

**ENTWICKLUNG UND WEITERENTWICKLUNG VON
INSTRUMENTEN ZUR ERFASSUNG DER
BEHANDLUNGSINTEGRITÄT**

DISSERTATION

zur Erlangung des Doktorgrades der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt dem Fachbereich Psychologie und Sportwissenschaften

der Goethe-Universität Frankfurt am Main

von

FLORIAN GRIKSCHKEIT

geboren am 16.06.1982 in Frankfurt am Main

Frankfurt 2015

vom Fachbereich Psychologie und Sportwissenschaften der Goethe-Universität
Frankfurt am Main als Dissertation angenommen

Dekan: Prof. Dr. Dr. Winfried Banzer

Erstgutachter: PD Dr. Florian Weck

Zweitgutachter: Prof. Dr. Sonja Rohrmann

Datum der Disputation: 13.01.2016

ERKLÄRUNG

Ich erkläre hiermit, dass ich mich bisher keiner Doktorprüfung unterzogen habe.

Frankfurt am Main, 14.10.2015

.....

(Florian Grikscheit)

EIDESSTATTLICHE VERSICHERUNG

Ich erkläre hiermit an Eides Statt, dass ich die vorgelegte Dissertation

**„Entwicklung und Weiterentwicklung von Instrumenten zur Erfassung der
Behandlungsintegrität“**

selbstständig angefertigt und mich anderer Hilfsmittel als der in ihr angegebenen nicht bedient habe, insbesondere, dass alle Entlehnungen aus anderen Schriften mit Angabe der betreffenden Schrift gekennzeichnet sind. Ich versichere, nicht die Hilfe einer kommerziellen Promotionsvermittlung in Anspruch genommen zu haben.

Frankfurt am Main, 14.10.2015

.....

(Florian Grikscheit)

ZUSAMMENFASSUNG

Die vorliegende kumulative Dissertation befasst sich mit der Erfassung der Behandlungsintegrität bestehend aus psychotherapeutischer Adhärenz, Kompetenz sowie der Behandlungsdifferenzierung im Rahmen der Psychotherapieforschung. Die Überprüfung, ob Behandlungen bzw. Interventionen so wie intendiert durchgeführt wurden, ist für die Sicherstellung valider Schlussfolgerungen aus einer klinischen Studie von hoher Relevanz.

Die erste Studie untersucht, ob die Erfassung der Behandlungsintegrität ökonomischer gestaltbar ist. Es zeigte sich, dass Beurteilungen der Adhärenz und Kompetenz basierend auf Sitzungssegmenten im Vergleich zu ganzen Sitzungen keine Unterschiede aufweisen hinsichtlich Reliabilität, Validität und Prädiktion des Behandlungserfolgs.

In der zweiten Studie wird die Entwicklung und Validierung einer Adhärenz- und Kompetenzskala vorgestellt. Diese Studie weist zudem auf die Verwendung im Rahmen der Aus- und Weiterbildung von Therapeuten hin.

Die dritte Studie zeigt, dass in Psychotherapiestudien die im Vergleich stehenden Behandlungsbedingungen gut voneinander unterscheidbar sein müssen. Für die Beschreibung der Behandlungsdifferenzierung und -spezifität wurde der Behandlungs-Spezifitäts-Index entwickelt, dessen Eignung bestätigt werden konnte.

Die vierte Studie überprüft, ob sich erfolgreiche von nicht erfolgreichen Therapien hinsichtlich der psychotherapeutischen Kompetenz, Adhärenz und psychotherapeutischen Beziehung unterscheiden. Es zeigte sich, dass Adhärenz eine Voraussetzung für kompetentes Vorgehen darstellt. Kompetenz beeinflusst die psychotherapeutische Beziehung maßgebend, die mitentscheidend für den (Miss-)Erfolg einer Behandlung zu sein scheint.

Insgesamt tragen die Ergebnisse zu einer differenzierteren, spezifischeren und ökonomischeren Erfassung der Behandlungsintegrität innerhalb der Psychotherapieforschung bei. Gleichzeitig erweitern sie den Fokus auf neue Ansätze für zukünftige Forschungen.

INHALTSVERZEICHNIS

1	Theoretischer Hintergrund, Ziele der Studien und Integration	1
1.1	Behandlungsintegrität im Rahmen der psychotherapeutischen Prozessforschung	2
1.1.1	Definition psychotherapeutische Adhärenz	3
1.1.2	Definition psychotherapeutische Kompetenz.....	4
1.1.3	Definition Behandlungsdifferenzierung.....	6
1.2	Methodisches Vorgehen bei der Evaluation von Behandlungsintegrität.....	7
1.2.1	Auswahl des Datenmaterials	8
1.2.2	Zusammenfassung der ersten Studie: „ <i>Assessing Treatment Integrity in Cognitive-Behavioral Therapy: Comparing Session Segments with Entire Sessions</i> “.....	11
1.3	Instrumente zur Erfassung der Behandlungsintegrität.....	13
1.3.1	Erfassung der psychotherapeutischen Kompetenz.....	13
1.3.2	Erfassung der psychotherapeutischen Adhärenz.....	15
1.3.3	Zusammenfassung der zweiten Studie: „ <i>Entwicklung und Validierung von Instrumenten zur Überprüfung der therapeutischen Adhärenz und Kompetenz bei der Behandlung der Panikstörung</i> “	17
1.3.4	Erfassung der Behandlungsdifferenzierung	18
1.3.5	Zusammenfassung der dritten Studie: „ <i>Evaluation of Treatment Integrity: The Treatment Specificity Index</i> “	21
1.4	Bedeutung der Behandlungsintegrität für den Behandlungserfolg	22
1.4.1	Zusammenfassung der vierten Studie: „ <i>Treatment failure in cognitive-behavioural therapy: Therapeutic alliance as a precondition for an adherent competent implementation of techniques</i> “	25
1.5	Gesamtdiskussion.....	27
1.5.1	Integration der Ergebnisse.....	27

1.5.2	Grenzen der Arbeit	28
1.5.3	Ausblick	30
1.6	Literaturverzeichnis.....	31
2	Publikationen.....	40
2.1	Weck, F., Grikscheit, F., Höfling, V. & Stangier, U. (2014). Assessing Treatment Integrity in Cognitive-Behavioral Therapy: Comparing Session Segments With Entire Sessions. <i>Behavior Therapy</i> , 45, 541-552.	41
2.2	Grikscheit, F., Lang, T., Kordt, A., Helbig-Lang, S., Geisler, K. & Weck, F., (2015a). Entwicklung und Validierung von Instrumenten zur Überprüfung der therapeutischen Adhärenz und Kompetenz bei der Behandlung der Panikstörung. <i>Zeitschrift für Klinische Psychologie und Psychotherapie</i> , 44, 1-9.	53
2.3	Grikscheit, F., Weck, W., Hautzinger, M., Heidenreich, T., Weigel, M., Rudari, V., Schimming, C. & Stangier, U. (2015). Evaluation of Treatment Integrity: The Treatment Specificity Index. <i>Journal of Clinical Psychology</i> , 71, 653-665.....	62
2.4	Weck, F., Grikscheit, F., Jakob, M., Höfling, V. & Stangier, U. (2015). Treatment failure in cognitive-behavioural therapy: Therapeutic alliance as a precondition for an adherent and competent implementation of techniques. <i>British Journal of Clinical Psychology</i> , 54, 91-108.	75
3	Curriculum Vitae.....	93

1 THEORETISCHER HINTERGRUND, ZIELE DER STUDIEN UND INTEGRATION

1.1 Behandlungsintegrität im Rahmen der psychotherapeutischen Prozessforschung

Zentraler Forschungsgegenstand der Psychotherapieforschung ist die Untersuchung der Wirksamkeit und Wirkungsweise psychotherapeutischer Verfahren (Comer & Kendel, 2013). Dabei wird im Wesentlichen versucht, über die gezielte Manipulation der unabhängigen Variable (in diesem Fall: Psychotherapie) und der anschließenden Messung der Auswirkungen auf die abhängige Variable (z.B. Schweregrad der Symptome eines Patienten) empirische Erkenntnisse zu gewinnen. Seit den Anfängen der Psychotherapieforschung in den 1920er Jahren fanden bedeutsame Entwicklungen in der Forschungsmethodik statt (vgl. Orlinsky, Rønnestad & Willutzki, 2004). Eine solche Neuerung stellte die Etablierung von Behandlungsmanualen in den 1980er Jahren dar (Luborsky & DeRubeis, 1984). Während zuvor eher naturalistisch gekennzeichnete Psychotherapien in Studien untersucht wurden, rückten mit dem Einzug der Manuale standardisierte Behandlungen im Rahmen von *randomisiert kontrollierten Studien* (RCT) in den Fokus der Psychotherapieforschung (vgl. Crits-Christoph, Gibbons & Mukherjee, 2013). Im Zuge dessen beschäftigte sich die Forschung vermehrt auch mit der Frage, ob die Manipulation der unabhängigen Variable tatsächlich wie beabsichtigt stattgefunden hat (vgl. Waltz, Addis, Koerner & Jacobson, 1993). Nur unter dieser Voraussetzung sind sinnvolle Schlussfolgerungen aus einer Studie möglich (Yeaton & Sechrest, 1981). Würden sich beispielsweise zwischen einer Behandlung A und einer Kontrollgruppe B widererwartend keine Unterschiede zeigen, könnte dies darauf zurückzuführen sein, dass in beiden Bedingungen von den Behandlungsvorgaben abgewichen wurde und somit eher ähnliche oder identische therapeutische Interventionen angewandt wurden (Pohl, Richter & Bohus, 2000). Große Unterschiede zwischen einer Behandlung X und einer Behandlung Y könnten nicht nur durch die unterschiedliche Wirksamkeit der Verfahren erklärt werden, sondern auch darauf zurückzuführen sein, dass die Therapeuten in

einer dieser Behandlungsbedingungen sehr viel besser trainiert wurden und daher die Behandlung sehr viel kompetenter durchgeführt wurde.

Die genannten Aspekte werden in der Psychotherapieforschung unter dem Begriff der *Behandlungsintegrität* (engl. *treatment integrity* oder *treatment fidelity*) zusammengefasst. Die Behandlungsintegrität beschreibt, inwieweit eine Behandlung so wie intendiert (d.h. gemäß den theoretischen und konzeptuellen Vorgaben) durchgeführt wurde (Perepletchikova & Kazdin, 2005; Pohl et al., 2000; Waltz et al., 1993; Yeaton & Sechrest, 1981). Dabei lassen sich unter der Behandlungsintegrität drei verschiedene Prozessmaße subsumieren (vgl. Abbildung 1): die *Adhärenz*, die psychotherapeutische *Kompetenz* und die *Behandlungsdifferenzierung* (Moncher & Prinz, 1991; Vermilyea, Barlow & O'Brien, 1984; Waltz et al., 1993; Yeaton & Sechrest, 1981). Die Behandlungsintegrität kann somit als ein wichtiger Einflussfaktor auf die Validität einer Studie gesehen werden (Bahr & Beck, 2009; Moncher & Prinz, 1991; Schlosser, 2002; Weck, Bohn, Ginzburg & Stangier, 2011b).

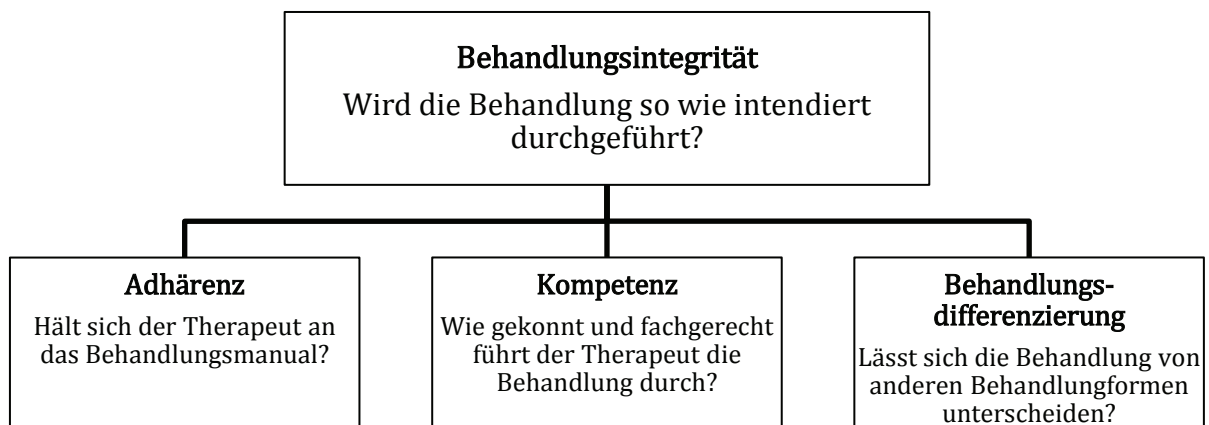


Abbildung 1 Behandlungsintegrität und deren Unterteilung nach Waltz et al. 1993 (vgl. Weck, 2013; S. 38).

1.1.1 Definition psychotherapeutische Adhärenz

Die Adhärenz (engl. adherence) oder Manualtreue beschreibt, inwieweit eine Behandlung oder eine einzelne Intervention gemäß den Vorgaben eines zugrundeliegenden

Manuals durchgeführt wurde. Ein Therapeut ließe sich als besonders adhärent beschreiben, wenn er sich beispielsweise an die genaue Reihenfolge von Übungen eines Manuals hält oder alle zur Verfügung stehenden Arbeitsblätter verwendet. Die Definition von Adhärenz schließt auch die Vermeidung von Interventionen aus anderen (nicht erwünschten) Therapieverfahren mit ein (Waltz et al., 1993). Demnach könnte man von einer Verringerung der Adhärenz einer verhaltenstherapeutischen Behandlung sprechen, wenn z.B. Interventionen aus der Gestalttherapie verwendet würden. Nach Waltz et al. (1993) ist es sinnvoll, Interventionen eines Behandlungsansatzes in vier unterschiedliche Kategorien einzuteilen, um die Adhärenz möglichst gut beschreiben zu können. Die erste Kategorie besteht aus den *spezifischen und essenziellen* (engl. *unique and essential*) Interventionen, d.h. solche, die charakteristisch und damit unverzichtbar für eine Behandlung sind und folglich nicht in einem anderen Behandlungsansatz verwendet werden (z.B. sokratischer Dialog bei einem kognitiven Therapieansatz). Die zweite Kategorie bilden die *essenziellen aber unspezifischen* (engl. *essential but not unique*) Interventionen, die von allgemeiner Natur sind und somit in mehreren Behandlungsansätzen zur Anwendung kommen (z.B. Etablierung einer therapeutischen Beziehung, Erstellung einer Tagesordnung). Interventionen, die *akzeptabel aber nicht notwendig* (engl. *acceptable but not necessary*) sind, können der dritten Kategorie zugeordnet werden (z.B. mit dem Patienten plaudern). Als eine letzte Kategorie können die *unerwünschten* (engl. *proscribed*) Interventionen aufgeführt werden. Als unerwünscht kann eine spezifische Intervention von Behandlung A in Behandlung B angesehen werden (z.B. eine Exposition in vivo innerhalb einer rein kognitiven Therapiebedingung).

1.1.2 Definition psychotherapeutischer Kompetenz

Das zweite Maß der Behandlungsintegrität ist die psychotherapeutische Kompetenz (engl. *competence*), mit deren Hilfe die kompetente Durchführung einer Behandlung oder

einer Interventionen beschrieben werden kann (Waltz et al., 1993). Unter einer kompetenten Durchführung ist zu verstehen, dass ein Therapeut bei der Implementierung einer Intervention verschiedene Kontextfaktoren berücksichtigt (Southam-Gerow & McLeod, 2013; Weck, 2013). Hierzu zählen unter anderem Patientenmerkmale (z.B. Störungsbild, Grad der Beeinträchtigung oder Lebensumstände), der aktuelle Stand der Therapie, der Zeitpunkt der Intervention und vieles mehr (vgl. Waltz et al., 1993). Eine Intervention wie beispielsweise ein sokratischer Dialog kann von einem Therapeuten weniger kompetent implementiert werden, dadurch dass er z.B. einen ungünstigen Zeitpunkt wählt (Auswahl der Intervention bzw. Zeitmanagement) oder er unangemessen viel Fachjargon verwendet (Art der Kommunikation). Insgesamt bedeutet eine kompetente Durchführung eine an der Situation und dem Patienten angepasste und somit eine möglichst zielführende und erfolgsversprechende Durchführung. Ergänzend dazu sei zu betonen, dass es sich in diesem Kontext nicht um die generelle Beschreibung der Kompetenz eines Therapeuten (im Sinne eines überdauernden Traits) handelt, sondern um die Beschreibung eines ganz spezifischen Verhaltens eines Therapeuten (Durchführungskompetenz) in einer bestimmten Situation und in Abhängigkeit eines bestimmten Behandlungsansatzes (Sharpless & Barber, 2009). Schließlich ist es auch denkbar, dass ein spezifisches therapeutisches Verhalten (z.B. Zeigen von Empathie) in einem Behandlungsansatz als besonders kompetent und in einem anderen Ansatz als genau das Gegenteil beurteilt werden würde (Waltz et al., 1993). Insgesamt wird deutlich, dass es sich bei der kompetenten Auswahl und Durchführung von psychotherapeutischen Interventionen um ein komplexes Geschehen handelt, bei dem nach Stangier (2015) diverse Informationsverarbeitungsprozesse (z.B. Wahrnehmung, Aufmerksamkeit, Vorstellungen und Erinnerungen) maßgeblich beteiligt sind. Ersichtlich wird außerdem, dass die Kompetenz und die Adhärenz nicht vollständig unabhängig voneinander betrachtet werden können. In gewisser Weise stellt die Adhärenz eine Grundlage

für die Kompetenz dar (vgl. Waltz et al., 1993). Allerdings bedeutet ein strenges adhärentes therapeutisches Vorgehen nicht zwangsläufig eine kompetente Durchführung einer Intervention (z.B. aufgrund der fehlenden Anpassung an den Patienten). Empirische Befunde verdeutlichen anhand von Korrelationen, die häufig über $r = .50$ liegen, den bedeutsamen Zusammenhang zwischen der Adhärenz und Kompetenz (vgl. Barber, Triffleman & Marmer, 2007). Trotz dieses hohen Zusammenhangs sollte jedoch nicht auf die separate Erfassung beider Konstrukte verzichtet werden, da von beiden ein unterschiedlicher prädiktiver Wert ausgehen kann (Barber et al., 2007).

1.1.3 Definition Behandlungsdifferenzierung

Die Behandlungsdifferenzierung oder Differenzierbarkeit (engl. treatment differentiation) spielt vor allem in Psychotherapiestudien eine Rolle, in denen zwei oder mehrere Behandlungsbedingungen miteinander verglichen werden. Die Behandlungsdifferenzierung ist ein Maß dafür, ob und inwiefern Behandlungsansätze voneinander unterscheidbar sind (Perepletchikova & Kazdin, 2005; Southam-Gerow & McLeod, 2013; Waltz et al., 1993). Dies stellt eine wichtige Voraussetzung für die Interpretierbarkeit der Ergebnisse einer Wirksamkeitsstudie dar. Eine niedrige Behandlungsdifferenzierung wäre beispielsweise in folgendem Szenario gegeben: In einer Studie soll Behandlung X mit Behandlung O verglichen werden. Beide Ansätze unterscheiden sich in ihren durchzuführenden Interventionen deutlich voneinander. Nachdem die jeweiligen Behandlungen durchgeführt wurden, stellt man fest, dass Behandlung X nicht mehr von O zu unterscheiden ist. Möglicherweise haben Therapeuten aus Behandlung O Interventionen aus Ansatz X oder umgekehrt durchgeführt. Somit wären Rückschlüsse auf die Wirksamkeit der beiden Ansätze nur noch unter Vorbehalt möglich.

1.2 Methodisches Vorgehen bei der Evaluation von Behandlungsintegrität

Die Erfassung der Behandlungsintegrität innerhalb der Psychotherapieforschung erfolgt meist auf unterschiedliche Art und Weise (Stangier, 2015; Webb, DeRubeis & Barber, 2010; Weck et al., 2011b). Als Datenbasis dienen häufig Videoaufzeichnungen (teilweise auch Audio- oder schriftliche Aufzeichnungen) von Therapiesitzungen, die im Kontext einer Studie durchgeführt werden. Diese Aufzeichnungen werden dann von unabhängigen Beurteilern (sog. *Ratern*) betrachtet und hinsichtlich der Behandlungsintegrität beurteilt. In einer Studie von Weck, Weigel, Richtberg und Stangier (2011c) zeigte sich, dass vor allem für die Erfassung von Kompetenz auf erfahrene Therapeuten zurückgegriffen werden sollte, anstatt die Beurteilungen durch Novizen (Studierende ohne klinische Erfahrungen) durchführen zu lassen. Für die Beurteilung der Behandlungsintegrität verwenden die Rater idealerweise standardisierte bzw. evaluierte Ratingskalen. Anhand von statistischen Analysen werden in einem weiteren Schritt die Beurteilungen der Rater miteinander verglichen. Hierbei strebt man eine möglichst hohe *Interraterreliabilität* an. Nach Portney und Watkins (2008) kann eine Interraterreliabilität ≥ 0.75 als gut interpretiert werden. Sie sind Voraussetzungen für weitere statistische Analysen, wie die Untersuchung des Zusammenhangs von Behandlungsintegrität und Behandlungserfolg (siehe Kapitel 1.4). Um solche hohe Interraterreliabilitäten zu erreichen, sollten die Beurteiler idealerweise in einem mehrstündigen Training in der Verwendung der Skalen zur Erfassung der Adhärenz und/oder Kompetenz geschult werden. In einer Studie erwies sich ein 10-stündiges Training als ausreichend um eine gute Interraterreliabilität zu erzielen (Weck, Hautzinger, Heidenreich & Stangier, 2010).

Auch wenn die Erfassung der Behandlungsintegrität im Rahmen einer Psychotherapiestudie sinnvoll und unerlässlich erscheint, zeigt sich, dass dies in der Praxis

häufig vernachlässigt wird (Moncher & Prinz, 1991). Perepletchikova, Treat und Kazdin (2007) analysierten randomisiert kontrollierte Studien aus insgesamt sechs renommierten internationalen Fachzeitschriften (*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* und *Journal of Clinical Psychiatry*) aus den Jahrgängen 2000-2004 hinsichtlich deren Evaluation der Behandlungsintegrität. Sie konnten nachweisen, dass lediglich bei 3.5% aller Studien die Behandlungsintegrität in angemessener Weise erfasst und analysiert wurde. Eine darauf aufbauende Forschungsarbeit benennt als Hauptgründe für die mangelnde Berücksichtigung, den Mangel an theoretischem Wissen, fehlende Richtlinien für die Implementierung der Behandlungsintegrität sowie den hohen Zeit- und Kostenaufwand (Perepletchikova, Hilt, Chereji & Kazdin, 2009). Folglich stellt sich die Frage, wie der Prozess der Evaluation kosteneffektiver gestaltet werden kann, um so die Quote der Erfassung zu erhöhen. Ein Aspekt, der nicht nur für die ökonomische Optimierung, sondern auch für das weitere Vorgehen bei der Messung der Behandlungsintegrität von entscheidender Bedeutung ist, ist die Auswahl des Datenmaterials.

1.2.1 Auswahl des Datenmaterials

Zur Beurteilung von Prozessmaßen wie Adhärenz und Kompetenz können schriftliche Behandlungsprotokolle (indirekte Einschätzung) oder aber auch schriftliche Transkriptionen von Therapiesitzungen sowie Ton- oder Videoaufzeichnungen (direkte Einschätzung) analysiert werden. Da Therapeuten jedoch zur Überschätzung der Behandlungsintegrität neigen, ist den direkten Einschätzungen mit einer möglichst hohen Informationsdichte, wie dies bei Videoaufnahmen der Fall ist, den Vorzug zu geben (Martino, Ball, Nich, Frankforter & Carrol, 2009; Waltz et al., 1993; Wickstrom, Jones, LaFleur & Witt, 1998).

Mit Ausnahme von spezifischen Fragestellungen, in denen eine ganz bestimmte Therapiesitzung oder Intervention im Fokus steht, sollten möglichst mehrere Therapiesitzungen für die Beurteilung der Behandlungsintegrität analysiert werden. Nach Dennhag, Gibbons, Barber, Gallop und Crits-Christoph (2012) benötigt man für eine hinreichende Beurteilung der Adhärenz und Kompetenz eines Therapeuten fünf bis zehn Therapiesitzungen. Dieses Vorgehen bedeutet jedoch einen hohen zeitlichen und damit auch finanziellen Aufwand.

Bei dem Vergleich des methodischen Vorgehens von Studien zur Prozessforschung wird deutlich, dass in den meisten Fällen Therapiesitzungen in ihrer vollständigen Länge beurteilt werden (z.B. Barber, Crits-Christoph & Luborksy, 1996; Feeley, DeRubeis & Gelfand, 1999; Huppert, Barlow, Gorman, Shear & Woods, 2006; Markowitz, Spielman, Scarvalone & Perry, 2000; Shaw et al., 1999; Zobel, Karim, Kech, Berger & Schramm, 2008). Für die Erreichung eines ökonomischeren Vorgehens wäre es günstig, lediglich repräsentative Ausschnitte einer Therapiesitzung für die Beurteilung der Behandlungsintegrität zu verwenden. Es liegen bereits Studien wie die von Luborsky, McLellan, Woody, O'Brien & Auerbach (1985) vor, die Sitzungsausschnitte für die Beurteilung der Adhärenz verwenden. Hier zeigten sich 15-minütige Ausschnitte einer Therapie als nutzbar für die Vorhersage des Behandlungserfolgs. Darüber hinaus existieren nur wenige Studien, die einen direkten Vergleich der Nützlichkeit von Ausschnitten und gesamten Sitzungen für die Erfassung von Prozessvariablen untersuchen (Bachrach et al., 1981; Mintz & Luborsky, 1971). In diesen beiden Studien wurden mehrere vier- bis fünfminütige Ausschnitte mit kompletten Sitzungen hinsichtlich verschiedener Prozessvariablen (u.a. Übertragungsprozesse, Empathie) in analytisch orientierten Therapien verglichen. Dabei wurde deutlich, dass sich die Beurteilungen basierend auf den Sitzungsausschnitten als weniger reliabel und valide im Vergleich zu den gesamten Sitzungen

erwiesen habe. Jedoch muss zum einen die sehr kurze Dauer der Ausschnitte und zum anderen die Übertragbarkeit auf die Behandlungsintegrität kritisch betrachtet werden.

Eine Arbeit von Weck, Bohn, Ginzburg und Stangier (2011a) weist auf die Nützlichkeit von Sitzungsausschnitten als Basis zur Beurteilung der Behandlungsintegrität hin. Im Rahmen dieser Studie wurden insgesamt 34 auf Video aufgezeichnete Therapiesitzungen einer kognitiven Therapie zur Behandlung der Sozialen Phobie in Drittel (erstes, mittleres und letztes Drittel) von einer durchschnittlichen Länge von ca. 20 Minuten unterteilt und mit der gesamten Therapiesitzung hinsichtlich der Adhärenz und Kompetenz verglichen. Dabei wiesen die Beurteilungen basierend auf dem mittleren Drittel zum einen eine zufriedenstellende Interraterreliabilität ($ICC_{\text{Adhärenz}} = 0.81$; $ICC_{\text{Kompetenz}} = 0.71$) auf und zum anderen zeigten sich deutliche Korrelationen ($r_{\text{Adhärenz}} = 0.65$; $r_{\text{Kompetenz}} = 0.71$) mit den Beurteilungen auf Grundlage der gesamten Sitzung. Es konnte auch nachgewiesen werden, dass die Einschätzungen basierend auf dem mittleren Drittel ($r_{\text{Adhärenz}} = 0.55$; $r_{\text{Kompetenz}} = 0.45$) in vergleichbarer Höhe den Therapieerfolg vorhersagen wie Einschätzungen beruhend auf der gesamten Therapiesitzung. Aus den Ergebnissen kann die Schlussfolgerung gezogen werden, dass Beurteilungen basierend auf dem mittleren Drittel einer Therapiesitzung eine geeignete Annäherung an die Beurteilungen von Adhärenz und Kompetenz auf Basis einer gesamten Sitzung darstellen. Eine erhebliche methodische Einschränkung dieser Studie besteht jedoch darin, dass die Beurteilung der jeweiligen Sitzungsdrittel von den gleichen Ratern durchgeführt wurde. Zwar wurden die Sitzungsdrittel in permutierter Reihenfolge präsentiert, jedoch können Erinnerungseffekte, die zu einer Überschätzung der Reliabilität und Validität führen könnten, nicht ausgeschlossen werden. Auch lässt sich die Generalisierbarkeit der Ergebnisse in Frage stellen, da sich die genannte Studie nur auf die Behandlung der Sozialen Phobie bezog.

Aus den genannten Gründen war es Ziel der ersten Studie zu untersuchen, ob sich Sitzungsausschnitte auch im Zusammenhang von anderen Störungsbildern und nach Ausschluss von Erinnerungseffekten im Vergleich zu ganzen Sitzungen als nützlich erweisen. Die Verwendbarkeit von Sitzungsausschnitten würde eine deutliche Reduktion des Zeitaufwandes und der Kosten für den gesamten Ratingprozess bedeuten. Dies wiederum könnte erheblich dazu beitragen, dass die Evaluation der Behandlungsintegrität im Rahmen von Vergleichsstudien deutlich häufiger angewandt werden würde, als dies von Perepletchikova und Kazdin (2005) beschrieben wurde.

1.2.2 Zusammenfassung der ersten Studie: „Assessing Treatment Integrity in Cognitive-Behavioral Therapy: Comparing Session Segments with Entire Sessions“

Fragestellung. Diese Studie prüft, ob Beurteilungen basierend auf Sitzungsausschnitten gleichermaßen geeignet sind wie gesamte Therapiesitzungen, um die psychotherapeutische Adhärenz und Kompetenz reliabel und valide zu erfassen. Ebenfalls soll untersucht werden, ob sich Unterschiede hinsichtlich der Vorhersagbarkeit des Therapieerfolgs zwischen einem Sitzungssegment und einer gesamten Sitzung finden lassen.

Methode. Die Datengrundlage dieses Vergleichs lieferten 84 randomisiert ausgewählte Behandlungsvideos aus drei klinischen Studien (jeweils 28 Behandlungen pro Studie) zur Untersuchung von kognitiv-verhaltenstherapeutischen Behandlungsansätzen im Rahmen der Behandlung von Sozialer Phobie (Stangier, Schramm, Heidenreich, Berger & Clark, 2010), Hypochondrie (Weck, Neng, Richtberg, Jakob & Stangier, 2015b) und der Rückfallprophylaxe bei rezidivierender Depression (Stangier et al., 2013). Vier klinisch erfahrene und trainierte Rater nahmen die Beurteilung der Adhärenz und Kompetenz anhand der Therapievideos vor. Jede Behandlungssitzung wurde jeweils von zwei Ratern anhand eines Sitzungssegments (mittleres Drittel der Gesamtsitzung) und von jeweils zwei Ratern

anhand der Gesamtsitzung beurteilt. Die Reihenfolge der Raterkombinationen sowie der Beurteilung von Sitzungssegmenten und vollständigen Sitzungen wurde permutiert angeordnet. Insgesamt konnten 336 Ratings generiert und statistisch ausgewertet werden.

Ergebnisse. Die Interraterreliabilitäten für die Adhärenz ($ICC_{[1,2]} = 0.72, p = < .001$) und Kompetenzbeurteilungen ($ICC_{[1,2]} = 0.72, p = < .001$) basierend auf Sitzungssegmenten erwiesen sich als zufriedenstellend. Es konnte eine Vergleichbarkeit der Reliabilität und Validität der Beurteilungen basierend auf den Sitzungssegmenten und gesamten Sitzungen nachgewiesen werden. Zwischen den Beurteilungen der Sitzungssegmente und vollständigen Sitzungen zeigten sich hohe Korrelationen bzgl. der Adhärenz ($r = 0.62, p < .001$) und für die Kompetenz ($r = 0.73, p < .001$). Ein signifikanter Zusammenhang zwischen Kompetenz und Behandlungserfolg konnte lediglich bei der Behandlung der Sozialen Phobie sowohl auf der Basis von Sitzungssegmenten ($r = 0.39, p < .05$) als auch auf Grundlage der gesamten Sitzung ($r = 0.46, p < .01$) gefunden werden. Zwar erwiesen sich die übrigen Zusammenhänge zwischen der Adhärenz / Kompetenz und dem Behandlungserfolg als nicht signifikant, jedoch zeigten sich keine Unterschiede darin, ob die Beurteilungen mittels Segmenten oder auf Basis der Gesamtsitzung erfolgten.

Schlussfolgerung. Es konnte gezeigt werden, dass Beurteilungen der Adhärenz und Kompetenz auf Grundlage von Sitzungsausschnitten (mittleres Drittel) adäquat durchgeführt werden können. Die so erzielten Beurteilungen entsprechen weitestgehend den Beurteilungen basierend auf ganzen Sitzungen. Somit stellen sie eine geeignete und ökonomische Alternative zu gesamten Sitzungen dar.

1.3 Instrumente zur Erfassung der Behandlungsintegrität

Für die Beurteilung von Behandlungsintegrität sollte auf reliable und valide Skalen zurückgegriffen werden. Dabei zeigen sich erhebliche Unterschiede vor allem hinsichtlich der Konstruktion zwischen Instrumenten zur Adhärenz- und der Kompetenzerfassung.

Im Unterschied zur Erfassung von Adhärenz und Kompetenz finden sich in der Literatur kaum Hinweise für ein adäquates methodisches Vorgehen zur Evaluation der Behandlungsdifferenzierung. Die wenigen beschriebenen Verfahren (z.B. Luborsky, et al., 1985; Zobel et al. 2008) weisen gewisse methodische und theoretische Mängel auf (siehe Kapitel 1.3.4).

