

Der „Fragebogen zur Evaluation von Psychotherapieverläufen“ (FEP-2): Validierungen und Manual

Wolfgang Lutz & Jan R. Böhnke

Zusammenfassung

Qualitätssicherungsmaßnahmen und das kontinuierliche Monitoring von Veränderungen bei psychotherapeutischen wie psychiatrischen Interventionen werden immer üblicher. Zur kontinuierlichen Messung des therapeutischen Fortschrittes sind Messinstrumente nötig, die verschiedene Vorbedingungen erfüllen. Da sie häufig erhoben werden, sollten sie kurz und kostengünstig sein. Trotzdem sollten sie aber auch ein breites Spektrum relevanter Aspekte erfassen und möglichst veränderungssensitiv sein. Der „Fragebogen zur Evaluation von Psychotherapieverläufen“ (FEP-2) ist ein solcher Bogen: er misst 4 Dimensionen therapeutischer Outcomes (Wohlbefinden, Symptombelastung, Inkongruenz und interpersonelle Probleme) mit 40 Items und ist außerdem unter einer „creative commons“-Lizenz verfügbar. In dieser Arbeit werden die Ergebnisse von Validierungsstudien vorgestellt sowie eine Dokumentation zur Verfügung gestellt, die die Verwendung des Fragebogens ermöglicht.

Schlüsselwörter: Veränderungsmessung, Qualitätssicherung, Fragebogenentwicklung, Therapieverlaufsforschung

The „Questionnaire for the Evaluation of Treatment Progress“ (FEP-2): Validation and user guide

Abstract

Quality assurance and the continuous monitoring of change and progress in psychotherapeutic and psychiatric interventions are becoming common practice. To be able to assess change in practice, special instruments are needed. Because of their frequent use, they should be short and affordable. However, they should cover a broad scope of relevant aspects and also be sensitive to change. The “Questionnaire for the Evaluation of Treatment Progress” (FEP-2) is such an instrument: it measures 4 dimensions of therapeutic outcome (well-being, symptom-distress, incongruence and interpersonal problems) with 40 items and is distributed under a “creative commons” licence. This article presents results from validation studies as well as a user guide for the questionnaire.

Keywords: assessment of change, quality assurance, questionnaire development, progress and outcome in psychotherapy

Möglichkeiten zur Etablierung funktionierender Wissenschaftler-Praktiker-Netzwerke und damit einem gelingenden beidseitigen Austausch stellen die laufende Evaluation psychotherapeutischer und psychiatrischer Maßnahmen sowie Qualitätssicherungsmaßnahmen dar. Die so erhobenen Daten stellen auf der einen Seite die Grundlage für eine am individuellen Fortschritt orientierten Forschung dar und liefern auf der anderen Seite konkrete Informationen über den Einzelfall, die eine weitere Auskunftquelle für den Therapeuten im Therapieprozess sind (Lambert, 2007; Lutz, 2002; Lutz, Lambert, et al., 2006; Lutz, 2007; Stulz & Lutz, 2007; Stulz, Lutz, Leach, Lucock & Barkham, 2007). Messinstrumente, die eine theoriegeleitete, ökonomische und veränderungssensitive Erhebung des Verlaufs wie der Ergebnisse der Therapie ermöglichen, sind somit eine wichtige Voraussetzung für solche Zusammenarbeiten (Barkham, et al., 2001; Borkovec, Echemendia, Ragusea & Ruiz, 2001; Howard, Moras, Brill, Martinovich & Lutz, 1996).

Eine Reihe von Übersetzungen entsprechender Instrumente aus dem englischsprachigen Raum sind mittlerweile auf Deutsch erhältlich. Lutz und Kollegen (Lutz, Tholen, Schürch & Berking, 2006) zeigen beispielsweise basierend auf bestehenden Instrumenten, wie eine Entwicklung von Kurzformen aus längeren Originalinstrumenten möglich ist. Diese könnten dann für die nötigen wiederholten Messungen im Therapieverlauf verwendet werden. Allerdings fehlen Instrumente, welche „public domain“ und damit für die klinisch-diagnostischen Anwender einfach zugänglich sind und günstig eingesetzt werden können (Lutz, et al., 2009). Ziel war es, eine multidimensionale, theoretisch fundierte und veränderungssensitive Erhebung des Therapieverlaufs zu erreichen, die gleichzeitig ökonomisch eingesetzt werden kann (Hill & Lambert, 2004). Mit dem „Fragebogen zur Evaluation von Psychotherapie-verläufen“ (FEP) werden drei zentrale klinisch-therapeutische Konzeptualisierungen therapeutischer Veränderungen erhoben: das Phasenmodell therapeutischer Veränderung (Howard, Lueger, Maling & Martinovich, 1993), das interpersonale Kreismodell (Horowitz, Rosenberg, Baer, Ureno & Villasenor, 1988; Kiesler, 1996) und das Inkongruenzkonzept (Grawe, 1998). Durch die so realisierte theoretische Breite hebt der FEP-2 sich von den bisher bestehenden Instrumenten zur Verlaufs- und Ergebnismessung ab: die Konzeptualisierungen sind in verschiedenen Therapieschulen anschlussfähig, weshalb sich der Fragebogen für viele Therapeuten eignet.

Das Phasenmodell

Das Phasenmodell therapeutischer Veränderungen (Grissom, Lyons & Lutz, 2002; Howard, et al., 1993; Howard, Lueger, Martinovich & Lutz, 1999; Stulz & Lutz, 2007) hat in Bezug auf

die Entwicklung von schulenübergreifenden Messinstrumenten im Bereich der Psychotherapie an Bedeutung gewonnen. Dieses Modell beschreibt eine Abfolge unterschiedlicher Veränderungsphasen im Verlauf des psychotherapeutischen Genesungsprozesses. Dabei werden drei aufeinanderfolgende Phasen unterschieden: Remoralisierungsphase (Wohlbefinden), Remediationsphase (Symptomatik) und Rehabilitationsphase (Funktionsfähigkeit/interpersonale Probleme). In der ersten Phase der Behandlung, der Remoralisierungsphase, verbessert sich das subjektive Wohlbefinden des Patienten. Positive Erwartungen und Hoffnungen ermöglichen es dem Patienten, sich frühzeitig in der Therapie besser zu fühlen und die Demoralisierung, welche ihn in die Therapie geführt hat, zu überwinden. Für manche Patienten kann dies bereits ausreichen, um eigene Bewältigungsmöglichkeiten zu aktivieren. Sie sind dann in der Lage, die Probleme, die sie in die Therapie brachten, zu lösen. Andere Patienten wechseln danach in eine zweite Phase der Therapie oder beginnen die Therapie gleich in dieser Phase, der Remediationsphase. In dieser Phase sind die Verbesserung der Symptomatik und/oder eine Lösung der Lebensprobleme, wegen derer die Patientin oder der Patient die Therapie ursprünglich aufgesucht hat, Ziel der Therapie. Ein Teil der Patienten wird danach noch in die sogenannte Rehabilitationsphase übergehen. Als Ziel dieser Phase kann die Wiederherstellung und/oder Verbesserung des allgemeinen Funktionsniveaus, das Verlernen dysfunktionaler Muster, kognitiver Schemata und Verhaltensweisen, sowie das Erlernen neuer Rollen bzw. eine Verbesserung interpersonaler Probleme gesehen werden (Grawe, 1998; 2004; Howard, et al., 1996; Howard, et al., 1999).

Neben diesem Modell gibt es weitere klinisch-therapeutische Konzeptualisierungen, die im Rahmen einer multidimensionalen Erhebung des Therapiefortschrittes einen wichtigen Beitrag liefern können. Dazu gehören Maße der interpersonalen Problemlage (Horowitz, et al., 1988; Kiesler, 1996) sowie Maße zur Erhebung der Inkongruenz bzw. zur Realisierung der wichtigsten motivationalen Ziele der Patienten (Grawe, 1998; 2004; Grosse Holtforth, Grawe & Tamcan, 2004).

