be used should be discussed and therapy videos should be assessed and discussed by the group, so as to achieve consensus on the assessment of the therapeutic adherence and competence. Studies have already shown 10 h of training to be sufficient to achieve good inter-rater reliability [Weck et al., 2010a]. But the complexity of the procedures used and of the treatments should be taken into account. Thus setting a criterion (e.g., raters must not differ from one another in their competence assessment at the end of the rater training by more than 2 points on a Likert Scale) helps to achieve sufficient inter-rater reliability.

Treatment Integrity and Treatment Outcome – Empirical Findings

The question arises whether adherent and competently conducted therapies also lead to improved therapeutic outcome. This question was pursued in a recently published meta-analysis [Webb et al., 2010]. No significant correlation was found between the treatment outcome and either adherence ($r = 0.02$) or competence ($r = 0.07$). Because of methodological variance, however, this result should be treated with caution. To clarify this point, we have looked at works from 1980 to 2009 in the *PsycINFO* and *Medline* databases, studying the relationship between adherence/competence and therapeutic outcome. We searched for *treatment integrity*, *manual adhe-*

rence, therapist adherence, therapist fidelity, treatment fidelity, process outcome relation, and therapist competence. So as not to increase the variance, we limited ourselves to works that used approaches based on cognitive behavioral therapy. Table 1 summarizes the results for adherence and table 2 those for competence. In most treatment studies, the patients suffered from depressive disorder. But there were also studies of anxiety disorders, substance disorders, eating disorders, and suicidal behavior.

For some studies that examine the impact of the therapist's adherence upon treatment outcomes, there are definitely no correlations [Godfrey et al., 2007; Huppert et al., 2001; Loeb et al., 2005]. A large proportion of the studies, however, show correlations between the treatment outcome and the therapists' adherence or competence. But these correlations are definite in only a few studies [DeRubeis and Feeley, 1990; Hogue et al., 2008; Kingdon et al., 1996; Kuyken et al., 2009; Trepka et al., 2004]. More frequently a correlation is demonstrated between treatment outcome and therapist adherence or competence for only some of the measures of outcome [Castonguay et al., 1996; Davidson et al., 2004; Feeley et al., 1999; Hoffart et al., 2005; Shaw et al., 1999] or occurs indirectly or in interaction with other variables [Bryant et al., 1999; Huppert et al., 2006]. The effect sizes of the correlations, which according to Webb et al. [2010] were reported as correlation coefficients (tables 1, 2), are in the low to moderate range.

On closer examination of the studies, it is clear that the monitoring of adherence and competence was based on very heterogeneous methodological approaches, and so the above-mentioned recommendations were frequently considered inadequate. While most studies monitored treatment integrity directly by video or audio recordings, there were significant differences in approach. Thus, some assessments are based on a single therapeutic session [e.g., Godfrey et al., 2007], while others considered almost every second session [e.g., Shaw et al., 1999]. In some there is random selection of sessions [e.g., Huppert et al., 2001], whereas in others there is consideration of a specific phase of therapy [e.g., Hogue et al., 2008] or session [e.g., Hoffart et al., 2005]. The number of raters varies from 1 [e.g., Trepka et al., 2004] to 13 [e.g., Feeley et al., 1999]. The qualification of the raters also shows significant variance. They range from students [e.g., DeRubeis and Feeley, 1990] to highly experienced therapists and supervisors [e.g., Kuykendall et al., 2009]. Some raters were trained to use the assessment instruments [e.g., Castonguay et al., 1996], but in most cases no details were provided about training of the raters. The limited comparability of the studies is also partly due to the different survey instruments. Thus, in some studies, adherence was measured using a single item [e.g., Huppert et al., 2001] and in others 50 were used [e.g., Loeb et al., 2005]. Only 2 studies on adherence used the same scales (Collaborative Study Psychotherapy Rating Scale). The studies on competence present a

somewhat more homogeneous picture in the use of scales, as 6 studies used the CTS and its revised version (CTS-R).