1.3.1 Erfassung der psychotherapeutischen Kompetenz

Vorteilhaft für die Erfassung der therapeutischen Kompetenz ist, dass auf bereits standardisierte und evaluierte Instrumente zurückgegriffen werden kann. Nach Weck (2013) existieren unterschiedliche Arten von Verfahren zur Kompetenzerfassung. Neben Verfahren, die Störungs- und Behandlungswissen evaluieren, kommen in der klinischen Forschung vor allem solche zum Einsatz, die die Durchführung von psychotherapeutischer Behandlung erfassen. Ein solches Instrument, welches häufig im Kontext der Kognitiven Verhaltenstherapie eingesetzt wird, ist die *Cognitive Therapy Scale* (CTS) von Young und Beck (1980) und in der deutschen Version von Weck et al. (2010). Diese Skala dient zur Einschätzung der psychotherapeutischen Kompetenz in der kognitiven Therapie. Die deutsche Version der CTS besteht aus insgesamt 14 Items die auf einer 7-stufigen Likertskala (0 = mangelhaft bis 6 = exzellent) beurteilt werden können. Für jede zweite Abstufung existieren Ankerbeschreibungen, um dem Beurteiler eine Hilfestellung zu ermöglichen (vgl. Abbildung 2). Die Items lassen sich in zwei Subskalen zusammenfassen. Die Skala „*allgemeine therapeutische Kompetenzen*“ umfasst die Items „Umgang mit

Problemen/Fragen/Einwänden“, „Klarheit der Kommunikation“, „interpersonelle Effektivität“, „Ressourcenaktivierung“, „Auswertung von Hausaufgaben“ als auch „Nutzung und Rückmeldung von Zusammenfassungen“. Die zweite Subskala „*sitzungsstrukturierende Kompetenzen*“ beinhaltet die Items „Tagesordnung“, „Zeitaspekte“, „geleitetes Entdecken“, „Fokus auf zentrale Kognitionen und Verhalten“, „Rational“, „Auswahl angemessener Strategien“, „angemessene Durchführung von Techniken“ sowie die „Vergabe von Hausaufgaben“. Des Weiteren lassen sich noch zwei ergänzende Items, ebenfalls auf einer 7-stufigen Likertskala, für die „Einstufung der Schwere der Behandlung“ und die „Gesamtkompetenz des Therapeuten“ beurteilen. Testanalytisch konnten für die

<p>5. Interpersonelle Effektivität Der Therapeut reagiert einfühlsam und wertschätzend auf die Problemdarstellung des Patienten. Ebenso ist das Beziehungsverhalten des Therapeuten durch Empathie und Wertschätzung gekennzeichnet. Der Therapeut ist in der Lage, dem Patienten Hoffnung auf einen Erfolg der Therapie zu vermitteln.</p>
<p>0 = mangelhaft Der Therapeut zeigt mangelhafte interpersonelle Fähigkeiten. Es ist nicht erkennbar, dass sich der Therapeut in die Lage des Patienten versetzen kann. Der Therapeut schüchtert den Patienten ein, ängstigt oder verunsichert ihn (z.B. wirkt der Therapeut dem Patienten gegenüber feindselig, herablassend oder in anderer Weise destruktiv). Der Patient beteiligt sich hierdurch nicht, verliert Vertrauen und/oder wird feindselig.</p>
<p>1 = gerade ausreichend</p>
<p>2 = mittelmäßig Der Stil des Therapeuten behindert zeitweise sein empathisches Verständnis. Der Patient zeigt wenig Vertrauen.</p>
<p>3 = zufriedenstellend</p>
<p>4 = gut Der Therapeut zeigt einen zufriedenstellenden Grad an Anteilnahme, Wertschätzung, Vertrauen und vermittelt dem Patienten Hoffnung auf Veränderung. Es sind keine bedeutsamen interpersonellen Probleme erkennbar. Der Therapeut ist in der Lage, explizite und implizite Mitteilungen in der Kommunikation mit dem Patienten zu verstehen.</p>
<p>5 = sehr gut</p>
<p>6 = exzellent Exzellente interpersonelle Effektivität. Der Therapeut zeigt einen optimalen Grad an Anteilnahme, Wertschätzung, Vertrauen und vermittelt dem Patienten in überzeugender Weise Hoffnung auf Veränderungen.</p>

Abbildung 2 Beispielim der deutschen Version der Cognitive Therapy Scale (CTS; Weck, et al., 2010) mit entsprechenden Ankerbeschreibungen

„allgemeinen therapeutischen Kompetenzen“ ($ICC_{[2,2]} = 0.85$; $\alpha = 0.71$) und „sitzungsstrukturierenden Kompetenzen“ ($ICC_{[2,2]} = 0.93$; $\alpha = 0.88$) gute bis sehr gute Beurteilungsübereinstimmungen und interne Konsistenzen nachgewiesen werden (Weck et al., 2010).

Auch wenn die CTS ein allgemein einsetzbares Instrument für die Kompetenzerfassung innerhalb einer kognitiv-verhaltenstherapeutischen Behandlung darstellt, kann es für die Klärung spezifischer Fragestellungen sinnvoll sein, dieses Instrument anzupassen. Daher entstanden in den letzten Jahren weitere Versionen durch detaillierte Beschreibungen des Therapeutenverhaltens und Spezifizierungen von Items u.a. für die Beurteilung der kognitiven Therapie der Sozialen Phobie (CTCS-SP; Consbruch, Clark & Stangier, 2012) und psychotischen Störungen (CTS-Psy; Haddock et al., 2001).

1.3.2 Erfassung der psychotherapeutischen Adhärenz

Im Unterschied zur Kompetenzerfassung kann bei der Evaluation der Adhärenz in der Regel nicht auf standardisierte Verfahren zurückgegriffen werden. Daher ist in vielen Fällen die Neuentwicklung einer Skala (bzw. Modifikation einer bestehenden Skala) notwendig. Dies ergibt sich aus dem Umstand, dass die Beurteilung der Adhärenz im hohen Maße von dem spezifischen Vorgehen einer Intervention oder eines Behandlungsansatzes abhängig ist. Idealerweise liegt einer zu untersuchenden Behandlung oder Intervention ein Manual zugrunde.

Ein Vorgehen, das sich für die Konstruktion einer solchen Adhärenzskala bewährte, besteht aus drei Schritten. Zunächst werden die Items anhand des zugrundeliegenden Behandlungsmanuals generiert. Hierbei erscheint es zum einen sinnvoll, eine Abstufung der Adhärenz anhand einer Likert-Skala (beispielsweise 0 = keine Manualtreue/nicht durchgeführt; 1 = geringe Manualtreue; 2 = moderate Manualtreue; 3 = hohe Manualtreue)

vorzunehmen und zum anderen das entsprechende Therapeutenverhalten möglichst detailliert zu beschreiben. Zur Sicherstellung, dass alle bedeutsamen Bestandteile einer Therapie oder einer Intervention ausreichend berücksichtigt wurden, wird in einem zweiten Schritt zu inhaltlichen Validierungen eine Expertenbefragung empfohlen (Weck et al., 2010). Dabei beurteilen die Experten die zuvor generierten Items hinsichtlich deren Relevanz („Für wie relevant halten Sie das Item?“) als auch bezüglich deren Eignung („Für wie geeignet halten Sie das Item?“). Über diese Beurteilungen hinaus können dann Anmerkungen und Verbesserungsvorschläge der Experten in einem dritten Schritten dann zur Überarbeitung der Skala dienen.

Wie die Studie von Perepletchikova et al. (2009) nachweisen konnte, ist ein Hinderungsgrund für die Erfassung der Behandlungsintegrität, dass keine für die jeweilige zu beantwortende Fragestellung geeigneten Instrumente vorhanden sind bzw. für die Konstruktion das entsprechende Wissen fehlt. Die Beantwortung von spezifischen Fragestellungen macht es daher notwendig, bestehende Instrumente anzupassen bzw. neu zu konstruieren, um die Behandlungsintegrität adäquat zu erfassen. Eine solche spezifische Frage lässt sich z.B. im Zusammenhang mit der Behandlung der Panikstörung benennen, für welche die benötigten Skalen zur Erfassung der Adhärenz und Kompetenz in der zweiten Studie konstruiert wurden. Bereits in mehreren Untersuchungen konnte die Exposition als eine zentrale und wirksame Intervention bei der Therapie der Panikstörung mit Agoraphobie beschrieben werden (Ruhmland & Margraf, 2001, Sánchez-Meca, Alcázar, Marín-Matínez & Gómez-Conesa, 2010). Dabei ist anzunehmen, dass der Expositionsvorbereitung (dies beinhaltet u.a. die Herleitung des Expositionsrationalis) eine hohe Bedeutung zukommt. Dazu liegen jedoch bislang noch keine empirischen Daten vor.

1.3.3 Zusammenfassung der zweiten Studie: „Entwicklung und Validierung von Instrumenten zur Überprüfung der therapeutischen Adhärenz und Kompetenz bei der Behandlung der Panikstörung“

Fragestellung. Ziel der vorliegenden Studie war die Entwicklung von zwei Instrumenten zur Evaluation der Adhärenz und Kompetenz bei der Vorbereitung von Expositionsübungen im Rahmen der Behandlung der Panikstörung mit Agoraphobie sowie deren statistische Validierung.

Methode. Die Datenbasis lieferte eine multizentrische, randomisiert kontrollierte Therapiestudie zur expositionsbasierten Behandlung der Panikstörung mit Agoraphobie. Auf Grundlage des Behandlungsmanuals (Lang, Helbig-Lang, Westphal, Gloster & Wittchen, 2012) wurden die Items der *Adhärenzskala Expositions Vorbereitung* (AS-EXVO) konstruiert. Die Konstruktion der *Kompetenzskala Expositions Vorbereitung* (KS-EXVO) orientierte sich an der Cognitive Therapy Scale (CTS; Weck et al., 2010; Young & Beck, 1980). Die Items der Skalen wurden vier Experten zur Inhaltsvalidierung vorgelegt. Die Validierung der Instrumente erfolgte auf der Basis von 84 randomisiert ausgewählten Behandlungsvideos.

Ergebnisse. Aufgrund der durchgeführten Expertenbefragung und der daran anschließenden Anpassung der Items kann von einer hohen Inhaltsvalidität der Skalen ausgegangen werden. Die statistische Auswertung zeigte eine hohe Interraterreliabilität für die AS-EXVO (zwischen $ICC_{[2,2]} = 0.62, p < .001$ und $ICC_{[2,2]} = 0.95, p < .001$) und für die KS-EXVO ($ICC_{[2,2]} = 0.77, p < .001$ und $ICC_{[2,2]} = 0.88, p < .001$). Es konnten interne Konsistenzen (Cronbachs α) für die Adhärenzskala von $\alpha = .82$ und für die Kompetenzskala von $\alpha = .71$ nachgewiesen werden. Für die Adhärenzskala wurde faktoranalytisch eine einfaktorielle Lösung gefunden. Für beide Verfahren wurde deren konvergente und diskriminante Validität bestätigt.

Schlussfolgerung. Es konnte gezeigt werden, dass es sich bei den konstruierten Verfahren zur Erfassung der Adhärenz (AS-EXVO) und Kompetenz (KS-EXVO) bei der Vorbereitung von Expositionsübungen bei dem Störungsbild der Panikstörung mit Agoraphobie um reliable als auch valide Instrumente handelt. Beide Verfahren können nicht nur in der klinischen Forschung, sondern auch in der Aus- und Weiterbildung von Therapeuten sinnvolle Anwendung finden.

1.3.4 Erfassung der Behandlungsdifferenzierung

Eine hohe Behandlungsdifferenzierung zu erreichen setzt voraus, dass die jeweiligen Behandlungsbedingungen einer Vergleichsstudie manualgetreu durchgeführt werden. Dabei ist es nicht nur von Bedeutung, ob eine Behandlungsbedingung „A“ hauptsächlich aus den Interventionen, die im Behandlungsmanual „A“ beschrieben sind, besteht (*erwünschte Interventionen*), sondern auch, ob möglichst wenig Interventionen aus Behandlung „B“ (*unerwünschte Intervention*) eingesetzt werden. Um dieses Verhältnis zu beschreiben, wurde

Purity Index (PI) = $\frac{\text{erwünschte Interventionen}}{\text{erwünschte Interventionen} + \text{unerwünschte Interventionen}}$

Beispiel: Berechnung des Purity Index für die Behandlung A und B anhand einer Adhärenzskala bestehend aus vier Items. Die Items „a₁“ und „a₂“ repräsentieren erwünschte Intervention aus Behandlung A und die Items „b₁“ und „b₂“ solche aus Behandlung B.

Behandlung	Item „a ₁ “	Item „a ₂ “	Item „b ₁ “	Item „b ₂ “
A	2	1	0	1
B	1	1	2	2

$$Purity\ Index_{(A)} = \frac{a_1 + a_2}{(a_1 + a_2) + (b_1 + b_2)} = \frac{2 + 1}{(2 + 1) + (0 + 1)} = 0.75$$

$$Purity\ Index_{(B)} = \frac{b_1 + b_2}{(b_1 + b_2) + (a_1 + a_2)} = \frac{2 + 2}{(2 + 2) + (1 + 1)} = 0.67$$

Abbildung 3 Formel des Purity Index nach Luborsky et al. (1985) und Beispielrechnung

in einigen Studien der sogenannte *Purity Index* (siehe Abbildung 3) oder auch *Reinheitsindex* berechnet (vgl. Luborsky et al., 1985; Stangier et al., 2013; Zobel, et al., 2008). Insgesamt kann die so beschreibbare Reinheit einer Behandlung als eine Unterform der Adhärenz angesehen werden. Eine Behandlung kann somit als besonders „rein“ interpretiert werden, sofern der Index sich einem Wert von 1 annähert.

Weisen nun alle Behandlungsbedingungen einer Vergleichsstudie eine hohe Reinheit auf, so ließe dies den Schluss zu, dass eine hohe Behandlungsdifferenzierung vorliegt. Jedoch liegt bei dieser Vorgehensweise eine Einschränkung dadurch vor, dass die Klassifikation der Interventionen in „erwünscht“ und „unerwünscht“ nicht ausreichend präzise ist. Unter den erwünschten Interventionen werden bei der Berechnung des Purity Index sowohl allgemeine Interventionen (z.B. Erstellen von Hausaufgaben) mit für die Behandlung spezifischen

Beispiel: Berechnung des Purity Index für die Behandlung A und B anhand einer Adhärenzskala, bestehend aus sechs Items. Die Items „a1“, „a2“ spiegeln spezifische Intervention der Behandlung A wider und die Items „b1“, „b2“ spezifischen Intervention der Behandlung B. Die beiden Items „a/b1“, „a/b2“ repräsentieren allgemeine Intervention, die sowohl in A und in B erwünscht sind.

Behandlung	Item „a1“	Item „a2“	Item „a/b1“	Item „a/b2“	Item „b1“	Item „b2“
A	2	2	0	0	2	0
B	2	0	2	2	0	0

$$Purity\ Index = \frac{\text{erwünschte I.}}{\text{erwünschte I.} + \text{unerwünschte I.}} = \frac{(\text{spezifische I.} + \text{allgemeine I.})}{(\text{spezifische I.} + \text{allgemeine I.}) + \text{unerwünschte I.}}$$

$$Purity\ Index_{(A)} = \frac{(a_1 + a_2) + (a/b_1 + a/b_2)}{(a_1 + a_2 + a/b_1 + a/b_2) + (b_1 + b_2)} = \frac{(2 + 2) + (0 + 0)}{(2 + 2) + (0 + 0) + (2 + 0)} = 0.67$$

$$Purity\ Index_{(B)} = \frac{(b_1 + b_2) + (a/b_1 + a/b_2)}{(b_1 + b_2 + a/b_1 + a/b_2) + (a_1 + a_2)} = \frac{(0 + 0) + (2 + 2)}{(0 + 0) + (2 + 2) + (2 + 0)} = 0.67$$

Ergebnis: Der Purity Index ist für Behandlung A und B identisch. Jedoch wurden in Behandlung A keine allgemeinen Interventionen, sondern vor allem spezifische Interventionen und in Behandlung B keine charakteristischen Interventionen, sondern hauptsächlich allgemeine Interventionen angewandt.

Abbildung 4 Beispielrechnung für die Verdeutlichung der Einschränkungen des Purity Index

Interventionen (z.B. Exposition in sensu im Rahmen eines Expositionsansatzes) zusammengefasst. Dadurch ergibt sich die Möglichkeit, dass im Rahmen einer Vergleichsstudie zwei Behandlungsbedingungen nicht wie beabsichtigt hauptsächlich aus spezifischen Interventionen, sondern zum Großteil aus allgemeinen Interventionen bestehen. In diesem Fall wäre der Purity Index genauso hoch, wie wenn die Behandlungen aus spezifischen Interventionen besteht. Folglich könnte dies zu einer Überschätzung der Behandlungsdifferenzierung führen. (vgl. Abbildung 4).

Auf Basis der oben dargestellten Unterscheidungen von Interventionsformen nach Waltz et al. (1993), kann für die Erreichung einer möglichst hohen Behandlungsdifferenzierung die Schlussfolgerung abgeleitet werden, dass Behandlungen idealerweise hauptsächlich aus spezifisch/essenziellen, einigen essenziellen/unspezifischen, möglichst wenigen akzeptablen/nicht notwendigen und keinen unerwünschten Interventionen zusammengesetzt sind. Für eine genauere Beschreibung der Behandlungsdifferenzierung ist es daher entscheidend nicht nur die Reinheit einer Behandlung als Grundlage zu verwenden, sondern auch die *Behandlungsspezifität* einzubeziehen. Nach Frank, Kupfer, Wagner, McEachran und Cornes (1991) kann die Behandlungsspezifität definiert werden als das Ausmaß der in einer Behandlung vorhandenen charakteristischen Interventionen. Wie die Reinheit kann somit auch die Spezifität als eine Unterkategorie der Adhärenz angesehen werden. Da jedoch für die Evaluation der Spezifität keine geeigneten Vorgehensweisen existieren, war es Ziel der dritten Studie, durch die Entwicklung des *Behandlungs-Spezifitäts-Index* (engl. Treatment Specificity Index) diese methodische Lücke zu schließen und damit die Evaluation der Behandlungsdifferenzierung zu verbessern (siehe Abbildung 5). Erreicht der Spezifitäts-Index einen Wert von 0.50 oder mehr, kann angenommen werden, dass eine Behandlung aus mindestens gleichvielen bzw. mehr spezifischen/charakteristischen

Intervention im Verhältnis zu den übrigen Interventionen besteht, woraus eine Behandlungsspezifität geschlussfolgert werden kann.

$$\text{Behandlungs – Spezifitäts – Index} = \frac{\text{spezifische I.}}{(\text{spezifische I.} + \text{allgemeine I.} + \text{unerwünschte I.})}$$

Abbildung 5 Formel des Behandlungs-Spezifitäts-Index; spezifische Interventionen (I.) entsprechen hier den „spezifisch/essenziellen“ Interventionen und allgemeine Interventionen entsprechen den „essenziellen/unspezifischen“ Interventionen gemäß der Kategorisierung nach Waltz et al. (1993). Vergleiche hierzu auch Kapitel 1.1.1.

1.3.5 Zusammenfassung der dritten Studie: „*Evaluation of Treatment Integrity: The Treatment Specificity Index*“

Fragestellung. Das vorrangige Ziel der Studie war die Entwicklung einer Methodik, um die Behandlungsdifferenzierung innerhalb einer Vergleichsstudie in geeigneter Weise beschreiben zu können. Hierfür wurde der Behandlungs-Spezifitäts-Index (BSI) als Ergänzung zum bereits vorhandenen Purity Index (PI) entwickelt.

Methode. Grundlage der Studie war die Reanalyse von Adhärenzratings (Weck et al., 2012; 2013c) einer randomisiert kontrollierten Studie zur Untersuchung der Rückfallprophylaxe bei rezidivierender Depression (Stangier et al., 2013). Innerhalb dieser Studie wurde eine kognitiv-behaviorale Therapie (CBT; Risch, Stangier, Heidenreich, & Hautzinger, 2012) mit einer aktiven Kontrollbedingung in Form eines psychoedukativen Behandlungsansatzes (PE; Hautzinger, Schlösser, & Barocka, 2006) verglichen. Aus beiden Bedingungen wurden je 40 Behandlungen randomisiert ausgewählt und jeweils eine Therapiesitzung anhand einer Videoaufzeichnung hinsichtlich der Adhärenz mittels einer Ratingskala ausgewertet. Auf Grundlage dieser Adhärenzratings wurde der Purity und der Spezifitäts-Index berechnet.

Ergebnisse. Bei der Kalkulation des Purity Index konnte kein signifikanter Unterschied ($t[78] = -1.4; p = .18$) zwischen der CBT-Bedingung ($PI_{CBT} = 0.98$) und der PE-Bedingung ($PI_{PE} = 0.99$) gefunden werden. Hinsichtlich der Behandlungsspezifität zeigte sich, dass der Spezifitäts-Index der PE-Bedingung ($BSI_{PE} = 0.53$) signifikant höher ($t[78] = 4.4; p < .001; d = 0.97$) als der der CTB-Bedingung ($BSI_{CBT} = 0.40$) ist. Weiter konnte gezeigt werden, dass während die PE-Bedingung signifikant mehr ($t[39] = 3.4; p < .001; d = 0.63$) spezifische Interventionen ($M = 1.63; SD = 0.37$) als allgemeine Interventionen ($M = 1.4; SD = 0.36$) beinhaltet, so die CBT-Bedingung aus signifikant mehr ($t[39] = -5.25; p < .001; d = 0.98$) allgemeinen ($M = 1.37; SD = 0.45$) als spezifischen ($M = 0.94; SD = 0.43$) Interventionen zusammensetzt. Darüber hinaus konnte eine signifikante Beziehung zwischen der Symptomveränderung eines Patienten vor der zu beurteilenden Therapiesitzung und der Behandlungsspezifität identifiziert werden.

Schlussfolgerungen. Aufgrund der Ergebnisse kann eine Einschränkung in der Behandlungsspezifität der CBT-Bedingung angenommen werden. Es konnte gezeigt werden, dass eine ausschließliche Betrachtung des Purity Index zu einer Überschätzung der Behandlungsdifferenzierung führen würde und daher für sich genommen nicht ausreichend ist. Die Kalkulation des Behandlungs-Spezifitäts-Index stellt eine sinnvolle Ergänzung zur Beurteilung von Psychotherapiestudien dar und ist eine geeignete Methode, um potentielle Einschränkungen in deren Validität zu erkennen.

1.4 Bedeutung der Behandlungsintegrität für den Behandlungserfolg

Wie bereits erläutert, ist die Erfassung der Behandlungsintegrität vor allem für die Sicherstellung der Validität einer Wirksamkeitsstudie von Relevanz. Eine Reihe von Studien aus der Prozessforschung geht jedoch auch der Fragestellung nach, ob die psychotherapeutische Adhärenz und Kompetenz Einfluss auf den Erfolg einer Behandlung

haben (z.B. Feeley et al., 1999; Perepletchikova & Kazdin, 2005; Strunk, Brotman, DeRubeis & Hollon, 2010; Weck et al., 2013c). Forschungsarbeiten wie die von Ginzburg et al. (2012), in der eine hochsignifikante Korrelation ($r = 0.68$) zwischen der Kompetenz und dem Erfolg einer kognitiv-behavioralen Behandlung der Sozialen Phobie gefunden werden konnte, stützen die intuitive Annahme, dass je kompetenter ein Therapeut die gewählten Interventionen durchführt, diese umso erfolgreicher sind. Während in Bezug zur Adhärenz und deren Einfluss auf das Behandlungsergebnis häufig keine signifikanten Effekte gefunden werden (z.B. Loeb et al., 2005, Weck et al., 2013c), weisen die Ergebnisse einiger Studien jedoch auf einen kurvilinearen Zusammenhang hin (vgl. Crits-Christoph et al., 1999; Stangier, Consbruch & Clark, 2014). Demnach ist anzunehmen, dass eine mittlere Ausprägung, die einem flexiblen therapeutischen Vorgehen entspricht, besonders günstig für den Therapieerfolg ist (Barber et al. 2006). Insgesamt erscheint die empirische Datenlage jedoch wenig eindeutig.

Daher wurde der Zusammenhang zwischen Adhärenz bzw. Kompetenz und Behandlungserfolg von Webb et al. (2010) in einer Metaanalyse untersucht. Dabei zeigte sich weder für die Adhärenz ($r = 0.02$) noch für die Kompetenz ($r = 0.07$) eine signifikante Korrelation mit dem Behandlungserfolg. Lediglich im Rahmen der Therapie von depressiven Störungen, nicht jedoch bei anderen Störungsbildern (z.B. Essstörungen, Panikstörung, Substanzabhängigkeit), konnte eine Beziehung zwischen dem Behandlungserfolg und der psychotherapeutischen Kompetenz ($r = 0.28$) festgestellt werden. Es lassen sich jedoch einige methodische Mängel dieser Metaanalyse aufführen, sodass die Ergebnisse unter Vorbehalt betrachtet werden müssen. Eine Einschränkung besteht darin, dass sich von den 36 eingeschlossenen Forschungsarbeiten 32 Studien auf den Adhärenz-Outcome-Zusammenhang bezogen und lediglich 17 Studien zum Kompetenz-Outcome-Zusammenhang analysiert werden konnten. Ebenfalls muss die Erhebung der Prozessmaße

im Rahmen der einbezogenen Studien kritisch beurteilt werden (Simons, Rozek & Serrano, 2013; Weck et al., 2011b). Aufgrund des Mangels an geeigneten Studien wurden auch solche eingeschlossen, die lediglich einzelne Komponenten der Adhärenz und Kompetenz erfassen (Dennhag et al., 2012).

Eine weitere Metaanalyse (Zarafonitis-Müller, Kurh & Bechdorf, 2014) weist zwei wesentliche Unterschiede zu der von Webb et al. (2010) auf. Zum einen schloss diese ausschließlich kognitiv-verhaltenstherapeutische Behandlungsstudien in ihre Untersuchung ein und zum anderen konnte ein signifikanter Gesamteffekt ($r = 0.24$) zwischen der psychotherapeutischen Kompetenz und Therapieerfolg nachgewiesen werden. Ein bedeutsamer Zusammenhang hinsichtlich der Adhärenz konnte hier ebenfalls nicht gezeigt werden. Allerdings muss auch hier aufgrund der geringen Anzahl von lediglich 13 Studien, die Aussagekraft der Ergebnisse als eingeschränkt betrachtet werden.

Eine Auffälligkeit der Studien, die den Zusammenhang zwischen therapeutischer Kompetenz/Adhärenz und dem Behandlungserfolg untersuchen, liegt darin, dass diese in der Regel ihre Schlussfolgerungen auf der Basis von nur wenigen Sitzungen ziehen und sich dabei meist auf den globalen Therapieerfolg beziehen. Da die empirisch nachgewiesenen Effektstärken bzgl. des Zusammenhangs zwischen Behandlungsintegrität und Therapieerfolg in der Regel eher gering ausfallen, erscheint es nach Weck (2013) sinnvoll, den Einfluss von Kompetenz/Adhärenz nicht nur global für den Behandlungserfolg zu untersuchen, sondern sich auf spezifischere Fragestellungen zu fokussieren. Eine solche Spezifizierung kann darin bestehen, dass die Adhärenz/Kompetenz in erfolgreichen Behandlungen (klinisch signifikante Verbesserungen der Symptomatik bzw. stabile Remissionen) mit solchen Therapien verglichen wird, in denen es zu einem Misserfolg (klinisch signifikante Verschlechterung der Symptomatik) gekommen ist. Der Einbezug von therapeutischen Misserfolgen, die nach Lutz, Kosfelder & Joormann (2004) nur wenig untersucht sind, hat eine stärkere

Kontrastierung zur Folge, sodass mit deutlichen Gruppenunterschieden gerechnet werden kann. Hintergrund dieser Überlegungen ist die Annahme, dass bei einem gewissen Ausmaß der therapeutischen Kompetenz (ggf. auch der Adhärenz), die Auswirkungen der Kompetenz auf den Behandlungserfolg sehr viel weniger ins Gewicht fallen als wenn die Kompetenz (bzw. Adhärenz) nur im geringen Maße vorhanden ist, wodurch mit sehr viel mehr Misserfolgen gerechnet werden muss. Die Frage, ob sich die therapeutische Adhärenz bzw. Kompetenz in Behandlungen, die erfolgreich verliefen, von solchen unterscheiden, in denen es zu Misserfolgen kam, ist Forschungsgegenstand der vierten Studie.

1.4.1 Zusammenfassung der vierten Studie: „*Treatment failure in cognitive-behavioural therapy: Therapeutic alliance as a precondition for an adherent and competent implementation of techniques*”

Fragestellung. Ziel dieser Studie war die Überprüfung, ob sich die psychotherapeutische Kompetenz/Adhärenz von erfolgreichen Therapien zu denjenigen Behandlungen unterscheiden, in denen es zu Misserfolgen gekommen ist. Darüber hinaus sollten weitere Einflussfaktoren wie die psychotherapeutische Beziehung und andere Patientenmerkmale wie die Patientenschwierigkeit mit in die Analysen einbezogen werden.

Methode. Die hier reanalysierten Daten stammen aus drei randomisiert kontrollierten Therapiestudien, die jeweils die Wirksamkeit eines kognitiv-verhaltenstherapeutischen Behandlungsansatzes bei Patienten mit remittierter Depression (Stangier et., 2013), Sozialer Phobie (Stangier et al., 2010) sowie Hypochondrie (Weck et al., 2015b) untersuchten. Aus diesen drei Studien wurden insgesamt 61 Patientenbehandlungen ausgewählt, die entweder als Therapieerfolg (N = 43) oder als Misserfolg (N = 18) anhand von bestimmten Kriterien (u.a. signifikant bedeutsame Symptomverbesserung bzw. -verschlechterung in den relevanten Erfolgsmaßen wie BDI-II, vgl. Jacobson & Truax, 1991) eingestuft wurden. Von den auf

diese Weise eingeschlossenen Patientenbehandlungen wurden jeweils die ersten drei, auf Video aufgezeichneten, Therapiesitzungen von zwei klinisch erfahrenen und trainierten Beurteilern hinsichtlich der Adhärenz, Kompetenz und der therapeutischen Beziehung beurteilt.

Ergebnisse. Adhärenz ($F_{[1,59]} = 4.51; p = .04; g = .58$) sowie die psychotherapeutische Beziehung ($F_{[1,59]} = 5.34; p = .02; g = .62$), nicht jedoch die Kompetenz ($F_{[1,59]} = 1.38; p = .24$), weisen in erfolgreichen Behandlungen signifikant höhere Werte auf als in solchen mit Misserfolg. Die therapeutische Beziehung konnte als ein Moderator für den Zusammenhang zwischen der Adhärenz und dem Behandlungserfolg identifiziert werden, so dass angenommen werden kann, dass je besser die therapeutische Beziehung desto stärker ist der Effekt der Adhärenz auf den Behandlungserfolg ($r = 0.24; p < .05$). Ebenfalls konnte gezeigt werden, dass es sich bei der therapeutischen Beziehung um einen signifikanten Mediator ($r = 0.26; 95\% \text{ CI } [.15; .38]$) zwischen der Kompetenz und dem Behandlungserfolg handelt. Eine höhere Adhärenz beeinflusst den Behandlungserfolg positiv, wenn es durch Kompetenz und psychotherapeutische Beziehung mediiert wird ($r = 0.16; 95\% \text{ CI } [.01; .31]$). Darüber hinaus konnte gezeigt werden, dass die Erfassung der Adhärenz, der Kompetenz sowie der psychotherapeutischen Behandlung sich über die Sitzungen 1-3 als stabil erweisen. Ebenfalls ließ sich nachweisen, dass die psychotherapeutische Beziehung die Adhärenz und die Kompetenz in der darauffolgenden Behandlungssitzung beeinflusst.

Schlussfolgerung: Die therapeutische Beziehung scheint ein wichtiger Prädiktor für Misserfolg in kognitiv-verhaltenstherapeutischen Behandlungen zu sein. Des Weiteren zeigte sich, dass die Beziehung zwischen Therapeut und Patient eine Voraussetzung für die adhärente und kompetente Implementierung von therapeutischen Techniken ist.

1.5 Gesamtdiskussion

1.5.1 Integration der Ergebnisse

Die durchgeführten und dargestellten Studien konnten einen wichtigen Beitrag für die Erfassung der Behandlungsintegrität (bestehend aus der psychotherapeutische Adhärenz, Kompetenz und Behandlungsdifferenzierung) leisten.

Da die Evaluation der Behandlungsintegrität stets mit einem hohen Zeit- und Kostenaufwand verbunden ist, erscheint es von besonderem Nutzen die Dauer der zu betrachtenden Behandlungssitzungen zu reduzieren. Die Ergebnisse des Vergleichs von besonders vielversprechenden Sitzungssegmenten (mittleres Sitzungsdrittel) mit vollständigen Behandlungssitzungen konnten zeigen, dass auch auf Basis von Ausschnitten reliable und vergleichbare Beurteilungen der Adhärenz und Kompetenz vorgenommen werden können (Weck, Grikscheit, Höfling & Stangier, 2014).

Neben dem hohen Aufwand ist das mangelnde theoretische Wissen ein weiterer Grund dafür, dass die Behandlungsintegrität nur in einem geringen Teil von randomisiert kontrollierten Studien evaluiert wird (vgl. Perplechikova et al., 2009). Durch die (Weiter-)Entwicklung der Adhärenz- (AS-EXVO) und Kompetenzskala (KS-EXVO) konnten nicht nur reliable und valide Instrumente für die Erfassung der Behandlungsintegrität der Expositionsvorbereitung bei dem Störungsbild der Panikstörung mit Agoraphobie entwickelt, sondern auch ein Beispiel für die Konstruktion und Anwendung solcher Skalen dargestellt werden (Grikscheit et al., 2015a).

Im Vergleich zur Adhärenz und Kompetenz stellte die Behandlungsdifferenzierung, jenes Prozessmaß der Behandlungsintegrität dar, welches bisher am wenigsten untersucht und beschrieben wurde. Durch die Entwicklung des Behandlungs-Spezifitäts-Index konnte eine geeignete und leicht anzuwendende Methodik zur Evaluation der Behandlungsdifferenzierung vorgestellt werden (Grikscheit et al. 2015b). Auch wenn kein

Zusammenhang zwischen der Spezifität bzw. der Reinheit einer Behandlung und deren Erfolg nachgewiesen werden konnte, verdeutlichen die Ergebnisse der Studie, dass durch die Kalkulation des Spezifitäts-Index in Ergänzung zu dem bereits bekannten Purity Index die interne Validität von randomisiert kontrollierten Studien umfassender beschrieben werden kann.

Durch die Analyse der Behandlungsintegrität innerhalb von erfolgreichen Behandlungen und solchen, in denen es zu Misserfolgen gekommen ist, konnten weitere, wenn auch unerwartete Erkenntnisse gewonnen werden (Weck, Grikscheit, Jakob, Höfling & Stangier, 2015a). Überraschenderweise konnte für die Adhärenz und die psychotherapeutische Beziehung, nicht jedoch für die Kompetenz, ein bedeutsamer Einfluss auf den (Miss-)Erfolg einer Behandlung nachgewiesen werden. Dies steht im Kontrast zu vorherigen Studien, in denen eine stärkere Beeinflussung des Behandlungserfolgs durch die Kompetenz als durch die Adhärenz festgestellt wurde (vgl. Barber et al, 1996; Ginzburg et al., 2012). Jedoch wurde in diesen Studien eine sehr viel generelle Betrachtung des Behandlungserfolgs vorgenommen, als dies hier der Fall gewesen ist. Anzumerken ist auch, dass eine indirekte Assoziation der Kompetenz mit dem (Miss-)Erfolg nachgewiesen werden konnte, da die psychotherapeutische Beziehung als ein bedeutsamer Mediator, vergleichbar mit vorherigen Studien (Barber et al., 1996; Despland et al, 2009), identifiziert werden konnte. Die Ergebnisse verdeutlichen, dass die Adhärenz eine wichtige Voraussetzung für die Kompetenz, diese wiederum eine Voraussetzung für die psychotherapeutische Beziehung ist, welche wiederum deutlichen Einfluss auf den Behandlungserfolg besitzt.

1.5.2 Grenzen der Arbeit

Bei der vorliegenden Arbeit lassen sich jedoch auch einige methodische Einschränkungen aufführen.

Erstens, die drei Studien zur Untersuchung der Sitzungssegmente, zur Entwicklung des Behandlungs-Spezifitäts-Index sowie zur Analyse von (Miss-)Erfolgen in Behandlungen basieren unter anderem auf Daten einer randomisiert kontrollierten Studie zur Rückfallprophylaxe der rezidivierenden Depression (Stangier et al., 2013). Hierbei ist hervorzuheben, dass die dort behandelten Patienten zum Zeitpunkt des Einschlusses in die Studie sich diagnostisch in der Remission befanden. Daraus ergibt sich, dass im Vergleich zu den ebenfalls verwendeten Daten aus den randomisiert kontrollierten Studien zur Behandlung der Sozialen Phobie (Stangier et al., 2010) und Hypochondrie (Weck et al., 2015b) ein anderes Erfolgsmaß verwendet wurde. Hierbei stand nicht eine deutliche Symptomreduktion im Vordergrund, sondern eine möglichst lange Erhaltung der Remission und somit eine Verhinderung eines Rückfalls.

Zweitens, die Studien zur Untersuchung der Sitzungssegmente sowie zur Entwicklung des Behandlungs-Spezifitäts-Index basieren auf der Beurteilung lediglich einer randomisiert ausgewählten Therapiesitzung. Empirische Befunde verdeutlichen, dass dadurch die Generalisierbarkeit auf die gesamte Behandlung eingeschränkt ist (Dennhag et al., 2012). Für eine umfassendere Einschätzung der Behandlungsspezifität als auch der Adhärenz und Kompetenz im Zusammenhang mit Sitzungsausschnitten sollten in weiteren Untersuchungen mehrere Sitzungen verwendet werden.

Drittens, die vorgestellten Studien fokussieren lediglich die Kognitive Verhaltenstherapie. Dadurch muss die Generalisierbarkeit auf andere Behandlungsmethoden als eingeschränkt angenommen werden. Daher erscheinen weitere Untersuchungen anhand verschiedener Behandlungsmethoden und Störungsbildern sinnvoll.