Interpersonale Probleme

Interpersonalen Verhaltensweisen stellen eine wichtige Ergebnisdimension von Psychotherapie dar und werden in zahlreichen Studien berücksichtigt (Lutz & Grawe, 2006). Hierbei wird insbesondere das auf Leary (1957) zurückgehende „Circumplex-Modell interpersonalen Verhaltens“ diskutiert. Dieses Kreismodell ordnet interpersonale Verhaltensweisen und Probleme in einem zweidimensionalen semantischen Raum an, der durch zwei orthogonal zueinander stehende Achsen definiert wird: eine horizontale Achse mit der Dimension „Affiliation“

und den beiden Polen „Liebe/Zuneigung“ und „Hass/Ablehnung“ und eine vertikale Achse mit der Dimension „Kontrolle“ und den beiden Polen „Dominanz / Autonomie“ und „Submission / Abhängigkeit“ (Fiedler, 2001). Interpersonale Probleme werden im Phasenmodell konzeptionell auf der Veränderungsstufe „Rehabilitation“ gesehen (Howard, et al., 1996; Lutz & Grawe, 2006).

Inkongruenz

Inkongruenz wird von Grawe (1998; 2004) als Ausmaß der Nichtbefriedigung motivationaler Ziele beschrieben. Um eigene Grundbedürfnisse zu befriedigen bzw. sich vor Frustrationen zu schützen, entwickeln Menschen unter dem Einfluss konkreter Lebensbedingungen spezifische motivationale Ziele, die einen zentralen Einfluss auf ihr konkretes Verhalten ausüben. Eine hohe Inkongruenz wird dabei als eine wesentliche Ursache für die Entstehung und Aufrechterhaltung psychischer Störungen angesehen und umgekehrt können psychische Störungen auch Ursachen einer wahrgenommenen Inkongruenz sein. Entsprechend kann die Veränderung der Inkongruenz ebenfalls als ein wichtiges therapeutisches Ziel und untersuchbare Ergebnisdimension von Psychotherapie betrachtet werden. Da sich das Konzept auf verschiedene Bereiche bezieht, wird eine gewisse Überlappung mit den übrigen Dimensionen erwartet. So kann Inkongruenz im Bezug auf das Wohlbefinden oder die Beschwerden, jedoch auch im Zusammenhang mit interpersonalem Erleben und Verhalten auftreten.

Der Fragebogen zur Erhebung von Psychotherapieverläufen (FEP-2)

Ziel bei der Entwicklung war es, multidimensional, aber ökonomisch den Verlauf von Psychotherapie abzubilden. Daher wurde angestrebt, die Anzahl der Items niedrig zu halten, vergleichbar zu anderen Instrumenten (Barkham, et al., 2001; Grissom, et al., 2002; Lambert, et al., 2004; Lutz, Tholen, et al., 2006). Die Basis der Itemgenerierung waren Vorstudien, in denen Items aus veränderungsrelevanten Bereichen aus der Literatur abgeleitet wurden (Lutz, Tholen, et al., 2006). Hauptkriterien waren die Veränderungssensitivität sowie die inhaltliche Zuordnung zu relevanten Störungsbildern im ambulanten Setting.

Skalenkonstruktion

Die drei Phasen des Phasenmodells werden im FEP-2 durch 40 Fragen repräsentiert, die sich wiederum zu vier theoretischen Skalen zusammenfassen lassen: Die Skala Wohlbefinden (sieben Items) bildet die Veränderungen in der „Remoralisierungsphase“ ab. Gemäß Phasen-

modell handelt es sich hier um Fragen des subjektiven Wohlbefindens, bei denen Verbesserungen relativ frühzeitig im Therapieverlauf erwartet werden.

Veränderungen in der „Remediationsphase“ werden von der Skala Beschwerden erfasst, welche elf Items zu typischen Symptomen insbesondere der Angst- und affektiven Störungen enthält. Diese beiden Störungsgruppen haben die höchste Prävalenz in der Bevölkerung und entsprechend auch in den Behandlungseinrichtungen, für welche der Fragebogen gedacht ist (Wittchen, Nelson & Lachner, 1998; Wittchen, Lieb, Wunderlich & Schuster, 1999). Aus diesem Grund wurden die Fragen auf die entsprechenden Symptome beschränkt, um diese häufigen Beschwerden gut abbilden zu können. Für die beiden Symptomgruppen können optional zwei Subskalen (Ängstlichkeit und Depressivität) berechnet werden.

Für Patienten, welche die „Rehabilitationsphase“ erreichen, werden zu einem späteren Zeitpunkt im Therapieverlauf Veränderungen auf der Skala Interpersonale Beziehung (zwölf Items) erwartet. Ziel dieser Skala ist es, interpersonale Schwierigkeiten gemäß dem Kreis- bzw. Circumplexmodell zu erfassen (Horowitz, et al., 1988; Kiesler, 1996). Die Items können zu den Dimensionen „introvertiert / scheu“, „ausnutzbar / nachgiebig“, „konkurrierend / unterstützend“ und „selbstunsicher / zurückhaltend“ zusammengefasst werden, welche sich entsprechend auf dem interpersonalen Kreismodell abbilden lassen (Lutz, 2002).

Durch die Skala Kongruenz (zehn Items) wird das Inkongruenzkonzept (Grawe, 1998) operationalisiert. Es wird erhoben, inwiefern wichtige motivationale Ziele vom Patienten umgesetzt werden können. Dabei wird zwischen Annäherungs- und Vermeidungszielen (motivationale Schemata) unterschieden, die angestrebt werden oder deren Eintreten verhindert werden soll. Neben dem Skalenwert Kongruenz lassen sich deshalb auch die Dimensionen Annäherung und Vermeidung separat berechnen. Die Items zur Kongruenz können nicht eindeutig einer der drei Phasen im Phasenmodell zugeordnet werden. Vielmehr wird angenommen, dass sich einzelne Aspekte der (In-)Kongruenz früh verändern (z.B. Inkongruenz bezüglich Lebensqualität), andere erst im späteren Verlauf (z.B. Inkongruenz bezüglich interpersonaler Faktoren).

Skalenauswertung

Die Skalen des FEP-2 werden durch die Addition der einzelnen Itemwerte und anschließende Division durch die Anzahl bearbeiteter Items gebildet. Dabei empfehlen wir, dass ein Skalenwert nur dann gebildet werden sollte, wenn mindestens 80 % der Items beantwortet sind. Bei den aus wenigen Items zusammengesetzten Skalen wird dadurch eine Mittelwertbildung aus einer Auswahl von Items verhindert, die nicht repräsentativ für das Belastungsbild des Patien-

ten ist. Vergleichbare Instrumente verwenden ähnliche Kriterien (Lutz & Howard, 1997). Neben den vier Subskalen kann nach den gleichen Regeln auch ein Gesamtwert psychischer Beeinträchtigung gebildet werden. Für differenziertere Auswertungen können die Subskalen Beschwerden, interpersonale Probleme und Kongruenz in die jeweiligen inhaltlichen Bereiche aufgeteilt werden. Die Beantwortung der Items erfolgt auf einer fünfstufigen Likertskala (von 1 = nie bis 5 = sehr oft). Von den vierzig Items sind 13 Items positiv formuliert (1, 2, 5, 9, 14, 15, 19, 21, 22, 27, 29, 30 und 39) und müssen vor der Skalenberechnung umgepolt werden (Umkodierung folgendermassen: 1 = 5; 2 = 4 ; 3 = 3; 4 = 2; 5 = 1). Höhere Werte bedeuten dann jeweils eine stärkere Belastung.

Die psychische Gesamtbelastung wird als Summe der Antworten aller 40 Items geteilt durch die Anzahl der ausgefüllten Items berechnet. Sofern weniger als 32 Items ausgefüllt wurden, kann der Wert nicht mehr als verlässlich angesehen werden (80 %-Regel; s.o.).