The heterogeneous methodology not only diminishes the comparability of these studies, but also the validity of the correlations found between therapeutic outcome and therapeutic adherence or competence. It is also problematic for the comparability of the studies that important information is lacking, such as about the raters, their training, the scales used to measure adherence and competence, the selection of therapeutic videos, or the effect sizes.

Discussion and Perspective

The present paper gives recommendations for the implementation, measurement, and evaluation of treatment integrity. It can be stated that, both for the research context and for psychotherapeutic care, there exist appropriate instruments for monitoring treatment integrity. More use should be made of these, within the framework of research studies, but also of education, training, and supervision to verify and assure treatment integrity.

Looking at the academic situation shows clearly that the low level of consideration of treatment integrity in the research context is in stark contrast to its importance. It was shown that in previous studies, no consistent standards were used. The heterogeneity of the studies that investigate the correlations between treatment integrity and treatment outcome, means that the studies are quite difficult to compare. Accordingly, recommendations, such as those given in the present study, are that greater consideration should be given in future investigations to make sure that comparative analyses (e.g., meta-analyses) can lead to more meaningful results. Therefore, the presentation of the methods used should be careful and complete.

It should be noted, however, that the above-mentioned studies on the correlation between treatment integrity and treatment outcome yielded quite promising results. Although some studies showed no correlation between treatment integrity and treatment outcome, many others did suggest a direct or indirect relationship between them.

In summary, implementation, measurement, and evaluation of treatment integrity are important tasks for psychotherapeutic research. However, in order to also improve the education and training of psychotherapists, the most promising instruments should be found to verify and ensure treatment integrity.

Disclosure Statement

The authors declare that they have no conflicts of interest regarding this work.

Table 1. Studies that examine the correlation between therapist adherence and treatment outcome

Study	Diagnosis, n	Monitoring of adherence	Results	Effect sizes, r
Bryant et al., 1999	depression, 26	rater: 1 doctoral candidate in clinical psychology; ratings: 2nd, 12th, and 18th session; instrument: Therapist Homework Assignment Competency Scale (THACS) developed by the authors	correlations were shown between the regular checking of homework and the patient's homework compliance. The completion of homework was in turn correlated with treatment outcome.	0.39 ^b
Castonguay et al., 1996	depression, 30	raters: 3 professional psychotherapists in training; ratings: 3 10-min segments of a randomly selected transcribed audio recording from the 4th and 7th sessions; instrument: Coding system of Therapists Focus (CSTF) [Goldfried et al., 1989]	results were mixed. But one partial result was that overly strict adherence was associated with relationship problems, with an increase of depressive symptoms according to a measure of depression.	-0.03 ^a
DeRubeis and Feeley, 1990	depression, 25	raters: 2 psychology students, no precise information available on the training of the raters; ratings: 4 sessions of all patients (between the 2nd and 12th sessions), randomized selection; instrument: Collaborative Study Psychotherapy Rating Scale (CSPRS) [Hollon et al., 1988]	an adherent method measured in the 2nd session predicts symptom improvement on the Beck Depression Inventory (BDI) [Beck et al., 1961].	0.24 ^a
Feeley et al., 1999	depression, 32	raters: 13 trained psychology students; ratings: 3 sessions (2nd session as well as 1 randomly selected from the 3rd and 4th quarters of treatment); instrument: CSPRS	for a subscale of the CSPRS, there was a correlation with the reduction of depressive symptoms (however there was no correlation for the other subscale).	-0.03 ^a
Godfrey et al., 2007	chronic fatigue, 71	raters: 3 experienced behavioral therapists and 1 rater; ratings: 71 audiotapes (every 3rd meeting of all patients was assessed by 2 raters); instrument: specially developed Primary Care Therapy Process Rating Scale	the adherence of the therapists had no effect on the treatment outcome.	0.00 ^d
Hogue et al., 2008	drug use among adolescents, 136	raters: 7 experienced behavioral therapists; ratings: videotapes of phase 1 (the first 2 tapes from sessions 1-5) and phase 2 (3 randomly chosen, consecutive tapes from session 6); instrument: Therapist Behavior Rating Scale-Competence (TBRS-C)	higher adherence was associated with reduced drug use. A medium adherence level had a positive effect on the reduction of behavioral problems, whereas a low or high level did not.	0.21 ^a
Huppert et al., 2001	panic disorder, 183	raters: trained raters, no detailed information about the experience or training of the raters; ratings: 330 randomly selected tapes; instrument: no information	successful and less successful therapists were not different in their adherence.	0.04 ^c
Huppert et al., 2006	panic disorder, 205	raters: apprentice psychotherapists already trained in the treatment method used; ratings: at least 3 audiotapes; instrument: self-developed scale [Barlow et al., 2000]	adherence alone had no significant effect on the treatment outcome. Adherence and motivation for treatment interacted, however, and were predictors of treatment outcome.	-0.24 ^a
Loeb et al., 2005	bulimia, 81	raters: 3 psychotherapists in training, no details about their experience or training; ratings: 3 audiotapes each per patient from the 6th, 12th, and 18th sessions (total of 243 tapes); instrument: Minnesota Therapy Rating Scale (MTRS) [DeRubeis et al., 1982]	no correlation was shown between therapist adherence and treatment outcome.	-0.19 ^a