Viertens, die in dieser Arbeit vorgestellten Analysen erfolgten auf Basis von randomisiert kontrollierten Studien, die ein hohes Maß an Standardisierung aufwiesen. Hierdurch können Einschränkungen in der Varianz der analysierten Prozessmaße

angenommen werden. Offen bleibt dadurch die Übertragbarkeit der Ergebnisse auf Studien mit geringerer Standardisierung.

Letztlich stellt die Untersuchung der Behandlungsintegrität immer nur einen eingeschränkten Blick auf das therapeutische Geschehen und auf die Wirkfaktoren einer Behandlung dar. Wie bisherige Studien (z.B. Crits-Christoph et al. 2013; Webb et al., 2010) und auch die Ergebnisse der eigenen Studie zum (Miss-)Erfolg von Behandlung verdeutlichen, muss eine komplexe Interaktion zwischen Patienten- und Therapeutenvariablen (wie die psychotherapeutische Beziehung) sowie von allgemeinen und spezifischen Faktoren angenommen werden.

1.5.3 Ausblick

Aufgrund dessen, dass die Prädiktion des Behandlungserfolgs durch die Adhärenz, Kompetenz und Behandlungsdifferenzierung nur eingeschränkt nachgewiesen wurde, erscheint es gelegentlich so als würde die Behandlungsintegrität immer wieder in den Hintergrund der Prozessforschung rücken. Die vorgestellte Arbeit verdeutlicht jedoch vor allem die Notwendigkeit der Erfassung der Behandlungsintegrität für die Sicherstellung der Validität einer Wirksamkeitsstudie. Da weiterhin davon auszugehen ist, dass die Evaluation der Behandlungsintegrität in der Psychotherapieforschung nur eingeschränkt stattfindet, wäre es wünschenswert, wenn die hier vorgestellten Instrumente und Methodiken einen Beitrag dazu leisten, dass zukünftig Studien die Adhärenz, Kompetenz und auch die Behandlungsdifferenzierung vermehrt erfassen. Außerdem wäre es erstrebenswert, dass die Erfassung der Behandlungsintegrität nicht mehr nur im Rahmen der Psychotherapieforschung, sondern im Rahmen der Supervision eingesetzt werden würde.

1.6 Literaturverzeichnis

- Bachrach, H., Curtis, H., Escoll, P., Graff, H., Huxster, H., Ottenberg, P. & Pulver, S., (1981). Units of observation and perspectives on the psychoanalytic process. *British Journal of Medical Psychology*, 54, 25-33. doi: 10.1111/j.2044-8341.1981.tb01466.x
- Bahr, S. S. & Beck, A. T. (2009). Treatment integrity of studies that compare short-term psychodynamic psychotherapy with cognitive-behavior therapy. *Clinical Psychology: Science and Practice*, 16, 370-378. doi: 10.1111/j.1468-2850.2009.01176.x
- 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 and Clinical Psychology*, 64, 619-622. doi: 10.1037/0022-006x.64.3.619
- Barber, J. P., Gallop, R., Crits-Christoph, P., Frank, A., Thase, M. E., Weiss, R. D., Gibbons, M. B. C. (2006). The role of therapist adherence, therapist competence, and alliance in predicting outcome of individual drug counseling: Results from the National Institute Drug Abuse Collaborative Cocaine Treatment Study. *Psychotherapy Research*, 16, 229-240. doi: 10.1080/10503300500288951
- Barber, J. P., Triffelman, E. & Marmer, C. (2007). Considerations in treatment integrity: implications and recommendations for PTSD research. *Journal of Traumatic Stress*, 20, 793-805. doi: 10.1002/jts.20295
- Comer, J. S. & Kendall, P. C. (2013). Methodology, Design, and Evaluation in Psychotherapy Research. In M. J. Lambert (Ed.), *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change*. (6th ed.) (S. 21-48). New York: Wiley.
- Consbruch, K., Clark, D. M. & Stangier, U. (2012). Assessing therapeutic competence in cognitive therapy for social phobia: psychometric properties of the cognitive therapy competence scale for social phobia (CTCS-SP). *Behavioural and Cognitive Psychotherapy*, 40, 149-161. doi: 10.1017/S1352465811000506

- Crits-Christoph, P., Gibbons, M. B. C. & Mukherjee, D. (2013). Psychotherapy process-outcome research. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change*. (6th ed.) (S. 298-340). New York: Wiley.
- Crits-Christoph, P., Siqueland, L., Blaine, J., Frank, A., Luborsky, L., Luborsky, L., et al. (1999). Psychosocial treatments for cocaine dependence: National Institute on Drug Abuse Collaborative Cocaine Treatment Study. *Archives of General Psychiatry*, *56*, 493-502.
- Dennhag, I., Gibbons, M. B. C., Barber, J. P., Gallop, R. & Crits-Christoph, P. (2012). How many treatment sessions and patients are needed to create a stable score of adherence and competence in the treatment of cocaine dependence? *Psychotherapy Research*, *22*, 475-488. doi: 10.1080/10503307.2012.674790
- Despland, J.-N., de Roten, Y., Drapeau, M., Currat, T., Beretta, V. & Kramer, U. (2009). The role of alliance in the relationship between therapist competence and outcome in brief psychodynamic psychotherapy. *The Journal of Nervous and Mental Disease*, *197*, 362-367. doi: 10.1097/NMD.0b013e3181a20849
- Feeley, M., DeRubeis, R. J. & Gelfand, L. A. (1999). The temporal relation of adherence and alliance to symptom change in cognitive therapy for depression. *Journal of Consulting and Clinical Psychology*, *67*, 578-582. doi: 10.1037/0022-006X.67.4.578
- Frank, E., Kupfer, D. J., Wagner, E. F., McEachran, A. B. & Cornes, C. (1991). Efficacy of interpersonal psychotherapy as a maintenance treatment of recurrent depression: contributing factors. *Archives of General Psychiatry*, *48*, 1053-1059. doi:10.1001/archpsyc.1991.01810360017002
- Ginzburg, D. M., Bohn, C., Höfling, V., Weck, F., Clark, D. M. & Stangier, U. (2012). Treatment specific competence predicts outcome in cognitive therapy for social

- anxiety disorder. *Behaviour Research and Therapy*, 50, 747-752. doi: 10.1016/j.brat.2012.09.001
- Grikscheit, F., Lang, T., Kordt, A., Helbig-Lang, S., Geisler, K. & Weck, F. (2015a). Entwicklung und Validierung von Instrumenten zur Überprüfung der therapeutischen Adhärenz und Kompetenz bei der Behandlung der Panikstörung. *Zeitschrift für Klinische Psychologie und Psychotherapie*, 44, 1-9. doi: 10.1026/1616-3443/a00030
- Grikscheit, F., Weck, F., Hautzinger, M., Heidenreich, T., Weigel, M., Rudari, V., Schimming, C. & Stangier, U. (2015b). Evaluation of Treatment Integrity: The Treatment Specificity Index. *Journal of Clinical Psychology*, 71, 653-665. doi: 10.1026/1616-3443/a000309
- Haddock, G., Devane, S., Bradshaw, T., McGovern, J., Tarrrier, N., Kinderman, P. et al. (2001). An investigation into the psychometric properties of the Cognitive Therapy Scale for Psychosis (CTS-Psy). *Behavioural and Cognitive Psychotherapy*, 29, 221-233. doi: 10.1017/S1352465801002089
- Hautzinger, S., Schlösser, R. & Barocka, A. (2006). Manualisierte Psychoedukation (MAPE): Anleitung für Therapeuten und Materialien für Patienten. Unveröffentlichtes Manuskript, University of Tübingen and University of Jena, Germany.
- Huppert, J. D., Barlow, D. H., Gorman, J.M. Shear, M. K. & Woods, S.W. (2006). The interaction of motivation and therapist adherence predicts outcome in cognitive behavioral therapy for panic disorder: preliminary findings. *Cognitive and Behavioral Practice*, 13, 198-204. doi:10.1016/j.cbpra.2005.10.001
- Jacobson, N. S. & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59, 12-19. doi: 10.1037//0022-006X.59.1.12

- Lang, T., Helbig-Lang, S., Westphal, D., Gloster, A. T. & Wittchen, H.-U. (2012). *Expositionsbasierte Therapie der Panikstörung und Agoraphobie. Ein Behandlungsmanual*. Göttingen: Hogrefe.
- Loeb, K. L., Wilson, G. T., Labouvie, E., Pratt, E., Hayaki, J., Walsh, B. T. et al. (2005). Therapeutic alliance and treatment adherence in two interventions for bulimia nervosa: a study of process and outcome. *Journal of Consulting and Clinical Psychology, 73*, 1097-1107. doi: 10.1037/0022-006X.73.6.1097
- Luborsky, L. & DeRubeis, R. J. (1984). The use of psychotherapy treatment manuals: a small revolution in psychotherapy research style. *Clinical Psychology Review, 4*, 5-14. doi: 10.1016/0272-7358(84)90034-5
- 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. doi: 10.1001/archpsyc.1985.01790290084010
- Lutz, W., Kosfelder, J. & Joormann, J. (2004). *Misserfolge und Abbrüche in der Psychotherapie: Erkennen - Vermeiden – Vorbeugen*. Bern: Verlag Hans Huber.
- Markowitz, J. C., Spielman, L. A., Scarvalone, P. A. & Perry, S. W. (2000). Psychotherapy adherence of therapists treating HIVpositive patients with depressive symptoms. *Journal of Psychotherapy Practice and Research, 9*, 75-80.
- Martino, S., Ball, S., Nich, C., Frankforter, T. L. & Carrol, K. M. (2009). Correspondence of motivational enhancement treatment integrity ratings among therapists, supervisors, and observers. *Psychotherapy Research, 19*, 181-193. doi: 10.1080/10503300802688460
- Mintz, J. & Luborsky, L. (1971). Segments versus whole sessions: Which is the better unit for psychotherapy process research? *Journal of Abnormal Psychology, 78*, 180-191. doi: 10.1037/h0031969

- Moncher, F. J. & Prinz, R. J. (1991). Treatment fidelity in outcome studies. *Clinical Psychology Review, 11*, 247-266. doi: 10.1016/0272-7358(91)90103-2
- Orlinsky, D. E., Rønnestad, M. H. & Willutzki, U. (2004). Fifty years of psychotherapy process-outcome research: continuity and change. In M. J. Lambert (Ed.), *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change* (5th ed.) (S. 307 – 389). New York: Wiley.
- Perepletchikova, F., Hilt, L. M., Chereji, E. & Kazdin, A. E. (2009). Barriers to implementing treatment integrity procedures: survey of treatment outcome researchers. *Journal of Consulting and Clinical Psychology, 77*, 212-218. doi: 10.1177/1534508411432466
- Perepletchikova, F. & Kazdin, A. E. (2005). Treatment integrity and therapeutic change: issues and research recommendations. *Clinical Psychology: Science and Practice, 12*, 365-383. doi: 10.1093/clipsy.bpi045
- 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. doi: 10.1037/0022-006X.75.6.829
- Pohl, U., Richter, H. & Bohus, M. (2000). Neuere Entwicklungen zur systematischen Überprüfung des therapeutischen Vorgehens: Skalen zur Erfassung von adherence und competence. *Psychiatrie, Psychotherapie, Medizinische Psychologie, 50*, 22-28.
- Portney, L. G. & Watkins, M. P. (2008). *Foundations of clinical research: applications to practice* (3rd ed.) London: Pearson Education.
- Risch, A. K., Stangier, U., Heidenreich, T. & Hautzinger, M. (2012). *Kognitive Erhaltungstherapie bei rezidivierender Depression: Rückfälle verhindern, psychische Gesundheit erhalten*. Berlin, Heidelberg, New York: Springer.

- Ruhmland, M. & Margraf, J. (2001). Effektivität psychologischer Therapien von Panik und Agoraphobie. Meta-Analysen auf Störungsebene. *Verhaltenstherapie, 11*, 41-53. doi: 10.1159/000050323
- Sánchez-Meca, J., Alcázar, A. I., Marín-Maténez, F. Gómez-Conesa, A. (2010). Psychological treatment of panic disorder with or without agoraphobia: a meta-analysis. *Clinical Psychology Review, 30*, 37-50. doi: 10.1016/j.cpr.2009.08.011.
- Schlosser, R. W. (2002). On the importance of being earnest about treatment integrity. *Augmentative and Alternative Communication, 18*, 36-44. doi: 10.1080/aac.18.1.36.44
- Sharpless, B. A. & Barber, J. P. (2009). A conceptual and empirical review of the meaning, measuring, development, and teaching of intervention competence in clinical psychology. *Clinical Psychology Review, 29*, 47-56. doi: 10.1016/j.cpr.2008.09.008
- Shaw, B. F., Elkin, I., Yamaguchi, J., Olmsted, M., Vallis, T. M., Dobson, K. S. et al., (1999). Therapist competence ratings in relation to clinical outcome in cognitive therapy of depression. *Journal of Consulting and Clinical Psychology, 67*, 837-846. doi: 10.1037/a0019631
- Simons, A. D., Rozek, D. C. & Serrano, J. L. (2013). Wanted: Reliable and valid measures for the science of cognitive behavioral therapy dissemination and implementation. *Clinical Psychology: Science and Practice, 20*, 181-194. doi: 10.1111/cpsp.12033
- Southam-Gerow, M. A. & McLeod, B. D. (2013). Advances in applying treatment integrity research for dissemination and implementation science: introduction to special issue. *Clinical Psychology: Science and Practice, 20*, 1–13. doi: 10.1111/cpsp.12019
- Stangier, U. (2015). Psychotherapeutische Kompetenzen in der kognitiven Verhaltenstherapie. *Psychotherapeut, 60*, 193-198. doi: 10.1007/s00278-015-0022-y
- Stangier, U, Consbruch K. v. & Clark, D. M. (2014). *Effective components of cognitive therapy for social anxiety disorder*. Paper presented on the 44th Congress of the

European Association for Behavioural & Cognitive Therapies, The Hague, The Netherlands 10-13 September.

- Stangier, U., Hilling, C., Heidenreich, T., Risch, A. K., Barocka, A., Schlösser, R., et al. (2013). Maintenance cognitive-behavioral therapy and manualized psychoeducation in the treatment of recurrent depression: a multicenter prospective randomized controlled trial. *The American Journal of Psychiatry*, *170*, 624-632. doi: 10.1176/appi.ajp.2013.12060734
- Stangier, U., Schramm, E., Heidenreich, T., Berger, M. & Clark, D. M., (2010). Cognitive therapy vs. Interpersonal psychotherapy in social anxiety disorder: A randomized controlled trial. *Archives of General Psychiatry*, *68*, 692-700. doi: 10.1001/archgenpsychiatry.2011.67
- Strunk, D. R., Brotman, M. A., DeRubeis, R. J. & Hollon, S. D. (2010). Therapist competence in cognitive therapy for depression: predicting subsequent symptom change. *Journal of Consulting and Clinical Psychology*, *78*, 429-437. doi: 10.1037/a001963
- Vermilyea, B. B., Barlow, D. H. & O'Brien, G. T. (1984). The importance of assessing treatment integrity: an example in the anxiety disorders. *Journal of Behavioral Assessment*, *6*, 1-11. doi: 10.1007/BF01321456
- 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. doi:10.1037/0022-006X.61.4.620
- Webb, C. A., DeRubeis, R. J. & Barber, J. P. (2010). Therapist adherence / competence and treatment outcome: A meta-analytic review. *Journal of Consulting and Clinical Psychology*, *78*, 200-211. doi:10.1037/a0018912

- Weck, F. (2013). *Psychotherapeutische Kompetenzen: Theorien – Erfassung – Förderung*. Heidelberg: Springer.
- Weck, F., Bohn, C., Ginzburg, D. M. & Stangier, U. (2011a). Assessment of adherence and competence in cognitive therapy: comparing session segments with entire session. *Psychotherapy Research*, *21*, 658-668. doi: 10.1080/10503307.2011.602751
- Weck, F., Bohn, C., Ginzburg, D. M. & Stangier, U. (2011b). Behandlungsintegrität: Implementierung, Messung, Evaluation und Zusammenhänge zum Therapieerfolg. *Verhaltenstherapie*, *21*, 99-107. doi: 10.1159/000328840
- Weck, F., Grikscheit, F., Höfling, V. & Stangier, U. (2014). Assessing Treatment Integrity in Cognitive-Behavioral Therapy: Comparing Session Segments With Entire Sessions. *Behavior Therapy*, *45*, 541-552. doi: 10.1016/j.beth.2014.03.003
- Weck, F., Grikscheit, F., Jakob, M., Höfling, V. & Stangier, U. (2015a). Treatment failure in cognitive-behavioural therapy: Therapeutic alliance as a precondition for an adherent and competent implementation of techniques. *British Journal of Clinical Psychology*, *54*, 91-108. doi: 10.1111/bjc.12063
- Weck, F., Hautzinger, M., Heidenreich, T. & Stangier, U. (2010). Erfassung psychotherapeutischer Kompetenzen: Validierung einer deutschsprachigen Version der Cognitive Therapy Scale. *Zeitschrift für Klinische Psychologie und Psychotherapie*, *29*, 244-250. doi: 10.1026/1616-3443/a000055
- Weck, F., Neng, J. M. B., Richtberg, S., Jakob, M. & Stangier, U. (2015b). Cognitive therapy versus exposure therapy for hypochondriasis (health anxiety): A randomized controlled trial. *Journal of Consulting and Clinical Psychology*, *83*, 665-676. doi: 10.1037/ccp0000013
- Weck, F., Rudari, V., Hilling, C., Hautzinger, M., Heidenreich, T. & Stangier, U. (2013c). Relapses in recurrent depression one year after maintenance cognitive-behavioral

- therapy: the role of therapist adherence, competence, and the therapeutic alliance. *Psychiatry Research*, 210, 140-145. doi: 10.1016/j.psychres.2013.05.036
- Weck, F., Weigel, M., Hautzinger, M., Barocka, A., Schlösser, R. G. M. & Stangier, U. (2012). Relapses in recurrent depression 1 year after psychoeducational treatment: the role of therapist adherence and competence, and the therapeutic alliance. *Psychiatry Research*, 195, 51-55. doi:10.1016/j.psychres.2011.07.025
- Weck, F., Weigel, M., Richtberg, S. & Stangier, U. (2011c). Reliability of adherence and competence assessment in psychoeducational treatment: influence of clinical experience. *Journal of Nervous and Mental Disease*, 199, 983-986. doi: 10.1097/NMD.0b013e3182392da1
- Wickstrom, K. F., Jones, K. M., LaFleur, L. H. & Witt, J. C. (1998). An analysis of treatment integrity in school-based behavioral consultation. *School Psychology Quarterly*, 13, 141-154. doi: 10.1097/NMD.0b013e3182392da1
- Yeaton, W. H. & Sechrest, L. (1981). Critical dimensions in the choice and maintenance of successful treatment: strength, integrity, and effectiveness. *Journal of Consulting and Clinical Psychology*, 49, 156-167. doi: 10.1037/0022-006X.49.2.156
- Young, J. & Beck, A. T. (1980). *Cognitive Therapy Scale rating manual*. Unpublished manuscript, Center for Cognitive Therapy, Philadelphia, PA.
- Zarafonitis-Müller, S., Kuhr, K. A. & Bechdorf, A. (2014). Der Zusammenhang der Therapeutenkompetenz und Adhärenz zum Therapieerfolg in der Kognitiven Verhaltenstherapie – metaanalytische Ergebnisse. *Fortschritte der Neurologie – Psychiatrie*, 82, 502-510. doi: 10.1055/s-0034-1366798
- Zobel, I., Karim, A., Kech, S., Berger, M. & Schramm, E. (2008). Wie adhärenz wird Clinical Management in randomisierten, kontrollierten Studien durchgeführt?. *Psychotherapie, Psychosomatik, medizinische Psychologie*, 58, 395-402. doi:10.1055/s-2008-1067361

2 PUBLIKATIONEN

Assessing Treatment Integrity in Cognitive-Behavioral Therapy: Comparing Session Segments With Entire Sessions

Florian Weck
Florian Grikscheit
Volkmar Höfling
Ulrich Stangier
University of Frankfurt

The evaluation of treatment integrity (therapist adherence and competence) is a necessary condition to ensure the internal and external validity of psychotherapy research. However, the evaluation process is associated with high costs, because therapy sessions must be rated by experienced clinicians. It is debatable whether rating session segments is an adequate alternative to rating entire sessions. Four judges evaluated treatment integrity (i.e., therapist adherence and competence) in 84 randomly selected videotapes of cognitive-behavioral therapy for major depressive disorder, social anxiety disorder, and hypochondriasis (from three different treatment outcome studies). In each case, two judges provided ratings based on entire therapy sessions and two on session segments only (i.e., the middle third of the entire sessions). Interrater reliability of adherence and competence evaluations proved satisfactory for ratings based on segments and the level of reliability did not differ from ratings based on entire sessions. Ratings of treatment integrity that were based on entire sessions and session segments were strongly correlated ($r = .62$ for adherence and $r = .73$ for competence). The relationship between treatment integrity and outcome was comparable for ratings based on session segments and those based on entire sessions. However, significant relationships between therapist competence and therapy outcome were only found in the

treatment of social anxiety disorder. Ratings based on segments proved to be adequate for the evaluation of treatment integrity. The findings demonstrate that session segments are an adequate and cost-effective alternative to entire sessions for the evaluation of therapist adherence and competence.

Keywords: adherence; assessment; competence; session segments; treatment integrity

Treatment integrity refers to the degree to which a treatment is delivered as intended (Yeaton & Sechrest, 1981). In psychotherapy research, treatment integrity prevails when the therapist demonstrates a high level of adherence and competence. Accordingly, adherence is defined as the extent to which a therapist employs interventions as described in the treatment manual, and competence is defined as the extent to which the therapist implements these interventions in a skillful manner (Waltz, Addis, Koerner, & Jacobson, 1993).

A high level of treatment integrity is a precondition for ensuring the experimental validity of psychotherapy treatment studies (Perepletchikova & Kazdin, 2005; Schlosser, 2002; Weck, Bohn, Ginzburg, & Stangier, 2011b). Thus, internal validity is ensured when changes in the dependent variable (treatment outcome) can be attributed to the independent variable (the intervention), which is not the case when treatment is not implemented as intended. Furthermore, external validity is ensured when the findings are generalizable, which does not occur when the independent variable (the intervention) is

This research was supported by Grant WE 4654/4-1 from the German Research Foundation. We would like to thank Dr. Evelyn Behar for her detailed and helpful suggestions regarding style and content in an earlier draft of this manuscript.

Address correspondence to Florian Weck, Ph.D., Department of Clinical Psychology and Psychotherapy, University of Frankfurt, Varrentrappstrasse 40–42, Frankfurt D-60486, Germany; e-mail: weck@psych.uni-frankfurt.de.

0005-7894/45/541-552/\$1.00/0

© 2014 Association for Behavioral and Cognitive Therapies. Published by Elsevier Ltd. All rights reserved.

not replicable because it was not implemented as intended.

In contrast to the general importance of treatment integrity, its evaluation has been neglected in psychotherapy research. The assessment of treatment integrity was found to be inadequate in 89% of randomized controlled trials (RCTs) published in the six most influential psychological and psychiatric journals from 2000 to 2004 (Perepletchikova, Treat, & Kazdin, 2007). An adequate assessment of treatment integrity would involve the evaluation of both therapist adherence and therapist competence, an independent evaluation of therapists' competence in applying the treatment protocol to the clients, and the presentation of data on the validity and reliability of used integrity measures. The considerable time input and high costs associated with the assessment process are particularly important barriers to assessing treatment integrity (Perepletchikova, Hilt, Chereji, & Kazdin, 2009).

For the evaluation of treatment integrity, direct assessment methods, such as information from audiotapes or videotapes evaluated by independent judges, are preferable to indirect methods, such as therapist self-reports, because therapists tend to overestimate their level of treatment integrity (Martino, Ball, Nich, Frankforter & Carrol, 2009). Judges evaluating treatment integrity must also be experienced therapists to provide adequate assessment (Weck, Hilling, Schermelleh-Engel, Rudari, & Stangier, 2011). Moreover, it has been shown that 5–10 treatment sessions per client are needed to achieve sufficient generalizability for ratings of adherence and competence (Dennhag, Gibbons, Barber, Gallop, & Crits-Christoph, 2012). This scenario leads to the high costs associated with the evaluation of treatment integrity.

It is questionable whether it is necessary to consider entire treatment sessions, from start to finish, or whether a segment of the treatment session would be sufficient to ensure an adequate evaluation of treatment integrity. Earlier studies have shown that process variables (e.g., empathy, amount of transference) assessed on the basis of segments were less reliable and valid than process variables assessed on the basis of entire sessions (Bachrach et al., 1981; Mintz & Luborsky, 1971). However, those studies used only very short segments (4- and 5-minute segments) for the evaluation of process variables and do not refer to treatment integrity at all.

One recent study directly compared the reliability and validity of adherence and competence ratings based on session segments with those based on entire sessions (Weck, Bohn, Ginzburg, & Stangier, 2011a). Accordingly, session segments entailed the first-, middle-, and final third of the entire session, with all

of those thirds lasting a mean of 20 minutes. Therapist adherence and competence in 34 videotaped therapy sessions of cognitive therapy for social anxiety disorder were evaluated by four independent judges. Two judges provided ratings based on entire therapy sessions and two judges based their ratings on all three session segments. The aggregated adherence and the aggregated competence evaluation based on the second segments (middle third of the entire sessions) yielded an interrater reliability ($ICC_{\text{adherence}} = .81$, $ICC_{\text{competence}} = .71$) and correlation with therapy outcome ($r_{\text{adherence}} = .55$, $r_{\text{competence}} = .45$) that were comparable to those of the entire sessions. Moreover, the evaluation based on the second segment correlated strongly with those based on entire sessions ($r_{\text{adherence}} = .65$, $r_{\text{competence}} = .71$). In contrast, ratings based on the first and third segment showed partly lower interrater reliability, lower or mostly insignificant relationships with therapy outcome, and lower correlations with the ratings based on entire sessions. The results imply that ratings based on the second segment seem to constitute a good approximation of ratings based on entire sessions. However, the generalizability of the findings was limited, because only the treatment of social anxiety disorder was examined, and only 10 therapists were employed. Moreover, judges who evaluated the second segment also evaluated the first and the final segments (the segments were presented in permuted order). Therefore, the reliability and validity of ratings based on the second segment could sometimes be overestimated, because judges could remember, in two thirds of the cases, information about the therapeutic process in the other segments. Therefore, further research that addresses these previous limitations and uses a larger sampling of treatment outcome investigations is necessary to demonstrate the suitability of session segments for the evaluation of treatment integrity.

The aim of the current study was to compare adherence and competence ratings based on session segments (i.e., the middle third segment of the entire session) to ratings based on entire sessions both in terms of (a) reliability and (b) relationship to treatment outcome. As such, cognitive-behavioral therapy (CBT), with a cognitive focus on three different disorders from three major diagnostic categories, were considered: mood disorder (i.e., major depressive disorder), anxiety disorder (i.e., social anxiety disorder), and somatoform disorder (i.e., hypochondriasis). We hypothesized that ratings of therapist adherence and competence that were based on segments would (a) be equivalent to and (b) strongly correlated ($r > .5$; according to Cohen, 1992) with ratings of therapist adherence and competence based on entire sessions (Hypothesis 1). We also hypothesized that ratings of

therapist adherence and competence that were based on segments would yield equivalent reliability to ratings based on entire sessions (Hypothesis 2). Finally, we hypothesized that adherence- and competence-outcome relationships would be equivalent for ratings based on segments versus entire sessions (Hypothesis 3).

Method

STUDY DESIGN

The assessment of treatment integrity was conducted in the context of three RCTs. The treatment studies included CBT, and addressed recurrent major depressive disorder (Stangier et al., 2013; registered under ISRCTN 81212636), social anxiety disorder (Stangier, Schramm, Heidenreich, Berger, & Clark, 2010), and hypochondriasis (Weck, Neng, & Stangier, 2013b; see also Weck, Neng, & Stangier, 2013a; registered under NCT01119469). Although the treatment studies for major depressive disorder and social anxiety disorder have been completed, the follow-up assessment for the hypochondriasis study is ongoing (with recruitment and treatment of participants completed). Diagnostic interviews for all three studies were conducted by trained, experienced, and licensed psychotherapists using the Structured Clinical Interview for DSM-IV (SCID-I and SCID-II; First, Gibbon, Spitzer, Williams, & Benjamin, 1997; First, Spitzer, Gibbon, & Williams,

1997; German version: Wittchen & Fydrich, 1997). Diagnosticians were trained in a 2-day workshop on how to administer the SCID. The German version of the SCID demonstrated good interrater reliability (percentage agreement/Yule's coefficient) for major depressive disorder (85%/.70), social anxiety disorder (89%/.67), and somatoform disorders (94%/.67; Wittchen et al., 1991). All three RCTs evidenced efficacy in treating the three disorders (Stangier et al., 2010, 2013; Weck et al., 2013b). An overview of the three treatment studies that form the basis for the current study is given in Table 1.

PROCEDURE

In the current study, only the CBT conditions (i.e., cognitive-behavioral therapy and cognitive therapy) of the three RCTs were considered for the evaluation of therapist adherence and competence. The control conditions (i.e., manualized psychoeducation, interpersonal psychotherapy, exposure therapy) differ significantly from these CBT approaches with a cognitive focus and were not considered because they could not be evaluated with the same adherence and competence scales. CBT did not include any combinations with active control conditions (e.g., CBT plus interpersonal psychotherapy), but only "pure" CBT. A total of 84 cases needed to be considered (by a determined

Table 1
Description of the Three Randomized Controlled Trials Which Form the Basis for the Evaluation of Treatment Integrity

	Studies		
	Stangier et al. (2013) Registered Under ISRCTN 81212636	Stangier, Schramm, Heidenreich, Berger, and Clark (2010) Not Registered	Weck, Neng, and Stangier (2013b) Registered Under NCT01119469
Diagnosis	Recurrent major depressive disorder (currently remitted)	Social anxiety disorder	Hypochondriasis
Central aims of the studies	Reduction of the risk of relapse	Reduction of social anxiety symptoms	Reduction of hypochondriacal symptoms
Sample sizes	180	117	84
Treatments (<i>n</i>)	Maintenance cognitive-behavioral therapy (90); manualized psychoeducation (90)	Cognitive therapy (38); interpersonal psychotherapy (38)	Cognitive therapy (38); exposure therapy (37)
Treatments extent	16 sessions	16 sessions	12 sessions
Exclusion criteria	Organic mental disorders, disorders caused by psychotropic substances, schizophrenia, schizoaffective disorder, bipolar disorder, borderline personality disorder, mental retardation, and acute suicidality	Psychosis, current substance dependency or abuse, Axis II personality disorders from Cluster A or B, severe depression, and acute suicidality	Major medical illness, diagnosis of substance dependence, schizophrenia, schizoaffective disorder, bipolar disorder, and acute suicidality
Outcome measures	Hamilton Rating Scale for Depression	Liebowitz Social Anxiety Scale	Yale-Brown Obsessive Compulsive Scale for Hypochondriasis

α error of .05 and a statistical power of $1-\beta = .80$) to identify moderate correlations between measures of treatment integrity and outcome measures (our power analysis followed the procedure of Faul, Erdfelder, Lang, & Buchner, 2007). Therefore, 28 cases were chosen from each of the three treatment studies, and one videotape was randomly selected from each case for the evaluation of therapists' adherence and competence. For all 84 videotapes, the middle segment was selected. Session segments had a mean duration of 18.4 minutes ($SD = 4.3$ minutes). In each case, two judges evaluated the treatment integrity of entire sessions and two evaluated the treatment integrity of the middle sessions. The assignment of judges was made in permuted order, so that all combinations of judges (i.e., Judge 1 with 2, Judge 1 with 3, Judge 1 with 4, Judge 2 with 3, Judge 2 with 4, and Judge 3 with 4) evaluated equal numbers of middle segments and entire sessions.

PARTICIPANTS

Clients

Sociodemographic and comorbidity data for the sample appears in Table 2.

Therapists

The 84 clients were treated by 50 therapists (39 female and 11 male). The average age of the therapists was 31.6 ($SD = 5.7$, range: 25–51 years). All therapists have a master's degree in clinical psychology. Fifteen (30.0%) are licensed psychotherapists and 35 were in psychotherapy training. On average, therapists had 2.9 years ($SD = 3.5$ years, range = 0–20 years) of clinical experience after their master's degree in clinical psychology. All therapists were trained via workshops for at least 20 hours in the underlying theory and practical elements of the treatments. Therapists received regular supervision at least once per month by licensed psychotherapists who have a Ph.D., at least 8 years of clinical experience, and are experts

in the administered treatment (e.g., publications in the field, author of a relevant treatment manual). In the current study, the average number of clients treated by therapists was 1.7 ($SD = 1.2$, range = 1–6). Thirty-one therapists treated 1 client, 12 treated 2 clients, 2 treated 3 clients, 3 treated 4 clients, 1 treated 5 clients, and 1 treated 6 clients.

Judges

Adherence and competence ratings were provided by four judges. All are master's-level clinical psychologists (three are licensed psychotherapists and one was at the end of psychological training) who had a mean of 6.3 years ($SD = 3.1$, range = 3.3–10.7) of clinical experience. All judges were trained in CBT for recurrent depression, social anxiety disorder, and hypochondriasis (at least a 2-day workshop for each disorder) and had practical experience in treating those clients. The judges received up-to-date versions of the three treatment manuals before they underwent rater training and were asked to study them intensively. Judges completed an 18-hour training course on how to use the adherence and competence scales. Training involved practicing using the two integrity scales to rate sessions from the three treatment outcome studies that were unique from the 84 cases selected for this study. During the training, all four judges practiced coding equivalent numbers of sessions and session segments.

COGNITIVE-BEHAVIORAL TREATMENTS

Maintenance CBT for Recurrent Major Depressive Disorder

In this investigation, therapy was aimed at preventing the recurrence of depressive episodes and included the following components: (a) cognitive case conceptualization, assessment of individual psychological risk factors for past recurrences, and deriving a psychological model of recurrent depression; (b) mindfulness meditation exercises; (c) modification of dysfunctional cognitions and beliefs; (d) behavioral

Table 2
Sociodemographic and Comorbidity Data for the Study Sample

	Recurrent Major Depressive Disorder ($n = 28$)	Social Anxiety Disorder ($n = 28$)	Hypochondriasis ($n = 28$)	Whole Sample ($N = 84$)
Sociodemographic Data				
Female (%)	21 (75.0)	13 (46.4)	16 (57.1)	50 (59.5)
Age (SD)	50.5 (12.3)	35.6 (9.6)	37.4 (10.3)	41.2 (12.6)
Caucasians (%)	28 (100.0)	28 (100.0)	27 (96.4)	83 (98.8)
Comorbid Disorders				
Comorbid Axis I disorder (%)	12 (42.9)	16 (57.1)	14 (50.0)	42 (50.0)
Comorbid Axis II disorder (%)	6 (21.4%)	12 (42.9)	1 (5.3)	19 (22.6)

activation, behavioral experiments, and stress testing; (e) enhancement of cognitions and reinforcement of behaviors that increase psychological well-being; and (f) relapse prevention. The intervention is described in detail in a treatment manual (Risch, Stangier, Heidenreich, & Hautzinger, 2012).

Cognitive Therapy for Social Anxiety Disorder

In this investigation, therapy included the following components: (a) establishing an idiosyncratic model based on a generic model of information processing in social anxiety disorder (Clark & Wells, 1995); (b) conducting role-play-based behavioral experiments to illustrate the adverse influence of self-focused attention, safety behaviors, and anxiety; (c) conducting attention-focused interventions (see Wells, 1990) to externalize self-focused attention; (d) restructuring distorted self-imagery using video feedback and cognitive preparation; (e) receiving survey feedback dealing with critical issues such as blushing and sweating; and (f) examining negative beliefs in anxiety-provoking social situations while abandoning safety behaviors and adopting an external focus in the context of behavioral experiments. The intervention is described in detail in Stangier, Clark, and Ehlers (2006).

