

Experimentelle Konstruktion eines Kurzfragebogens zur direkten Veränderungsmessung psychotherapeutischer Effekte im Befinden

Günter Krampen

Zusammenfassung. Befunde aus zwei Studien zur Entwicklung eines Kurzinstruments zur direkten Veränderungsmessung psychotherapeutischer Effekte werden dargestellt. Die Fragebogenkonstruktion fokussiert Veränderungen im Bereich des Befindens. An Studie I waren 150 Patienten von 14 niedergelassenen Psychotherapeuten beteiligt. Sie wurden nach dem Zufall einer Therapie- oder einer Wartelistekontrollgruppe zugewiesen. Nach dreimonatiger Psychotherapie bzw. Wartezeit beantworteten alle den aus zwölf Komparativaussagen bestehenden „Veränderungsfragebogen zum Spannungserleben und Befinden – Psychotherapie“ (VFE-PT). Die Änderungssensitivität des Kurzfragebogens für Psychotherapieeffekte wird im Vergleich zur unbehandelten Kontrollgruppe auf Item- und Skalenniveau empirisch bestätigt. Studie II folgt einem naturalistischen Design, nach dem 275 erwachsene ambulante Patienten den VFE-PT 12 bis 20 Wochen nach Therapiebeginn ausgefüllt haben. Die Befunde bestätigen die guten Itemparameter und interne Konsistenz des Verfahrens. Die konvergente Validität des VFE-PT wird durch bedeutsame Korrelationen zu Indikatoren klinisch relevanter Veränderungen aus der indirekten Veränderungsmessung bestätigt.

Schlüsselwörter: Psychotherapieergebnisse, direkte Veränderungsmessung, Therapieerfolgskontrolle, ambulante Psychotherapie, Psychometrie, psychologische Messung, experimenteller Plan, Reliabilität, Validität

Experimental construction of a short questionnaire for direct measurement of psychotherapeutic outcomes in well-being

Abstract. Results of two studies on the construction of a short questionnaire for the direct measurement of psychotherapeutic outcomes are presented. Items focus on changes in behaviour and experience with reference to well-being. Study I investigated 150 outpatients of 14 experienced psychotherapists. Patients were randomized to a therapy group or a waiting list control group. After three months of waiting or psychotherapy, subjects completed the “Questionnaire on Personal Changes” consisting of twelve phrased as comparisons overtime items. Group comparisons confirm the change sensitivity of the items and the scale for psychotherapeutic effects in contrast to the untreated control group. Study II followed a naturalistic design in which 275 adult psychotherapy outpatients completed the questionnaire 12–20 weeks after the start of therapy. Results confirm the good psychometric properties of scale with reference to item parameters and internal consistency. Convergent validity of the scale is empirically supported by significant correlations to indicators of psychotherapeutic outcomes from indirect measurements of change determined by pretest-posttest-differences.

Key words: direct measurement of change, psychotherapeutic outcomes, treatment effectiveness evaluation, outpatient psychotherapy, psychometrics, psychological assessment, experimental design, test reliability, test validity

Die Methodik der direkten Veränderungsmessung, bei der Patienten selbst für einen definierten Zeitraum (etwa vom Therapiebeginn bis zum Erhebungszeitpunkt während oder beim Abschluss einer Therapie oder katamnestisch) subjektiv Veränderungen des Verhaltens und Erlebens einschätzen, wird inzwischen aufgrund ihrer Ökonomie, einfachen Durchführbarkeit und einiger methodologischer Vorteile ergänzend zur Methodik der indirekten Veränderungsmessung (über Vortest-Nachtest-Differenzwerte) in der Psychotherapieforschung anerkannt. Dies zeigt sich etwa darin, dass dieser Ansatz zum Standardinhalt moderner Übersichtsbeiträge und Einführungen zur klinisch-psychologischen Diagnostik und Evaluation geworden ist (vgl. etwa Fydrich, 2006; Reinecker-Hecht & Baumann, 2005; Stieglitz, 2003; Wittchen & Hoyer, 2006). Gleichwohl wird die direkte Veränderungsmessung auf-

grund von Bedenken, die sich auf Vermutungen über ihre potenziell stärkere Anfälligkeit für Gedächtniseffekte und Antworttendenzen beziehen, gegenüber der indirekten Messung von Veränderungen über Vortest-Nachtest-Differenzwerte skeptischer beurteilt und etwa als „kleine Lösung“ (Fydrich, 2006, S. 529) für den Fall empfohlen, dass keine Vortestdaten erhoben wurden. Zwar nicht übersehen, aber minimiert werden dabei die bereits von Bereiter (1962, 1963) pointiert dargestellten methodenimmanenten Probleme der indirekten Veränderungsmessung, die sich auf statistische und Reliabilitäts-bedingte Artefakte durch die Regression zur Mitte, das Reliabilitäts-Validitäts-Dilemma und das Physikalismus-Subjektivismus-Problem beziehen. Fazit ist in der Regel, dass beide Ansätze der Veränderungsmessung nicht äquivalent sind, da sie zum Teil zu unterschiedlichen Ergebnissen führen, und daher

als komplementäre Methoden kombiniert eingesetzt werden sollten, um wahre Veränderungen bei Psychotherapiepatienten möglichst valide abschätzen zu können.