^aEffect sizes could be carried over from the meta-analysis by Webb et al. [2010].

^bEffect sizes were in the original studies as correlations and could be carried over.

^cEffect sizes were not submitted as correlations and were therefore calculated on the basis of other statistical parameters (F- or t-values) [Rustenbach, 2003].

^dEffect sizes were submitted neither as correlations nor in the form of other statistical parameters. Following the procedure of Webb et al. [2010], the effect sizes were conservatively set to 0 if it was reported that no significant correlations existed.

Table 2. Studies that examine the correlation between the competence of the therapist and treatment outcome

Study	Diagnosis, n	Monitoring of competence	Results	Effect sizes, r
Bryant et al., 1999	depression, 26	rater: 1 doctoral candidate in clinical psychology; ratings: 2nd, 12th, and 18th session; instrument: Cognitive Therapy Scale (CTS) and Therapist Homework Assignment Competency Scale (THACS) developed by the authors	there was a correlation between overall therapeutic competence and doing the homework assignments, which in turn were correlated with therapeutic outcome.	0.32 ^b
Davidson et al., 2004	self-harm, parasuicidal behavior, 161	raters: not available; ratings: 49 randomly selected audiotapes; instrument: a competence scale developed by the authors: Manual Assisted Cognitive Therapy Rating Scale (MACT Rating Scale)	there were correlations between therapeutic competence and the reduction of depressive symptoms (but only in the assessment of others, not in self-assessment).	0.21 ^c
Hoffart et al., 2005	panic disorder, personality accentuation in Cluster C, 35	raters: 2 psychologists; ratings: 3 sessions with each patient; instrument: CTS	therapeutic competence did not predict global treatment success, but was correlated with a reduction in patients' dysfunctional schemata.	-0.04 ^a
Huppert et al., 2001	panic disorder, 183	raters: trained raters, no information about them or their training; ratings: 526 randomly selected tapes; instrument: no information	successful therapists exhibited a higher level of competence than less successful therapists.	0.13 ^c
Kingdon et al., 1996	dysthymia, panic disorder, generalized anxiety disorder, 70	rater: an experienced therapist and supervisor; ratings: global assessment of competence by the supervisor (without video- or audiotapes); instrument: CTS	therapists rated as competent were more successful than therapists who were rated as less competent or for whom the rating was uncertain (particularly regarding the reduction of depressive symptoms).	0.18 ^c
Kuyken et al., 2009	depression, 69	rater: 1 very experienced therapist; ratings: no further information; instrument: Evaluation of Therapist's Behavior Form (ETBF) and Patient's Report of Therapy Form (PRTF), scales developed for the study	a higher level of therapeutic competence led to a better treatment outcome, independent of the patients' comorbid disorders.	0.32 ^a
Shaw et al., 1999	depression, 36	rater: 2 experienced therapists and supervisors; ratings: 9 out of 20 therapy sessions; instrument: CTS	a more competent approach led to more significant reductions in depressive symptoms. However, the correlations were not consistent for all measures of depression.	0.14 ^a
Trepka et al., 2004	depression, 30	rater: 1 experienced therapist and expert in cognitive therapy; ratings: 1 randomly selected session between the 3rd and penultimate meeting; instrument: CTS	more competent therapists had more successful therapies and fewer discontinuations of treatment.	0.28 ^a