Cognitive Therapy for Hypochondriasis

In this investigation, therapy included the following components: (a) establishing an idiosyncratic model based on information processing in hypochondriasis; (b) psychoeducation (e.g., information about the autonomic nervous system); (c) attention-focused interventions (see Wells, 1990) to reduce selective attention to bodily sensations; (d) cognitive restructuring of dysfunctional health-related automatic thoughts and beliefs; (e) behavioral experiments to demonstrate the negative aspects of reassurance and avoidance behavior; (f) imagery rescripting for the modification of intrusive images; and (g) relapse prevention. The intervention is described in detail in a treatment manual (see Weck, 2014).

OUTCOME MEASURES

In all three RCTs, internationally established and disorder-specific interviews were conducted in order to evaluate symptoms at pre- and posttreatment. The evaluation was conducted in each study by experienced master's-level clinicians who were blind to the treatment conditions. In the study on recurrent depression, the Hamilton Rating Scale for Depression (HRSD-17; Hamilton, 1960; German version: Baumann, 1976) was used and demonstrated high internal consistency in the current study (Cronbach's $\alpha = .85$ at posttreatment). In the study on social anxiety disorder, the Liebowitz Social Anxiety Scale

(LSAS; Liebowitz, 1987; German version: Stangier & Heidenreich, 2005) was used and showed high internal consistency in the current study (Cronbach's $\alpha = .90$ at posttreatment). In the study on hypochondriasis, the Yale-Brown Obsessive Compulsive Scale for Hypochondriasis (H-YBOCS; German version: Weck, Gropalis, Neng, & Witthöft, 2013) was used and demonstrated high internal consistency in the current study (Cronbach's $\alpha = .87$ at posttreatment).

PROCESS MEASURES

Adherence Assessment

We developed the Cognitive-Behavioral Therapy Adherence Scale (CBT-AS) to assess the level of therapist adherence to the three different treatment manuals. The items of the CBT-AS were developed on the basis of the Cognitive-Behavioral Maintenance-Adherence Scale (CBMT-AS; Weck, Hilling, et al., 2011), which evaluated adherence to the treatment manual for recurrent depression. The interrater reliability of the CBMT-AS in former studies was $ICC_{(2,2)} = .83-.99$ for the single items and $ICC_{(2,2)} = .72-.93$ for the mean score of all items (Weck, Hilling, et al., 2011). The wording of the CBMT-AS was adapted and expanded to also evaluate the manualized interventions for social anxiety disorder and hypochondriasis. We included only items that allowed an evaluation of all three treatments. The CBT-AS consists of nine items that were evaluated in the current study: (a) time management, (b) use of materials, (c) application of treatment content, (d) consideration of general principles of CBT, (e) development of an individual cognitive-behavioral model, (f) modification of automatic thoughts and dysfunctional behavior, (g) assigning homework, (h) relapse prevention, and (i) the exclusion of nonadherent techniques (e.g., interventions of psychodynamic therapy). During the development of the CBT-AS and training rating phase, a 3-point rating scale (0 = *not adherent*, 1 = *partly adherent*, and 2 = *adherent*) proved to be most appropriate. For all items of the CBT-AS, descriptions of the category "not adherent" (e.g., Item 2 [use of materials]: "The therapist does not use any materials and thus rates the interrelationships without the treatment material. That is, checklists, tables, protocols or diagrams of models and so on are not used, although the relevant main things are discussed") and "adherent" (e.g., Item 2 [use of materials]: "The therapist uses the appropriate material or gives it to the clients") were provided. A copy of the CBT-AS is available from the first author upon request.

Competence Assessment

Therapist competence was evaluated with the Cognitive Therapy Scale (CTS; Young & Beck,

1980; German version: Weck, Hautzinger, Heidenreich, & Stangier, 2010). The CTS was used to assess the level of therapist competence. The German version of the CTS contains 14 items that evaluate the level of therapist competence in conducting specific components of cognitive therapy: (a) agenda setting, (b) dealing with problems/questions/objections, (c) clarity of communication, (d) pacing and efficient use of time, (e) interpersonal effectiveness, (f) resource activation, (g) reviewing previously set homework, (h) using feedback and summaries, (i) guided discovery, (j) focus on central cognitions and behavior, (k) rationale, (l) selecting appropriate strategies, (m) appropriate implementation of techniques, and (n) assigning homework. The response format is a 7-point rating scale (0 = *poor*, 1 = *barely adequate*, 2 = *mediocre*, 3 = *satisfactory*, 4 = *good*, 5 = *very good*, and 6 = *excellent*). For all items of the CTS, descriptions of the categories “poor,” “mediocre,” “good,” and “excellent” are provided (e.g., for an “excellent” rating on agenda setting: “Therapist worked with clients to set an appropriate agenda with target problems that were suitable for the time available. Therapist established priorities and then followed the agenda”). The German version of the CTS has previously demonstrated satisfactory to good interrater reliability for the different items, $ICC_{(2,2)} = .66-.95$, and good interrater reliability for the mean score on all 14 items, $ICC_{(2,2)} = .90$ (Weck et al., 2010).

STATISTICAL ANALYSIS

Interrater Reliability

Interrater reliability was analyzed using intraclass correlation coefficients (ICCs), using Model 1, $ICC_{(1,n)}$ (see Shrout & Fleiss, 1979). ICCs were at all times calculated from the mean of two ratings, $ICC_{(1,2)}$. A 95% confidence interval was used to determine statistical significance. In line with Portney and Watkins (2008), ICCs between .50 and .74 were considered moderate and values larger than .75 were considered good, with ICCs between .00 and .49 considered unsatisfactory. For the comparison of different ICCs, Fisher Z transformations were used (see McGraw & Wong, 1996).

Outcome Standardization

The different outcome measures (HRSD-17, LSAS, and H-YBOCS) were standardized by Z transformations because the three measures each have different ranges. Moreover, standardization is necessary because for two of the studies (social anxiety disorder and hypochondriasis), the aim and definition of success was the reduction of symptoms, whereas the other study (recurrent major depressive disorder) aimed to reduce the risk of relapse.

Path-Analysis Models Within Treatments

The aim of the present study was to demonstrate the validity of therapist adherence and competence ratings based on session segments compared with those ratings based on entire sessions. Therefore, for each treatment, two path-analysis models were specified with therapist adherence (CBT-AS) and competence (CTS) as predictor variables and the standardized outcome as the criterion variable. One model was a one-group model that assumed no differences between session segments and entire sessions, whereas the other model was a two-group model comparing the session segments with entire sessions. The respective two path-analysis models within each treatment were evaluated with regard to model fit to decide whether the one- or two-group path-analysis model should be regarded as superior.

Path-Analysis Models Across Treatments

Furthermore, both for session segments and entire sessions, two path-analysis models across the three therapy treatments were specified, with therapist adherence (CBT-AS) and therapist competence (CTS) as predictor variables, and the standardized outcome measure as the criterion variable. Additionally, the correlations between the predictor variables were estimated. The first path-analysis model (one-group) was evaluated with equality constraints for the parameters, assuming no differences between the three treatments. The second model was a multi-sample path-analysis one (three-group) with separate parameter estimations for each treatment. The two path-analysis models were evaluated with respect to model fit to decide whether the one- or three-group path analysis model should be regarded as the superior solution. Finally, the parameter estimates of the superior model based on session segments can be compared with the parameter estimates of the model based on entire sessions.

Data Analysis and Model Fit Evaluation

Data were analyzed using Mplus (Version 6; Muthén & Muthén, 2010), applying the maximum likelihood estimator (ML), because the variables did not deviate substantially from a normal distribution (Kline, 2005). For model fit evaluation, the χ^2 value and the degrees of freedom (*df*) were reported, as well as the root mean square error of approximation (RMSEA), the comparative fit index (CFI), the standardized root mean square residual (SRMR), and the Akaike information criterion (AIC). We followed generally accepted standards concerning model fit (Schermelleh-Engel, Moosbrugger, & Müller, 2003). Specifically, the range of values between 0 and 2 for the ratio of the χ^2 value and the *df* indicate a good model fit, whereas

the range of values between 2 and 3 for this ratio indicate only an acceptable model fit. RMSEA values less than .05 indicated a good model fit, and those less than .08 indicated an acceptable model fit. CFI values ranging between .95 and 1.00 indicate a good model fit, whereas values between .90 and .95 signify an acceptable fit. SRMR values should be below .08.

Results

INTERRATER RELIABILITY AND CORRELATION OF RATINGS BASED ON SESSION SEGMENTS AND ENTIRE SESSIONS (HYPOTHESIS 1)

Adherence Ratings

Table 3 presents interrater reliabilities for the CBT-AS for ratings based on session segments and entire sessions. Item 7 (assigning homework) was only assessable in the entire session and not in the session segment. For most of the CBT-AS items, interrater reliabilities were satisfactory to good for both the segment- and entire-session-based ratings.

No significant differences were found between the mean adherence scores based on session segments and on entire sessions, $F(1, 66) = .02, p > .10$. A strong correlation was found between the mean adherence score (Items 1–9 of the CBT-AS) based on session segments and entire sessions ($r = .62, p < .001$).

Competence Ratings

Table 4 presents interrater reliabilities of the CTS for ratings based on session segments and entire sessions. Items 1 (agenda setting), 7 (reviewing previously set homework), and 14 (assigning homework) were only assessable in entire sessions. For

most CTS items, interrater reliabilities were satisfactory to good for both the segment- and entire-session-based ratings.

The mean competence score based on session segments was significantly lower than the mean competence score based on entire sessions, $F(1, 166) = 5.10, p < .03$. However, differences in the mean score between segment- and entire-session ratings were attributable mainly to the three items of the CTS (i.e., agenda setting, reviewing previously set homework, and assigning homework), which could not be evaluated in the session segments. When those items of the CTS were not considered for the calculation of the mean competence scores, the data were almost identical for ratings based on session segments ($M = 4.07, SD = .73$) and for entire sessions ($M = 4.10, SD = .77$) and did not differ significantly from each other, $F(1, 166) = 0.70, p > .10$. A strong correlation was also found for the mean competence scores (Items 1–14 of the CTS) based both on session segments and on entire sessions ($r = .73, p < .001$).

COMPARISON OF THE RELIABILITY OF RATINGS BASED ON SESSION SEGMENTS AND ENTIRE SESSIONS (HYPOTHESIS 2)

Adherence Ratings

For Items 8 (relapse prevention; $Z = 3.71, p < .01$) and 9 (exclusion of nonadherent techniques; $Z = 2.19, p < .05$), segment-based ratings had significantly higher ICCs than entire session-based ratings. No significant differences were found between the interrater reliabilities of the mean adherence scores (Items 1–9) based on session segments, $ICC_{(1,2)} = .72$, versus entire sessions, $ICC_{(1,2)} = .79, Z = 1.05, p > .10$.

Table 3

Interrater Reliabilities for Segment-Based and Entire Session-Based Adherence Ratings, Based on the Cognitive-Behavioral Therapy Adherence Scale (CBT-AS)

	Session Segment		Entire Session	
	<i>M</i> (<i>SD</i>)	<i>ICC</i> _(1,2)	<i>M</i> (<i>SD</i>)	<i>ICC</i> _(1,2)
1. Time management	1.59 (0.67)	.80***	1.59 (0.66)	.74***
2. Use of materials	1.37 (0.87)	.79***	1.70 (0.62)	.74***
3. Application of treatment content	1.89 (0.39)	.69***	1.92 (0.31)	.71***
4. Consideration of general principles of cognitive-behavioral therapy	1.89 (0.35)	.41*	1.86 (0.37)	.49**
5. Development of an individual cognitive-behavioral model	1.31 (0.85)	.83***	1.55 (0.69)	.76***
6. Modification of automatic thoughts and dysfunctional behavior	1.67 (0.65)	.74***	1.57 (0.76)	.76***
7. Assigning homework	-	-	1.62 (0.69)	.79***
8. Relapse prevention	0.77 (0.82)	.85** ^a	0.89 (0.85)	.59*
9. Exclusion of nonadherent techniques	1.95 (0.26)	.63*** ^a	1.92 (0.32)	.38***
CBT-AS mean score (Items 1–6, 8, 9)	1.72 (0.29)	.72***	1.74 (0.31)	.72***
CBT-AS mean score (Items 1–9)	-	-	1.72 (0.30)	.79***

Note. ICC = intraclass correlation coefficient.

* $p < .05$, ** $p < .01$, *** $p < .001$.

^a = ICCs of segment-based ratings were significantly higher than entire session-based ratings.

Table 4

Interrater Reliabilities for Segment-Based and Session-Based Competence Ratings, Based on the Cognitive Therapy Scale (CTS)

	Session Segment		Entire Session	
	<i>M</i> (<i>SD</i>)	ICC _(1,2)	<i>M</i> (<i>SD</i>)	ICC _(1,2)
1. Agenda setting	-	-	1.88 (1.75)	.77***
2. Dealing with problems/questions/objections	4.28 (0.82)	.66***	4.24 (0.82)	.50**
3. Clarity of communication	4.39 (0.80)	.61***	4.32 (0.80)	.47**
4. Pacing and efficient use of time	4.07 (1.05)	.63***	3.57 (1.19)	.70***
5. Interpersonal effectiveness	4.26 (0.91)	.62***	4.43 (0.91)	.66***
6. Resource activation	3.52 (1.04)	.63***	3.88 (1.08)	.78***
7. Reviewing previously set homework	-	-	2.75 (1.82)	.77***
8. Use of feedback and summaries	3.95 (0.92)	.45**	3.95 (0.93)	.54***
9. Guided discovery	3.82 (1.02)	.48**	3.83 (0.93)	.60**
10. Focus on central cognitions and behavior	4.15 (0.93)	.70***	4.17 (0.97)	.72***
11. Rationale	3.86 (0.97)	.50**	4.13 (1.03)	.63***
12. Selection of appropriate strategies	4.43 (1.00)	.62***	4.60 (1.00)	.55***
13. Appropriate implementation of techniques	4.04 (0.93)	.64***	4.01 (0.95)	.72***
14. Assigning homework	-	-	3.29 (1.50)	.73***
CTS mean score (Items 2–6, 8–13)	4.07 (0.72)	.72***	4.10 (0.77)	.76***
CTS mean score (Items 1–14)	-	-	3.81 (0.77) ^a	.77***

Note. ICC = intraclass correlation coefficient.

* $p < .05$, ** $p < .01$, *** $p < .001$.

^a = This score differed significantly ($p < .03$) from the segment-based CTS mean score (Items 2–6, 8–13).

Competence Ratings

For none of the items were interrater reliabilities of the ratings based on session segments significantly different than for the ratings based on entire sessions (all $ps > .10$). For the mean competence score (Items 1–14), no significant differences were found regarding the interrater reliability for ratings based on session segments, $ICC_{(1,2)} = .72$, and ratings based on entire sessions, $ICC_{(1,2)} = .77$, $Z = 0.72$, $p > .10$.

COMPARISON OF THE ADHERENCE- AND COMPETENCE-OUTCOME RELATIONSHIP OF RATINGS BASED ON SESSION SEGMENTS AND ENTIRE SESSIONS (HYPOTHESIS 3)

Analyses Within Treatments

For the recurrent major depressive disorder treatment, the two-group model ($\chi^2 = 0.00$, $df = 2$, $AIC = 282.30$) did not yield a superior model fit relative to the one-group model ($\chi^2 = 0.00$, $df = 1$, $AIC = 272.64$). The same was true for the two-group model ($\chi^2 = 2.32$, $df = 2$, $AIC = 252.83$) and the one-group model ($\chi^2 = 2.54$, $df = 1$, $AIC = 248.41$) for the social anxiety disorder treatment, and for the two-group model ($\chi^2 = 1.05$, $df = 2$, $AIC = 278.48$) and the one-group model ($\chi^2 = 0.64$, $df = 1$, $AIC = 276.52$) for the hypochondriasis treatment.

Analyses Across Treatments

The one-group path-analysis model yielded a good fit for session segments ($\chi^2 = 0.14$, $df = 1$,

$RMSEA = .00$, $CFI = 1.00$, $SRMR = .01$, $AIC = 406.14$) and for entire sessions ($\chi^2 = 1.23$, $df = 1$, $RMSEA = .03$, $CFI = .99$, $SRMR = .03$, $AIC = 417.75$). The multisample path (three-group) models also yielded a good model fit for session segments ($\chi^2 = 2.86$, $df = 3$, $RMSEA = .00$, $CFI = 1.00$, $SRMR = .05$, $AIC = 402.10$) and for entire sessions ($\chi^2 = 2.76$, $df = 3$, $RMSEA = .00$, $CFI = 1.00$, $SRMR = .05$, $AIC = 413.67$) and were slightly superior to the one-group solutions with regard to the AIC. Figure 1 shows the parameter estimates for the multisample path model based on session segments and for the model based on entire sessions. There was no relationship between therapist adherence and therapy outcome in any of the analyzed models.

The parameter estimates of the relationship between therapist competence and therapy outcome differed between the three treatments (see Figure 1). For the subsample of clients with social anxiety disorder, a significant relationship between the CTS and therapy outcome was found for both the segment- and entire session-based ratings. Overall, the pattern of parameter estimates remained the same, regardless of whether the ratings were based on session segments or on entire sessions.

Discussion

The aim of the current study was to investigate the suitability of ratings based on session segments for the evaluation of treatment integrity. Ratings of therapist adherence and competence based on session segments

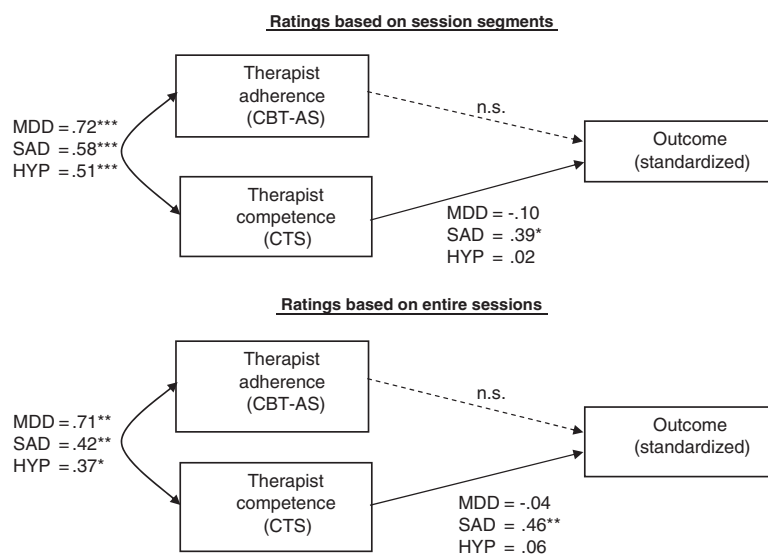


FIGURE 1 Completely standardized solution for the multisample path-analysis model (three-groups) for ratings based on sessions segments (top) and for those based on entire sessions (bottom). CBT-AS = Cognitive-Behavioral Therapy Adherence Scale; CTS = Cognitive Therapy Scale; MDD = maintenance cognitive-behavioral therapy for recurrent major depressive disorder; SAD = cognitive therapy for social anxiety disorder; HYP = cognitive therapy for hypochondriasis. * $p < .05$, ** $p < .01$, *** $p < .001$.

(i.e., the middle third of the sessions) were compared with those based on entire sessions. The results revealed that ratings of therapist adherence and competence based on segments were (a) equivalent to and (b) strongly correlated with ratings based on entire sessions (Hypothesis 1). The reliability values for ratings of therapist adherence and competence that were based on segments were not significantly different from the reliability found for ratings based on entire sessions (Hypothesis 2). Moreover, the adherence- and competence-outcome relationship of ratings based on session segments and entire sessions were comparable, and significant relationships between therapist competence and therapy outcome were only found in the treatment of social anxiety disorder (Hypothesis 3).

The current study addressed several limitations of a former study comparing the reliability and validity of adherence and competence ratings based on session segments and entire sessions (Weck et al., 2011a). First, the findings in the current study are not only limited to the treatment of social anxiety disorder, but are also based on treatments for major depressive disorder and hypochondriasis. Second, a larger number of therapists was used in the current study than in the earlier study, thereby increasing generalizability of findings. Third, in the earlier investigation, the same two judges always evaluated treatment integrity of session segments, whereas two other judges evaluated treatment integrity of entire ses-

sions. Thus, it was possible that individual differences between the raters affected the results. In contrast, all four judges in the current study evaluated equal numbers of both session segments and entire sessions. Consequently, comparisons between ratings based on session segments and entire sessions could no longer be confounded by different characteristics of the judges. Fourth, in the study by Weck et al. (2011a), judges evaluated all three thirds of the therapy sessions to compare the reliability and validity of ratings based on different segments to ratings based on entire sessions. Thus, it was possible that the judges remembered information from other segments, which may have led to an overestimation of the reliability and validity of segment-based ratings. In the current study, only the most promising session segment (i.e., the middle third of the session) was considered and judges had no information about the remaining content of the therapy session. Thus, contamination effects could be excluded.

The relationship between treatment integrity and therapy outcome was comparable for ratings based on session segments and those based on entire sessions. Therefore, our results support the suitability and methodological viability of segment-based ratings in comparison to entire session-based ratings. With the exception of therapist competence predicting outcome in social anxiety disorder, we generally found a nonsignificant relationship between treatment integrity and outcome. Similarly, a meta-

analysis including mostly studies that examined only one or a few therapy sessions did not find a significant adherence- or competence-outcome relationship (Webb, DeRubeis, & Barber, 2010). Empirical research has established that 5 to 10 treatment sessions must be considered to create a stable measurement of treatment integrity (Dennhag et al., 2012), which could account for our failure to find a relationship between treatment integrity and therapy outcome. Nevertheless, our findings indicate that rating session segments could be a useful means of assessing treatment integrity across a larger number of treatment sessions with reduced effort and cost.

Correlations between treatment integrity and therapy outcome were significant for the subsample of clients with social anxiety disorder, but not for the subsample of clients with depression or hypochondriasis. The relationship between treatment integrity and outcome has also been found to be inconsistent in previous studies (Webb et al., 2010), raising the possibility that different methodological approaches (either to the evaluation of treatment integrity or the evaluation of treatment outcome) might influence findings regarding these relationships. One advantage of the current study is the use of identical methods across the three samples for the assessment of treatment integrity. Therefore, characteristics unique to social anxiety disorder might be responsible for the emergence of this relationship, and future investigations may wish to explore what those specific characteristics might be.

Whereas mean adherence and mean competence ratings were assessed with satisfactory to good interrater reliability, some items from the adherence (e.g., Items 4, 8, and 9) and competence (e.g., Items 3, 8, and 9) scales did not yield satisfactory interrater reliabilities. This finding differs from previous investigations in which almost all adherence and competence items were assessed with good interrater reliability (e.g., Weck, Hilling, et al., 2011). However, in the current study, it was not the interrater reliability of the same two judges that was computed, but the interrater reliability of four alternating judges, which were combined in permuted order (i.e., Judge 1 with 2, Judge 1 with 3, Judge 1 with 4, Judge 2 with 3, Judge 2 with 4, and Judge 3 with 4). This study design likely led to lower interrater reliabilities due to the fact that consensus within six rater combinations must be achieved. Furthermore, we had to use ICC Model 1, which is more restrictive than Model 2 (see Shrout & Fleiss, 1979). Thus, it may be that interrater reliabilities were underestimated in the current study.

We also found that two of the adherence items (i.e., relapse prevention and exclusion of nonad-

herent techniques) evidenced significantly higher interrater reliability when ratings were based on session segments compared with when ratings were based on entire sessions. Therefore, the use of session segments may even be advantageous compared with the use of entire sessions for the evaluation of some indices of treatment integrity. It might be that when evaluating a longer videotape sequence, raters find it more difficult to remember and process all of the information relative to evaluating shorter segments.

Some limitations of this investigation should be taken into consideration. First, treatment integrity was only evaluated in CBT; thus, the results may not be generalizable to other (e.g., psychodynamic) therapeutic approaches. Because differences have been found between CBT and other approaches (e.g., supportive expressive therapy and individual counseling) with respect to the stability of adherence and competence evaluations (Dennhag et al., 2012), differences might also emerge for different treatment approaches with respect to the suitability of session segments for the evaluation process.

Second, we only took one treatment session into account for the assessment of treatment integrity. Therefore, it remains unclear whether rating only one therapy session is adequate for the assessment of treatment integrity. Third, it is questionable whether the CTS is appropriate for assessing competence in all three treatments we examined because the CTS was originally developed to assess therapist competence in the treatment of depression. However, the CTS has also been used to assess therapist competence in the treatment of psychosis, social anxiety disorder, antisocial personality disorder, generalized anxiety disorder, and across a range of other disorders (see Muse & McManus, 2013). Most items in the CTS were relatively broadly formulated (e.g., agenda setting, clarity of communication, using feedback and summaries, selecting appropriate strategies, assigning homework) and only one item (guided discovery) has a uniquely cognitive focus. Therefore, we believe that the CTS is also adequate for evaluating other treatment strategies, such as the mindfulness meditation exercises that were part of maintenance CBT for recurrent major depressive disorder. However, the CTS could not address all possible prescribed treatment components and should be adapted for very specific treatment strategies. Relatedly, treatments in the current study clearly had more of a cognitive focus than a behavioral one; thus, our results may be more relevant for therapies with such a cognitive focus.

Fourth, the assessments were conducted in the context of highly standardized treatment studies, which could lead to high adherence scores with low variance. This limited variance may have restricted the

likelihood of finding correlations between treatment integrity and outcome. Therefore, in addition to addressing these issues in highly controlled clinical trials, less standardized settings should also be utilized in future studies. Moreover, therapists' adherence was evaluated with a 3-point scale format, which may have further restricted variance; less restricted scale formats should thus be considered in future studies. Finally, even though the reliability of segment-based ratings seems to be comparable to entire session ratings, it is important to remember that this does not speak to the *validity* of those ratings. If the primary goal of an integrity check is to ensure that a therapist did not use specific nonadherent techniques, it would be better to consider a whole therapy session instead of only a third. Likewise, because specific aspects of therapist adherence and competence were not assessable in the middle segments (i.e., agenda setting, assigning homework, and reviewing homework), when such aspects are of interest (e.g., Bryant, Simons, & Thase, 1999; Weck, Richtberg, Esch, Höfling, & Stangier, 2013), entire sessions or a later segment should be considered for the evaluation process to ensure that this behavior is observable. Of course, one must weigh such advantages against the cost of coding entire sessions.

In summary, the current study provides further empirical evidence of the suitability of session segments for the evaluation of treatment integrity. Ratings based on videotapes from the middle third of therapy sessions demonstrated comparable reliability and comparable relationships with treatment outcome to entire session-based ratings. These results have important implications for psychotherapy research, because the consideration of session segments, in contrast to entire sessions, leads to a considerable reduction in both the time needed for and the costs associated with the evaluation of treatment integrity. Hopefully, this investigation will influence current practice and lead to a more frequent assessment of treatment integrity in psychotherapy research, and consequently ensure greater internal and external validity in treatment studies.

Conflict of Interest Statement

The authors declare that there are no conflicts of interest.

References

- Bachrach, H., Curtis, H., Escoll, P., Graff, H., Huxster, H., Ottenberg, P., & Pulver, S. (1981). Units of observation and perspectives on the psychoanalytic process. *British Journal of Medical Psychology*, *54*, 25–33. <http://dx.doi.org/10.1111/j.2044-8341.1981.tb01466.x>
- Baumann, U. (1976). *Methodische Untersuchungen zur Hamilton-Depression-Skala* [Methodologic studies of the Hamilton Depression Scale for depression]. *Archiv für Psychiatrie und Nervenkrankheiten*, *222*, 359–375.
- Bryant, M. J., Simons, A. D., & Thase, M. E. (1999). Therapist skill and patient variables in homework compliance: Controlling an uncontrolled variable in cognitive therapy outcome research. *Cognitive Therapy and Research*, *23*, 381–399. <http://dx.doi.org/10.1023/A:1018703901116>
- Clark, D. M., & Wells, A. (1995). A cognitive model of social phobia. In R. G. Heimberg, M. Liebowitz, D. Hope, & F. Schneider (Eds.), *Social phobia: Diagnosis, assessment, and treatment* (pp. 69–93). New York, NY: Guilford Press.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159.
- Dennhag, I., Gibbons, M. B. C., Barber, J. P., Gallop, R., & Crits-Christoph, P. (2012). How many treatment sessions and patients are needed to create a stable score of adherence and competence in treatment of cocaine dependence? *Psychotherapy Research*, *22*, 475–488. <http://dx.doi.org/10.1080/10503307.2012.674790>
- Faul, F., Erdfelder, E., Lang, A. -G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *29*, 175–191. <http://dx.doi.org/10.3758/BF03193146>
- First, M. B., Gibbon, M., Spitzer, R. L., Williams, J. B. W., & Benjamin, L. S. (1997). *Structured Clinical Interview for DSM-IV Axis II Personality Disorders (SCID-II)*. Washington, DC: American Psychiatric Press.
- First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (1997). *Structured Clinical Interview for DSM-IV Axis I Disorder (SCID-I)*. Washington, DC: American Psychiatric Press.
- Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry*, *23*, 56–62. <http://dx.doi.org/10.1136/jnnp.23.1.56>
- Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York, NY: Guilford Press. <http://dx.doi.org/10.1177/1049731509336986>
- Liebowitz, M. R. (1987). Social phobia. *Modern Problems of Pharmacopsychiatry*, *22*, 141–173.
- Martino, S., Ball, S., Nich, C., Frankforter, T. L., & Carrol, K. M. (2009). Correspondence of motivational enhancement treatment integrity ratings among therapists, supervisors, and observers. *Psychotherapy Research*, *19*, 181–193. <http://dx.doi.org/10.1080/10503300802688460>
- McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, *1*, 30–46. <http://dx.doi.org/10.1037/1082-989X.1.4.390>
- Mintz, J., & Luborsky, L. (1971). Segments versus whole sessions: Which is the better unit for psychotherapy process research? *Journal of Abnormal Psychology*, *78*, 180–191. <http://dx.doi.org/10.1037/h0031969>
- Muse, K., & McManus, F. (2013). A systematic review of methods for assessing competence in cognitive-behavioural therapy. *Clinical Psychology Review*, *33*, 484–499. <http://dx.doi.org/10.1016/j.cpr.2013.01.010>
- Muthén, L. K., & Muthén, B. O. (2010). *Mplus user's guide* (6th ed.). Los Angeles, CA: Author.
- Perepletchikova, F., Hilt, L. M., Chereji, E., & Kazdin, A. E. (2009). Barriers to implementing treatment integrity procedures: Survey of treatment outcome researchers. *Journal of Consulting and Clinical Psychology*, *77*, 212–218. <http://dx.doi.org/10.1037/a0015232>
- Perepletchikova, F., & Kazdin, A. E. (2005). Treatment integrity and therapeutic change: Issues and research recommendations. *Clinical Psychology: Science and Practice*, *12*, 365–383. <http://dx.doi.org/10.1093/clipsy/bpi045>

- 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. <http://dx.doi.org/10.1037/0022-006X.75.6.829>
- Portney, L. G., & Watkins, M. P. (2008). *Foundations of clinical research: Applications to practice* (3rd ed.). London, UK: Pearson Education.
- Risch, A. K., Stangier, U., Heidenreich, T., & Hautzinger, M. (2012). *Kognitive Erhaltungstherapie bei rezidivierender depression: Rückfälle verhindern, psychische Gesundheit erhalten* [Cognitive behavioral maintenance therapy for recurrent depression: Preventing relapse, maintaining mental health]. Heidelberg, Germany: Springer.
- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Test of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online, 8*, 23–74.
- Schlosser, R. W. (2002). On the importance of being earnest about treatment integrity. *Augmentative and Alternative Communication, 18*, 36–44. <http://dx.doi.org/10.1080/714043395>
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420–428. <http://dx.doi.org/10.1037//0033-2909.86.2.420>
- Stangier, U., Clark, D. M., & Ehlers, A. (2006). *Soziale phobie* [Social phobia]. Göttingen, Germany: Hogrefe.
- Stangier, U., & Heidenreich, T. (2005). Liebowitz social anxiety scale. In *Collegium internationale psychiatriae scalarum*. Göttingen, Germany: Beltz Test.
- Stangier, U., Hilling, C., Heidenreich, T., Risch, A. K., Barocka, A., Schlösser, R., . . . Hautzinger, M. (2013). Maintenance cognitive-behavioral therapy and manualized psychoeducation in the treatment of recurrent depression: A multicenter prospective randomized controlled trial. *American Journal of Psychiatry, 170*, 624–632. <http://dx.doi.org/10.1176/appi.ajp.2013.12060734>
- Stangier, U., Schramm, E., Heidenreich, T., Berger, M., & Clark, D. M. (2010). Cognitive therapy vs interpersonal psychotherapy in social anxiety disorder: A randomized controlled trial. *Archives of General Psychiatry, 68*, 692–700. <http://dx.doi.org/10.1001/archgenpsychiatry.2011.67>
- Waltz, J. A., 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. <http://dx.doi.org/10.1037/0022-006X.61.4.620>
- Webb, C. A., DeRubeis, R. J., & Barber, J. P. (2010). Therapist adherence/competence and treatment outcome: A meta-analytic review. *Journal of Consulting and Clinical Psychology, 78*, 200–211. <http://dx.doi.org/10.1037/a0018912>
- Weck, F., Bohn, C., Ginzburg, D. M., & Stangier, U. (2011a). Assessment of adherence and competence in cognitive therapy: Comparing session segments with entire session. *Psychotherapy Research, 21*, 658–669. <http://dx.doi.org/10.1080/10503307.2011.602751>
- Weck, F., Bohn, C., Ginzburg, D. M., & Stangier, U. (2011b). Treatment integrity: Implementation, assessment, evaluation, and correlations with outcome. *Verhaltenstherapie, 21*, 99–107. <http://dx.doi.org/10.1159/000328840>
- Weck, F., Gropalis, M., Neng, J. M. B., & Witthöft, M. (2013). The German version of the H-YBOCS for the assessment of hypochondriacal cognitions and behaviors: Development, reliability and validity. *International Journal of Behavioral Medicine, 20*, 618–626. <http://dx.doi.org/10.1007/s12529-012-9276-8>
- Weck, F., Hautzinger, M., Heidenreich, T., & Stangier, U. (2010). Erfassung psychotherapeutischer kompetenz: Validierung einer deutschsprachigen version der Cognitive Therapy Scale [Assessing psychotherapeutic competencies: Validation of a German version of the Cognitive Therapy Scale]. *Zeitschrift für Klinische Psychologie und Psychotherapie, 39*, 244–250. <http://dx.doi.org/10.1026/1616-3443/a000055>
- Weck, F., Hilling, C., Schermelleh-Engel, K., Rudari, V., & Stangier, U. (2011). Reliability of adherence and competence assessment in cognitive behavioral therapy: Influence of clinical experience. *Journal of Nervous and Mental Disease, 199*, 276–279. <http://dx.doi.org/10.1097/NMD.0b013e3182124617>
- Weck, F., Neng, J. M. B., & Stangier, U. (2013a). The effects of attention training on the perception of bodily sensations in patients with hypochondriasis: A randomized controlled pilot trial. *Cognitive Therapy and Research, 37*, 514–520. <http://dx.doi.org/10.1007/s10608-012-9482-3>
- Weck, F., Neng, J. M. B., & Stangier, U. (2013b). *Effektivität von kognitiver therapie vs. expositionstherapie bei hypochondrie: Erste ergebnisse einer randomisierten kontrollierten studie* [Efficacy of cognitive therapy vs exposure therapy for hypochondriasis: Preliminary results of a randomized controlled trial]. Paper presented at the Congress of the German Psychological Society, Trier, Germany.
- Weck, F., Richtberg, S., Esch, S., Höfling, V., & Stangier, U. (2013). The relationship between therapist competence and homework compliance in maintenance cognitive therapy for recurrent depression: Secondary analysis of a randomized trial. *Behavior Therapy, 44*, 162–172. <http://dx.doi.org/10.1016/j.beth.2012.09.004>
- Weck, F. (2014). Treatment of mental hypochondriasis: a case report. *Psychiatric Quarterly, 85*, 57–64. <http://dx.doi.org/10.1007/s11126-013-9270-6>
- Wells, A. (1990). Panic disorder in association with relaxation induced anxiety: An attentional training approach to treatment. *Behavior Therapy, 21*, 273–280.
- Wittchen, H. -U., & Fydrich, T. (1997). *Strukturiertes Klinisches Interview für DSM-IV. Manual zum SKID-I und SKID-II* [Structured Clinical Interview for DSM-IV. Manual for the SCID-I and SCID-II]. Göttingen, Germany: Hogrefe.
- Wittchen, H. -U., Zaudig, M., Spengler, P., Mombour, W., Hiller, W., Essau, C. A., . . . Williams, J. (1991). Wie zuverlässig ist operationalisierte Diagnostik?: Die test-retest reliabilität des Strukturierten Klinischen Interviews für DSM-III-R [How reliable are operationalized diagnoses?: The test-retest reliability of the Structured Clinical Interview for DSM-III-R]. *Zeitschrift für Klinische Psychologie, 20*, 136–153.
- Yeaton, W. H., & Sechrest, L. (1981). Critical dimensions in the choice and maintenance of successful treatment: Strength, integrity, and effectiveness. *Journal of Consulting and Clinical Psychology, 49*, 156–167. <http://dx.doi.org/10.1037/0022-006X.49.2.156>
- Young, J., & Beck, A. T. (1980). *Cognitive Therapy Scale rating manual*. Unpublished manuscript, Center for Cognitive Therapy, Philadelphia, PA.