Der Gesamtwert „Wohlbefinden“ berechnet sich aus der Summe aller Items der „Wohlbefinden“-Subskala (1, 5, 9, 14, 22, 30, 39) geteilt durch die Anzahl der ausgefüllten Items. Sofern weniger als sechs Items ausgefüllt wurden, sollte der Wert nicht mehr berechnet werden.

Der Gesamtwert für „Beschwerden / Allgemeine Symptombelastung“ ergibt sich als Summe der zu dieser Subskala gehörenden Items (4, 6, 7, 8, 10, 12, 23, 24, 31, 34, 38) geteilt durch die Anzahl der ausgefüllten Items. Sofern weniger als neun Items ausgefüllt wurden, ist dieser Wert als fehlend zu betrachten. Hier können auch die Subskalen Ängstlichkeit (Items 4, 8, 12 und 23; alle Items müssen ausgefüllt sein) und Depressivität (Items 7, 10, 24, 31, 34, 38; fünf Items müssen ausgefüllt sein) berechnet werden

Bei der Skala „Beziehung“ lässt sich zunächst auch ein Gesamtwert „Interpersonale Probleme“ bestimmen als Summe aller zugeordneten Items (11, 13, 16, 18, 20, 25, 26, 28, 32, 33, 36, 40) geteilt durch die Anzahl der ausgefüllten Items. Sofern weniger als zehn Items ausgefüllt wurden, ist der Gesamtwert als fehlend zu betrachten. Gemäß des Kreismodells lassen sich noch vier Subskalen bilden, bei denen jeweils alle Items ausgefüllt sein müssen, sonst gilt dieser Subskalenwert als fehlend (80 %-Regel, s.o.). Die Skalenwerte werden jeweils als Summe geteilt durch die Anzahl der beantworteten Items berechnet. Die erste Skala ist „introvertiert/scheu“ (Items 28, 36, 40), die zweite Skala „ausnutzbar/nachgiebig“ (Items 13, 20, 26), die dritte Skala „konkurrierend/unterstützend“ (Items 11, 16, 33) und die vierte Skala „selbstunsicher/zurückhaltend“ aus den Items 18, 25 und 32.

Als letzte Dimension bleibt die Skala „Kongruenz“. Bei ihr kann zunächst ein Gesamtwert als Summe der Antworten auf alle zugeordneten Items (2, 3, 15, 17, 19, 21, 27, 29, 35, 37) geteilt durch die Anzahl der ausgefüllten Items berechnet werden. Sofern weniger als acht Items

ausgefüllt wurden, ist der Gesamtwert als fehlend zu betrachten. Für die Dimension Kongruenz gibt es zwei Subskalen. Die erste fasst die „Annäherungsziele“ (Items 2, 15, 19, 21, 27, 29; alle müssen ausgefüllt sein) und die zweite die „Vermeidungsziele“ (Items 3, 17, 35, 37; alle müssen ausgefüllt sein). Für beide Subskalen wird der Wert als Summe geteilt durch die Zahl der Items berechnet.

Im Folgenden werden die Ergebnisse mehrerer Untersuchungen berichtet, die die Güte und Struktur des Fragebogens untersuchten.

Psychometrische Kennwerte des „Fragebogens zur Evaluation von Psychotherapieverläufen“

Studie I (Lutz et al., 2009)

Diese Studie sollte die Umsetzung der dargelegten grundlegenden Prinzipien des Fragebogens untersuchen. Neben der Validierung der einzelnen Konstrukte sollten cut offs für ambulante vs. nicht-belastete Patienten und der reliable Veränderungsindex (reliable change index, Jacobson & Truax, 1991)) berechnet werden. Dazu wurden die Daten einer nicht-klinischen, einer studentischen und einer ambulanten Therapiepatienten-Stichprobe verwertet. Die Ergebnisse sind im Detail bei (Lutz, et al., 2009) nachzulesen.

Nicht-klinische Stichprobe

Für die nicht-klinische Stichprobe lagen Daten von 274 Personen vor, davon 53.3 % Frauen. Diese Stichprobe wurde im Rahmen von mehreren Diplomarbeiten am Institut für Psychologie der Universität Bern erhoben (Boixet, 2005; Jermann, 2006). Das Durchschnittsalter der geschichteten Stichprobe betrug 35.3 Jahre ($SD = 12.9$) bei einem Range von 15 bis 74 Jahren. 54.4 % gaben an ledig zu sein, 40.1 % waren verheiratet. 40.9 % hatten einen Gymnasialabschluss, 38 % einen Sekundarschulabschluss, die restlichen hatten einen niedrigeren oder einen anderen Abschluss. Bei den Berufsabschlüssen hatten je ein Drittel einen Universitätsabschluss (34.3 %) oder eine Lehre (30.3 %). 64.4 % hatten zum Zeitpunkt der Befragung eine Vollzeitanstellung, 32.2 % arbeiteten Teilzeit.

Studentische Stichprobe

Die studentische Stichprobe bestand aus Psychologiestudierenden und wurde im Rahmen einer obligatorischen Veranstaltung erhoben. Der Fragebogen wurde an alle Anwesenden verteilt, ca. 1 % der Studierenden lehnte eine Teilnahme ab. 96 Personen haben einen ausge-

füllten Bogen abgegeben, davon 81.5 % Frauen. Das Durchschnittsalter betrug 25.3 Jahre ($SD = 5.3$) bei einer Bandbreite von 20 bis 62 Jahren. Die meisten waren noch ledig (92.7 %) und gaben als Beruf Student/Studentin an (91.7 %).

Ambulante Stichprobe

Die Daten der ambulanten Stichprobe ($N = 184$) wurden in der Poliklinischen Psychotherapieambulanz der Universität Osnabrück erhoben. Der Fragebogen wird dort von allen Patienten einmal pro Quartal ausgefüllt. Der Frauenanteil betrug 61.6 %. Das Durchschnittsalter lag in dieser Stichprobe bei 37.4 Jahren ($SD = 12.3$, 17 bis 72 Jahre). 53.4 % hatten die Behandlung aufgrund einer affektiven Störung aufgesucht, bei 23.3 % lag eine Angststörung vor (nach ICD-10). Bei 10 % lag eine Diagnose für eine komorbide Störung vor.

Untersuchungsinstrumente zur Prüfung der externen Validität

Zur Prüfung der konvergenten Validität wurden vier Fragebogen hinzugezogen, die als valide Messinstrumente für die intendierten Dimensionen gelten und bereits in zahlreichen Studien zum Einsatz kamen. Das Beck Depressions-Inventar (BDI, Hautzinger, Bailer, Worall & Keller, 1994) ist ein Selbstbeurteilungsinstrument zur Erfassung der Schwere depressiver Symptomatik mit 21 Items. Es ist vielfältig in Psychotherapieergebnisstudien eingesetzt worden. Die psychometrischen Kennwerte wurden anhand unterschiedlicher klinischer Stichproben ermittelt und zeigen hohe Werte bezüglich Reliabilität und Validität.

Der OQ-45 (Lambert, et al., 2004) misst den Therapiefortschritt der Patienten und kann zur Verlaufs- und Ergebnismessung eingesetzt werden. Der Fragebogen hat insgesamt 45 Items und deckt die Bereiche Symptombelastung (25 Items), Interpersonale Beziehungen (elf Items) und soziale Rolle (neun Items) ab. Nebst den drei Skalen kann auch ein Gesamtwert der allgemeinen psychischen Beeinträchtigung gebildet werden. Der OQ-45 ist veränderungssensitiv und hat in der englischen Originalversion eine hohe Reliabilität (Cronbachs Alpha für den Gesamtwert: .93) und Validität (Korrelationen mit Angst- und Depressionsskalen von .64 bis .88) (Lambert, et al., 2004). Verwendet wurde die deutsche Version (Lambert, Hannover, Nisslmüller, Richard & Kordy, 2002), die ähnliche Werte aufweist.