Die empirische Befundlage zur konvergenten Validität der Methoden zur direkten und indirekten Veränderungsmessung im klinisch-psychologischen Bereich ist nicht sehr breit und dabei nicht ohne Widersprüche. Zielke und Kopf-Mehnert (1978) berichten so für den „Veränderungsfragebogen des Erlebens und Verhaltens“ (VEV) in einer Stichprobe von 45 ambulanten Patienten nach neunwöchiger klient-zentrierter Psychotherapie Validitätskoeffizienten für Vortest-Nachtest-Differenzen in klinisch relevanten Skalen (zur Symptombelastung, neurotischen Tendenz, Fehlanpassung, Psychasthenie und Angst), die im Mittel bei um $r_M = .58$ liegen. Die gemeinsame Varianz beider Methoden zur Abschätzung psychotherapeutischer Effekte variiert zwischen 30 % und 40 %. Etwa 40 % gemeinsame Varianz werden auch von Stieglitz und Baumann (2001) berichtet, was durch einige Befunde aus der Evaluationsforschung zu stationären psychosomatischen Rehabilitationsmaßnahmen (vgl. etwa Nübling, Steffanowski, Wittmann & Schmidt, 2004; Schmid, 2008; Schmidt, Steffanowski, Nübling, Lichtenberg & Wittmann, 2003) bestätigt wird. Bleßmann (2004) konnte dagegen – allerdings für Krankheitsverläufe bei chronischen Erkrankungen – für direkte und indirekte Veränderungsmaße lediglich eine durchschnittliche Korrelation von $r = .20$ (mithin nur 4 % gemeinsame Varianz) feststellen, was sie dazu führt, beiden Methoden der Veränderungsmessung die Vergleichbarkeit (und konvergente Validität) abzusprechen. Auch Raspe, Weber, Voigt, Kosinski und Petras (1997), Kohlmann und Raspe (1998) sowie Michalak, Kosfelder, Meyer und Schulte (2003) berichten über ähnlich mäßige Übereinstimmungen zwischen direkten und indirekten Veränderungsmaßen.

Neben reinen Forschungsinstrumenten ohne Normierung (siehe etwa Kriebel, Paar, Schmitz-Buhl & Raatz, 2001; Stapel, Zielke & Hoff-Emden, 2005; Zielke & Kopf-Mehnert, 2001a) liegt im deutschen Sprachraum für die direkte Veränderungsmessung psychotherapeutischer Effekte bislang lediglich der aus 42 Items bestehende VEV (Zielke & Kopf-Mehnert, 1978) vor. Der VEV wurde spezifisch „aus der Veränderungstheorie der Gesprächspsychotherapie und dem Erfahrungskonzept von Gesprächspsychotherapeuten“ (Zielke & Kopf-Mehnert, 1978, S. 9) abgeleitet, wird seit längerem aber auch in anderen psychotherapeutischen Kontexten angewendet (siehe hierzu etwa Fydrich, 2006; Zielke & Kopf-Mehnert, 2001b). Die Veränderungsnormen des VEV basieren auf einer Stichprobe von 90 „mehrfach beeinträchtigten“ (Zielke, 1978, S. 89) ambulanten Gesprächspsychotherapie-Patienten, von denen 45 nach einer Therapiedauer von neun Wochen (mit wöchentlichen Therapiesitzungen) und 45 nach einer Wartezeit von neun Wochen den VEV beantwortet haben. Therapie- und Wartegruppe sind nicht randomisiert, aber *post hoc* nach dem Alter sowie den Persönlichkeitsmerkmalen Neurotizismus, Extraversion und Intelligenz parallelisiert worden. Der VEV differenziert sowohl auf Item- als auch auf Skalenniveau statistisch und prak-

tisch signifikant zwischen beiden Gruppen, was seine Brauchbarkeit für die Veränderungsmessung in der klient-zentrierten Psychotherapie bestätigt. Die Veränderungsnormen des VEV beziehen sich auf kritische Grenzwerte des VEV-Skalenwerts, die anhand der statistischen Parameter (Mittelwert, Standardabweichung, Reliabilitätskoeffizient, Standardmessfehler) in der nicht behandelten Wartegruppe berechnet wurden.