RECEIVED: June 25, 2013

ACCEPTED: March 3, 2014

Available online 14 March 2014

Entwicklung und Validierung von Instrumenten zur Überprüfung der therapeutischen Adhärenz und Kompetenz bei der Behandlung der Panikstörung

Florian Grikscheit¹, Thomas Lang², Anne Kordt², Sylvia Helbig-Lang³,
Kira Geisler² und Florian Weck⁴

¹Abteilung Klinische Psychologie und Psychotherapie, Goethe-Universität Frankfurt

²Christoph-Dornier-Stiftung für Klinische Psychologie, Institut Bremen

³Abteilung Klinische Psychologie und Psychotherapie, Universität Hamburg

⁴Abteilung Klinische Psychologie, Psychotherapie und Experimentelle Psychopathologie,
Universität Mainz

Zusammenfassung. Theoretischer Hintergrund: Die Güte der Expositionsvorbereitung wird als wichtiger Faktor für die erfolgreiche Behandlung der Panikstörung mit Agoraphobie gesehen. Jedoch fehlt es an geeigneten Instrumenten, um dies zu untersuchen. **Fragestellung:** Ziel der Studie ist die Konstruktion und Validierung von Skalen zur Erfassung der Adhärenz und Kompetenz bei der Vorbereitung der Exposition. **Methode:** Vier Experten wurden zur Inhaltsvalidierung herangezogen. Die weitere Überprüfung der Reliabilität und Validität erfolgte auf der Basis von 84 psychotherapeutischen Behandlungen von Patienten mit Panikstörung und Agoraphobie. **Ergebnisse:** Die Adhärenz- und Kompetenzskala wiesen eine hohe Interraterreliabilität auf und es zeigten sich deutliche Hinweise für deren Validität. **Schlussfolgerungen:** Die entwickelten Verfahren können sowohl in der klinischen Forschung, als auch in der Aus- und Weiterbildung von Therapeuten sinnvoll eingesetzt werden. **Schlüsselwörter:** Adhärenz, psychotherapeutische Kompetenz, Inhaltsvalidität, Beurteilerübereinstimmung, Faktorenstruktur

Development and Validation of Instruments for the Evaluation of Therapeutic Adherence and Competence in the Treatment of Panic Disorders

Abstract. Background: The quality of preparing exposure interventions can be considered an important factor for the successful treatment of panic disorder with agoraphobia. However, appropriate instruments are missing to investigate this. **Objective:** The study's aim is the construction and validation of scales to evaluate the therapeutic adherence and competence of preparing exposure interventions. **Method:** Four experts were used for content validation. The further verification of reliability and validity was made on the basis of 84 psychotherapeutic treatments of patients with panic disorder and agoraphobia. **Results:** The interrater agreement for the adherence and competence scale was high and there were clear indications for their validity. **Conclusion:** The developed instruments can be used meaningfully in clinical research as well as for therapist training. **Keywords:** adherence, therapeutic competence, content validity, interrater agreement, factor structure

Die Panikstörung mit Agoraphobie ist in der klinischen Praxis ein häufig zu behandelndes Störungsbild. Nach dem Bundesgesundheits-Survey (Jacobi et al. 2004) ist für die Panikstörung eine kombinierte Lebenszeitprävalenz von 3.9% anzunehmen. Studien belegen, dass bei Ausbleiben einer Behandlung ein ungünstiger und chronifizierter Verlauf zu erwarten ist (vgl. Batelaan et al.,

2010; Wittchen, 1991). Für die Behandlung gilt die Kognitive Verhaltenstherapie (KVT) als Standard, deren Effektivität mehrfach bestätigt werden konnte (vgl. Heinrichs, Alpers & Gerlach, 2009; Ruhmland & Margraf, 2001). Darüber hinaus wurde auch für einzelne kognitiv-verhaltenstherapeutische Interventionen die Wirksamkeit belegt (vgl. Lang, Helbig-Lang & Petermann, 2009; Sán-

chez-Meca, Alcázar, Marín-Matínez & Gómez-Conesa, 2010). Demnach können interozeptive Expositionsübungen und vor allem die Exposition in vivo als zentrale Interventionen angesehen werden (vgl. Lang et al., 2009, 2013; Heinrichs et al., 2009). Es ist anzunehmen, dass für die erfolgreiche Durchführung der Exposition in vivo die adäquate Vorbereitung (u. a. Ableitung des Expositionsrationals, verständliche Erklärung der Expositionsregeln, Motivationale Klärung) eine entscheidende Rolle spielt (vgl. Lang, Helbig-Lang, Westphal, Gloster & Wittchen, 2012; Neudeck, 2005). In diesem Zusammenhang stellt sich die Frage, inwieweit Unterschiede in der adhärennten (d. h. manualtreuen) und kompetenten Umsetzung von expositions vorbereitenden Interventionen durch den Therapeuten einen Einfluss auf Faktoren wie die Bereitschaft des Patienten zur Durchführung der Konfrontation, Abbruchraten oder der Symptomreduktion haben (vgl. Waltz, Addis, Koerner & Jacobson, 1993; Weck, Bohn, Ginzburg & Stangier, 2011). Empirische Nachweise zur Bedeutung der Expositions vorbereitung stehen jedoch noch aus. Um solche Fragestellungen untersuchen zu können, werden geeignete Adhärenz- und Kompetenzskalen benötigt. Dass ein Mangel an solchen Instrumenten besteht, zeigen Studien, die den Effekt von therapeutischer Adhärenz bzw. Kompetenz auf den Erfolg von Expositionsbehandlung bei der Panikstörung untersuchen (vgl. Huppert, Barlow, Gorman, Shear & Woods, 2006; Huppert et al. 2001). Beide der genannten Arbeiten weisen methodische Mängel in der Erfassung der Adhärenz und der Kompetenz auf. Kritisch anzumerken ist, dass in der früheren Studie von Huppert et al. (2001) sowohl die Adhärenz als auch die Kompetenz jeweils nur mit einem Globalitem erfasst wurden. In der nachfolgenden Studie (Huppert et al., 2006) wurde zwar ein umfassenderes Instrument für die Erfassung der Adhärenz verwendet, jedoch ohne Aussagen über die Reliabilität anzugeben. Die Erfassung der Kompetenz blieb darüber hinaus in dieser Studie unberücksichtigt.

Um gezielt den Einfluss von expositions vorbereitenden Interventionen in der Panikbehandlung zu untersuchen, ist eine Festlegung dessen notwendig, was unter einer adhärennten und kompetenten Durchführung zu verstehen ist. Für diesen Zweck bedarf es eines etablierten und evaluierten Behandlungsmanuals, in dem eine detaillierte Beschreibung der Expositions vorbereitung enthalten ist. Basierend auf einem solchen Manual können Skalen zur Erfassung der therapeutischen Adhärenz und Kompetenz konstruiert bzw. Anpassungen von bereits vorhandenen Instrumenten vorgenommen werden. Für die Kompetenzerfassung stellt die Cognitive Therapy Scale (CTS; Young & Beck, 1980; deutsche Übersetzung und Validierung von Weck, Hautzinger, Heidenreich & Stangier, 2010) ein geeignetes Instrument dar, das ursprünglich für die Behandlung der Depression entwickelt wurde, inzwischen jedoch bei einer großen Bandbreite von Störungen eingesetzt wird (z. B. Muse & McManus, 2013). Für die Abbildung von kompetenten therapeutischen Verhalten bezüglich exposi-

tions vorbereitender Interventionen erscheint eine Erweiterung der CTS jedoch sinnvoll.

Die auf diese Weise konstruierten Instrumente zur Erfassung der Adhärenz und Kompetenz werden jedoch nicht nur in der klinischen Forschung benötigt, sondern könnten auch in der Aus- und Weiterbildung von Psychotherapeuten sinnvoll eingesetzt werden (Weck, Bohn et al., 2011; Weck, 2013). Beispielweise kann die Evaluation der therapeutischen Kompetenz und Adhärenz von Therapiesitzungen (z. B. im Rahmen der Supervision) dazu beitragen, die Ausführung der angewandten Interventionen zu erlernen oder zu verbessern und somit die Verbreitung von evidenzbasierten Therapieansätzen begünstigen.

Zielsetzung

Ziele der vorliegenden Studie waren: (1) Die Entwicklung zweier Instrumente zur Erfassung der Adhärenz und Kompetenz bei der Vorbereitung von Expositionsübungen im Rahmen der Behandlung der Panikstörung mit Agoraphobie. (2) Analyse der Inhaltsvalidität der entwickelten Items anhand einer Expertenbefragung. (3) Überprüfung der Interraterreliabilitäten und der internen Konsistenz sowie (4) der Konstruktvalidität beider Instrumente. (5) Die Bestimmung der faktoriellen Struktur der Adhärenzskala.

Methode

Beschreibung der Behandlungsstudie

Die Datenbasis der vorliegenden Arbeit liefert eine multizentrische, randomisiert kontrollierte Therapiestudie zur expositions basierten Behandlung der Panikstörung mit Agoraphobie (registriert unter NCT01323556; Lang, Gerlach, Kirchner, Alpers & Hamm, 2013). Die Studie zielt darauf ab, die Effekte einer Angststeigerungsbedingung mithilfe von interozeptiven Übungen während einer Exposition in vivo zu untersuchen. Einschlusskriterien waren das Vorliegen einer Panikstörung mit Agoraphobie (nach DSM-IV; Saß, Wittchen & Zaudig, 1996), Alter zwischen 18–65 Jahren und ein CGI-Score (Clinician Global Impression Scale; Busner & Targum, 2007) von ≥ 4 . Insgesamt wurden 125 Patienten mit Panikstörung und Agoraphobie in die Behandlungsstudie eingeschlossen.

Die Behandlung, basierend auf dem Manual von Lang et al. (2012), welches für die Studie adaptiert wurde (Lang, Helbig, Westphal, Gloster & Wittchen, 2006), umfasste jeweils 12 Therapie- und zwei Boostersitzungen. Während der Fokus der Sitzungen 1 bis 3 auf der Psychoedukation zum Thema Angst und Panik lag, wurde in den Sitzungen 4 und 5 der Schwerpunkt auf die Reduktion der Angst vor Körperempfindungen mit Hilfe von interozeptiven Exposi-

tionsübungen gelegt. Die 6. Sitzung war auf die Vorbereitung der Exposition in vivo ausgerichtet. Aufgrund dessen wurde sie unter Beratung von Autoren des Manuals ausgewählt, um die adhärenz- und kompetente Durchführung der expositionsvorbereitenden Interventionen zu beurteilen. Der entsprechende Abschnitt des Manuals diente dabei als Grundlage für die Konstruktion der beiden Instrumente. Nach diesem Manual ist für die 6. Sitzung vorgesehen, die interozeptiven Expositionsübungen erneut durchzuführen und eine Schlussfolgerung daraus abzuleiten, eine situationale Angsthierarchie zu erstellen, die individuellen Vermeidungsverhaltensweisen des Patienten zu wiederholen und als Hauptintervention ein Gedankenexperiment durchzuführen, aus dem schließlich das Expositionsrational abgeleitet werden soll. Abschließend ist eine ausführliche Besprechung der Expositionsregeln vorgesehen. Darauf aufbauend sollten in den folgenden Sitzungen verschiedene Expositionsübungen in vivo (z. B. Bus fahren) durchgeführt werden.

Stichproben und Untersuchungsablauf

Patienten. Von den 125 Patienten lagen für 110 (91 %) Behandlungen Videoaufzeichnungen der 6. Therapiesitzung vor. Insgesamt wurden 84 (67 %) der Behandlungen randomisiert ausgewählt und in der vorliegenden Studie berücksichtigt (eine Anzahl von 84 Behandlungen wurde auf der Basis einer Poweranalyse ausgewählt, da in weiteren Analysen, die nicht Gegenstand der vorliegenden Arbeit sind, Zusammenhänge zwischen therapeutischen Kompetenzen und Merkmalen der Exposition untersucht werden sollen). Von diesen 84 Patienten waren 53 (63 %) weiblich und das Durchschnittsalter betrug 33,9 Jahre ($SD = 10,0$ Jahre; Spanne = 19–60). Von den Patienten wiesen 66 % eine komorbide Störung auf (hauptsächlich Affektive Störungen und Angststörung).

Therapeuten. Die Behandlung der Patienten wurde von 33 Therapeuten durchgeführt. Diese waren zu 82 % weiblich ($N = 27$) und wiesen ein durchschnittliches Alter von 28,5 Jahren ($SD = 2,9$ Jahre; Spanne 25–37) auf. Die Therapeuten verfügten über eine mittlere klinische Erfahrung von 2,4 Jahre ($SD = 1,7$ Jahre; Spanne = 4–79 Monate). Insgesamt behandelte jeder der Therapeuten im Schnitt innerhalb der Studie 2,5 Patienten ($SD = 1,5$; Spanne = 1–7). Die Therapeuten erhielten vor Beginn der Behandlungen ein 3-tägiges Manualtraining. Nur die Therapeuten wurden für die Studie zugelassen, die eine Zertifizierung zum Studientherapeuten erhielten. Hierfür musste im Rahmen eines Rollenspiels ein Gedankenexperiment zur Herleitung des Expositionsrational durchgeführt werden und von der Studienleitung als fachgerecht eingestuft werden.

Rater und deren Training. Die ausgewählten Videoaufzeichnungen wurden von zwei unabhängigen und erfahrenen Ratern beurteilt. Eine Beurteilerin (weiblich,

31 Jahre und 5,5 Jahre klinische Erfahrung) war promovierte und approbierte Psychotherapeutin, der zweite Beurteiler (männlich, 32 Jahre und 4,5 Jahre klinische Erfahrung) befand sich am Ende der Ausbildung zum Psychologischen Psychotherapeuten. Beide Beurteiler verfügten über Erfahrungen in der Beurteilung der Adhärenz und Kompetenz im Rahmen vorheriger Studien. Zudem erhielten sie ein 9-stündiges Training, in dem die Verwendung der Skalen erläutert und anhand von Trainingsvideos (welche nicht in Datenanalyse eingingen) geübt wurde. Um das zu beurteilende therapeutische Verhalten besser einschätzen zu können, wurden das zugrundeliegende Behandlungsmanual und speziell die korrekte Durchführung der in Sitzung 6 zu implementierenden Intervention in einem weiteren 4-stündigen Training von einem Autor des Behandlungsmanuals (T.L.) verdeutlicht. Die Beurteilungen der Videobänder fanden unabhängig voneinander statt. Um einem Auseinanderdriften der Ratings entgegenzuwirken, verglichen die Rater ihre Einschätzungen von zehn Videos (ca. jedes 5. beurteilte Video) ohne diese jedoch nachträglich zu verändern. Die Rater hatten keine Informationen über den Therapieerfolg der einzelnen Behandlungen.

Die Untersuchungsinstrumente

Konstruktion der Adhärenzskala Expositionsvorbereitung (AS-EXVO). Die AS-EXVO wurde entwickelt um die Adhärenz bei der expositionsbasierten Therapie bei der Panikstörung mit Agoraphobie zu erfassen (Lang, et al. 2006). Auf Basis der in dem Behandlungsmanual für die 6. Therapiesitzung vorgesehenen Interventionen wurden insgesamt 14 Items konstruiert. Ergänzt wurde ein Item zur Erfassung der zeitlichen Strukturierung der Therapiesitzung („Zeitmanagement“), ein weiteres, um den Einsatz von Interventionen aus anderen Therapierichtungen zu erheben sowie ein Item zur Gesamteinschätzung der Manualtreue. Jedes Item enthielt eine Beschreibung der Hauptmerkmale der jeweiligen zu beurteilenden Intervention. Als Antwortformat wurde eine vierstufige Likert-Skala gewählt (0 = keine Manualtreue/nicht durchgeführt; 1 = geringe Manualtreue; 2 = moderate Manualtreue; 3 = hohe Manualtreue). Für die jeweiligen Endpunkte liegen Ankerbeschreibungen vor. Zur Einschätzung des Globalitems wurde ein 7-stufiges Antwortformat (0 = gar nicht manualtreu; 2 = größere Abweichungen; 4 = kleinere Abweichungen; 6 = sehr manualtreu) gewählt. Die 17 konstruierten Items (siehe Tabelle 1) wurden vier Experten (T.L., S.H.-L., K.G. & A.K.) schriftlich und unabhängig voneinander vorgelegt. Die Experten waren Autoren des Behandlungsmanuals bzw. waren an dessen Entwicklung beteiligt. Darüber hinaus verfügten sie über eine langjährige klinische Erfahrung auf dem Gebiet der Verhaltenstherapie sowie der Behandlung der Panikstörung. Die Experten beurteilten jedes Item hinsichtlich dessen Relevanz (0 = gar nicht relevant; 1 = ein

wenig relevant; 2 = ziemlich relevant; 3 = sehr relevant) als auch bezüglich dessen Eignung (0 = gar nicht geeignet; 1 = ein wenig geeignet; 2 = ziemlich geeignet; 3 = sehr geeignet). Darüber hinaus konnten die Experten weitere Anmerkungen zur Verbesserung der Items im Freitext abgeben, sowie ihre Einschätzung darüber, ob Items zur Erfassung der Adhärenz fehlten.

Kompetenzskala Expositions Vorbereitung (KS-EXVO).

Um die Kompetenz des Therapeuten bei der Vorbereitung der Exposition in vivo zu erfassen, wurden zwei Items generiert: (1) „Durchführung interozeptive Expositionsübung“ sowie (2) „Durchführung Gedankenexperiment“. Der strukturelle Aufbau der konstruierten Items entsprach denen der Cognitive Therapy Scale (CTS; Young & Beck, 1980; Weck et al. 2010). Die Items sind beurteilbar anhand eines 7-stufigen Antwortformates, welches von „gar nicht kompetent“ bis „sehr kompetent“ reicht. Für jede zweite Abstufung liegen Beschreibungen des Therapeutenverhaltens als Anker vor. Das Vorgehen der inhaltlichen Validierung war identisch mit der Expertenbefragung hinsichtlich der Adhärenzskala. Aufgrund der inhaltlichen Anmerkungen der Experten wurde ein weiteres Item (3) „Ableitung Expositionsrationale“ konstruiert.

Cognitive Therapy Scale (CTS). Zur Bestimmung der konvergenten Validität wurde die deutsche Version der Cognitive Therapy Scale (CTS; Weck et al., 2010) eingesetzt. Da ein gewisses Ausmaß an Adhärenz als eine Voraussetzung für therapeutische Kompetenz angesehen werden kann, ist ein substantieller Zusammenhang zwischen beiden Konstrukten zu erwarten (vgl. Waltz et al., 1993, Barber, Triffleman & Marmer, 2007). Die CTS besteht aus 14 Items zur Erfassung von kognitiv-verhaltenstherapeutischen Basiskompetenzen und einem Item zur globalen Einschätzung der therapeutischen Kompetenz. Die Reliabilität der CTS für den Gesamtscore (Item 1–14) ist mit einer Beurteilerübereinstimmung von $ICC_{(2,2)} = .90$, $p < .001$ als sehr hoch einzustufen (vgl. Weck et al., 2010). Die interne Konsistenz ist mit einem Cronbachs $\alpha = .86$ ebenfalls als sehr gut zu bewerten.

Interpersonelles Verhalten der Patienten. Zur Untersuchung der diskriminanten Validität wurde die Assessment Form of Patient Interpersonal Behavior (AFPIB; Richtberg, Jakob, Höfling & Weck, 2014) verwendet, die das interpersonelle Verhalten der Patienten in einer Behandlungssitzung erfasst. Das interpersonelle Verhalten der Patienten wird als weitgehend unabhängig von der Erfassung der Adhärenz und Kompetenz des Therapeuten gesehen und wird deshalb zur Bestimmung der diskriminanten Validität herangezogen. Hintergrund für diese Annahme ist, dass unter therapeutischer Kompetenz die Fähigkeit des Therapeuten verstanden wird, Interventionen selbst bei unterschiedlichem (und auch problematischem) interpersonellen Verhalten eines Patienten in einer kompetenten und manualtreuen Art und Weise durchzuführen (vgl. Waltz et al., 1993). Diese Skala besteht aus zehn

Items, wobei jedes Item Verhaltensweisen als Gegensatzpaare beschreibt (z.B. Feindseligkeit vs. Freundlichkeit, Unterbrechungen vs. Aufmerksames Zuhören). Das Instrument weist eine hohe interne Konsistenz (Cronbachs $\alpha = .91$) als auch eine gute Beurteilerübereinstimmung von $ICC_{(2,2)} = .81$, $p < .001$ auf (vgl. Richtberg et al., 2014). Die Konstrukt- und Kriteriumsvalidität konnte in der Validierungsstudie nachgewiesen werden.

Datenanalyse

Zur Bestimmung der Beurteilerübereinstimmung wurden Intraklassenkorrelationen ($ICC_{(2,2)}$) unter der Verwendung des unjustierten Modells berechnet (Shrout & Fleiss, 1979). Nach den Empfehlungen von Portney und Watkins (2008) lassen sich ICC's von $\geq .75$ als gut ansehen. Für die Bestimmung der internen Konsistenz wurde Cronbachs α verwendet. Die faktorielle Struktur der AS-EXVO wurde mittels einer exploratorischen Faktorenanalyse untersucht. Um die Eignung der Daten für die Durchführung einer Faktorenanalyse zu überprüfen, wurde der Bartlett-Test auf Sphärizität durchgeführt sowie der Kaiser-Meyer-Lövin Koeffizient (KMO) berechnet. Nach Bühner (2006) gelten KMO -Koeffizienten < 0.5 als inakzeptabel, < 0.6 als schlecht, < 0.7 als mäßig, < 0.8 als mittelmäßig, < 0.9 als gut und darüber hinaus als sehr gut. Als Extraktionsmethode wurde eine Hauptachsenanalyse gewählt. Um die Anzahl der Faktoren zu bestimmen, wurde eine Parallelanalyse durchgeführt. Hierbei werden nur Faktoren extrahiert, deren empirisch beobachtete Eigenwerte über dem Eigenwerteverlauf einer zufällig erzeugten Datenmatrix liegen (Horn, 1965; O'Conner, 2000). Für die Beurteilung der Konstruktvalidität wurden Produkt-Moment-Korrelationskoeffizienten zwischen der AS-EXVO bzw. KS-EXVO und der CTS und AFPIB berechnet. Alle Analysen wurden mit SPSS 19 durchgeführt.

Ergebnisse

Expertenbefragung

Die quantitativen Ergebnisse der Expertenbefragung für die Items der AS-EXVO und der KS-EXVO sind Tabelle 1 zu entnehmen. Um die Eignung der Items zu verbessern wurden die inhaltlichen Anmerkungen und Verbesserungen bezüglich deren Formulierungen und Ankerbeschreibungen berücksichtigt. Aufgrund der inhaltlichen Anmerkungen der Experten wurde die KS-EXVO um ein weiteres Item „Ableitung Expositionsrationale“ ergänzt.

Beurteilerübereinstimmung

Tabelle 2 zeigt die ICCs für die Items der AS-EXVO, der KS-EXVO sowie der Gesamtwerte der CTS und des

Tabelle 1. Einschätzung der Items der AS-EXVO und KS-EXVO hinsichtlich deren Relevanz und Eignung (Skalierung von 0 = gar nicht relevant/geeignet bis 3 = sehr relevant/geeignet) im Rahmen der Expertenbefragung

Item	Relevanz <i>M (SD)</i>	Eignung <i>M (SD)</i>
AS-EXVO		
1. Tagesordnung	1.25 (0.50)	1.75 (0.50)
2. Hausaufgabenbesprechung	1.75 (0.96)	1.75 (0.50)
3. Durchführung interozeptive Expositionsübung	2.50 (0.58)	2.00 (0.00)
4. Nachbesprechung der interozeptiven Expositionsübungen	2.25 (0.50)	2.00 (0.00)
5. Feststellung: Wiederholung verändert Toleranz und Angstinduktion	2.00 (0.82)	2.25 (0.50)
6. Aufstellen situationale Angsthierarchie	2.50 (0.58)	2.00 (0.00)
7. Auswahl der individuellen Expositionsübungen ^a	1.75 (0.96)	1.50 (0.58)
8. Wiederholung Vermeidungsverhaltensweisen	1.75 (0.96)	2.00 (0.00)
9. Einführung und Vorbereitung des Gedankenexperiments	3.00 (0.00)	1.00 (0.00)
10. Durchführung des Gedankenexperiments ^b	3.00 (0.00)	0.75 (0.96)
11. Schlussfolgerungen aus Gedankenexperiment	2.50 (1.00)	1.00 (0.82)
12. Ableitung des Behandlungsrationalis aus Gedankenexperiment ^c	2.75 (0.50)	1.50 (0.58)
13. Besprechen der Expositionsregeln	2.00 (0.82)	2.00 (0.00)
14. Hausaufgabenvergabe	2.00 (0.82)	1.75 (0.50)
15. Zeitmanagement ^d	2.00 (1.16)	1.50 (1.00)
16. Interventionen aus anderen Therapierichtungen ^e	2.67 (0.58)	0.67 (1.16)
17. Gesamteinschätzung der Manualtreue	–	–
KS-EXVO		
1. Durchführung Expositionsübung	3.00 (0.00)	2.50 (0.58)
2. Durchführung Gedankenexperiment	3.00 (0.00)	1.75 (0.96)

Anmerkungen: ^a wurde nach Expertenbefragung in Item 6 integriert. ^b wurde aufgrund einer fehlerhaften Formulierung, entsprechend den Anmerkungen der Experten, umformuliert. ^c wurde nach Expertenbefragung in Item 11 integriert. ^d wurde nach Expertenbefragung ausgeschlossen. ^e die Polung des Antwortformats dieses Items enthielt einen Fehler, was zu einer geringen Einstufung der Eignung führte. Dieser Fehler wurde behoben.

AFPIB. Für die Items der AS-EXVO finden sich Interraterreliabilitäten zwischen $ICC_{(2,2)} = 0.62$, $p < .001$ und $ICC_{(2,2)} = 0.95$, $p < .001$ sowie eine Intraklassenkorrelation des Mittelwerts aller Items (Gesamtskala) von $ICC_{(2,2)} = 0.93$, $p < .001$. Für die Items der KS-EXVO kann die Reliabilität mit Werten zwischen $ICC_{(2,2)} = .77$, $p < .001$ und $ICC_{(2,2)} = .88$, $p < .001$ als hoch bis sehr hoch angesehen werden.

Faktorenanalyse und interne Konsistenz

Der Kaiser-Meyer-Olkin-Koeffizient (KMO) von .74 und der signifikante Bartlett-Test auf Sphärizität ($\chi^2 = 358.15$, $df = 78$, $p < .000$) deuten auf geeignete Daten hin. Für die ersten vier Faktoren ergab sich ein Eigenwerteverlauf von 4.58, 1.57, 1.17 und 1.12. Aufgrund des Scree-Tests und der Parallelanalyse wurde eine einfaktorische Faktorenlösung gewählt. Ein Faktor klärt dabei 35.3% der Varianz auf. Die Faktorenladungen sind Tabelle 3 zu entnehmen. Die interne Konsistenz (Cronbachs α) für die Adhärenzskala (AS-EXVO) beträgt $\alpha = .82$ für die KS-EXVO $\alpha = .71$.

Konvergente und Diskriminante Validität

Die Korrelationen zwischen psychotherapeutischer Kompetenz (CTS) und der AS-EXVO ($r = .86$; $p < .01$) sowie mit der KS-EXVO ($r = .84$; $p < .01$) erweisen sich entsprechend der konvergenten Validität als hoch. Für den Zusammenhang zwischen der AS-EXVO und der KS-EXVO findet sich eine Korrelation von $r = .79$ ($p < .01$).

Im Sinne der diskriminanten Validität zeigen sich niedrige Korrelation zwischen dem interpersonellen Verhalten des Patienten (AFPIB) und der Adhärenzskala (AS-EXVO) von $r = .27$ ($p < .01$) sowie mit den Kompetenzitems (KS-EXVO) von $r = .34$ ($p < .01$).

Diskussion

Das Ziel der Studie war die Entwicklung von Skalen zur Erfassung der Adhärenz (AS-EXVO) und Kompetenz (KS-EXVO) bei der Vorbereitung von Expositionsübungen bei dem Störungsbild der Panikstörung mit Agoraphobie. In der vorliegenden Arbeit konnten Hinweise für die Reliabilität und Validität der beiden Instrumente gefunden werden.

Tabelle 2. Interklassenkorrelation und Mittelwerte für die Items der Adhärenzskala, deren Gesamtwert, die Items der Kompetenzskala, sowie den Gesamtwert der CTS und des AFPIB

Item	<i>N</i>	<i>M (SD)</i>	<i>ICC (2,2)</i>
Adhärenzskala (AS-EXVO)	83	2.56 (0.85)	.94***
1. Tagesordnung			
2. Hausaufgabenbesprechung	84	2.48 (0.55)	.74***
3. Durchführung interozeptive Expositionsübung	83	2.71 (0.68)	.95***
4. Nachbesprechung der interozeptiven Expositionsübungen	83	2.56 (0.57)	.76***
5. Feststellung: Wiederholung verändert Toleranz und Angstinduktion	83	2.36 (0.74)	.87***
6. Aufstellen situationale Angsthierarchie	83	2.77 (0.39)	.75***
7. Wiederholung Vermeidungsverhaltensweisen	82	2.56 (0.70)	.85***
8. Einführung und Vorbereitung des Gedankenexperiments	82	2.68 (0.45)	.63***
9. Durchführung des Gedankenexperiments	82	2.80 (0.37)	.71***
10. Schlussfolgerungen aus Gedankenexperiment	82	2.63 (0.50)	.84***
11. Besprechen der Expositionsregeln	82	2.65 (0.43)	.62***
12. Hausaufgabenvergabe	82	2.53 (0.53)	.79***
13. Interventionen aus anderen Therapierichtungen	82	2.98 (0.13)	.66***
14. Gesamteinschätzung der Manualtreue	84	4.14 (0.82)	.86***
Gesamtadhärenz (Items 1–13)	84	2.64 (0.31)	.93***
Kompetenzskala (KS-EXVO)			
1. Durchführung Expositionsübung	82	3.93 (0.88)	.88***
2. Durchführung Gedankenexperiment	82	4.12 (0.59)	.77***
3. Ableitung Expositionsrationale	82	3.85(0.66)	.80***
Gesamtkompetenz (Items 1–3)	83	3.97 (0.56)	.88***
CTS Gesamtwert	84	3.86 (0.48)	.94***
AFPIB Gesamtwert	84	2.75 (0.44)	.78***

Anmerkung: * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$.

Tabelle 3. Ergebnisse der Faktorenanalyse des Adhärenzskala AS-EXVO (Hauptachsenanalyse)

	Faktorladungen
1. Tagesordnung	.40
2. Hausaufgabenbesprechung	.64
3. Durchführung interozeptive Expositionsübung	.35
4. Nachbesprechung der interozeptiven Expositionsübungen	.60
5. Feststellung: Wiederholung verändert Toleranz und Angstinduktion	.63
6. Aufstellen situationale Angsthierarchie	.44
7. Wiederholung Vermeidungsverhaltensweisen	.50
8. Einführung und Vorbereitung des Gedankenexperiments	.71
9. Durchführung des Gedankenexperiments	.62
10. Schlussfolgerungen aus Gedankenexperiment	.71
11. Besprechen der Expositionsregeln	.61
12. Hausaufgabenvergabe	.44
13. Interventionen aus anderen Therapierichtungen	.29

Erfassung der Adhärenz und die Konstruktion der AS-EXVO

Bei der Konstruktion der AS-EXVO konnte auf ein ausführliches Behandlungsmanual zurückgegriffen werden, in dem die Durchführung der einzelnen Interventionen genau beschrieben wurde. Dieser Umstand sowie die

durchgeführte Befragung von vier Experten lassen auf eine hohe inhaltliche Validität des Instrumentes schließen. Um mit Hilfe einer Adhärenzskala Unterschiede in der Durchführung von umfangreichen Interventionen (wie z. B. eines Gedankenexperimentes) überhaupt messbar machen zu können, ist es unerlässlich, dass ausführliche Beschreibungen vorliegen, die der intendierten Durchführung entsprechen.

Die Reliabilität der AS-EXVO lässt sich anhand der exzellenten Beurteilerübereinstimmung der Gesamtskala belegen. Auch für die meisten der einzelnen Items zeigen sich sehr hohe Übereinstimmungen. Die Konstruktvalidität der AS-EXVO kann einerseits durch die hohe Korrelation mit der Kompetenz (konvergente Validität) und andererseits mit der ebenfalls erwartungsgemäßen niedrigeren Korrelation (diskriminante Validität) mit dem Interaktionsverhalten des Patienten (AFPIB) bestätigt werden.

Die Ergebnisse der explorativen Faktorenanalyse weisen auf die einfaktorielle Faktorenstruktur der AS-EXVO hin. Jedoch zeigt sich, dass die beiden Items „Durchführung interozeptiver Expositionsübung“ und „Interventionen aus anderen Therapierichtungen“ nur geringe Faktorladungen aufweisen. Besonders überraschend erscheint dieser Umstand im Falle des erstgenannten Items. Eine Erklärung hierfür könnte jedoch darin liegen, dass sich die interozeptiven Expositionsübungen in der Durchführung deutlich von den anderen Interventionen unterscheiden. Eine adhärenzte Umsetzung dieser Intervention besteht darin, gemeinsam mit dem Patienten vier interozeptive Expositionsübungen (u. a. Luft anhalten, auf der Stelle drehen, auf der Stelle rennen, Kopf schütteln) zwei Mal hintereinander durchzuführen. Demnach unterscheidet sich diese Intervention von allen anderen zum einen in der sehr (körperlich) aktiven Beteiligung des durchführenden Therapeuten und zum anderen darin, dass es sich hierbei um eine tatsächliche Exposition handelt. Dennoch ist davon auszugehen, dass die interozeptiven Übungen als auch die daraus abgeleiteten Schlussfolgerungen (v. a. Wiederholung verändert die Angsttoleranz und die Gefahrenbewertung) einen Beitrag zur Vorbereitung der Exposition in vivo leisten. So kann der Patient beispielsweise bereits von den gemachten Erfahrung während der interozeptiven Exposition auf die Exposition in vivo schlussfolgern.

Hinsichtlich der konstruierten Items ist kritisch anzumerken, dass lediglich für die Endpunkte des Antwortformates (0 = keine Manualtreue und 3 = hohe Manualtreue) Ankerbeschreibungen vorliegen. Dem ist jedoch entgegenzusetzen, dass bei der Durchführung einer komplexen Intervention unzählige mögliche Abweichungen vom Behandlungsmanual denkbar sind. Die Beschreibung dieser Abweichungen in Form von Verhaltensankern für die Abstufung ist daher nur schwierig zu realisieren. Aus diesem Grunde ist es ratsam, bei der Evaluation von therapeutischer Adhärenz und Kompetenz auf erfahrene Beurteiler zurückzugreifen (vgl. Weck, Hilling, Schermelleh-Engel, Rudari & Stangier, 2011) und ausführliche Ratertrainings durchzuführen.