Das Treatment Evaluation and Management System TEaM (Grissom, et al., 2002) wird vor allem im US-amerikanischen Raum neben dem OQ-45 (Lambert, et al., 2004) als Instrument zur Evaluation von Psychotherapie herangezogen. Der TEaM-Fragebogen beinhaltet insgesamt 58 Items und erfasst dabei die drei Dimensionen des Phasenmodells therapeutischer Veränderungen (Howard, et al., 1993): subjektives Wohlbefinden (drei Items), Symptomatik

(29 Items) und Funktionsbeeinträchtigung (13 Items). Die einzelnen Skalen weisen in der englischen Originalversion interne Konsistenzen (Cronbachs Alpha) von .82 (Wohlbefinden) bis .90 (Symptome) und für den Gesamtwert von .88 auf (Grissom, et al., 2002). Diese konnten auch in der deutschen Übersetzung erhoben an einer 197 Personen umfassenden studentischen Stichprobe approximativ erhalten werden (zwischen .75 in der Skala Wohlbefinden und .90 in der Skala Symptome, Gesamtwert = .91).

Für die Abgrenzung von somatischen Beschwerden (diskriminante Validität) wurde der Fragebogen zum Gesundheitszustand SF12 (Bullinger & Kirchberger, 1998) erhoben. Er ist ein Instrument zur subjektiven Selbsteinschätzung der Gesundheit. Es handelt sich dabei um eine Kurzform des SF-36 mit den beiden Bereichen körperliche und psychische Gesundheit. Dabei misst der SF-12 gemäß amerikanischen Studien (Bullinger & Kirchberger, 1998) dasselbe Spektrum körperlicher und psychischer Gesundheit wie sie mit dem SF-36 erfasst wird. Die deutsche Version wurde an verschiedenen Stichproben mit insgesamt über 4000 Fällen normiert. Aufgrund der verhaltensnahen Formulierung der Items ist das Instrument eher als Maß der subjektiven Gesundheit aus Patientensicht zu verstehen.

Ergebnisse

Der varianzanalytische Vergleich der Werte der drei Stichproben zeigte, dass sich die klinische Stichprobe auf allen Skalen sowie dem Gesamtwert des FEP-2 signifikant von den beiden nicht-klinischen Stichproben unterscheidet und höhere Belastungswerte aufweist (kontrolliert für Alter und Geschlecht). Auf der Skala Wohlbefinden sowie beim Gesamtwert unterscheiden sich zusätzlich die Studierenden signifikant von der nicht-klinischen Stichprobe aus der allgemeinen Bevölkerung. Die Reliabilität wurde durch die interne Konsistenz (Cronbachs Alpha) und die Test-Retest-Korrelation geprüft. Die internen Konsistenzen sind mit Werten zwischen .74 und .94 durchwegs hoch (Lutz, et al., 2009).

Die Retest-Korrelation des FEP-2 wurde nur für die beiden nicht-klinischen Stichproben ermittelt. Für die zusammengefügte Stichprobe ($N = 269$) ergab sich über einen Zeitraum von bis zu acht Wochen (mittlerer Abstand zwischen den Messungen = 3.1 Wochen, $SD = 1.4$) ein Wert von $r_{tt} = .69$ für den Gesamtwert psychische Beeinträchtigung. Für die Studierendenstichprobe allein wurde ein engeres Retest-Zeitintervall definiert. Für diese Teilstichprobe ($N = 59$) ergibt sich für ein Intervall von einer Woche ein Wert von $r_{tt} = .77$ für den Gesamtwert. Um die Validität des „Fragebogens zur Evaluation von Psychotherapieverläufen“ (FEP-2) zu prüfen, wurden die Subskalen sowie der Gesamtwert der nicht-klinischen Bevölkerungsstichprobe mit den Instrumenten (BDI, SF12, TEaM, OQ-45) korreliert. Die Korrelationen des

Gesamtwertes des FEP-2 liegen mit den Gesamtwerten für TEaM, SF12 Psyche und OQ-45 allesamt in einem akzeptablen Rahmen (zwischen $r = .71$ bei SF12 Psyche und $r = 0.81$ bei OQ-45). Auch die FEP-2-Subskalen Wohlbefinden und Beschwerden korrelieren jeweils am deutlichsten mit den korrespondierenden Prüfinstrumenten ($r = .74$ mit Symptome TEaM und $r = .77$ Symptome OQ-45). Die durchweg niedrige Korrelation des FEP-2 und seiner Subskalen mit dem SF12-Körper ($r = .13$ mit FEP-2-Beschwerden und Beziehung bis $r = .22$ mit FEP-2-Wohlbefinden) zeigen, dass der FEP-2 nicht primär auf die körperliche, sondern lediglich die psychische Symptomatik abzielt. Zusätzlich fällt auf, dass die FEP-2-Subskala Beziehung die niedrigsten Korrelationen mit den Prüfskalen hat (.39 bis .61, ohne SF12 Körper) und ihre höchsten Korrelate sind nicht TEaM Funktion, OQ-45 Interpersonal und Soziale Rolle (Lutz, et al., 2009).

Dies zeigt, dass mit dem FEP-2 Informationen in verschiedenen Bereichen erhoben werden können, wie dies mit den verschiedenen Dimensionen intendiert wurde. Der Gesamtwert, gebildet aus den verschiedenen Bereichen, erlaubt zudem die Abbildung des allgemeinen psychischen Wohlbefindens.

Veränderungssensitivität

Ein Instrument zur Verlaufsmessung muss sich als veränderungssensitiv erweisen. Dies wurde anhand von Effektstärken zwischen Vorher- und Zwischenerhebungsmessungen sowie Vorher- und Abschlussmessungen geprüft (Mittelwertsdifferenzen geteilt durch die Streuung zu Therapiebeginn). Zur Effektstärkenberechnung wurden Teilstichproben der klinischen Stichprobe verwendet, für die bereits Zwischenerhebungen bzw. Abschlussmessungen vorlagen (Personen mit 10-20 Sitzungen bei der Zwischenmessung, $N = 144$; Personen mit 10-76 Sitzungen bei der Abschlussmessung, $N = 45$). Dabei zeigt sich, dass bei den Patienten zur Zwischenerhebung in der Wohlbefindens-Skala (Effektstärke / $ES = .47$) und der Kongruenz-Skala ($ES = .44$) bereits eine deutliche Verbesserung stattfindet und die Beschwerden sich leicht verbessern ($ES = .27$), während die Skala Beziehung bezüglich dieser frühen Zwischenmessung weitgehend unverändert bleibt ($ES = .07$). Dieses Muster ist mit den erwarteten Veränderungen nach dem Phasenmodell therapeutischer Veränderung konsistent, wonach sich zu Therapiebeginn zunächst das Wohlbefinden verbessert (Remoralisierung), die Symptomreduktion (Remediation) und die Verbesserung des Funktionsniveaus in relevanten Lebensbereichen (Rehabilitation) sich dagegen erst in einer späteren Therapiephase abspielt (Howard, et al., 1999). Die bereits früh gebesserten Bereiche haben auch bei den beendeten Therapien die stärksten Veränderungen. Die Effektstärken nehmen insgesamt stark zu (Wohlbefinden,

$ES = 1.66$; Beschwerden, $ES = 0.90$; Kongruenz $ES = 1.80$) und der Bereich interpersonale Beziehung kommt für beendete Therapien auf eine Effektstärke von 0.5. Dies erklärt sich dadurch, dass ein Teil der Patienten mit abgeschlossener Therapie die Rehabilitationsphase im Phasenmodell erreicht hat und sich entsprechend in relevanten Lebensbereichen verbessern konnte. Werden die Effektstärken für die nicht-klinische Stichprobe berechnet, ergeben sich für die einzelnen Skalen Werte zwischen .13 und .27. Diese Werte sind deutlich niedriger als für die abgeschlossenen Therapien.