^aEffect sizes could be carried over from the meta-analysis by Webb et al. [2010].

^bEffect sizes were in the original studies as correlations and could be carried over.

^cEffect sizes were not submitted as correlations and were therefore calculated on the basis of other statistical parameters (F- or t-values) [Rustenbach, 2003].

References

- Bachrach H, Curtis H, Escoll P, et al.: Units of observation and perspectives on the psychoanalytic process. *Br J Med Psychol* 1981;54:25–33.
- Barber JP, Triffleman E, Marmer C: Considerations in treatment integrity: implications and recommendations for PTSD research. *J Trauma Stress* 2007;20:793–805.
- Barlow DH, Gorman JM, Shear MK, Woods SW: Cognitive-behavioral therapy, imipramine, or their combination for panic disorder: a randomized controlled trial. *JAMA* 2000;283:2529–2536.
- Beck AT, Ward CH, Mendelson M, et al.: An inventory for measuring depression. *Arch Gen Psychiatry* 1961;4:561–571.
- Bell AJ, Borrelli B, Resnick B, et al.: Enhancing treatment fidelity in health behavior change studies: best practices and recommendations from the NIH Behavior Change Consorting. *Health Psychol* 2004;23:443–451.
- Bryant MJ, Simons AD, Thase ME: Therapist skill and patient variables in homework compliance: controlling an uncontrolled variable in cognitive therapy outcome research. *Cognit Ther Res* 1999;23:381–399.
- Carroll KM, Nuro KF: One size cannot fit all: a stage model for psychotherapy manual development. *Clin Psychol Sci Prac* 2002;9:396–406.
- Castonguay LG, Goldfried MR, Wiser S, et al.: Predicting the effect of cognitive therapy for depression: a study of unique and common factors. *J Consult Clin Psychol* 1996;64:497–504.
- Chevron ES, Rounsaville BJ: Evaluating the clinical skills of psychotherapists: a comparison of techniques. *Arch Gen Psychiatry* 1983;40:1129–1132.
- Davidson K, Scott J, Schmidt U, et al.: Therapist competence and clinical outcome in the prevention of parasuicide by manual assisted cognitive behavior therapy trial: the POPMACT study. *Psychol Med* 2004;34:855–863.
- DeRubeis RJ, Feeley M: Determinants of change in cognitive therapy for depression. *Cognit Ther Res* 1990;14:469–482.
- DeRubeis RJ, Hollon SD, Evans SD, et al.: Can psychotherapies for depression be discriminated? A systematic investigation of cognitive therapy and interpersonal therapy. *J Consult Clin Psychol* 1982;50:744–756.
- Ehrhardt KE, Barnett DW, Lentz FE, et al.: Innovative methodology in ecological consultation: use of scripts to promote treatment acceptability and integrity. *Sch Psychol Q* 1996;2:149–168.
- Feeley M, DeRubeis RJ, Gelfand LA: The temporal relation of adherence and alliance to symptom change in cognitive therapy for depression. *J Consult Clin Psychol* 1999;67:578–582.
- Godfrey E, Chaider T, Ridsdale L, et al.: Investigating the 'active ingredients' of cognitive behaviour therapy and counseling for patients with chronic fatigue in primary care: developing a new process measure to assess treatment fidelity and predict outcome. *Br J Clin Psychol* 2007;46:253–272.
- Goldfried MR, Newman CN, Hayes AM: The Coding System of Therapeutic Focus. Unpublished training manual, State University of New York at Stony Brook, 1989.