Erfassung der Kompetenz und die Konstruktion der KS-EXVO

Die drei Items der KS-EXVO wurden konstruiert, um die therapeutische Kompetenz bei der Implementierung ex-

positionsvorbereitender Intervention genauer zu erfassen. Die Cognitive Therapy Scale (CTS), die der Konstruktion als Grundlage diente, ist ein Instrument, das bei verschiedenen psychischen Störungen und deren kognitiv-verhaltenstherapeutischen Behandlung eingesetzt wird. Die Kompetenz bei der Durchführung von Techniken wird jedoch bei der CTS nur durch ein Item abgedeckt („Angemessene Durchführung von Techniken“). Dieses Item wurde von den hier befragten Experten jedoch nur als mittelmäßig relevant ($M = 1.75$; $SD = 1.5$) und wenig geeignet ($M = 0.75$; $SD = 0.50$) für die Abbildung von Unterschieden in der Durchführung der Expositionsvorbereitung eingestuft. Ist es für die zu beantwortende Fragestellung einer klinischen Studie von besonderem Interesse, die kompetente Durchführung einer spezifischen Intervention zu erfassen, ist die Konstruktion weiterer spezifischer Items unerlässlich. Allerdings muss sich die inkrementelle Validität der KS-EXVO im Vergleich zu einer allgemeinen Kompetenzskala wie der CTS im Hinblick auf ein Kriterium wie z. B. Therapieerfolg in nachfolgenden Studien empirisch noch nachgewiesen werden. Ähnliche Modifikationen und Erweiterungen der CTS wurden in der Vergangenheit bereits in Bezug auf die Behandlung der Sozialen Phobie (CTCS-SP; Consbruch, Clark & Stangier, 2012) und psychotischer Störungen (CTS-Psy; Haddock et al., 2001) vorgestellt.

Die inhaltliche Validität der ergänzend konstruierten Items (KS-EXVO) kann durch das zugrundeliegende Manual sowie aufgrund der durchgeführten Expertenbefragung angenommen werden. Die hohen bis sehr hohen Beurteilerübereinstimmungen sowie die gute interne Konsistenz belegen die Reliabilität der KS-EXVO.

Erwartungsgemäß weist die KS-EXVO eine hohe Korrelation mit der Cognitive Therapy Scale (CTS) auf, was darauf schließen lässt, dass im Sinne der konvergenten Validität das gemeinsame zugrundeliegende Konstrukt der therapeutischen Kompetenz erfasst wird. Die Konstruktvalidität wird ergänzend durch die niedrige Korrelation (diskriminante Validität) der KS-EXVO mit dem interpersonellen Verhalten der Patienten (AFPIB) belegt.

Einschränkungen

Als Einschränkung kann die sehr homogene Therapeutenstichprobe gesehen werden. Aufgrund des Vorgehens bei der Therapeutenauswahl (nur Therapeuten, die ein Mindestmaß an Adhärenz zeigten) ist davon auszugehen, dass alle an der Studie beteiligten Therapeuten ein hohes Maß an Adhärenz hatten, was zu einer Einschränkung der Varianz führte.

In diesem Zusammenhang kann auch die hohe Standardisierung der Studie als Einschränkung gesehen werden. So stellt sich die Frage, ob die Skalen zur Erfassung der Adhärenz und Kompetenz auch in einem weniger

standardisierten Behandlungssetting eine vergleichbare Güte aufweist.

Eine weitere Einschränkung betrifft den Umstand, dass jeder Therapeut im Durchschnitt 2.5 Patienten behandelte. Betrachten die Rater nun mehrere Behandlungen eines Therapeuten kann nicht ausgeschlossen werden, dass die Beurteilungen der einzelnen Videos durch Erinnerungseffekte beeinflusst sind. Um solche Effekte zu minimieren, wurde die Reihenfolge der zu beurteilenden Videos den Ratern vorgegeben und so gewählt, dass die einzelnen zu beurteilenden Behandlungen desselben Therapeuten nicht hintereinander, sondern mit maximalen Abstand geratet wurden.

Kritisch anzumerken ist zuletzt, dass die KS-EXVO nur einen Teil der Panikbehandlung fokussiert. Aufgrund der Expertenbefragung ist davon auszugehen, dass die Expositionsvorbereitung ein zentraler, jedoch auch wenig untersuchter Bestandteil ist, für deren Erfassung spezifische Instrumente wie die KS-EXVO als Ergänzung zu einer allgemeinen Kompetenzskala wie der CTS benötigt werden.

Ausblick

Diese Studie zeigt, dass es bei der AS-EXVO und der KS-EXVO um zwei Verfahren handelt, die expositionsvorbereitende Intervention in der Behandlung der Panikstörung hinsichtlich der therapeutischen Adhärenz und Kompetenz valide und reliabel erfassen. Im Rahmen von klinischen Studien ist ein hohes Maß an Behandlungsintegrität (Adhärenz und Kompetenz) notwendig, um die interne und externe Validität von Studien sicherzustellen (Schlosser, 2002). Dies kann mit Hilfe der Skalen überprüft werden. Zudem können die Skalen für das Training und die Auswahl von Studientherapeuten eingesetzt werden, um geeignete Therapeuten auszuwählen. Aber auch außerhalb der klinischen Forschung erscheint es sinnvoll, die vorgestellte Adhärenz- und Kompetenzskala, beispielsweise in der Aus- und Weiterbildung von Therapeuten, einzusetzen. So können sie beispielsweise im Rahmen von Supervision oder zur Überprüfung eines Therapeutentrainings eingesetzt werden.

Literatur

- Barber, J. P., Triffleman, E. & Marmer, C., (2007). Considerations in treatment integrity: implications and recommendations for PTSD research. *Journal of Traumatic Stress, 20*, 793–805.
- Batelaan, N. M., de Graaf, R., Penninx, B. W. J. H., van Balkom, A. J. L. M., Vollebergh, W. A. M. & Beekman, A. T. F. (2010). The 2-year prognosis of panic episodes in the general population. *Psychological Medicine, 40*, 147–157.
- Bühner, M. (2006). *Einführung in die Test- und Fragebogenkonstruktion*. München: Pearson.
- Busner, J. & Targum, S. D. (2007). The Clinical Global Impressions Scale: applying a research tool in clinical practice. *Psychiatry, 4*, 28–37.
- Consrubch, K. v., Clark, D. M. & Stangier, U. (2012). Assessing therapeutic competence in cognitive therapy for social phobia: psychometric properties of the Cognitive Therapy Competence Scale for Social Phobia (CTCS-SP). *Behavioural and Cognitive Psychotherapy, 40*, 149–161.
- Haddock, G., Devane, S., Bradshaw, T., McGovern, J., Tarrier, N., Kinderman, P. et al. (2001). An investigation into the psychometric properties of the Cognitive Therapy Scale for Psychosis (CTS-Psy). *Behavioural and Cognitive Psychotherapy, 29*, 221–233.
- Heinrichs, N., Alpers, G. W. & Gerlach, A. L. (2009). *Evidenzbasierte Leitlinie zur Psychotherapie der Panikstörung und Agoraphobie*. Göttingen: Hogrefe.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*, 179–185.
- Huppert, J. D., Barlow, D. H., Gorman, J. M. Shear, M. K. & Woods, S. W. (2006). The interaction of motivation and therapist adherence predicts outcome in cognitive behavioral therapy for panic disorder: preliminary findings. *Cognitive and Behavioral Practice, 13*, 198–204.
- Huppert, J. D., Bufka, L. F., Barlow, D. H., Gorman, J. M., Shear, M. K. & Woods, S. W. (2001). Therapists, therapist variables, and cognitive-behavioral therapy outcome in a multicenter trial for panic disorder. *Journal of Consulting and Clinical Psychology, 69*, 747–755.
- Jacobi, F., Wittchen, H. U., Höltling, C., Höfler, M., Pfister, H., Müller, N. et al. (2004). Prevalence, comorbidity and correlates of mental disorders in the general population: results from the German Health Interview and Examination Survey (GHS). *Psychological Medicine, 34*, 597–611.
- Lang, T., Helbig-Lang, S., Gloster, A. T., Richter, J., Hamm, A. O. Fehm, L. et al. (2013). *Fear activation and reduction of fear during exposure in panic disorder and agoraphobia – An investigation of the Emotional Processing Theory*. Manuscript submitted for publication.
- Lang, T., Helbig-Lang, S. & Petermann, P. (2009). Was wirkt in der Kognitiven Verhaltenstherapie der Panikstörung und Agoraphobie – Ein systematisches Review. *Zeitschrift für Psychiatrie Psychologie und Psychotherapie, 57*, 161–175.
- Lang, T., Helbig, S., Westphal, D., Gloster, A. T. & Wittchen, H.-U. (2006). *Studienmanual – Panikstörung mit Agoraphobie*. Unveröffentlichtes Manuskript, Universität Bremen.
- Lang, T., Helbig-Lang, S., Westphal, D., Gloster, A. T. & Wittchen, H.-U. (2012). *Expositionsbasierte Therapie der Panikstörung und Agoraphobie. Ein Behandlungsmanual*. Göttingen: Hogrefe.
- Lang, T., Gerlach, A. L., Kirchner, T., Alpers, G. W. & Hamm, A., University Medicine Greifswald. Mechanisms of Panic Disorders Treatment. Bethesda (MD): National Library of Medicine (US). 2000-. Retrieved 2014 Jul 28, from <http://clinicaltrials.gov/show/NCT01323556> NLM Identifier: NCT01323556
- Muse, K., & McManus, F. (2013). A systematic review of methods for assessing competence in cognitive-behavioural therapy. *Clinical Psychology Review, 33*, 484–499.
- Neudeck, P. (2005). „Bedrohliche Hasen“ – Konfrontationstherapie bei Panikstörungen. In P. Neudeck & H.-U. Wittchen (Hrsg.). *Konfrontationstherapie bei psychischen Störungen*. Göttingen: Hogrefe.

- O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, & Computers*, 32, 396–402.
- Portney, L. G. & Watkins, M. P. (2008). *Foundations of clinical research: applications to practice* (3rd ed.) London: Pearson Education.
- Richtberg, S., Jakob, M., Höfling, V. & Weck, F. (2014). Assessment of patient interpersonal behavior: Development and validation of a rating scale. *Psychotherapy Research*. doi: 10.1080/10503307.2014.947391
- Ruhmland, M. & Margraf, J. (2001). Effektivität psychologischer Therapien von Panik und Agoraphobie. Meta-Analyse auf Störungsebene. *Verhaltenstherapie*, 11, 41–53.
- Sánchez-Meca, J., Alcázar, A. I., Marín-Matínez, F. & Gómez-Conesa, A. (2010). Psychological treatment of panic disorder with or without agoraphobia: a meta-analysis. *Clinical Psychology Review*, 30, 37–50.
- Saß, H., Wittchen, H. & Zaudig, M. (1996): *Diagnostisches und Statistisches Manual Psychischer Störungen DSM-IV*. Hogrefe, Göttingen.
- Schlosser, R. W. (2002). On the importance of being earnest about treatment integrity. *Augmentative and Alternative Communication*, 18, 36–44.
- Shrout, P. E. & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, 86, 420–428.
- Waltz, J. A., 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.
- Weck, F. (2013). *Psychotherapeutische Kompetenzen: Theorien, Erfassung, Förderung*. Heidelberg: Springer.
- Weck, F., Bohn, C., Ginzburg, D. M. & Stangier, U. (2011). Behandlungsintegrität: Implementierung, Messung, Evaluation und Zusammenhänge zum Therapieerfolg. *Verhaltenstherapie*, 21, 99–107.
- Weck, F., Hautzinger, M., Heidenreich, T. & Stangier, U. (2010). Erfassung psychotherapeutischer Kompetenzen: Validierung einer deutschsprachigen Version der Cognitive Therapy Scale. *Zeitschrift für Klinische Psychologie und Psychotherapie*, 29, 244–250.
- Weck, F., Hilling, C., Schermelleh-Engel, K., Rudari, V. & Stangier, U. (2011). Reliability of adherence and competence assessment in cognitive behavioral therapy: influence of clinical experience. *Journal of Nervous and Mental Disease*, 199, 276–279.
- Wittchen, H.-U. (1991). Der Langzeitverlauf unbehandelter Angst-Störungen: Wie häufig sind Spontanremissionen? *Verhaltenstherapie*, 1, 273–282.
- Young, J. & Beck, A. T. (1980). *Cognitive Therapy Scale rating manual*. Unpublished manuscript, Center of Cognitive Therapy, Philadelphia, PA.

Dipl.-Psych. Florian Grikscheit

Abteilung Klinische Psychologie und Psychotherapie
Goethe-Universität Frankfurt
Varrentrappstr. 40–42
60486 Frankfurt
E-Mail: grikscheit@psych.uni-frankfurt.de

Evaluation of Treatment Integrity: The Treatment Specificity Index

Florian Grikscheit,¹ Florian Weck,² Martin Hautzinger,³ Thomas Heidenreich,⁴
Maria Weigel,¹ Visar Rudari,¹ Christine Schimming,¹ and Ulrich Stangier¹

¹University of Frankfurt

²University of Mainz

³University of Tuebingen

⁴University of Applied Sciences Esslingen

Objective: To differentiate between treatments in a randomized controlled trial (RCT), therapies need to display not only high purity but also high specificity. The aim of this study was to demonstrate the benefits of calculating a treatment specificity index for the evaluation of treatment differentiation.

Method: Based on an RCT of relapse prevention in depression, comparing a cognitive with a psychoeducational treatment, the specificity and the purity index were calculated. **Results:** As indicated by the specificity index, both conditions differed in their levels of implemented specific and common interventions. A significant relationship was found between symptom change before a therapy session and treatment specificity. **Conclusions:** The specificity index is an appropriate method for enhancing the internal validity of RCTs in evaluating treatment integrity. © 2015 Wiley Periodicals, Inc. *J. Clin. Psychol.* 00:1–13, 2015.

Keywords: treatment specificity index; treatment integrity; treatment differentiation; purity index; adherence

The “dodo bird verdict” (Luborsky, Singer, & Luborsky, 1975) opened the debate of whether common factors among all therapies or specific elements used by a given school of therapy produce symptom change (Reisner, 2005). One prerequisite for interpreting a randomized controlled trial (RCT) and for drawing meaningful conclusions on the effect of interventions is that the compared treatment conditions are distinguishable from each other. Only then it can be assumed that the hypothesis of an RCT will be tested adequately (Bhar & Beck, 2009). Accordingly, treatment differentiation, which is beside adherence and therapeutic competence, a component of treatment integrity, defines whether and to what extent treatment conditions differ from each other (Southam-Gerow & McLeod, 2013).

Treatment Purity

To ensure a high differentiability of treatments, observed in an RCT, it is necessary that they were implemented according to their protocol (treatment adherence). However, for the evaluation of treatment differentiation, it is important to assess not only whether treatment condition “A” mostly comprises interventions described in protocol “A” (prescribed interventions), but also whether interventions of treatment condition “B” were applied by the therapists in treatment condition “A” (proscribed interventions) and vice versa.

For this purpose, the *purity index* was calculated in several studies (e.g., Luborsky, McLellan, Woody, O’Brian, & Auerbach, 1985; Stangier et al., 2013; Zobel, Karim, Kech, Berger, & Schramm, 2008). This index is the ratio of prescribed interventions for a treatment condition to the sum of all interventions (prescribed and proscribed interventions). Following this definition, treatment purity can be considered as a component of adherence. A treatment can be interpreted as “pure” when the purity index is tending toward 1. If each condition of an RCT shows a high degree of purity, then this might suggest a high level of treatment differentiation. However, this

Please address correspondence to: Florian Grikscheit, Goethe University Frankfurt, Institute for Psychology, Department of Clinical Psychology and Psychotherapy, Varrentrappstr. 40–42, D-60486 Frankfurt, Germany. E-mail: grikscheit@psych.uni-frankfurt.de

method has one limitation: The classification of prescribed interventions is insufficiently precise, because interventions that are specific to one treatment condition (e.g., exposure in sensu in an exposition approach) and common interventions (e.g., setting treatment goals, which could be part of more than one treatment approach) were mixed together.

Thus, it is possible that two conditions compared in an RCT do not mostly comprise desired specific interventions, rather mainly common interventions. In such cases, the purity index will reveal the same level as a treatment mainly comprising specific interventions, and as a result treatment differentiation could be overestimated.

Treatment Specificity

In addition to the purity, a high specificity is also necessary to achieve a high level of treatment differentiation. Treatment specificity, which could also be regarded as part of adherence, is defined as the level of implemented interventions that are unique to a particular treatment approach (e.g., Frank, Kupfer, Wagner, McEachran, & Cornes, 1991). To make reliable statements about treatment specificity, it is necessary to perform a more detailed analysis of the composition of treatments. Accordingly, Waltz, Addis, Koerner, & Jacobson (1993) suggested distinguishing between four types of interventions. First, there are *unique and essential* interventions, which include specified interventions (e.g., forming a contingency contract in behavioral therapy) for the appropriate treatment, which are not used in another approach being tested.

The second type refers to *essential but not unique* interventions, which could also be used in another condition (e.g., establish a therapeutic alliance). The third type refers to interventions that are *acceptable but not necessary* for a treatment (e.g., chatting with the patient). Fourth, there are *proscribed* interventions, which are based on the belief that a unique intervention of treatment “A” is an unwanted intervention in treatment “B” (e.g., focus on defense mechanisms in behavioral therapy). For an ideal treatment differentiation, the treatments mostly comprise *unique/essential*, some *essential/not unique*, a few *acceptable/not necessary* and no *proscribed* interventions.

Furthermore, we used this categorization of interventions to calculate an index for evaluating specificity. This *treatment specificity index* can be defined by the ratio of unique/essential interventions to all interventions (unique/essential, essential/not unique, and proscribed). From a specificity index equal to or exceeding 0.50, unique/essential interventions were used equal or more often than other interventions, and consequently treatment specificity could be assumed.

Thus, the purity (use of more prescribed than of proscribed interventions) and the specificity of a treatment (use of more unique than of essential/not unique interventions) are prerequisites of treatment differentiation. Indeed, it can be assumed that a considerable part of RCTs did not adequately address the evaluation of treatment differentiation (Bhar & Beck, 2009). For comparative studies, like the RCT of Stangier et al. (2013), which calculated the purity index, but without drawing conclusions on the treatment differentiation, the calculation of the treatment specificity index might be useful to gain further information on the differentiability of treatments and on the validity of the study.

Relationship Between Specificity and Symptom Change

The specificity index may also help to enhance our understanding of the relationship between adherence and outcome. Several studies (e.g., Webb, DeRubeis, & Barber, 2010; Weck, Bohn, Ginzburg, & Stangier, 2011) found only a weak or no association between adherence and posttherapy outcome. An explanation could be that adherence as one process measure is too global, so that it seems reasonable to examine the relationship between outcome and treatment specificity, as well as purity, as two separate elements of adherence. While Luborsky et al. (1985) found a positive correlation between the purity index and outcome, no or only insufficient data were available concerning the relationship between treatment specificity and outcome.

Another reason may be that the association between adherence and outcome is too distal. Therefore, Boswell et al. (2013) has recommended analyzing the influence of adherence not on posttherapy outcome, but subsequent symptom change (i.e., from one session to the next).

According to this, the empirical findings are also inconsistent. Whereas Strunk, Cooper, Ryan, DeRubeis, & Hollon (2012) demonstrated that adherence to different concrete therapy techniques predicted subsequent symptom change, Webb et al. (2012) revealed a lack of such a relationship. Both studies were based on a sample including patients with a major depressive disorder treated with cognitive therapy. To gain further insights, it seemed reasonable to examine the influence of purity and treatment specificity on subsequent symptom change.

Uncertainty also exists about the factors that affect adherence and thus treatment specificity as well as purity. It can be assumed that a patient's symptom improvement before a therapy session may encourage therapist's adherence. Accordingly, Boswell et al. (2013) has recommended evaluating the prediction of pre-session symptom change. Following the study of Loeb et al. (2005), which found that prior symptom change was related to subsequent adherence in the middle part of treatment, we analyzed the relationship between symptom change and both purity and treatment specificity in the next session.

Purpose of This Study

The first aim of the current study was to demonstrate, by reanalyzing the adherence ratings of an RCT by Stangier et al. (2013), that the treatment specificity index can provide further information on treatment differentiation. We proposed that the specificity index might be useful for drawing conclusions on the validity of treatment comparison trials. A second aim of the study was to validate whether the extent of treatment specificity and purity has an influence on subsequent symptom change. Third, we wanted to examine the patient's symptom change immediately before a session as a factor that affects purity and treatment specificity.

Method

Design

The data base for the current study included adherence ratings (Weck et al., 2012; Weck et al., 2013) of an RCT of relapse prevention in depression (Stangier et al., 2013). The main objective of this RCT was to compare the effects of a maintenance cognitive-behavioral therapy (M-CBT) condition (Risch, Stangier, Heidenreich, & Hautzinger, 2012) with manualized psychoeducation (M-PE; Hautzinger, Schlösser, & Barocka, 2006), as an active control condition, concerning the reduction of relapse risk for remitted outpatients with recurrent depression. Patients who had been diagnosed with at least three previous episodes and currently had a stable level of remission were randomized to both treatment conditions.

In addition to M-CBT (57.5%) or M-PE (62.5%), patients continued their pharmacological treatment (treatment as usual) if they had already received an antidepressant medication before the initiation of the study. In this RCT by Stangier et al. (2013), a total of 180 patients were treated in both conditions in 16 sessions. All therapy sessions were recorded on videotape. According to the recommendations of Waltz et al. (1993), several strategies were used to ensure the implementation of treatment integrity: A detailed treatment manual exists for both conditions. The therapists of M-CBT (45-hour workshop plus a fresh-up workshop) and M-PE (2-day workshop) received intensive training and accompanying video-based supervision.

As a main result, it could be shown that both treatment conditions were not significantly different in their effectiveness of relapse prevention in patients with fewer earlier depressive episodes. With regard to our research questions, 40 treatments for each condition were randomly selected. For each of the treatments, one videotaped session was also randomly selected and evaluated with respect to treatment adherence.

M-PE

M-PE (Hautzinger et al., 2006) conveys to the patient an understanding of the symptoms, causes, and course of depression, as well as information on medication, side effects, and risk factors for triggering relapse. The treatment's basis is a psychobiological model of depression.

Table 1
Patient's and Therapist's Characteristics for M-PE and M-CBT Treatment Condition

	M-PE (<i>n</i> = 40)	M-CBT (<i>n</i> = 40)
Patient's characteristics		
Female (%)	28 (70)	25 (62.5)
Age (<i>SD</i> , range)	48 (12.7, 25-70)	51.1 (11.7, 22 - 72)
BDI baseline (<i>M</i> , <i>SD</i>)	12.9 (9.53)	15.2 (11.0)
Any Axis I comorbidity, no (%)	20 (50.0)	14 (35.0)
Any Axis II comorbidity, no (%)	12 (30.0)	8 (20.0)
Current antidepressant use, no (%)	25 (62.5)	23 (57.5)
Therapist's characteristics		
	(<i>n</i> = 22)	(<i>n</i> = 20)
Female (%)	17 (77.3)	15 (75)
Age (<i>SD</i> , range)	32.2 (5.4, 24 - 44)	31.6 (6.5, 25 - 50)
Experience (<i>SD</i> , range)	1.3 (1.9, 0-5)	2.3 (4.6, 0-20)
Clinical psychologists	19 (86.4 %)	19 (95%)
Psychiatrists	3	1

Note. M-PE = manualized psychoeducation; M-CBT = maintenance cognitive-behavioral therapy; *SD* = standard deviation; *M* = mean.

Besides providing psychoeducation, one main aspect of this condition is facilitating treatment motivation and medication compliance. Each session (duration of 20 minutes) is divided into a patient-specific and an educational part. The therapists were instructed to not only focus on education but also listen to the patient, give active advice based on the individual needs of the patient, allow space for expressing concerns and personal experiences, and reinforce the patient's resources. The manual proscribed any use of specific psychotherapeutic techniques, such as the analysis of dysfunctional cognitions, problem-solving, skill training, and analysis of negative life events or role plays.

M-CBT

The aim of M-CBT is to establish cognitive and behavioral skills to prevent a relapse. M-CBT comprises various approaches and combines interventions of cognitive-behavioral therapy (CBT; Bockting et al., 2005), with elements of continuation phase cognitive therapy (CCT; Jarrett et al., 2001), well-being therapy (WBT; Fava, Rafanelli, Grandi, Conti, & Belluardo, 1998), and mindfulness-based cognitive therapy (MBCT; Teasdale et al., 2000). Main interventions of this treatment condition include the derivation of a psychological model, mindfulness meditation exercises, modification of dysfunctional cognitions, behavioral experiments and stress testing, increasing psychological well-being. The therapist can implement the interventions flexibly, according to the patient's individual problems and needs. The duration of the 16 sessions is between 50 and 60 minutes. Therapists were instructed to assign and review homework or to use work sheets in each session. These interventions are furthermore described in a treatment manual (Risich et al., 2012).

Participants

Patients. Patients' characteristics of the 40 analyzed video tapes of M-PE and M-CBT treatment (*n* = 80) conditions are shown in Table 1. Accordingly, no significant differences between the patients of M-CBT and M-PE were found: sex ($p = .23$), age ($p = .36$), BDI (Beck Depression Inventory; Beck, Ward, Mendelson, Moch, & Erbaugh, 1961) baseline ($p = .53$), number of any Axis I ($p = .14$) and Axis II ($p = .10$) comorbidity, and current use of antidepressant ($p = .04$). The present subsample did not differ significantly from the original one with regard to all abovementioned characteristics.

Therapists. The 40 patients in the M-PE condition were treated by 22 therapists and in the M-CBT condition by 20 therapists. Further therapist characteristics are shown in Table 1. Both groups of therapists showed no significant differences in sex ($p = .86$), age ($p = .73$), experience ($p = .44$), or profession ($p = .34$).

Raters. The videotapes were rated by four raters (two for each condition). All raters were clinical psychologists, with clinical experience ranging between 2 and 7 years (M-PE ratings) and 2 and 5 years (M-CBT ratings). Both raters for each condition were familiar with the corresponding manual and received a 10-hour training course.

Measures

M-PE/CBT-Adherence Scale (M-PE/CBT-AS). With the aim of evaluating the extent of interventions applied by the therapists, regardless of treatment condition, the M-PE/CBT-AS, which comprises 24 items, was constructed. This scale represents a combination of the M-PE-AS (Weck, Weigel, Richtberg, & Stangier, 2011), based on the treatment protocol for manualized active psychoeducation (Hautzinger et al., 2006), and the M-CBT-AS (Weck, Hilling, Schermelleh-Engel, Rudari, & Stangier, 2011), based on the manual for cognitive behavioral therapy (Risch et al., 2012).

The M-PE/CBT-AS comprises seven items that are specific to M-PE, 10 items specific to M-CBT, five common items, and two additional items. One item assesses the implementation of proscribed intervention of other therapeutic approaches and one reflecting global adherence. The response format is a three-point rating scale ranging from 0 (*not adherent*) to 2 (*adherent*). All items comprised a short description of the characteristic features of the various interventions and an anchor description, which is to simplify the judgment. Three experts (M.H., T.H., & M.H.) evaluated all the items to ensure content validity. The three experts, who also developed the underlying manuals, have many years of experience in cognitive-behavioral therapy in treating depression, and psychotherapy research.

The inter-rater reliability ranged from good, intraclass correlation coefficient (ICC) (2,2) = 0.81, $p < 0.001$, for the M-PE condition, to very good, ICC (2,2) = 0.95, $p < 0.001$, for the M-CBT condition (Stangier et al., 2013). The items of the M-PE/CBT-AS were assigned to three different scales (unique/essential; essential/not unique and proscribed), conducted by the three experts (see Table 2), in accordance with the recommendation of Waltz et al. (1993). The fourth scale (acceptable but not necessary intervention) was not applied, because this scale does not provide any additional information on neither purity nor treatment specificity. None of the constructed items fits into this scale.

Symptom change. Before each therapy session, the patients' current symptomatology was evaluated by means of a five-item short questionnaire. To assess depressive symptoms, the underlying RCT used a short questionnaire comprising the following items: "Please decide on how much the following discomfort bothered you over the past seven days (including today)"; "Lack of energy or slowing down, agitation or thinking"; "Melancholy"; "Feeling of disinterest"; "Feeling that everything is exhausting"; and "Feeling worthless". All items were rated on a 4-point Likert scale ranging from 0 (*not at all*) to 3 (*intense/I could hardly bear it*).

The prior symptom change was determined by calculating the difference between the prior score (the session before) and the current session score. Subsequent symptom change was calculated by the difference between the current session score and the subsequent symptom score. In both cases, a positive score indicates symptom reduction, with higher scores indicating a larger change. The possible range of symptom change spans the values -15 to +15. Based on the high correlation (M-PE: $r = .72$, $p < 0.001$; M-CBT: $r = .69$, $p < 0.001$) with the BDI (Beck, et al. 1961), criterion validity of the short questionnaire can be assumed. The instrument had high internal consistencies (Cronbach's α) for M-PE ($\alpha = .86$) and for M-CBT ($\alpha = .88$).

Data Analysis

Following Luborsky et al. (1985), we calculated the purity index for both treatment conditions, which is the ratio of the proscribed interventions (unique/essential and essential/not unique)

Table 2

M-PE/CBT-AS Items With Assignments to the Three Scales (Unique/Essential, Essential/Not Unique, Proscribed)

M-PE	Scale	M-CBT
<ul style="list-style-type: none"> ● Adherence to session structure ● Role allocation ● Reference to biopsychological model ● Focus on medication intake ● Working on current problems ● Information gathering ● Psychoeducative approach 	Unique / essential	<ul style="list-style-type: none"> ● Reviewing of homework ● Encouraging self-monitoring ● Elaborating the individualized cognitive model of depression ● Identifying automatic thoughts and core beliefs ● Modifying automatic thoughts and core beliefs with cognitive techniques ● Behavioral experiments and test of resilience ● Setting of homework ● Interventions of acceptance and commitment therapy ● Interventions of wellbeing therapy ● Listed optional interventions
<ul style="list-style-type: none"> ● Agenda ● Time management ● Use of materials ● Listening and responding to questions and problems of understanding ● Relapse prevention 	Essential/not unique	<ul style="list-style-type: none"> ● Agenda ● Time management ● Use of materials ● Listening and responding to questions and problems of understanding ● Relapse prevention
<ul style="list-style-type: none"> ● Reviewing of homework ● Encouraging self-monitoring ● Elaborating the individualized cognitive model of depression ● Identifying automatic thoughts and core beliefs ● Modifying automatic thoughts and core beliefs with cognitive techniques ● Behavioral experiments and test of resilience ● Setting of homework ● Interventions of acceptance and commitment therapy ● Interventions of wellbeing therapy ● Listed optional interventions ● Exclusion of nonadherent techniques 	Proscribed	<ul style="list-style-type: none"> ● Adherence to session structure ● Role allocation ● Reference to biopsychological model ● Focus on medication intake ● Working on current problems ● Information gathering ● Psychoeducative approach ● Exclusion of nonadherent techniques

Note. M-PE = manualized psychoeducation; M-CBT = maintenance cognitive-behavioral treatment condition.

to all interventions (prescribed and proscribed). The treatment specificity index was calculated by the ratio of unique/essential interventions to all interventions. In addition to comparing the means for both treatment conditions, effect sizes (Cohen's *d*) were calculated.

ICCs (model 2, ICC [2,n]; see Shrout & Fleiss, 1979), using a two-way random model with an absolute agreement definition, were calculated to analyze the inter-rater reliabilities of the three treatment scales (unique/essential, essential/not unique, and proscribed), the purity index, the treatment specificity index, and the mean adherence score for each treatment condition. As suggested by Portney and Watkins (2009), ICCs $\geq .75$ indicate high reliabilities, ICCs between .75 and .50 moderate, and ICCs $< .50$ indicate poor reliabilities. We performed all analyses with SPSS (version 20). To take into account multiple testing, correlations between purity and

Table 3
Interrater reliabilities of the Unique/Essential, Essential/Not Unique, and Proscribed Scales, Mean Adherence Score, Purity Index, and Treatment Specificity Index for M-PE and M-CBT Treatment Conditions

	Interrater reliability M-PE (n = 40)		Interrater reliability M-CBT (n = 40)	
	M (SD)	ICC (2,2)	M (SD)	ICC (2,2)
Unique/essential	1.63 (.37)	.84***	0.94 (.43)	.90***
Essential/not unique	1.40 (.36)	.72***	1.37 (.45)	.92***
Proscribed	0.03 (.11)	.85***	0.06 (.16)	.68***
Mean adherence score (items 1–24) ^a	0.86 (.17)	.87***	0.94 (.24)	.91***
Purity index	0.99 (.03)	.86***	0.98 (.05)	.83***
Treatment specificity index	0.53 (.10)	.76***	0.40 (.16)	.88***

Note. M-PE = manualized psychoeducation; M-CBT = maintenance cognitive-behavioral treatment condition; ICC = intraclass correlation coefficient.

^aIn M-CBT except item 10–12.

*p < 0.05. **p < 0.01. ***p < 0.001.

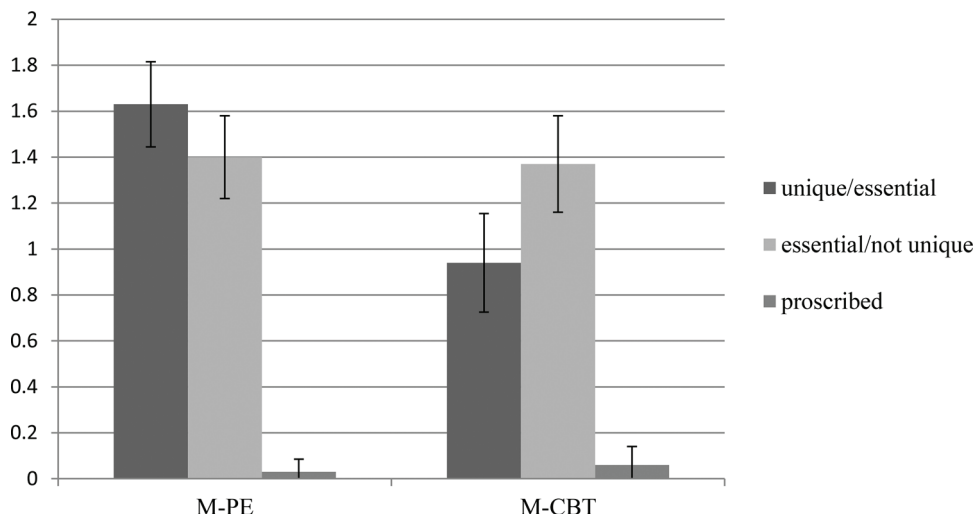


Figure 1. Means of unique/essential, essential/not unique, and proscribed scale for M-PE and M-CBT.

Note. Error bars represent standard error. M-PE = manualized psychoeducation; M-CBT = maintenance cognitive-behavioral treatment condition.

treatment specificity with prior and subsequent symptom change were calculated with Bonferroni adjustment.

Results

Interrater Reliability

ICCs for each of the three treatment scales, for purity and the treatment specificity indices of both conditions, are reported in Table 3.

Purity Index

Figure 1 shows the different degrees of the unique/essential, essential/not unique and proscribed scales for M-PE and M-CBT treatment. In both conditions, the values of proscribed interventions

(unique/essential and essential/not unique interventions) were significantly higher than the proscribed interventions. This is also evident in a high purity index for both conditions. The purity index for M-PE was 0.99 (standard deviation [*SD*] = 0.03; *range*: 0.82–1.00) and 0.98 (*SD* = 0.05; *range*: 0.80–1.00) for M-CBT. Both purity indices did not differ significantly from each other, $t(78) = -1.36$; $p = .178$.

Treatment Specificity

In accordance with our first aim, we calculated the treatment specificity index for both conditions. This index reached a value of 0.53 (*SD* = .10; *range*: .20 - .75) for M-PE treatment. In this condition, the most frequently used interventions were unique/essential (mean [*M*] = 1.63; *SD* = .37) ones. The therapists for M-PE used significantly less, $t(39) = 3.40$; $p < .001$; $d = 0.63$, essential/not unique ($M = 1.40$; *SD* = .36) and least, $t(39) = 22.36$; $p < .001$; $d = -5.15$, proscribed ($M = 0.03$; *SD* = .11) interventions.