Faktorenstruktur

Vier theoretische Faktoren (Wohlbefinden, Beschwerden, Interpersonale Beziehung und Inkongruenz) sind im Fragebogen angelegt. Zur Überprüfung der Faktorenstruktur wurde eine Hauptkomponentenanalyse mit der zusammengeführten Gesamtstichprobe ($N = 554$) durchgeführt. In der unrotierten Lösung ergab sich eine Hauptkomponente mit 42 % Varianzaufklärung, was die Berechnung eines Gesamtwerts rechtfertigt. Da eine starke Komponente „allgemeine Psychopathologie“ in der Regel die differenziertere theoretische Konzeption überlagert (Halstead, Leach & Rust, 2007; Lutz, et al., 2009), kann nicht von einer vollständigen Unabhängigkeit der vier Bereiche des Fragebogens ausgegangen werden. In einer konfirmatorischen Faktorenanalyse wurde getestet, ob sich vier korrelierte Faktoren rechtfertigen lassen. Dazu wurden die vier Konstrukte Wohlbefinden, Beschwerden, Interpersonale Probleme und Inkongruenz als latente Variablen modelliert, die alle mit einander korreliert waren und jeweils nur ihre Skalenitems als Indikatoren hatten (konfirmatorische Faktorenanalyse; z.B. Loehlin, 2004). Die Ergebnisse der mit Mplus (Muthén & Muthén, 1998-2004) durchgeführten Analyse sind in Tabelle 1 zusammengefasst. Beide Modelle passen nach dem reinen Modelltest nicht (1 Faktor: $\chi^2(df = 702; N = 480) = 2639.41, p < 0.001$; vier Faktoren: $\chi^2(df = 696; N = 480) = 2083.41, p < 0.001$), doch die in Tabelle 1 aufgeführten vergleichenden Fit-Indizes zeigen eine deutliche Verbesserung der Modellpassung durch die Annahme der vier postulierten Faktoren.

Abbildung 1 verdeutlicht durch die Angabe der Korrelationen zwischen den vier latenten Konstrukten, dass sie nur bedingt als einzelne Entitäten gesehen werden können. Insbesondere die Inkongruenz korreliert stark mit den anderen drei Konstrukten.

Aus diesen Ergebnissen (und der in Lutz et al., 2009 berichteten explorativen Faktorenanalyse) schließen wir, dass die verlässlichste Aussage über den Zustand des Patienten durch den Gesamtwert getroffen wird und die vier Subdimensionen brauchbare Zusatzinformationen liefern können.

Tabelle 1
Fit-Indizes der Ein- und Vier-Faktoren-Lösung für den FEP-2

	ein Faktor	vier Faktoren
χ^2/df	3.76	2.99
CFI	0.83	0.88
AIC // BIC	46776.90 // 47102.78	46232.90 // 46583.85
RMSEA [90 % Konfidenzintervall]	0.08 [0.07; 0.08]	0.06 [0.06; 0.07]
SRMR	0.06	0.05

Anmerkungen. Item 35 wurde ausgelassen, da es zu hohe Korrelationen mit verschiedenen Faktoren aufwies. $N = 480$.

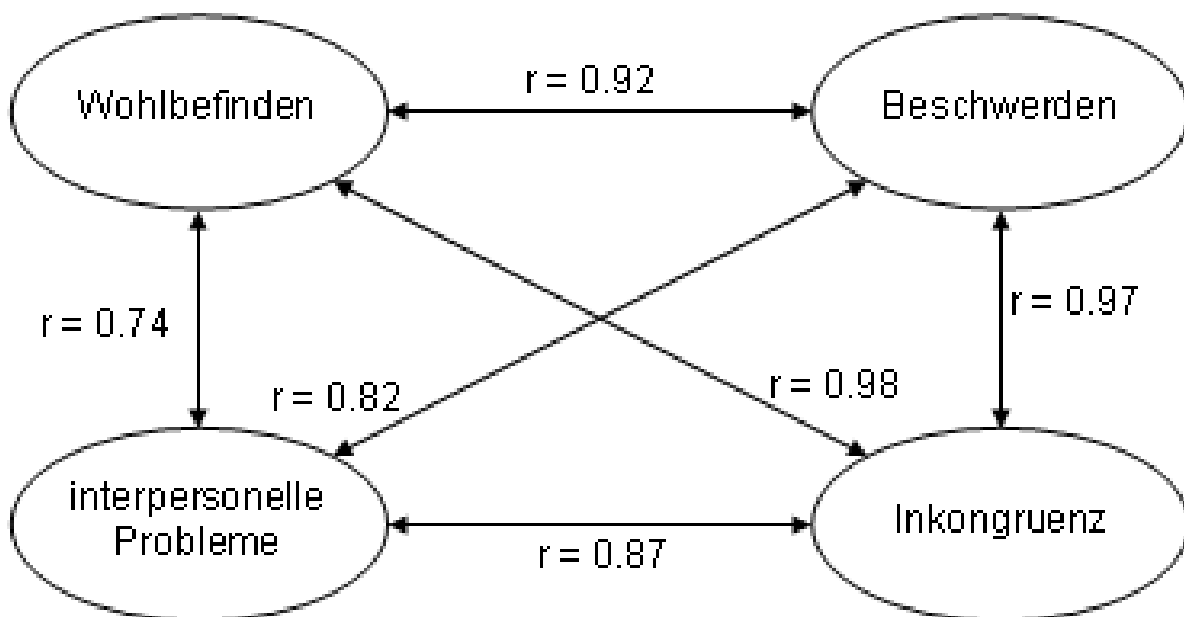


Abbildung 1. Korrelationen zwischen den latenten Faktoren in der konfirmatorischen Faktorenanalyse des FEP-2 (ohne Item 35; s. Anm. Tabelle 1).

Reliable und klinisch bedeutsame Veränderung

Für den Einsatz bei der Messung von Veränderungen ist es notwendig, dass das Instrument ermitteln kann, ob und inwieweit sich der Patient zwischen Therapiebeginn und Sitzung t verbessert, nicht verändert oder verschlechtert hat. Dazu kann entweder der absolute Therapiefortschritt in Relation zum Messfehler bzw. der Reliabilität des Instrumentes betrachtet werden, oder aber die Verbesserung im Vergleich zur nicht-klinischen Stichprobe. Für ersteres

hat sich die Berechnung der klinisch bedeutsamen Veränderung in der Literatur durchgesetzt (RC, reliable change oder reliable Veränderung, z.B. Jacobson & Truax, 1991; Wise, 2004). Eine Veränderung gilt dann als reliabel, wenn die absolute Veränderung in Rohwerten größer ist als der RC. Eine Beschwerdenzunahme wird als reliabel verschlechtert beurteilt, während eine Beschwerdenabnahme im gleichen Umfang als reliable Verbesserung gilt. Die RCs reichen von 0.72 (Gesamtwert) bis 1.20 (Wohlbefinden; vgl. auch Tabelle 2).

Für die klinisch relevante Veränderung sollte zusätzlich der Patientenrohwert wahrscheinlicher aus der nicht-klinischen Population stammen. Den cut-off Wert bildet dabei die Grenze zwischen klinischer und Normpopulation. Werte unterhalb des cut-off gehören mit größerer Wahrscheinlichkeit zur nicht-klinischen als zur klinischen Stichprobe (Jacobson & Truax, 1991). Hat der Rohwert eines Patienten den cut-off unterschritten, gilt er als nicht mehr klinisch bedeutsam beeinträchtigt. Vorläufige Ergebnisse ohne die Erhebung einer repräsentativen Normstichprobe zeigen, dass bei der Gesamtskala ein Mittelwert von 2.45 als cut-off zwischen auffälligen Werten (oberhalb dieses Wertes) und nicht-auffälligen herangezogen werden kann (vgl. auch Tabelle 2).