- Grawe K: *Psychologische Therapie*. Bern, Hans Huber, 1998.
- Gresham FM, MacMillan DL, Beebe-Frankenberger ME, et al.: Treatment integrity in learning disabilities interventions research: Do we really know how treatments are implemented? *Learn Disabil Res Pract* 2000;15:198–205.
- Hoffart A, Sexton H, Nordahl HM, et al.: Connection between patient and therapist and therapist's competence in schema-focused therapy of personality problems. *Psychother Res* 2005;15:409–419.
- Hogue A, Henderson CE, Dauber S, et al.: Treatment adherence, competence, and outcome in individual and family therapy for adolescent behaviour problems. *J Consult Clin Psychol* 2008;76:544–555.
- Hollon SD, Evans MD, Auerbach A, et al.: Development of a system for rating therapies for depression: Differentiating cognitive therapy, interpersonal psychotherapy and clinical management pharmacotherapy. Unpublished manuscript, Nashville, Vanderbilt University, 1988.
- Huppert JD, Bufka LF, Barlow DH, et al.: Therapists, therapist variables, and cognitive-behavioral therapy outcome in a multicenter trial for panic disorder. *J Consult Clin Psychol* 2001;5:747–755.
- Huppert JD, Gorman JM, Shear MK, et al.: The interaction of motivation and therapist adherence predicts outcome in cognitive behavioral therapy for panic disorder: preliminary findings. *Cogn Behav Pract* 2006;13:198–204.
- Kingdon D, Tyrer P, Seivewright N, et al.: The Nottingham study of neurotic disorder: influence of cognitive therapists on outcome. *Br J Psychiatry* 1996;169:93–97.
- Kuyken W, Tsvirikos D: Therapist competence, comorbidity and cognitive-behavioral therapy for depression. *Psychother Psychosom* 2009;78:42–48.
- Linden M, Langhoff C: Verhaltenstherapie-Kompetenz-Checkliste: Kompetenzerfassung, Qualitätssicherung und Supervision. *Psychotherapeut*, 2011; 55:477–484.
- Linden M, Langhoff C, Rotter M: Definition, operationalization and quality assurance of psychotherapy. An investigation with the behavior therapy-competence-checklist (BTCC). *Psychiatr Danub* 2007a;19:308–316.
- Linden M, Müller C, Milew D: Das Mehrebenen-Modell psychotherapeutischer Kompetenz. *Verhaltenstherapie* 2007b;17:52–59.
- Loeb KL, Wilson GT, Labouvie E, et al.: Therapeutic alliance and treatment adherence in two interventions for bulimia nervosa: a study of process and outcome. *J Consult Clin Psychol* 2005;73:1097–1106.
- Mintz J, Luborsky L: Segments versus whole sessions: which is the better unit for psychotherapy process research? *J Abnorm Psychol* 1971;78:180–191.
- Perepletchikova F, Hilt LM, Chereji E, et al.: Barriers to implement treatment integrity procedures: survey of treatment outcome researchers 2009;77:212–218.
- Perepletchikova F, Kazdin AE: Treatment integrity and therapeutic change: issues and research recommendations. *Clin Psychol* 2005;12:365–383.
- Perepletchikova F, Treat TA, Kazdin AE: Treatment integrity in psychotherapy research: analysis of the studies and examination of the associated factors. *J Consult Clin Psychol* 2007;75:829–841.
- Pohl U, Richter H, Bohus M: Neuere Entwicklungen zur systematischen Überprüfung des therapeutischen Vorgehens: Skalen zur Erfassung von adherence und competence. *Psychother Psychosom Med Psychol* 2000;50:22–28.