The treatment specificity index for M-CBT yielded a significantly smaller, $t(78) = -4.42$; $p < .001$; $d = 0.97$, value of 0.40 (*SD* = .16; *range*: .00–1.00) than in M-PE. In contrast to M-PE, in the M-CBT treatment condition, the most frequent interventions were essential/not unique ($M = 1.37$; *SD* = .45) and there were significantly less, $t(39) = -5.25$; $p < .001$; $d = 0.98$, unique/essential ($M = 0.94$; *SD* = .43) interventions and the least, $t(39) = 12.27$; $p < .001$; $d = -2.71$, proscribed ones ($M = 0.06$; *SD* = .16). With regard to the implementation of essential/not unique interventions, no significant difference, $t(78) = 0.33$; $p = .33$, between M-PE and M-CBT could be found. An inferential statistical comparison between M-PE and M-CBT concerning the scales was possible only for the essential/not unique scale, because this scale comprises the same items for both treatment conditions.

Prediction of Subsequent Symptom Change by Purity and Treatment Specificity Index

For each treatment condition, the correlations between the purity index and treatment specificity index with subsequent symptom change were calculated. In both treatment conditions, a significant relationship with the subsequent symptom change could not be found for the purity index (M-PE: $r = -.28$, $p = \text{n.s.}$; M-CBT: $r = .26$, $p = \text{n.s.}$) or the treatment specificity index (M-PE: $r = -.05$, $p = \text{n.s.}$; M-CBT: $r = -.04$, $p = \text{n.s.}$).

Prediction of Purity and Specificity by Prior Symptom Change

Only in the M-PE condition could a positive significant correlation between the purity index ($r = .52$, $p < .05$) and the treatment specificity index ($r = .45$, $p < .05$) with prior symptom change be shown. A greater symptom improvement was associated with a higher ratio of unique interventions and a higher ratio of proscribed interventions to all applied interventions. In the M-CBT condition, no significant influence of prior symptom change, on neither the purity ($r = .26$, $p = \text{n.s.}$) nor the specificity index ($r = .35$, $p = \text{n.s.}$), could be revealed.

Discussion

The aim of the present study was to examine the benefit of calculating the specificity index for the evaluation of treatment differentiation and to investigate the relationship between treatment specificity and both prior and subsequent symptom change. The results revealed that in addition to the purity index, the treatment specificity index provides further information for evaluating treatment differentiation and that it is an appropriate method for enhancing internal validity of an RCT (first aim). With regard to session outcome, a relationship could not be demonstrated for the purity index or the specificity index (second aim). However, we found that in the M-PE condition, prior symptom deterioration is associated with decreased treatment specificity in the following session (third aim).

Benefit of Treatment Specificity Index for the Evaluation of Treatment Differentiation

The purity index, which was very close to a value of 1 in both conditions, indicated a high degree of prescribed interventions, compared to proscribed ones. Based on this finding, one might expect a high level of treatment differentiation, but the calculation of the specificity indices revealed an important difference between both conditions. Not only did the indices of M-PE and M-CBT differ significantly, but only one of them exceeded a value of 0.50, which indicates that the degree of treatment specificity was acceptable, but not high.

It must be noted, that the interpretation of a specificity index as equal to or above a value of 0.50 should not be misunderstood as an indication of “better“ therapy, but as a characteristic that increases the differentiability from other therapies. In contrast to M-PE, the specificity index of M-CBT fell below that point, which demonstrated that this treatment condition comprises more common interventions than of specific ones. This deviation from the prescribed implementation of the M-CBT treatment condition can be interpreted as a limitation of treatment differentiation and consequently of internal validity. These results revealed that an evaluation of treatment differentiation that is based solely on the purity index entails a risk of overestimation. The additional calculation of the treatment specificity index represents a reasonable and appropriate method for a realistic evaluation of treatment differentiation.

The specificity index also provides some improvements in relation to other existing methods. Previous studies (e.g., Luborsky et al., 1985; Zobel et al., 2008) evaluated treatment specificity by assessing only prescribed interventions for each treatment condition. This leads to a focus on the differences between treatments, while possible similarities were not considered. Because of the distinction of specific and common (essential/not unique) interventions in the calculation of the specificity index, a possible overlap between treatments is taken into account. Waltz et al. (1993) pointed out that otherwise, important information may be lost, because “it might create the illusion of treatment distinctiveness when, in fact, the similarities were more important than the differences” (p. 626).

Furthermore, the different levels of specificity in both treatment conditions could explain the unexpected findings of the main study (Stangier et al., 2013), namely, the M-PE and M-CBT showed no significant differences in their effectiveness. One possible reason for the lower specificity of the M-CBT condition could be the different complexity of both treatment conditions; M-CBT treatment comprises considerably more interventions and their implementation is more challenging than M-PE. The finding that a more complex intervention reduces treatment integrity has been reported before (e.g., Leichsenring, 2011).

Subsequent Symptom Change

The second aim was to examine the influence of treatment specificity and purity on session outcome with regard to the recommendations made in several studies (e.g., Boswell et al., 2013; Strunk et al., 2012). Neither in M-PE nor in the M-CBT condition a significant correlation was found between the purity and the treatment specificity index with subsequent symptom change. Thus, these data support the findings of Webb et al. (2012), that adherence had little impact on symptom change.

However, it should be noted that, in contrast to the abovementioned studies, the sample for the current study consisted of patients with a recurrent depressive disorder who met the remission criteria. Potential therapist adherence was less important for the symptom deterioration of this subsample of patients with remitted depression. Despite the missing predictive value of treatment specificity with respect to session outcome, the advantage of high specificity should be considered in relation to an increased differentiability of treatment conditions compared in an RCT.

Prior Symptom Change

As a third aim, we analyzed the influence of prior symptom change on specificity and purity in both treatment conditions. Particularly in M-PE, significant correlations between prior symptom

change and the treatment specificity index, as well as purity, could be demonstrated. Although the bidirectionality of this relationship should be considered, due to the chronological order of the process, these data indicate that in the M-PE condition, the more the patient's depressive symptoms deteriorated before a session, the less prescribed and specific were the interventions the therapists used. This confirms the results of Loeb et al. (2005) and provides evidence that this relationship is significant not only for the middle part of a treatment but also across all phases.

Nevertheless, the question arose as to why there was a correlation between prior symptom change and both indices only in M-PE and not in M-CBT. An explanation is that in the psychoeducative approach, the worse the patient's symptoms become, the more a therapist used proscribed interventions (indicated by the purity index). Furthermore, therapists might employ less specific interventions (indicated by the treatment specificity index), which, in this condition, means the application of psychotherapeutic methods. This was not the case in the M-CBT condition. Many psychotherapeutic interventions are allowed (except, for example, a focus on medication intake or reference to the biopsychological model) and available for a therapist to respond to symptom deterioration or to a crisis. In summary, it can be assumed that in a considerably more complex treatment approach like M-CBT, the selection and the implementation of interventions depend less on the symptom change of a patient before the session and probably more on other factors like therapeutic competence.

Limitations

Some limitations of this study should be noted. First, because of the fact that the current article is a reanalysis of an RCT, only existing data could be used. To evaluate symptom change for each randomly selected video session, only data from a short questionnaire were available and not from a well-established instrument (e.g., BDI). However, the short questionnaire demonstrated good reliability and strong correlations with the BDI.

Second, patients of both M-PE and M-CBT conditions additionally received a pharmacological treatment (as usual). A potential influence of this medication on symptom change was not considered in the data analysis. However, it can be assumed that the potential influence of medication would be similar in both groups. In a preliminary analysis, no significant differences between patients of both M-PE and M-CBT conditions were found, according to their use of antidepressant medication.

Third, the evaluation of purity and treatment specificity was based on only one randomized selected session of each analyzed therapy. However, empirical research has demonstrated that 5 to 10 treatment sessions are necessary to achieve satisfactory dependability (Dennhag, Gibbons, Barber, Gallop, & Crits-Christoph, 2012). Therefore, it can be assumed that a generalization of purity and treatment specificity to the entire treatment on the basis of a single session may not be representative. For a more comprehensive estimation of this process measures, further research should take more treatment sessions into account.

Fourth, one aim of the study was to examine the relationship between (prior/subsequent) symptom change and both indices. It should be taken into account that this represents a limited view of psychotherapy and the factors that affect effectivity. Supported by recent literature (e.g., Webb et al., 2010, Crits-Christoph, Gibbons, & Mukherjee, 2013), a complex interaction between patient and therapeutic factors as well as specific and common treatment ingredients can be assumed. Thus, the possibility cannot be excluded that an unconsidered third variable accounts for the observed effects.

Fifth, critically, it should be noted that in the context of RCTs, even experts are not always in agreement on how the treatments should be conducted. In such cases treatment integrity must be questioned (Barber, 2009). To establish a more consistent understanding, it might be important to consider which interventions of a certain treatment are specific and which are common. The current article attempts to increase awareness of this important issue.

Sixth, it is important to note that the current study is based on an RCT that focuses on relapse prevention. The underlying sample comprised patients who were in remission at the time

of inclusion to the study. For a higher level of generalization of the treatment specificity index usability, future research with different psychotherapeutic approaches (for example treatment of current depression) and disorders is necessary.

Conclusion

The study revealed that calculating the treatment specificity index is a method that can be used in a given RCT to evaluate whether one treatment is different from another. To increase our understanding of treatment effects in clinical research, the specificity index may help in parsing out common from specific therapeutic components. Future RCTs should ensure that the specific (unique) interventions in each treatment condition are implemented as intended, with the aims of achieving a high treatment differentiation and preventing restrictions in internal validity.

Furthermore, for the execution of further clinical trials, we recommend evaluating the compared treatment conditions with an adherence scale, which considers unique/essential, essential/not unique, and proscribed interventions for each treatment condition. Accordingly, we also suggest calculating the treatment specificity index in addition to the purity index. The current study indicates that the evaluation of treatment differentiation using these indices is simple and economical and yields further important information.

References

- Barber, J. P. (2009). Toward a working through of some core conflicts in psychotherapy research. *Psychotherapy Research*, 19, 1–12. doi:10.1080/10503300802609680
- Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, 4, 561–571. doi:10.1001/archpsyc.1961.01710120031004
- Bhar, S. S., & Beck, A. T. (2009). Treatment integrity of studies that compare short-term psychodynamic psychotherapy with cognitive-behavior therapy. *Clinical Psychology: Science and Practice*, 16, 370–378. doi:10.1111/j.1468-2850.2009.01176.x
- Bockting, C. L. H., Schene, A. H., Spinhoven, P., Koeter, M. W. J., Wouters, L. F., Huyser, J., & Kamphuis, J. H. (2005). Preventing relapse/recurrence in recurrent depression with cognitive therapy: a randomized controlled trial. *Journal of Consulting and Clinical Psychology*, 73, 647–657. doi:10.1037/0022-006X.73.4.647
- Boswell, J. F., Gallagher, M. W., Sauer-Zavala, S. E., Bullis, J., Gorman, J. M., Shear, M. K., Woods, S., & Barlow, D. H. (2013). Patient characteristics and variability in adherence and competence in cognitive-behavioral therapy for panic disorder. *Journal of Consulting and Clinical Psychology*, 81, 443–454. doi:10.1037/a0031437
- Crits-Christoph, P., Gibbons, M. B. C., & Mukherjee, D. (2013). Psychotherapy process-outcome research. Bergin and Garfield's handbook of psychotherapy and behavior change. (6th ed.) (S.298-340.) New York: Wiley & Sons.
- Dennhag, I., Gibbons, M. B. C., Barber, J. P., Gallop, R., & Crits-Christoph, P. (2012). How many treatment sessions and patients are needed to create a stable score of adherence and competence in treatment of cocaine dependence? *Psychotherapy Research*, 22, 475–488. doi:10.1080/10503307.2012.674790
- Fava, G. A., Rafanelli, C., Grandi, S., Conti, S., & Belluardo, P. (1998). Prevention of recurrent depression with cognitive behavioral therapy: preliminary findings. *Archives of General Psychiatry*, 55, 816–820. doi:10.1001/archpsyc.55.9.816
- Frank, E., Kupfer, D. J., Wagner, E. F., McEachran, A. B., & Cornes, C. (1991). Efficacy of interpersonal psychotherapy as a maintenance treatment of recurrent depression: contributing factors. *Archives of General Psychiatry*, 48, 1053–1059. doi:10.1001/archpsyc.1991.01810360017002
- Hautzinger, S., Schlösser, R., & Barocka, A. (2006). Manualisierte Psychoedukation (MAPE): Anleitung für Therapeuten und Materialien für Patienten [Manualized psychoeducation (MAPE): introduction for the therapist and material for the patient]. Unpublished manuscript, University of Tübingen and University of Jena, Germany.
- Jarrett, R. B., Kraft, D., Doyle, J., Foster, B. M., Eaves, G. G., & Silver, P. C. (2001). Preventing recurrent depression using cognitive therapy with and without a continuation phase: a randomized clinical trial. *Archives of General Psychiatry*, 58, 381–388. doi:10.1001/archpsyc.58.4.381

- Leichsenring, F., Salzer, S., Hilsenroth, M. J., Leibing, E., Leweke, F., & Rabung, S. (2011). Treatment Integrity: An unresolved issue in psychotherapy research. *Current Psychiatry Reviews*, 7, 313–321. doi:10.2174/157340011797928259
- Loeb, K. L., Wilson, G. T., Labouvie, E., Pratt, E. M., Hayaki, J., Walsh, B. T. . . . Fairburn, C. G. (2005). Therapeutic alliance and treatment adherence in two interventions for bulimia nervosa: a study of process and outcome. *Journal of Consulting and Clinical Psychology*, 73, 1097–1107. doi:10.1037/0022-006X.73.6.1097
- 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. doi:10.1001/archpsyc.1985.01790290084010
- Luborsky, L., Singer, B., & Luborsky, L. (1975). Comparative studies of psychotherapies: Is it true that "everyone has von and all must have prizes"? *Archives of General Psychiatry*, 32, 995–1008. doi:10.1001/archpsyc.1975.01760260059004
- Portney, L. G., & Watkins, M. P. (2009). *Foundations of clinical research: applications to practice* (3rd ed.). London: Pearson.
- Reisner, A. D. (2005). The common factors, empirically validated treatments, and recovery models of therapeutic change. *The Psychological Record*, 55, 377–399.
- Risch, A. K., Stangier, U., Heidenreich, T., & Hautzinger, M. (2012). Kognitive Erhaltungstherapie bei rezidivierender Depression: Rückfälle verhindern, psychische Gesundheit erhalten. [Cognitive maintenance therapy for recurrent depression: preventing relapse, promoting mental health]. Berlin, Heidelberg, New York: Springer.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlation: uses in assessing rater reliability. *Psychological Bulletin*, 86, 420–428. doi:10.1037/0033-2909.86.2.420
- Southam-Gerow, M. A., & McLeod, B. D. (2013). Advances in applying treatment integrity research for dissemination and implementation science: introduction to special issue. *Clinical Psychology: Science and Practice*, 20, 1–13. doi:10.1111/cpsp.12019
- Stangier, U., Hilling, C., Heidenreich, T., Risch, A. K., Barocka, A., Schlösser, R., . . . Hautzinger, M. (2013). Maintenance cognitive-behavioral therapy and manualized psychoeducation in the treatment of recurrent depression: a multicenter prospective randomized controlled trial. *The American Journal of Psychiatry*, 170, 624–632. doi:10.1176/appi.ajp.2013.12060734
- Strunk, D. R., Cooper, A. A., Ryan, E. T., DeRubeis, R. J., & Hollon, S. D. (2012) The process of change in cognitive therapy for depression when combined with antidepressant medication: Predictors of early intercession symptom gains. *Journal of Consulting and Clinical Psychology*, 80, 730–738. doi:10.1016/j.brat.2010.03.011
- Teasdale, J. D., Segal, Z. V., Williams, J. M., Ridgeway, V. A., Soulsby, J. M., & Lau, M. A. (2000). Prevention of relapse/recurrence in major depression by mindfulness-based cognitive therapy. *Journal of Consulting and Clinical Psychology*, 68, 615–623. doi:10.1037/0022-006X.68.4.615
- 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. doi:10.1037/0022-006X.61.4.620
- Webb, C. A., DeRubeis, R. J., & Barber, J. P. (2010). Therapist adherence/competence and treatment outcome: A meta-analytic review. *Journal of Consulting and Clinical Psychology*, 78, 200–211. doi:10.1037/a0018912
- Webb, C. A., DeRubeis, R. J., Dimidjian, S., Hollon, S. D., Amsterdam, J. D., & Shelton, R. C. (2012). Predictors of patient cognitive therapy skills and symptom change in two randomized clinical trials: the role of therapist adherence and the therapeutic alliance. *Journal of Consulting and Clinical Psychology*, 80, 373–381. doi:10.1037/a0027663
- Weck, F., Hilling, C., Schermelleh-Engel, K., Rudari, V., & Stangier, U. (2011). Reliability of adherence and competence assessment in cognitive behavioral therapy. *The Journal of Nervous and Mental Disease*, 199, 276–279. doi:10.1097/NMD.0b013e3182124617
- Weck, F., Rudari, V., Hilling, C., Hautzinger, M., Heidenreich, T., & Stangier, U. (2013). Relapses in recurrent depression one year after maintenance cognitive-behavioral therapy: the role of therapist adherence, competence, and the therapeutic alliance. *Psychiatry Research*, 210, 140–145. doi:10.1016/j.psychres.2013.05.036
- Weck, F., Weigel, M., Hautzinger, M., Barocka, A., Schlösser, R. G. M., & Stangier, U. (2012). Relapses in recurrent depression 1 year after psychoeducational treatment: the role of therapist adherence and competence, and the therapeutic alliance. *Psychiatry Research*, 195, 51–55. doi:10.1016/j.psychres.2011.07.025

- Weck, F., Weigel, M., Richtberg, S., & Stangier, U. (2011). Reliability of adherence and competence assessment in psychoeducational treatment: influence of clinical experience. *Journal of Nervous and Mental Disease*, 199, 983–986. doi:10.1097/NMD.0b013e3182392da1
- Zobel, I., Karim, A., Kech, S., Berger, M., & Schramm, E. (2008). Wie adhärent wird Clinical Management in randomisierten, kontrollierten Studien durchgeführt? [The Adherence of Clinical Management in a Randomized Controlled Trial]. *Psychotherapie, Psychosomatik, medizinische Psychologie*, 58, 395–402. doi:10.1055/s-2008-1067361



Treatment failure in cognitive-behavioural therapy: Therapeutic alliance as a precondition for an adherent and competent implementation of techniques

Florian Weck*, Florian Grikscheit, Marion Jakob, Volkmar Höfling and Ulrich Stangier

Department of Clinical Psychology and Psychotherapy, Goethe-University, Frankfurt, Germany

Objectives. Treatment failure is a common phenomenon, but little is known about the reasons. Therapeutic alliance, therapist adherence, and therapist competence are considered important aspects of treatment success and formed the focus of the current investigation.

Design. Three randomized controlled trials for the treatment of depression, social phobia, and hypochondriasis were the basis of the current study.

Methods. The role of therapeutic alliance, as well as therapist adherence and competence, were investigated in 61 patients, which were classified either as treatment failure or as treatment success. Process variables were evaluated by independent raters on the basis of videotapes of the first three treatment sessions.

Results. Therapists' adherence and therapeutic alliance differed significantly between successful treatments and those classified as failures, whereas therapists' competence did not. In cross-sectional analysis, we found a moderating effect of adherence with alliance on treatment outcome, indicating that the better the therapeutic alliance, the stronger the effect of adherence on treatment outcome. Moreover, higher therapists' competence was found to affect treatment outcome positively, only mediated by therapeutic alliance. Higher therapists' adherence affected treatment outcome positively, only mediated by the competence–alliance relationship. In additional longitudinal analyses, we found evidence that the therapeutic alliance within one session influences therapists' adherence and competence in the subsequent session, but not the other way around.

Conclusions. Therapeutic alliance proved to be an important variable for the prediction of treatment failure. Furthermore, in our longitudinal analyses, we found evidence that the therapeutic alliance is a precondition for the adherent and competent implementation of therapeutic techniques, which questions the results of our cross-sectional analysis and of previous research.

*Correspondence should be addressed to Florian Weck, Department of Clinical Psychology and Psychotherapy, Goethe-University, Varrentrappstrasse 40–42, Frankfurt D-60486, Germany (email: weck@psych.uni-frankfurt.de).

Practitioner points

Clinical implications

- Treatment failure is associated with a lower therapeutic alliance in cognitive-behavioural treatment.
- Therapeutic alliance seems to be an important precondition for the adherent and competent implementation of therapeutic techniques.
- Therapeutic alliance should be monitored during psychotherapeutic treatment.

Cautions or limitations

- Results are limited to cognitive-behavioural therapy and may not be representative for other treatment approaches.
- Process analyses are based on highly standardized randomized controlled trials and may not be generalizable to routine care.

As stated by Lambert (2013), ‘in general we can say that about two-thirds of adults who enter treatment in randomized controlled trials (RCTs) have a positive outcome in about 14 sessions, but about a third either show no benefit or worsen’ (p. 189). In other words, even though psychotherapy can be highly effective, treatment failure remains a common phenomenon. Nonetheless, the conditions leading to treatment failure are rarely investigated (Dimidjian & Hollon, 2011). A better understanding of these reasons would be highly desirable, because such findings would have important implications for the improvement of psychological treatment (Lambert, 2011).

Various essentially theoretical reasons for treatment failure have been proposed, including treatment *delivery factors* (Dimidjian & Hollon, 2011). One delivery factor is the *therapeutic alliance*, which is described as the collaborative and affective bond between the therapist and the patient (Luborsky, 1984). In a recent meta-analysis, moderate but consistent relationships between therapeutic alliance and outcome were found ($r = .28$; Horvath, Del Re, Flückiger, & Symonds, 2011). Moreover, therapeutic alliance proved to be an important predictor of treatment failure in psychotherapeutic treatment (Samstag, Batchelder, Muran, Safran, & Winston, 1998; Samstag *et al.*, 2008). However, in both studies, therapeutic alliance was not assessed by independent raters, but only by patients and therapists. Therefore, it is possible that the evaluation of therapeutic alliance was confounded by knowledge of therapy outcome.

As further delivery factors and reasons for treatment failure, therapist adherence and therapist competence can be considered. *Therapist adherence* is defined as the extent to which a therapist employs interventions as described in the treatment manual and *therapist competence* is defined as the extent to which the therapist implements these interventions skilfully and appropriately for the patients in question (Waltz, Addis, Koerner, & Jacobson, 1993). For example, Kuyken and Tsivrikos (2009) found that therapists’ competence was significantly associated with therapy outcome in a naturalistic cognitive-behavioural therapy (CBT) for major depression. A recent meta-analysis, however, revealed that neither therapist adherence nor therapist competence was a significant predictor of treatment outcome (Webb, DeRubeis, & Barber, 2010). Yet, it is important to note that the meta-analysis was based on methodologically very heterogeneous studies which evaluated therapist adherence and competence and often with inappropriate methodological approaches (Simons, Rozek, & Serrano, 2013; Weck, Bohn, Ginzburg, & Stangier, 2011). For example, one methodological concern is the fact that many of the studies included in the meta-analysis assessed adherence and competence on the basis of only one therapy session, which was clearly insufficient to generalize the evaluation (Dennhag, Gibbons, Barber, Gallop, & Crits-Christoph, 2012). Other concerns relate to low statistical power, the inadequacy of adherence/competence measures, and

the clinical inexperience of competence raters. Therefore, until now, no definitive statement on the adherence–outcome and competence–outcome relationship can be made, so that further studies are necessary.

The prediction of therapy outcome from therapeutic processes is not straightforward, but in fact a complex situation and various methodological concerns should be addressed (Crits-Christoph, Gibbons, & Mukherjee, 2013). One concern with regard to the causality of the process–outcome alliance. For example, it can be questioned whether a better alliance leads to a better therapy outcome or vice versa (Barber, 2009). In line with the latter possibility, it could be demonstrated that the alliance–outcome relationship in late therapy sessions ($r = .39$) is higher than that in early sessions ($r = .25$; Horvath *et al.*, 2011). One approach to addressing this methodological concern in process analysis is to take early therapy sessions into account, in order to predict subsequent therapy success. This should ensure that the impact of therapeutic processes on (therapeutic) success is investigated and not the other way round (DeRubeis & Feeley, 1990). Such an approach was recently chosen to analyse the adherence–outcome, competence–outcome and alliance–outcome relationships (Strunk, Brotman, & DeRubeis, 2010; Strunk, Brotman, DeRubeis, & Hollon 2010).

The relationship between alliance, adherence and competence has rarely been investigated and remains unclear (Barber, 2009). In empirical research, a close association between therapist adherence and competence was found (Barber, Triffelman, & Marmar, 2007), which conforms with the notion that adherence is a precondition for competence (Waltz *et al.*, 1993). Furthermore, both theoretical reflections and empirical research suggest that therapist competence is a precondition for a good therapeutic alliance (Barber, Crits-Christoph, & Luborsky, 1996; Despland *et al.*, 2009) and that a good therapeutic alliance is in turn the basis for a positive adherence–outcome relationship (Barber *et al.*, 2006; McLeod, Southam-Gerow, Tully, Rodríguez, & Smith, 2013). Accordingly, previous research considered linear and curvilinear relationships. However, the evaluations of alliance in those earlier studies were based only on patients' self-report measures, so that further research is needed.

The aim of the current study was to compare therapist adherence, therapist competence, and therapeutic alliance in both successful and unsuccessful treatments. In order to enhance the generalizability of this investigation, we considered cognitive-behavioural cases of three different disorders from three major diagnostic categories: mood disorders (i.e., major depression), anxiety disorders (i.e., social phobia), and somatoform disorders (i.e., hypochondriasis). Moreover, three early therapy sessions per patient were considered for the prediction of treatment failure, in order to reduce a potential confounding effect between the therapeutic processes and therapy outcome.

With our current study, we also wish to expand and address weaknesses of previous process analyses, in which we partly considered the same samples. In the study of Ginzburg *et al.* (2012), therapist competence (but not adherence) proved to be an important predictor of outcome in the treatment of social phobia, and in Weck, Rudari *et al.* (2013), therapeutic alliance (but not adherence or competence) was a predictor of therapy outcome in patients with major depression. However, in those previous studies, only one or two therapy sessions were considered and the selection of treatment sessions was randomized, which could overestimate the relationship between process variables and outcome, because later therapy sessions were also considered. Moreover, the assessment of process variables was conducted by different judges with different methods (i.e., different measures and different perspectives [self-report vs. independent judges]) and there was no focus on treatment failure/success. Therefore, in the current study,

process variables were evaluated by unitary methods and, furthermore, focused on the importance of those process variables on treatment failure.

We hypothesized that evaluations of therapist adherence, therapist competence, and therapeutic alliance would be higher in treatments classified as successful, compared to those classified as treatment failures (Hypothesis 1). With respect to the relationship between adherence, alliance, and outcome, we assumed that alliance would moderate the relationship between adherence and outcome, with a better alliance enhancing the relationship between adherence and outcome (Hypothesis 2). Furthermore, alliance was assumed to mediate the relationship between competence and outcome (Hypothesis 3), and the relationship between adherence and outcome was considered as mediated by competence and alliance (Hypothesis 4). Additionally, the relationships between adherence and alliance, and competence and alliance were examined from a longitudinal perspective. It was hypothesized that adherence, competence, and alliance would yield stability if assessed in subsequent sessions (Hypothesis 5a). Finally, we assumed that adherence and competence within one session would influence alliance in the subsequent session (Hypothesis 5b).

Method

Description of the primary studies

The current study is a secondary analysis of three RCTs. The three treatment studies referred to included CBT and addressed social phobia (Stangier, Schramm, Heidenreich, Berger, & Clark, 2010), hypochondriasis (Weck, Neng, & Stangier, 2013a, 2013b) and recurrent major depressive disorder (Stangier *et al.*, 2013) respectively (see Table 1). The protocols of all three studies were approved by institutional review boards and all participants gave their informed consent. Diagnoses in all three treatment studies were evaluated uniformly by trained and experienced diagnosticians, using the Structured Clinical Interview for DSM-IV (First, Spitzer, Gibbon, & Williams, 1997). While, in the studies treating social phobia and hypochondriasis, the reduction of social anxiety and hypochondriacal symptoms was intended, in the study treating recurrent major depression, the aim was to prevent recurrent depressive symptoms. In the current study, only the CBT approaches of the three RCTs were considered for the analysis of treatment success and failure. Data for the active control conditions (e.g., psychoeducation) were not used, in order to ensure homogeneity of the data, because data for all three RCTs should be considered in one analysis.

Determination of success and failure

The aim of the current study was to compare treatments which could be clearly classified as failure with treatments which can be clearly classified as successes. By using this approach, the current study differs from previous ones, which also considered patients with moderate treatment effects and which use continuous outcome variables. The current approach was chosen, because we intended to evaluate process variables by independent judges in three treatment sessions and because this procedure is associated with high costs. For the comparison of patients with treatment failure and with treatment success, we expected larger effects than for the same sample size of patients with the whole spectrum of treatment effects.

For every RCT, one clinical interview and one self-report measure were considered for the determination of treatment failure as opposed to success. For patients with major depression, the Hamilton Rating Scale for Depression (HRSD-17; Collegium Internationale

Table 1. Overview about the three randomized controlled trials which form the basis for the current study

	Studies		
	Stangier <i>et al.</i> (2013) registered under ISRCTN 81212636	Stangier <i>et al.</i> (2010) not registered	Weck <i>et al.</i> (2013a, 2013b) registered under NCT01119469
Diagnosis	Recurrent major depression (currently remitted)	Social phobia	Hypochondriasis
Central aims of the studies	Reduction of the risk of relapse	Reduction of social anxiety symptoms	Reduction of hypochondriacal symptoms
Whole sample sizes (CBT sample size)	180 (87)	117 (38)	84 (38)
Treatments	Maintenance CBT; manualized psychoeducation	Cognitive therapy; interpersonal psychotherapy	Cognitive therapy; exposure therapy
Treatments extent	16 sessions	16 sessions	12 sessions
Exclusion criteria	Organic mental disorders, disorders caused by psychotropic substances, schizophrenia, schizoaffective disorder, bipolar disorder, borderline personality disorder, mental retardation, and acute suicidality	Psychosis, current substance dependency or abuse, Axis II personality disorders from the dramatic or odd cluster, severe depression, and acute suicidality	Major medical illness, diagnosis of substance addiction, schizophrenia, schizoaffective disorder, bipolar disorder, and acute suicidality

Note. CBT = cognitive-behavioural therapy.

Psychiatriae Scalarum [CIPS], 1977; Hamilton, 1960) and the Beck Depression Inventory (BDI; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961; Hautzinger, Bailer, Worall, & Keller, 1994) was chosen, for patients with social phobia, the Liebowitz Social Anxiety Scale (Liebowitz, 1987; Stangier & Heidenreich, 2005) and the Social Phobia and Anxiety Inventory (Fydrich, 2002; Turner, Beidel, Dancu, & Stanley, 1989), and for patients with hypochondriasis, the Yale-Brown Obsessive Compulsive Scale for Hypochondriasis (Weck, Gropalis, Neng, & Witthöft, 2013) and the Illness Attitude Scales (Hiller, Rief, & Fichter, 2002; Kellner, 1986) were chosen. All measures used are internationally well established and have demonstrated their reliability and validity. For the definition of treatment failure and success, the suggestions of Jacobson and Truax (1991) were taken into account and changes larger than two standard deviations in the outcome measures were considered to constitute clinically significant change. Table 2 summarizes the criteria for success and failure for the current study, which are described additionally below in greater detail.

Treatment failure and treatment success

Drop outs, non-responses, and deteriorations were considered as *treatment failure*. However, only drop outs were taken into account as treatment failure when there were indicators of failure (e.g., dissatisfaction about the treatment process). Drop outs were not considered as treatment failure, when external causes were reasonable and thus the drop

Table 2. Definition of treatment success and treatment failure in the current study (criteria 1, 2, and 3 or criterion 4 must be fulfilled to indicate a treatment success or failure, respectively)

Treatment study	Treatment success	Treatment failure
Social phobia (Stangier et al., 2013)	<ol style="list-style-type: none"> 1. Diagnosis of social phobia is no longer present in the SCID 2. At least two <i>SD</i> symptom reduction in the LSAS (clinical interview) 3. At least two <i>SD</i> symptom reduction in the SPAI (self-report) 	<ol style="list-style-type: none"> 1. Diagnosis of social phobia is still present in the SCID 2. Less than one <i>SD</i> symptom reduction in the LSAS (clinical interview) 3. Less than one <i>SD</i> symptom reduction in the SPAI (self-report) 4. Drop out
Hypochondriasis (Weck et al., 2013a, 2013b)	<ol style="list-style-type: none"> 1. Diagnosis of hypochondriasis is not longer present in the SCID 2. At least two <i>SD</i> symptom reduction in the H-YBOCS (clinical interview) 3. At least two <i>SD</i> symptom reduction in the IAS (self-report) 	<ol style="list-style-type: none"> 1. Diagnosis of hypochondriasis is still present in the SCID 2. Less than one <i>SD</i> symptom reduction in the H-YBOCS (clinical interview) 3. Less than one <i>SD</i> symptom reduction in the IAS (self-report) 4. Drop out
Major depression (Stangier et al., 2010)	<ol style="list-style-type: none"> 1. Diagnosis of major depression is still not present in the SCID 2. Less than one <i>SD</i> symptom increase in the HRSD-17 (clinical interview) 3. Less than one <i>SD</i> symptom increase in the BDI (self-report) 	<ol style="list-style-type: none"> 1. Diagnosis of major depression is again present in the SCID 2. More than two <i>SD</i> symptom increase in the HRSD-17 (clinical interview) 3. More than two <i>SD</i> symptom increase in the BDI (self-report) 4. Drop out

Note. BDI = Beck Depression Inventory; H-YBOCS = Yale-Brown Obsessive Compulsive Scale for Hypochondriasis; HRSD-17 = Hamilton Rating Scale for Depression; IAS = Illness Attitude Scales; LSAS = Liebowitz Social Anxiety Scale; SCID = Structured Clinical Interview for DSM-IV; SPAI = Social Phobia and Anxiety Inventory.

out did not indicate that the treatment had failed (e.g., patient's move to another city for job reasons).

According to our definition, 43 patients who received CBT in the three treatment conditions were classified as treatment successes and 18 as treatment failures. These 61 patients participated in the current study, which constituted 37.42% of the whole sample of 163 patients who received CBT in the three treatment trials. We found no significant differences between the occurrence of treatment success and treatment failure between the therapists ($p > .10$).

Participants

Clients

In the current sample of 61 patients, the mean age was 45.0 years ($SD = 12.5$ years) and 34 (55.7%) were female. Nine patients had a primary diagnosis of social phobia, eleven of hypochondriasis, and forty-one of recurrent major depressive disorder. Twenty-eight (45.9%) of the patients had a comorbid axis-I-disorders and seventeen (27.9%) a comorbid axis-II-disorder.

Therapists

Patients in the current study were treated by 32 therapists (25 female). The average age of the therapists was 31.6 ($SD = 6.3$; range: 25–51 years). On average, they had 3.0 years ($SD = 3.6$ years; range: 0–20 years) of clinical experience. All therapists were trained in the relevant CBT and received regular supervision. In the current study, the average number of clients treated by therapists was 1.9 ($SD = 1.1$; range: 1–5). Fifteen therapists treated one patient, nine treated two patients, five treated three patients, two treated four patients, and one treated five patients.

Judges

Process measures (therapist adherence, therapist competence, and therapeutic alliance) were performed by two judges. Both were clinical psychologists and psychotherapists who had four and 5 years of clinical experience respectively. Judges were familiar with the applied treatment manuals and had completed a 26-hr training course on how to use process rating scales. Training involved practicing the rating of therapy sessions using the process scales. Twelve therapy sessions, which were not part of the current study, were rated during training. During the training, discrepancies in the ratings were discussed with an experienced trainer in order to reach consensus between the two judges. Judges were blind regarding therapy outcome.