Studie II (Schürch et al., in press)

Ein salientes Thema, das aus dieser ersten Untersuchung hervorgeht, ist die Frage der Dimensionalität. In einer Nachfolgearbeit (Schürch, Lutz & Böhnke, in press) wurde an derselben Stichprobe wie in der ersten Studie (Lutz, et al., 2009) untersucht, ob und wie sich problematische Antwortmuster mit der Hilfe des Rasch-Modells identifizieren lassen.

Das dazu verwendete Rasch-Modell ist ein Item Response Modell, das annimmt, dass alle Items eines Tests auf einer Dimension (hier oftmals als „Fähigkeit“ bezeichnet) messen und sich lediglich in ihrer Schwierigkeit unterscheiden, nicht aber in ihrer Trennschärfe (Rost, 2004). In der Arbeit wird gezeigt, dass sich nahezu alle vierzig Items auf einer Dimension anordnen lassen, die einen allgemeinen Belastungsfaktor misst. Lediglich fünf der Items scheinen (auch) andere Merkmale der Patienten zu messen. Diese Items sind „hatte ich Schlafprobleme“ (Item 6), „war ich selbstbeherrscht“ (Item 19), „gingen mir die Probleme anderer Menschen schnell zu nahe“ (Item 26), „war ich Teil einer erfüllten und intimen Beziehung“ (Item 27) und „mischte ich mich zu sehr in die Angelegenheiten anderer ein“ (Item 33). Dass sich diese fünf Items nicht auf der Dimension anordnen lassen, kann unterschiedliche Gründe haben. Plausibel für vier dieser Items ist, dass sie nicht (nur) die Dimension „Allgemeine Psychopathologie“ ansprechen. Schlafstörungen können auch andere Gründe als psychosoziale Belastung haben und die drei anderen Items könnten deutlicher durch die Person-Umwelt-Interaktion geprägt sein als das für die anderen Items des Tests der Fall ist.

Tabelle 2
Psychometrische Kennwerte des FEP-2 aus der neuen Stichprobe und zum Vergleich der Ergebnisse von Lutz et al. (2009).

Kennzahl	Datengrundlage	
	Lutz et al. (2009)	neues Datenset
α Wohlbefinden	0.92	0.87
α Beschwerden	0.92	0.92
α Beziehung	0.86	0.86
α Inkongruenz	0.86	0.83
α Gesamtwert	0.96	0.96
cut-off Wohlbefinden	2.90	2.93
cut-off Beschwerden	2.39	2.69
cut-off Beziehung	2.15	2.24
cut-off Inkongruenz	2.58	2.60
cut-off Gesamt	2.45	2.49
RC Wohlbefinden	1.20	1.15
RC Beschwerden	0.89	0.97
RC Beziehung	0.75	0.87
RCI Inkongruenz	0.92	0.94
RCI Gesamt	0.72	0.75

Anmerkungen. cut-off = $(s1*M2+s2*M1)/(s1+s2)$; RC = $1.96*S_{diff}$; $S_{diff} = \sqrt{2*SE^2}$;
 $SE = s1*\sqrt{(1-r_{tt})}$, SE = Standarderror, r_{tt} = test-retest Korrelation, aus Lutz et al. (2009) übernommen; s1 ist Standardabweichung der Bevölkerungsstichprobe, α = Cronbachs Alpha.

Studie III

Validierung des cut-offs für die Gesamtskala

Zur Validierung der oben dargestellten Kennwerte des FEP werden in verschiedenen Projekten derzeit neue Daten erhoben. Durch die Zusammenarbeit mit der Poliklinischen Psychotherapieambulanz Osnabrück wurde eine neue ambulante Patientenstichprobe erhoben und in einer Diplomarbeit (Jung, 2008) eine neue nicht-klinische Stichprobe erhoben.

Ambulante Stichprobe

Die verwendeten Daten von 129 Patient stammen alle von der Standardmessung des Status im ersten Quartal der Therapie. Der Frauenanteil betrug 64.3 %. Das Durchschnittsalter lag bei

41.7 Jahren ($SD = 12.2$, 19 bis 76 Jahre). Von den Patienten suchten 37.2 % die Behandlung aufgrund einer affektiven Störung auf, 17.8 % aufgrund einer Angststörung, 14.7 % aufgrund einer schweren Belastung oder Anpassungsstörung. Die restlichen Patienten litten nach ICD-10 an diversen anderen Störungen. 60.4 % wiesen mindestens eine weitere Diagnose auf.

Nicht-klinische Stichprobe

Die nicht-klinische Stichprobe ($N = 120$) ist eine Quota-Stichprobe, die im Westen der Bundesrepublik Deutschland erhoben wurde. Die zu erhebenden Quoten wurden aufgrund des scientific-use Files des Mikrozensus¹ bestimmt. Neben Geschlecht wurde nach Alter in drei Gruppen (18 bis 30 Jahre, 31 bis 50 Jahre und 51 bis 65 Jahre) und höchster allgemeiner Schulabschluß (Haupt-/ Volksschule, Realschule/Polytechnische Oberschule, Fachhochschulreife und Abitur) erhoben. Diese Stichprobe hatte einen Frauenanteil von 50 % und 95 % der Befragten waren Deutsche. Das mittlere Alter betrug 43.9 Jahre ($SD = 13.6$, 19 bis 87 Jahre). Der Familienstand der meisten war entweder ledig (37.5 %) oder verheiratet (38.3 %; 14.2 % waren geschieden). 71.6 % der Befragten lebten mit einem festen Partner zusammen (verheiratet wie unverheiratet; 16.7 % gaben an, längerfristig keinen Partner zu haben). Von den Befragten gaben 40 % an, einen Hauptschulabschluß zu haben, 31.7 % die mittlere Reife und 28.3 % das Abitur. Derzeit arbeiten 53.3 % der Befragten Vollzeit, 33.3 % Teilzeit und 9.2 % gaben an, arbeitslos zu sein.

Ergebnisse

Die Skala erweist sich weiterhin als sehr reliabel (Cronbachs Alpha von 0.96 bei $N = 220$). Als cut-off zwischen dem Spektrum der nicht-belasteten und dem Spektrum der belasteten Personen ergibt sich ein Wert von 2.49 für die Gesamtskala, was sehr nahe an dem Ergebnis der ersten Studie liegt (2.45). Der cut-off für die Abgrenzung zwischen nicht-belasteten und ambulanten PatientInnen ließ sich also replizieren. Auch die anderen psychometrischen Kennwerte lassen sich zumindest in der Größenordnung replizieren (s. Tabelle 2). Da die Kennwerte stichproben- und verteilungsabhängig sind, sind kleine Schwankungen zu erwarten, doch sind diese gering genug, um die Werte als verlässlich zu betrachten.

¹ Statistisches Bundesamt; „Campus-File“ über das Forschungsdatenzentrum der Statistischen Ämter des Bundes und der Länder (<http://www.forschungsdatenzentrum.de/>)

Diskussion

Die bisherigen Untersuchungen zeigen, dass der FEP-2 durch seine breite theoretische Fundierung mehrere zentrale Dimensionen psychotherapeutischer Veränderungen erfassen kann. Diese unterschiedlichen Komponenten psychischer Gesundheit bzw. psychischer Belastungen sind empirisch von einander abgrenzbar und die gleichzeitig angestrebte geringere Itemzahl als die Summe der verschiedenen Instrumente für die einzelnen Konzepte ermöglicht eine zeitsparende Erhebung des Status zu verschiedenen Zeitpunkten im Therapieverlauf. Sowohl eine Erfassung der einzelnen Symptombereiche, wie auch die Messung eines allgemeinen Faktors der psychischen Befindlichkeit (begründet durch die hohen mittleren Korrelationen der einzelnen Dimensionen) sind möglich. Der „public domain“-Status macht den Fragebogen darüber hinaus zu einem ökonomischen und dadurch breit einsetzbaren Instrument.