- Portney LG, Watkins MP: *Foundations of Clinical Research: Applications to Practice*, 3rd ed. Upper Saddle River, Pearson Prentice Hall, 2009.
- Rustenbach SJ: *Metaanalyse: Eine anwendungsorientierte Einführung*. Bern, Hans Huber, 2003.
- Shaw BF, Elkin I, Yamaguchi J, et al.: Therapist competence ratings in relation to clinical outcome in cognitive therapy of depression. *J Consult Clin Psychol* 1999;67:837–846.
- Sholomskas DE, Syracuse-Siewert G, Rounsaville BJ, et al.: We don't train in vain: a dissemination trial of three strategies of training clinicians in cognitive-behavioral therapy. *J Consult Clin Psychol* 2005; 73:106–115.
- Staats M, Bär T, Linden M: Messinstrumente der Therapeutencompliance in der Verhaltenstherapie. *Verhaltenstherapie* 2003;13:62–67.
- Sterling-Turner HE, Watson TS, Moore JW: The effects of direct training and treatment integrity on treatment outcomes in school consultation. *Sch Psychol Q* 2002;17:47–77.
- Trepka C, Rees A, Shapiro DA, et al.: Therapist competence and outcome of cognitive therapy for depression. *Cognit Ther Res* 2004;28:143–157.
- Vermilyea BB, Barlow DH, O'Brien GT: The importance of assessing treatment integrity: an example in the anxiety disorders 1984;6:1–11.
- Waltz J, Addis ME, Koerner K, et al.: Testing the integrity of a psychotherapy protocol: assessment of adherence and competence. *J Consult Clin Psychol* 1993;61:620–630.
- Webb CA, DeRubeis RJ, Barber JP: Therapist adherence/competence and treatment outcome: a meta-analytic review. *J Consult Clin Psychol* 2010;78:200–211.
- Weck F, Hautzinger M, Heidenreich T, et al.: Erfassung psychotherapeutischer Kompetenzen: Validierung einer deutschsprachigen Version der Cognitive Therapy Scale. *Z Klin Psychol Psychother* 2010a;39:244–250.
- Weck F, Bohn C, Ginzburg DM, et al.: Assessment of adherence and competence in cognitive therapy: comparing session segments with entire sessions. Submitted for publication, 2010b.
- Weck F, Hautzinger M, Heidenreich T, et al.: Psychoedukation bei depressiven Störungen: Erfassung von Interventionsmerkmalen und Behandlungskompetenzen. *Psychother Psychosom Med Psychol* 2011a;61:148–153.
- Weck F, Hilling C, Schermelleh-Engel K, et al.: Reliability of adherence and competence assessment in cognitive behavioral therapy: influence of clinical experience. *J Nerv Ment Dis* 2011b;199:276–279.
- Wickstrom KF, Jones KM, LaFleur LH, et al.: An analysis of treatment integrity in school-based behavioral consultation. *Sch Psychol Q* 1998;13:141–154.
- Wirtz M, Caspar F: *Beurteilerübereinstimmung und Beurteilerreliabilität*. Göttingen, Hogrefe, 2002.
- Yeaton WH, Sechrest L: Critical dimensions in the choice and maintenance of successful treatment: strength, integrity, and effectiveness. *J Consult Clin Psychol* 1981;49:156–167.
- Young J, Beck AT: The development of the Cognitive Therapy Scale. Unpublished manuscript, Philadelphia, Center for Cognitive Therapy, 1980.
- Zarbock G, Drews M, Bodansky A, et al.: The evaluation of supervision: construction of brief questionnaires for the supervisor and the supervisee. *Psychother Res* 2009;19:194–204.