Description of the cognitive-behavioural treatments

Maintenance CBT for recurrent major depression aimed at preventing recurrences of depressive episodes. All interventions of this treatment are described in detail in a treatment manual (Risch, Stangier, Heidenreich, & Hautzinger, 2012). Interventions of cognitive therapy for social phobia (Stangier, Clark, & Ehlers, 2006) and those of cognitive therapy for hypochondriasis (see Weck, 2014) are described in a treatment manual as well.

Measures

Cognitive-behavioural therapy adherence scale (CBT-AS)

The CBT-AS was developed on the basis of a scale for assessing therapist adherence in the treatment of major depression (Weck, Hilling, Schermelleh-Engel, Rudari, & Stangier, 2011). To address all three treatments in the current study, we expanded the adherence scale and included further items which were important for therapists' adherence in the treatment of social phobia and hypochondriasis. The treatment manuals of the three treatment studies served as the basis for generating the CBT-AS items used in the current study. The response format of the CBT-AS is a 3-point rating scale (not adherent, partly adherent, and adherent) and the scale includes 23 items: (1) agenda, (2) reviewing homework, (3) time management, (4) use of materials, (5) application of treatment content, (6) consideration of general principles of cognitive therapy, (7) development of an individual cognitive model, (8) identification of automatic thoughts, (9) modification of automatic thoughts, (10) behavioural experiments, (11) setting homework, (12) relapse prevention, and (13) the exclusion of non-adherent techniques. Furthermore, Items 14–23 evaluated specific aspects of therapists' adherence in the specific treatment studies. Items 14–16 address social and hypochondriasis (14 – identification of safety

behaviour, 15 – modification of safety behaviour, and 16 – identification/modification of attention processes), Item 17 addresses only hypochondriasis (psychoeducation), Items 18 and 19 address only social phobia, and Items 20–23 only major depression (20 – facilitation of self-monitoring, 21 – approaching questions and difficulty in understanding, 22 – acceptance and commitment therapy interventions and 23 – wellbeing-therapy interventions). In the current study, the interrater reliability for the CBT-AS mean score was very high (intraclass correlation coefficient – $ICC_{(3,2)} = .86; p < .001$).

Cognitive Therapy Scale (CTS)

Therapist competence was evaluated with the CTS (Weck, Hautzinger, Heidenreich, & Stangier, 2010; Young & Beck, 1980), which evaluates competence on a 7-point rating scale (poor, barely adequate, mediocre, satisfactory, good, very good, and excellent). The German version of the CTS contains 14 items which evaluate the level of therapist competence: (1) agenda, (2) dealing with problems/questions/objections, (3) clarity of communication, (4) pacing and efficient use of time, (5) interpersonal effectiveness, (6) resource activation, (7) reviewing previously set homework, (8) using feedback and summaries, (9) guided discovery, (10) focus on central cognitions and behaviour, (11) rationale, (12) selecting appropriate strategies, (13) appropriate implementation of techniques, and (14) setting homework. In the current study, the interrater reliability for the CTS mean was very high ($ICC_{(3,2)} = .92; p < .001$). The internal consistency (Cronbach's α) for all 14 items of the CTS was $\alpha = .95$.

Helping Alliance Questionnaire (HAQ)

Therapeutic alliance was evaluated with the HAQ (Bassler, Potratz, & Krauthauser, 1995; Luborsky, 1984). The questionnaire consists of 11 items which are answered on a 6-point Likert scale, ranging from 1 (*strongly disagree*) to 6 (*strongly agree*). For example, Item 9 of the HAQ reads: 'I feel I am working together with the therapist in a joint effort'. We developed an independent rater's version of the HAQ in German, the HAQ-R (Richtberg, Jakob, Höfling, & Weck, 2015), by rewording the items (e.g., 'I believe the patient is working together with the therapist in a joint effort'). In the current study, the interrater reliability for the HAQ-R mean was good ($ICC_{(3,2)} = .84; p < .001$). The internal consistency (Cronbach's α) for all 11 items of the HAQ-R was $\alpha = .96$.

Procedure

For all 61 patients, the first three therapy sessions were selected in order to evaluate therapists' adherence, therapists' competence, and the therapeutic relationship. A total of 175 videotapes were rated by the two independent judges. Although all treatment sessions should have been videotaped by the therapist, videotapes were not available for all treatment sessions because of a technical defect, video cameras being unavailable, therapist non-compliance, or early patient drop-out. Therefore, subsequent therapy sessions (but only for the first half of the therapy) were also considered for the analysis. Altogether, 50 (82.0%) videotapes of the first therapy sessions, 49 (80.3%) of the second therapy sessions, 44 (72.1%) of the third therapy session, 13 (21.3%) of the fourth therapy sessions, 12 (19.6%) of the fifth therapy sessions, and 4 (6.6%) videotapes of the sixth therapy sessions were taken into account.

Statistical analysis

Preliminary analysis

Interrater reliability was analysed with ICCs, using model 3 ($ICC_{(3,m)}$; Shrout & Fleiss, 1979). ICCs were at all times calculated from the mean of two judgments ($ICC_{(3,2)}$). Analysis of variance (ANOVA) was used to compare the mean values of the different outcome groups (treatments classified as success vs. failure). Categorical variables (e.g., sex) were analysed with chi-square tests. Comparison of process measures (CBT-AS, CTS, and HAQ-R) between treatments classified as success and failure were conducted by multivariate analysis of variance (MANOVA). Effect sizes were determined using Hedges' *g* (Hedges & Olkin, 1985).

Outcome models (OM)

Path analysis models were specified with treatment outcome (treatment success vs. treatment failure) as the categorical-dependent variable. For Outcome Model 1 (OM 1), alliance (HAQ-R) and adherence (CBT-AS) were considered as predictors (see Figure 1a). In addition to linear effects, quadratic effects for adherence and alliance were assumed, as well as a moderation effect of alliance and adherence. For Outcome Model 2 (OM 2), alliance (HAQ-R) was specified as the predictor variable mediating the effect between competence (CTS) and outcome (see Figure 1b). Additionally, adherence (CBT-AS) was specified with a mediating effect on therapy outcome via competence (CTS) and alliance (HAQ-R). 95% confidence intervals (CI) for indirect effects were computed (MacKinnon, Lockwood, & Williams, 2004). Correlations between predictor variables were estimated in OM 1 and OM 2.

Autoregressive models (AM)

Path analysis models were specified regressing session three variables (e.g., the CTS score of session three) on the respective session two and session one variables (e.g., on the CTS score of sessions two and one), and regressing session two variables on the respective session one variables (see Figure 2a,b). Furthermore, the residual variances and correlations of the residual variances of session three and session two were set free, thus accounting for systematic situational influences. Finally, crossover effects were estimated, assuming that, for example, competence (CTS) in session one would influence alliance (HAQ-R) in session two. Autoregressive Model 1 (AM 1) examined adherence (CBT-AS) and alliance (HAQ-R), whereas Autoregressive Model 2 (AM 2) examined competence (CTS) and alliance (HAQ-R).

Data analysis and model fit evaluation

Data were analysed using *Mplus* (Version 6; Muthén & Muthén, 2010), applying the maximum likelihood estimator. In order to evaluate the model fit AM 1 and AM 2, the χ^2 -value and the degrees of freedom (*df*) were reported, as well as the root mean square error of approximation (RMSEA), the comparative fit index (CFI), and the standardized root mean square residual (SRMR). Some standards exist concerning the model fit (Schermelleh-Engel, Moosbrugger, & Müller, 2003). Values below 2 for the ratio between the χ^2 -value and the *df* indicate a good model fit, and values below 3 for this ratio indicate an acceptable fit. RMSEA values less than .05 were found to indicate a good model fit and less than .08, an acceptable model fit. The CFI indicates a good model fit for values within the range 0.95 and 1.00, whereas values between .90 and .95 signify an acceptable fit.

SRMR values should remain below .10 for acceptable model fit. *Mplus* does not provide the usual fit statistic for models with categorical dependent variables.

Results

Preliminary analysis

Table 3 shows the socio-demographic data and comorbidity of the patients classified as treatment successes and treatment failures. No significant differences between success and failures were found regarding age, sex, educational level, number of comorbid axis-I-disorders, and number of comorbid axis-II-disorders (see Table 3).

Prediction of treatment success and treatment failure (Hypothesis 1)

For the prediction of treatment outcome (success vs. failure), mean scores of the adherence measure (CBT-AS), competence measure (CTS), and alliance measure (HAQ-R) over the three treatment sessions were computed. The means and standard deviations of these measures are given in Table 4 separately for treatments classified as successes or failures. Substantial relationships were found between the CBT-AS, the CTS, and the HAQ-R. The CBT-AS correlated $r = .62$ ($p < .001$) with the CTS and $r = .56$ ($p < .001$) with the HAQ-R. Correlations between the CTS and the HAQ-R were $r = .79$ ($p < .001$).

Table 3. Soziodemographic data and comorbidity of the patients classified as treatment success and failure

Measures	Treatments classified as success ($n = 43$)	Treatment classified as failure ($n = 18$)	Test value
Soziodemographic data			
Age (SD)	44.56 (12.31)	45.94 (13.10)	$F_{(1,59)} = 0.15$
Female (%)	24 (55.81)	10 (55.55)	$\chi^2_{(1)} = 0.00$
Patients with university entrance qualification (%)	16 (37.21)	9 (50.00)	$\chi^2_{(1)} = 3.64$
Comorbidity			
Comorbid Axis-I-disorder (%)	21 (48.83)	7 (38.89)	$\chi^2_{(1)} = 0.51$
Comorbid Axis-II-disorder (%)	11 (25.58)	6 (33.33)	$\chi^2_{(1)} = 0.38$

Table 4. Mean scores and standard deviations of the measures of therapists' adherence (CBT-AS), therapists' competence (CTS), and the therapeutic alliance (HAQ-R) for treatments classified as success and failure

Measures	Treatments classified as success ($n = 43$)	Treatment classified as failure ($n = 18$)	F-value $F_{(1,59)}$
Process measures (SD)			
CBT-AS	1.56 (0.19)	1.43 (0.29)	4.51*
CTS	3.14 (0.71)	2.91 (0.72)	1.38
HAQ-R	3.84 (0.58)	3.43 (0.76)	5.34*

Note. CBT-AS = cognitive-behavioural therapy adherence scale; CTS = Cognitive Therapy Scale; HAQ-R = Helping Alliance Questionnaire (rater version).

* $p < .05$.

The MANOVA for the two outcome groups (treatment success vs. treatment failure) revealed a significant group effect ($F_{(3,57)} = 57.00; p = .04$). Significant higher scores were found for the treatments classified as successes in comparison to those classified as failures for the CBT-AS ($F_{(1,59)} = 4.51; p = .04; g = .58$) and the HAQ-R ($F_{(1,59)} = 5.34; p = .02; g = .62$), but not for the CTS ($F_{(1,59)} = 1.38; p = .24$).

Outcome models (Hypotheses 2, 3 and 4)

Relationship between adherence, alliance, and outcome (OM 1)

The resulting path model with adherence (CBT-AS), alliance (HAQ-R) and treatment outcome is depicted in Figure 1a. Only the moderation of CBT-AS with HAQ-R yielded significant effects on the treatment outcome ($r = .24; p < .05$). Higher levels of therapy alliance were associated with a stronger relationship between adherence and treatment outcome.

Relationship between adherence, competence, alliance, and outcome (OM 2)

The resulting path model with adherence (CBT-AS), competence (CTS), alliance (HAQ-R) and treatment outcome is depicted in Figure 1b. HAQ-R has a small effect on treatment

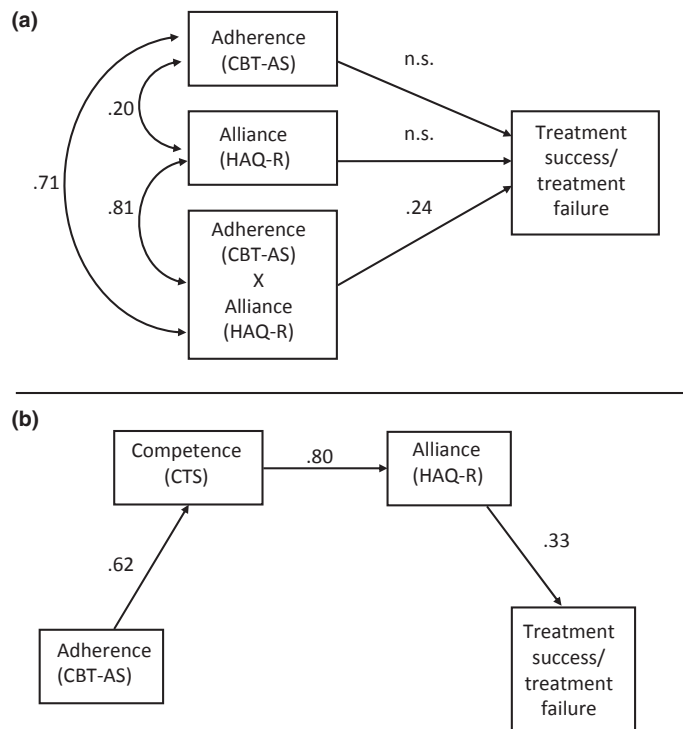


Figure 1. (a) Path diagram of the completely standardized solution of the Outcome Model 1 with treatment outcome as categorical dependent variable and adherence, therapeutic alliance, and the moderator term between alliance and adherence as predictors. (b) Path diagram of the completely standardized solution of Outcome Model 2 with treatment outcome as categorical dependent variable and adherence, competence, and therapeutic alliance as predictors.

outcome ($r = .33$; $p < .05$), whereas CTS has a large effect on HAQ-R ($r = .80$; $p < .001$), and CBT-AS on CTS ($r = .62$; $p < .001$). The indirect (mediation) effect of CTS on treatment outcome via HAQ-R is substantial at .26 (95% CI [.15; .38]). There were no indications of a significant direct effect of CTS on treatment outcome. The indirect (mediation) effect of CBT-AS on treatment outcome via CTS and HAQ-R is small, but differs significantly from zero at .16 (95% CI [.01; .31]).

Autoregressive models (Hypotheses 5a and 5b)

Relationship between adherence and alliance (AM 1)

Model AM 1 examining the adherence (CBT-AS) and alliance (HAQ-R) for three subsequent sessions is depicted in Figure 2a. This model showed good model fit ($\chi^2 = 2.77$; $df = 6$, RMSEA = .00; CFI = 1.00; SRMR = .05). The relationships between the CBT-AS scores in the three subsequent sessions were moderate ($r > .4$) and between the HAQ-R scores, high ($r > .7$). Additionally, there was a small direct effect from HAQ-R (session 1) on HAQ-R (session 3) with .18, but no direct effect from CBT-AS (session 1) on CBT-AS (session 3). These results can be interpreted as evidence of the stability of adherence and alliance assessment in subsequent sessions (Hypothesis 5a). The residual variances for CBT-AS and HAQ-R (sessions 2 and 3) differed from zero and yielded significant correlations with each other at the respective time point ($r > .4$) indicating that there were substantial situational influences. Finally, there was only one small crossover effect from HAQ-R (session 1) on CBT-AS (session 2), but no effects from CBT-AS on HAQ-R (Hypothesis 5b).

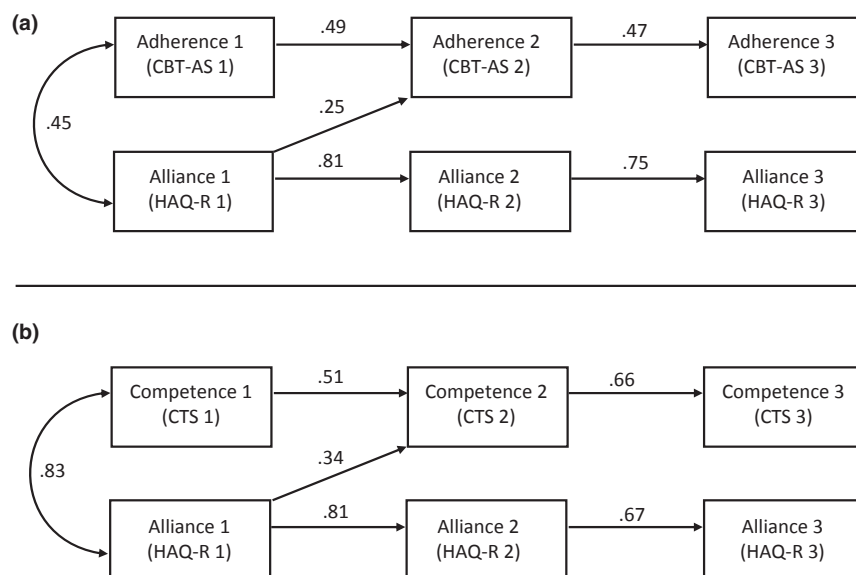


Figure 2. (a) Path diagram of the completely standardized solution of the Autoregressive Model 1 with adherence and alliance in three subsequent sessions. A small effect from Alliance 1 on Alliance 3 had been omitted. (b) Path diagram of the completely standardized solution of the Autoregressive Model 2 with competence and helping alliance in three subsequent sessions. Small effects from Competence 1 on Competence 3 and from Alliance 1 on Alliance 3 had been omitted.

Relationship between competence and alliance (AM 2)

Model AM 2 examining competence (CTS) and alliance (HAQ-R) for three subsequent sessions is illustrated in Figure 2b. This model showed a good model fit ($\chi^2 = 3.99$; $df = 5$, RMSEA = .00; CFI = 1.00; SRMR = .03). The relationships between the CTS scores in the three subsequent sessions were substantial ($r > .5$) and between the HAQ-R scores, high ($r > .7$). Additionally, there was a small direct effect from HAQ-R (session 1) on HAQ-R (session 3) with .26 and a small direct effect from CTS (session 1) on CTS (session 3) with .26. Therefore, this can also be interpreted as evidence of the stability of competence in subsequent sessions (Hypothesis 5a). The residual variances for CTS and HAQ-R (sessions 2 and 3) differed from zero and yielded significant correlations with each other at the respective time point ($r > .6$). Finally, there was only one moderate crossover effect from HAQ-R (session 1) on CTS (session 2), but no effects from CTS on HAQ-R (Hypothesis 5b).

Discussion

The aim of the current study was to investigate the role played by therapeutic adherence and competence, as well as by therapeutic alliance in treatment failure. We found significantly higher scores for therapists' adherence and therapeutic alliance, but not for therapists' competence in treatments classified as successes, in comparison to those classified as failures (Hypothesis 1). As expected, therapeutic alliance was found to be a significant moderator of the adherence–outcome relationship (Hypothesis 2). Moreover, we found therapeutic alliance to be a significant mediator between competence and outcome (Hypothesis 3), and competence and alliance to be a significant mediator between adherence and outcome (Hypothesis 4). Additionally, longitudinal perspectives reveal that assessments of adherence, competence, and alliance demonstrated substantial stability (Hypothesis 5a). Furthermore, we could not confirm by means of the longitudinal analysis that adherence and/or competence within one session would influence alliance in the subsequent session (Hypothesis 5b). Instead, there is evidence that the alliance influenced therapist adherence and competence in the subsequent session.

In the current study, therapist adherence and therapeutic alliance were associated moderately with success and failure in psychotherapeutic treatments. Moreover, alliance was a moderator of the adherence–outcome relationship, which was also found in one previous study (Barber *et al.*, 2006). However, the finding that therapist competence was not associated with treatment failure, was unexpected. In contrast, previous studies found stronger associations between therapist competence and outcome than between therapist adherence and outcome (Barber *et al.*, 1996; Ginzburg *et al.*, 2012). However, these earlier studies did not focus on treatment failure, but on outcome in general. On the other hand, in our cross-sectional analysis, therapist competence seems to be indirectly associated with treatment success and failure, because the alliance was found to be a significant mediator between competence and outcome. This relationship between therapist competence and therapeutic alliance was found in other studies as well (Barber *et al.*, 1996; Despland *et al.*, 2009). Moreover, in line with the theoretical assumptions (Waltz *et al.*, 1993), we found therapist adherence to be a precondition for competence, which is then a precondition for alliance, and in turn, a precondition for treatment success.

However, the longitudinal analysis of our data yielded different results to the cross-sectional analysis. Therapist adherence and/or competence within one session did not significantly influence therapeutic alliance in the subsequent session, but alliance

significantly influenced therapist adherence and competence. We believe that these longitudinal analyses have important advantages compared to the cross-sectional analyses, because they enable us to consider the therapy process of three consecutive therapy sessions and not only the mean scores of those sessions, which is clearly less precise. The differing results between cross-sectional and longitudinal analyses could be explained by different levels of precision of the considered data base. Because longitudinal analyses are more precise, the results can be considered as more important and reveal that therapeutic alliance seems to be more important for therapist adherence and competence, than the other way around. One possible explanation of this finding is that a good therapeutic alliance enables the therapist to implement therapeutic techniques more adherently and competently, and that a good therapeutic alliance leads to more successful therapy results. However, we also have to take into consideration that the relationship between therapeutic alliance and treatment failure/success might not be causal. Because an alternative explanation could be that a further variable, which we did not evaluate (e.g., patient motivation) is the cause of both a good therapeutic alliance and treatment success.

The current study has many strengths, including the consideration of three different RCTs and disorders, the consideration of three subsequent therapy sessions, the analysis of relevant process variables by two independent raters, the use of standardized and reliable rating scales, and several methodological aspects (e.g., consideration of quadratic relationships, confirmatory path analysis instead of regression analysis). However, several limitations should be taken into consideration as well. In our study, we addressed CBT and considered only a subsample of those treatments (37%) which demonstrated either failure or success. The reason for this procedure was grounded in the high costs of the rating process (2 judges \times 61 patients \times 3 videotapes = > 366 hr rating time) and the aims of our study (i.e., investigation of relevant variables for treatment failure). It is therefore important to note that our results are not directly comparable to studies which considered the whole spectrum of treatment outcome and are not generalizable to all patients who receive CBT.

In the current study, the waiting list control groups were not taken into consideration. This procedure implies that we expected positive effects only in the active treatment conditions (see Dunn & Bentall, 2007). However, positive effects could also have occurred in the waiting list control group. Novel statistical approaches (Dunn & Bentall, 2007; Dunn *et al.*, 2012) are able to address this issue, but these analyses are beyond the scope of this paper.

Moreover, the assessment of process variables was conducted in the context of three highly standardized RCTs. In the RCTs, therapists received intensive training, which can be considered as limiting the level of variance in therapists' competence. It is questionable whether different results would emerge in less standardized treatment studies.

Furthermore, the BDI was used as the self-report measure for patients with recurrent major depression. However, future studies should favor the revision of the BDI, because it more accurately represents the diagnostic criteria of the diagnosis (Beck & Steer, 1996).

Different process measures were rated by the same raters. Therefore, it can be questioned whether the simultaneous assessment of different processes decreases the quality of the data. However, in empirical research, it could be demonstrated that the simultaneous assessment of different instruments is possible without substantially lowering the quality of the data (Ulvenes *et al.*, 2012).

References

- Barber, J. P. (2009). Toward a working through of some core conflicts in psychotherapy research. *Psychotherapy Research, 19*, 1–12. doi:10.1080/10503300802609680
- 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 and Clinical Psychology, 64*, 619–622. doi:10.1037/0022-006X.64.3.619
- Barber, J. P., Gallop, R., Crits-Christoph, P., Frank, A., Thase, M. E., Weiss, R. D., & Gibbons, M. B. C. (2006). The role of therapist adherence, therapist competence, and alliance in predicting outcome of individual drug counseling: Results from the National Institute Drug Abuse Collaborative Cocaine Treatment Study. *Psychotherapy Research, 16*, 229–240. doi:10.1037/0736-9735.25.3.461
- Barber, J. P., Triffelman, E., & Marmer, C. (2007). Considerations in treatment integrity: Implications and recommendations for PTSD research. *Journal of Traumatic Stress, 20*, 793–805. doi:10.1002/jts.20295
- Bassler, M., Potratz, B., & Krauthauser, H. (1995). Der “Helping Alliance Questionnaire” (HAQ) von Luborsky [The “Helping Alliance Questionnaire” (HAQ) by Luborsky]. *Psychotherapeut, 40*, 23–32.
- Beck, A. T., & Steer, R. A. (1996). *Manual for the Beck Depression Inventory-2*. San Antonio, TX: Psychological Corporation.
- Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4*, 561–571. doi:10.1001/archpsyc.1961.01710120031004
- Collegium Internationale Psychiatriae Scalarum (CIPS, Ed.). (1977). *CIPS: Hamilton Depression Scale*. Weinheim, Germany: Beltz Test.
- Crits-Christoph, P., Gibbons, M. B. C., & Mukherjee, D. (2013). The efficacy and effectiveness of psychotherapy. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (6th ed., pp. 298–340). New York, NY: Wiley.
- Dennhag, I., Gibbons, M. B. C., Barber, J. P., Gallop, R., & Crits-Christoph, P. (2012). How many treatment sessions and patients are needed to create a stable score of adherence and competence in the treatment of cocaine dependence? *Psychotherapy Research, 22*, 475–488. doi:10.1080/10503307.2012.674790
- DeRubeis, R. J., & Feeley, M. (1990). Determinants of change in cognitive therapy for depression. *Cognitive Therapy and Research, 14*, 469–482. doi:10.1007/BF01172968
- Despland, J.-N., de Roten, Y., Drapeau, M., Currat, T., Beretta, V., & Kramer, U. (2009). The role of alliance in the relationship between therapist competence and outcome in brief psychodynamic psychotherapy. *The Journal of Nervous and Mental Disease, 197*, 362–367. doi:10.1097/NMD.0b013e3181a20849
- Dimidjian, S., & Hollon, S. D. (2011). What can be learned when empirical supported treatments fail? *Cognitive and Behavioral Practice, 18*, 303–305. doi:10.1077/22911000459
- Dunn, G., & Bentall, R. (2007). Modelling treatment-effect heterogeneity in randomized controlled trials of complex interventions (psychological treatments). *Statistics in Medicine, 26*, 4719–4745. doi:10.1002/sim.2891
- Dunn, G., Fowler, D., Rollinson, R., Freeman, D., Kuipers, E., Smith, B., . . . Bebbington, P. (2012). Effective elements of cognitive behaviour therapy for psychosis: Results of a novel type of subgroup analysis based on principal stratification. *Psychological Medicine, 42*, 1057–1068. doi:10.1017/S0033291711001954
- First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (1997). *Structured Clinical Interview for DSM-IV Axis I Disorder (SCID-I)*. Washington, DC: American Psychiatric Press.
- Fydrich, T. (2002). Soziale Phobie und Angst Inventar (SPAI) [Social phobia and anxiety inventory]. In E. Brähler, J. Schumacher, & B. Strauß (Eds.), *Diagnostische Verfahren in der Psychotherapie* [Diagnostic instruments in psychotherapy] (pp. 335–338). Göttingen, Germany: Hogrefe.

- Ginzburg, D. M., Bohn, C., Höfling, V., Weck, F., Clark, D. M., & Stangier, U. (2012). Treatment specific competence predicts outcome in cognitive therapy for social anxiety disorder. *Behaviour Research and Therapy*, *50*, 747–752. doi:10.1016/j.brat.2012.09.001
- Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry*, *23*, 56–62. doi:10.1136/jnnp.23.1.56
- Hautzinger, M., Bailer, M., Worall, H., & Keller, F. (1994). *Das Beck Depressionsinventar [The Beck Depression Inventory]*. Bern, Switzerland: Huber.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
- Hiller, W., Rief, W., & Fichter, M. M. (2002). Dimensional and categorical approaches to hypochondriasis. *Psychological Medicine*, *32*, 707–718. doi:10.1017/S003329170200552
- Horvath, A. O., Del Re, A. C., Flückiger, C., & Symonds, D. (2011). Alliance in individual psychotherapy. *Psychotherapy*, *48*, 9–16. doi:10.1037/a0022186
- Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, *59*, 12–19. doi:10.1037/0022-006X.59.1.12
- Kellner, R. (1986). *Somatization and hypochondriasis*. New York, NY: Praeger.
- Kuyken, W., & Tsivrikos, D. (2009). Therapist competence, comorbidity and cognitive-behavioral therapy for depression. *Psychotherapy and Psychosomatics*, *78*, 42–48. doi:10.1159/000172619
- Lambert, M. J. (2011). What have we learned about treatment failure in empirically supported treatments? Some suggestions for practice. *Cognitive and Behavioral Practice*, *18*, 413–420. doi:10.1016/j.cbpra.2011.02.002
- Lambert, M. J. (2013). The efficacy and effectiveness of psychotherapy. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (pp. 169–931). New York, NY: Wiley.
- Liebowitz, M. R. (1987). Social phobia. *Modern Problems of Pharmacopsychiatry*, *22*, 141–173.
- Luborsky, L. (1984). *Principles of psychoanalytic psychotherapy. Manual for supportive-expressive psychotherapy*. New York, NY: Basic Books.
- MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research*, *39*, 99–128. doi:10.1207/s15327906mbr3901_4
- McLeod, B. D., Southam-Gerow, M. A., Tully, C. B., Rodríguez, A., & Smith, M. M. (2013). Making a case for treatment integrity as a psychosocial treatment quality indicator for youth mental health care. *Clinical Psychology: Science and Practice*, *20*, 14–32. doi:10.1111/cpsp.12020
- Muthén, L. K., & Muthén, B. O. (2010). *Mplus user's guide* (6th ed.). Los Angeles, LA: The Author.
- Richtberg, S., Jakob, M., Höfling, V., & Weck, F. (2015). Assessment of patient interpersonal behavior: Development and validation of a rating scale. *Psychotherapy Research*. Advance online publication. doi:10.1080/10503307.2014.947391
- Risch, A. K., Stangier, U., Heidenreich, T., & Hautzinger, M. (2012). *Kognitive Erhaltungstherapie bei rezidivierender Depression: Rückfälle verhindern, psychische Gesundheit erhalten. [Cognitive behavioral maintenance therapy for recurrent depression: Preventing relapse, maintaining mental health]*. Heidelberg, Germany: Springer.
- Samstag, L. W., Batchelder, S. T., Muran, J. C., Safran, J. D., & Winston, A. (1998). Early identification of treatment failure in short-term psychotherapy: An assessment of therapeutic alliance and interpersonal behavior. *Journal of Psychotherapy Practice and Research*, *7*, 126–143.
- Samstag, L. W., Muran, J. C., Wachtel, P. L., Slade, A., Safran, J. D., & Winston, A. (2008). Evaluating negative process: A comparison of working alliance, interpersonal behavior, and narrative coherence among three psychotherapy outcome conditions. *American Journal of Psychotherapy*, *62*, 165–194.

- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Test of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, 8, 23–74.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86, 420–428.
- Simons, A. D., Rozek, D. C., & Serrano, J. L. (2013). Wanted: Reliable and valid measures for the science of cognitive behavioral therapy dissemination and implementation. *Clinical Psychology: Science and Practice*, 20, 181–194. doi:10.1111/cpsp.12033
- Stangier, U., Clark, D. M., & Ehlers, A. (2006). *Soziale Phobie [Social phobia]*. Göttingen, Germany: Hogrefe.
- Stangier, U., & Heidenreich, T. (2005). *Liebowitz Social Anxiety Scale*. In Collegium Internationale Psychiatricae Scalarum (Ed.). Göttingen, Germany: Beltz Test.
- Stangier, U., Hilling, C., Heidenreich, T., Risch, A. K., Barocka, A., Schlösser, R., . . . Hautzinger, M. (2013). Maintenance cognitive-behavioral therapy and manualized psychoeducation in the treatment of recurrent depression: A multicenter prospective randomized controlled trial. *American Journal of Psychiatry*, 170, 624–632. doi:10.1176/appi.ajp.2013.12060734
- Stangier, U., Schramm, E., Heidenreich, T., Berger, M., & Clark, D. M. (2010). Cognitive therapy vs interpersonal psychotherapy in social anxiety disorder: A randomized controlled trial. *Archives of General Psychiatry*, 68, 692–700. doi:10.1001/archgenpsychiatry.2011.67
- Strunk, D. R., Brotman, M. A., & DeRubeis, R. J. (2010). The process of change in cognitive therapy for depression: Predictors of early inter-session symptom gains. *Behavioral Research and Therapy*, 48, 599–606. doi:10.1016/j.brat.2010.03.011
- Strunk, D. R., Brotman, M. A., DeRubeis, R. J., & Hollon, S. D. (2010). Therapist competence in cognitive therapy for depression: Predicting subsequent symptom change. *Journal of Consulting and Clinical Psychology*, 78, 429–437. doi:10.1037/a0019631
- Turner, S. M., Beidel, D. C., Dancu, C. V., & Stanley, M. A. (1989). An empirically derived inventory to measure social fears and anxiety: The social phobia anxiety inventory. *Psychological Assessment*, 1, 35–40. doi:10.1037/1040-3590.1.1.35
- Ulvenes, P. G., Berggraf, L., Hoffart, A., Levy, R. A., Ablon, J. S., McCullough, L., & Wampold, B. E. (2012). Can two psychotherapy process measures be dependably rated simultaneously? A generalizability study. *Journal of Counseling Psychology*, 59, 638–644.
- Waltz, J. A., 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. doi:10.1037/0022-006X.61.4.620
- Webb, C. A., DeRubeis, R. J., & Barber, J. P. (2010). Therapist adherence/competence and treatment outcome: A meta-analytic review. *Journal of Consulting and Clinical Psychology*, 78, 200–211. doi:10.1037/a0018912
- Weck, F. (2014). Treatment of mental hypochondriasis: A case report. *Psychiatric Quarterly*, 85, 57–64. doi:10.1007/s11126-013-9270-6
- Weck, F., Bohn, C., Ginzburg, D. M., & Stangier, U. (2011). Treatment integrity: Implementation, assessment, evaluation, and correlations with outcome. *Verhaltenstherapie*, 21, 99–107. doi:10.1159/000328840
- Weck, F., Gropalis, M., Neng, J. M. B., & Witthöft, M. (2013). The German version of the H-YBOCS for the assessment of hypochondriacal cognitions and behaviors: Development, reliability and validity. *International Journal of Behavioral Medicine*, 20, 618–626. doi:10.1007/s12529-012-9276-8
- Weck, F., Hautzinger, M., Heidenreich, T., & Stangier, U. (2010). Erfassung psychotherapeutischer Kompetenz: Validierung einer deutschsprachigen Version der Cognitive Therapy Scale. [Assessing psychotherapeutic competencies: Validation of a German version of the Cognitive Therapy Scale.]. *Zeitschrift für Klinische Psychologie Und Psychotherapie*, 39, 244–250. doi:10.1026/1616-3443/a000055
- Weck, F., Hilling, C., Schermelleh-Engel, K., Rudari, V., & Stangier, U. (2011). Reliability of adherence and competence assessment in cognitive behavioral therapy: Influence of clinical

- experience. *Journal of Nervous and Mental Disease*, 199, 276–279. doi:10.1097/NMD.0b013e3182124617
- Weck, F., Neng, J. M. B., & Stangier, U. (2013a, May). *Effektivität von Kognitiver Therapie vs. Expositionstherapie bei Hypochondrie: Erste Ergebnisse einer randomisierten kontrollierten Studie [Efficacy of cognitive therapy vs exposure therapy for hypochondriasis: Preliminary results of a randomized controlled trial]*. Paper presented at the Congress of the German Psychological Society, Trier, Germany.
- Weck, F., Neng, J. M. B., & Stangier, U. (2013b). The effects of attention training on the perception of bodily sensations in patients with hypochondriasis: A randomized controlled pilot trial. *Cognitive Therapy Research*, 37, 514–520. doi:10.1007/s10608-012-9482-3
- Weck, F., Rudari, V., Hilling, C., Hautzinger, M., Heidenreich, T., Schermelleh-Engel, K., & Stangier, U. (2013). Relapses in recurrent depression 1 year after maintenance cognitive-behavioral therapy: The role of therapist adherence, competence, and the therapeutic alliance. *Psychiatry Research*, 210, 140–145. doi:10.1016/j.psychres.2013.05.036
- Young, J., & Beck, A. T. (1980). *Cognitive Therapy Scale rating manual*. Unpublished manuscript. Philadelphia, PA: Center for Cognitive Therapy.

Received 23 January 2014; revised version received 29 June 2014