In den statistischen Analysen zeigte sich, dass der Fragebogen als Gesamtskala wie auch in den vier zentralen Subdimensionen reliabel ist. Die psychometrischen Eigenschaften des Gesamtwertes sind mit denen ähnlicher Instrumente im englischsprachigen Raum vergleichbar (Barkham, et al., 2001; Grissom, et al., 2002; Lambert, et al., 2004). Aufgrund der Ergebnisse aus den Faktorenanalysen erscheint es sinnvoll, auch die Subskalen zu den Beschwerden separat zu berechnen. Der Vergleich zwischen der ambulanten klinischen Stichprobe sowie der an der Bevölkerung erhobenen Stichprobe und der Studierendenstichprobe zeigt, dass alle Dimensionen in der Lage sind, klinische Beschwerden zu identifizieren bzw. das Instrument deutlich erhöhte Werte bei Personen mit psychischen Beschwerden aufweist als bei nicht belasteten Personen. In der Validierung dieser Befunde an zwei neuen Stichproben ließen sich diese Eigenschaften replizieren.

Besonderer Wert wurde bei der Entwicklung auf die Veränderungssensitivität gelegt. Die entsprechenden Prüfwerte (Effektstärken) zeigen vergleichbare Werte wie weitaus umfangreichere und bereits länger eingeführte Instrumente, welche aber nicht eine entsprechende Ökonomie besitzen. Zum Beispiel wurden zu den Zwischenmesszeitpunkten nach zehn Sitzungen an der psychotherapeutischen Praxisstelle an der Universität Bern für ein vergleichbares Wohlbefindensmaß (Emotionalitätsinventar, EMI-B; Ullrich de Muynck & Ullrich, 1977) eine Vorher- Zwischenhebungs-Effektstärke von .57 und nach 20 Sitzungen für ein vergleichbares Beschwerdemaß (Brief Symptom Inventory, Franke, 2000) eine Vorher-Zwischenhebungs-Effektstärken von .47 festgestellt. Dies lässt sich durchaus mit dem hier untersuchten Instrument, für das die Zwischenhebungen zwischen der zehnten und der zwanzigsten Sitzung lagen, vergleichen. Die Vorher-Nachher-Effektstärken liegen mit Werten

von deutlich über Eins beim Wohlbefinden, der Inkongruenz und dem Gesamtwert im hohen Bereich und lassen auf eine sehr gute Veränderungssensitivität bei abgeschlossenen Therapien schließen.

Literatur

- Barkham, M., Margison, F., Leach, C., Lucock, M., Mellor-Clark, J., Evans, C., Benson, L., Connell, J., Audin, K. & McGrath, G. (2001). Service profiling and outcomes benchmarking using the CORE-OM: Toward practice-based evidence in the psychological therapies. *Journal of Consulting and Clinical Psychology, 69*, 184-196.
- Boixet, S., (2005). Die Skalen Wohlbefinden und Beziehung im Fragebogen zur Evaluation von Psychotherapieverläufen FEP-2. Konstruktion und erste Überprüfung psychometrischer Kennwerte. Unveröffentlichte Lizentiatsarbeit, Universität Bern, Bern.
- Borkovec, T. D., Echemendia, R. J., Ragusea, S. A. & Ruiz, M. (2001). The Pennsylvania Practice Research Network and future possibilities for clinically meaningful and scientifically rigorous psychotherapy effectiveness research. *Clinical Psychology: Science and Practice, 8*, 155-167.
- Bullinger, M. & Kirchberger, I., (1998). *SF-36. Fragebogen zum Gesundheitszustand*. Göttingen: Hogrefe.
- Fiedler, P., (2001). *Dissoziative Störungen und Konversion*. Weinheim: Beltz.
- Franke, G., (2000). *BSI. Brief Symptom Inventory — Deutsche Version. Manual*. Göttingen: Beltz.
- Grawe, K., (1998). *Psychologische Therapie*. Bern: Hogrefe.
- Grawe, K., (2004). *Neuropsychotherapie*. Göttingen: Hogrefe.
- Grissom, G. R., Lyons, J. S. & Lutz, W. (2002). Standing on the shoulders of a giant: Development of an outcome management system based on the dose model and phase model of psychotherapy. *Psychotherapy Research, 12*, 397-412.
- Grosse Holtforth, M., Grawe, K. & Tamcan, Ö., (2004). *Inkongruenzfragebogen. Manual*. Göttingen: Hogrefe.
- Halstead, J. E., Leach, C. & Rust, J. (2007). The development of a brief distress measure for the evaluation of psychotherapy and counseling (sPaCE). *Psychotherapy Research, 17*, 656-672.
- Hautzinger, M., Bailer, M., Worall, H. & Keller, F., (1994). *Beck-Depressions-Inventar (BDI)*. Bern: Huber.
- Hill, C. E. & Lambert, M. J., (2004). Methodological issues in studying psychotherapy processes and outcomes. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (pp. 84-135). New York: Wiley.
- Horowitz, L. M., Rosenberg, S. E., Baer, B. A., Ureno, G. & Villasenor, V. S. (1988). Inventory of interpersonal problems: Psychometric properties and clinical applications. *Journal of Consulting and Clinical Psychology, 56*, 885-892.
- Howard, K. I., Lueger, R. J., Maling, M. S. & Martinovich, Z. (1993). A phase model of psychotherapy outcome: Causal mediation of change. *Journal of Consulting and Clinical Psychology, 61*, 678-685.
- Howard, K. I., Lueger, R. J., Martinovich, Z. & Lutz, W., (1999). The cost-effectiveness of psychotherapy: Dose-response and phase models. In N. E. Miller & K. M. Magruder (Eds.), *Cost-effectiveness of psychotherapy: A guide for practitioners, researchers, and policymakers* (pp. 143-152). New York: Oxford University Press.
- Howard, K. I., Moras, K., Brill, P. L., Martinovich, Z. & Lutz, W. (1996). Evaluation of psychotherapy: Efficacy, effectiveness, and patient progress. *American Psychologist, 51*, 1059-1064.
- 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.
- Jermann, B., (2006). Die Skalen Beschwerden und Kongruenz im Fragebogen zur Evaluation von Psychotherapieverläufen (FEP-2). Konstruktion und erste Überprüfung psychometrischer Kennwerte. Unveröffentlichte Lizentiatsarbeit, Universität Bern, Bern.

- Jung, F., (2008). Allgemeine Selbstwirksamkeit: Zusammenhänge mit den Dimensionen des psychischen Wohlbefindens im Fragebogen zur Evaluation von Psychotherapieverläufen (FEP). Eine Untersuchung an einer annähernd bevölkerungsrepräsentativen Stichprobe. Unveröffentlichte Diplomarbeit, Universität Trier, Trier.
- Kiesler, D. J. (1996). From communications to interpersonal theory: A personal odyssey. *Journal of Personality Assessment*, 66, 267-282.
- Lambert, M. (2007). Presidential address: What we have learned from a decade of research aimed at improving psychotherapy outcome in routine care. *Psychotherapy Research*, 17, 1-14.
- Lambert, M. J., Hannöver, W., Nisslmüller, K., Richard, M. & Kordy, H. (2002). Fragebogen zum Ergebnis von Psychotherapie: Zur Reliabilität und Validität der deutschen Übersetzung des Outcome Questionnaire 45.2 (OQ-45.2). *Zeitschrift für Klinische Psychologie und Psychotherapie*, 31, 40-47.
- Lambert, M. J., Morton, J. J., Hatfield, D., Harmon, S. C., Reid, R. C. et al. (2004). *Administration and scoring manual for the OQ-45.2*. Orem, UT: American Professional Credentialing Service.
- Leary, T., (1957). *Interpersonal diagnosis of personality*. New York: Ronald Press.
- Loehlin, J. C., (2004). *Latent variable models. An introduction to factor, path, and structural equation analysis*. Mahwah, NJ: Lawrence Erlbaum.
- Lutz, W. (2002). Patient-focused psychotherapy research and individual treatment progress as scientific groundwork for an empirical based clinical practice. *Psychotherapy Research*, 12, 251-272.
- Lutz, W. (2007). Evidenzbasierung in der Psychotherapie: A bridge over troubeled water. *Zeitschrift für Neuropsychologie*, 18, 9-21.
- Lutz, W. & Grawe, K., (2006). Psychotherapieforschung: Grundlagen, Konzepte und neue Trends. In B. Strauss, F. Caspar & F. Hohagen (Hrsg.), *Lehrbuch der Psychotherapie* (S. 727-768). Göttingen: Hogrefe.
- Lutz, W. & Howard, K. I., (1997). Compass questionnaire validity checks. Unpublished Working Paper, Northwestern University, Evanston/Chicago.
- Lutz, W., Tholen, S., Schürch, E. & Berking, M. (2006). Die Entwicklung, Validierung und Reliabilität von Kurzformen gängiger psychometrischer Instrumente zur Evaluation des therapeutischen Fortschritts in der Psychotherapie und Psychiatrie. *Diagnostica*, 52, 11-25.
- Lutz, W., Lambert, M. J., Harmon, S. C., Tschitsaz, A., Schurch, E. & Stulz, N. (2006). The probability of treatment success, failure and duration- what can be learned from empirical data to support decision making in clinical practice? *Clinical Psychology & Psychotherapy*, 13, 223-232.
- Lutz, W., Schürch, E., Stulz, N., Böhnke, J. R., Schöttke, H., Rogner, J. & Wiedl, K. H. (2009). Entwicklung und psychometrische Kennwerte des Fragebogens zur Evaluation von Psychotherapieverläufen (FEP). *Diagnostica*, 55,
- Muthén, B. O. & Muthén, L. K., (1998-2004). *Mplus user's guide (3)*. Los Angeles, CA: Muthen & Muthen.
- Rost, J., (2004). *Lehrbuch Testtheorie — Testkonstruktion*. Bern: Huber.
- Schürch, E., Lutz, W. & Böhnke, J. R. (im Druck). Identifikation problematischer Antwortmuster im Fragebogen zur Evaluation von Psychotherapieverläufen mit Hilfe der Rasch-Analyse. *Zeitschrift für Klinische Psychologie und Psychotherapie*.
- Stulz, N. & Lutz, W. (2007). Multidimensional patterns of change in outpatient psychotherapy: The phase model revisited. *Journal of Clinical Psychology*, 63, 817-833.
- Stulz, N., Lutz, W., Leach, C., Lucock, M. & Barkham, M. (2007). Shapes of early change in psychotherapy under routine outpatient conditions. *Journal of Consulting and Clinical Psychology*, 75, 864-874.
- Ullrich de Muynck, R. & Ullrich, R., (1977). *Das Emotionalitätsinventar als Befindlichkeitsmaß. Testmanual EMI-B*. München: Pfeiffer.
- Wise, E. A. (2004). Methods for analyzing psychotherapy outcomes: A review of clinical significance, reliable change, and recommendations for future directions. *Journal of Personality Assessment*, 82, 50-59.
- Wittchen, H.-U., Nelson, C. B. & Lachner, G. (1998). Prevalence of mental disorders and psychosocial impairments in adolescents and young adults. *Psychological Medicine*, 28, 109-126.
- Wittchen, H.-U., Lieb, R., Wunderlich, U. & Schuster, P. (1999). Comorbidity in primary care: Presentation and consequences. *Journal of Clinical Psychiatry*, 60, 29-36.

Anschrift der Verfasser:

Prof. Dr. Wolfgang Lutz
Dipl.-Psych. Jan R. Böhnke
Fachbereich I - Psychologie
Universität Trier
D-54286 Trier
E-Mail: wolfgang.lutz@uni-trier.de

Fragebogen zur Evaluation von Psychotherapieerläufen (FEP-2)¹

Mit diesem Fragebogen würden wir gerne von Ihnen erfahren, wie Sie sich in der letzten Woche gefühlt und in bezug auf andere verhalten haben.

Bitte lesen Sie jede Frage aufmerksam durch und kreuzen Sie jeweils die für Sie zutreffendste Antwort an. Bitte überlegen Sie dabei nicht, ob eine Antwort einen guten Eindruck machen könnte, sondern antworten Sie so, wie es auf Ihre gegenwärtige Lebenssituation zutrifft.

Kreuzen Sie bitte pro Frage nur eine Antwort an. Alle Ihre Antworten werden selbstverständlich vertraulich behandelt.

nie	selten	manchmal	oft	sehr oft
1	2	3	4	5

In der letzten Woche...

1. fühlte ich mich wohl	1	2	3	4	5
2. ging ich vielen Interessen nach	1	2	3	4	5
3. fühlte ich mich ohnmächtig	1	2	3	4	5
4. war ich nervös	1	2	3	4	5
5. fühlte ich mich unbelastet und zufrieden	1	2	3	4	5
6. hatte ich Schlafprobleme	1	2	3	4	5
7. belastete mich meine Zukunftsaussicht	1	2	3	4	5
8. war ich ängstlich	1	2	3	4	5
9. war ich den an mich gestellten Anforderungen gewachsen	1	2	3	4	5
10. war ich sehr einsam und alleine	1	2	3	4	5
11. hatte ich Probleme, mich für andere zu freuen	1	2	3	4	5
12. war ich unter Anspannung und innerem Druck	1	2	3	4	5
13. war ich leicht von anderen zu überreden	1	2	3	4	5
14. war ich guter Dinge	1	2	3	4	5
15. war ich selbstsicher und selbstbewusst	1	2	3	4	5
16. hatte ich Probleme, andere zu unterstützen	1	2	3	4	5
17. fühlte ich mich von anderen nicht geschätzt	1	2	3	4	5
18. hatte ich Probleme, Aggressionen zu zeigen, wenn nötig	1	2	3	4	5

nie	selten	manchmal	oft	sehr oft
1	2	3	4	5

In der letzten Woche...

19. war ich selbstbeherrscht	1	2	3	4	5
20. war ich leicht von anderen auszunutzen	1	2	3	4	5
21. empfand ich einen Sinn in meinem Leben	1	2	3	4	5
22. war ich voller Kraft und Ausdauer	1	2	3	4	5
23. war ich panisch und voller Angst	1	2	3	4	5
24. fühlte ich mich ohne Wert	1	2	3	4	5
25. hatte ich Probleme, vertrauten Personen gegenüber Ärger zu äussern, wenn nötig	1	2	3	4	5
26. gingen mir die Probleme anderer Menschen schnell zu nahe	1	2	3	4	5
27. war ich Teil einer erfüllten und intimen Beziehung	1	2	3	4	5
28. hatte ich Probleme, auf andere Menschen zuzugehen	1	2	3	4	5
29. war ich unabhängig und frei	1	2	3	4	5
30. war ich voll innerer Ruhe	1	2	3	4	5
31. konnte ich mich für nichts begeistern	1	2	3	4	5
32. hatte ich Probleme, anderen negative Gefühle mitzuteilen	1	2	3	4	5
33. mischte ich mich zu sehr in die Angelegenheiten anderer ein	1	2	3	4	5
34. dachte ich daran, mir das Leben zu nehmen	1	2	3	4	5
35. fühlte ich mich ungenügend und unzureichend	1	2	3	4	5
36. hatte ich Probleme, anderen gegenüber offen zu sein und Gefühle zu zeigen	1	2	3	4	5
37. fühlte ich mich von anderen im Stich gelassen	1	2	3	4	5
38. war ich deprimiert und niedergeschlagen	1	2	3	4	5
39. war ich optimistisch	1	2	3	4	5
40. war ich schnell vor anderen verlegen	1	2	3	4	5