Im Folgenden wird über die experimentell ausgerichtete Konstruktion eines Kurzinstruments für die direkte Veränderungsmessung in der ambulanten Psychotherapie von Erwachsenen sowie seine Bewährung in der Versorgungsforschung (d. h., in der vierten Phase des 4-Phasen-Prüfmodells für klinische Interventionen; siehe etwa Krampen, Schui & Wiesenhütter, 2008) informiert. Empirisch geprüft wird die psychometrische Qualität der nur zwölf Items umfassenden Skala. Neben den klassischen Itemparametern, der internen Konsistenz der Skala sowie ihrer konvergenten (gegenüber Befunden aus der indirekten Veränderungsmessung) und diskriminanten Validität (gegenüber konstruktfernen Variablen und Antworttendenzen) steht die Änderungssensitivität des Verfahrens für psychotherapeutische Effekte auf der Item- und Skalenebene im Vordergrund, die über Vergleiche einer Therapie- und einer Warteliste-Kontrollgruppe bestimmt wird.

Methoden

Erhebungsinstrumente

Die Itementwicklung des *Veränderungsfragebogens zu Entspannungserleben und Befinden – Psychotherapie (VFE-PT)* geht ursprünglich auf die Indikations- und Effektbereiche des Autogenen Trainings (Krampen, 1991) zurück. Aus einem Pool von 24 Items erwiesen sich dort insgesamt zwölf Items als sensitiv für die Effekte systematischer Entspannungsverfahren in klinisch-psychologischen Interventionskontexten. Jeweils zwei der Items markieren die störungsunspezifischen Zielbereiche (1) der psychischen und körperlichen Entspannungs- und Erholungsfähigkeiten (Veränderungen des Verhaltens; siehe Item-Nr. 1 und 2 in Tabelle 1), (2) der emotionalen Stabilisierung (Veränderungen des Erlebens; Item-Nr. 4 und 5), (3) der Selbstregulation willkürlicher und autonomer Funktionen (Veränderungen des Verhaltens; Item-Nr. 6 und 7), (4) der verbesserten Nutzung eigener Fähigkeiten und Leistungssteigerung (Veränderungen des Verhaltens; Item-Nr. 8 und 9), (5) des Wohlbefindens und der verbesserten Schmerzbewältigung (Veränderungen des Erlebens; Item-Nr. 3 und 11) sowie (6) der erhöhten Selbstwirksamkeit und -kontrolle (Veränderungen des Erlebens; Item-Nr. 10 und 12). In der Instruktion wird zunächst der Zeitpunkt (Datum) des Beginns der Behandlung bzw. der Wartezeit explizit vorgegeben. Unter Bezug auf diesen Zeitpunkt und den Vergleichszeitraum wird sodann instruiert: „Überlegen Sie, was Sie zu diesem Zeitpunkt machten und wie Sie sich fühlten. Bitte prüfen Sie bei den nachfolgenden Aussagen immer, ob sich bei Ihnen seitdem – also

im Verlauf der Behandlung (bzw. der Wartezeit) – eine Veränderung in die eine oder andere Richtung vollzogen hat. Schätzen Sie diese Veränderung jeweils anhand der folgenden Skala ein: +3 (starke positive Änderung), +2 (mittlere positive Änderung), +1 (schwache positive Änderung), 0 (keine Änderung), -1 (schwache negative Änderung), -2 (mittlere negative Änderung), -3 (starke negative Änderung).“

Nach den Taxonomien interventionsübergreifender Wirk- und Effektfaktoren von Psychotherapien (vgl. etwa Grawe, 1995, 1998) können die Items einer mehr oder weniger erfolgreichen *aktiven Hilfe zur Problembewältigung*, mithin der Problembewältigungsperspektive zugeordnet werden. Als Indikatoren einer im subjektiven Zeitvergleich verbesserten, verschlechterten oder unveränderten Fähigkeit, Lebens- und Handlungsprobleme zu bewältigen, erstrecken sie sich auf ein breiteres Spektrum des Verhaltens und Erlebens mit klinischer Relevanz. Sie betreffen direkt einen veränderten (verbesserten) Umgang mit einer Vielzahl zuvor als schwierig und/oder problematisch erlebten Lebenssituationen: Situationen, in denen man früher besorgt und/oder unausgeglichen war, nicht gut schlafen konnte, Medikamente nahm, keine Ausdauer zeigte, sich nicht konzentrieren konnte, sich krank und/oder insgesamt unwohl fühlte, nicht gelassen bleiben konnte, mit den Belastungen nicht zurecht kam (siehe Tabelle 1). Die Items beziehen sich auf störungsunspecifische Veränderungen im Befinden, die mit einer erhöhten Selbstwirksamkeit, dem Dreh- und Angelpunkt erfolgreicher psychologischer Interventionsprozesse und -ergebnisse nach der Theorie zur *self efficacy* von Bandura (1977), in Zusammenhang gebracht werden können. Dies steht in Beziehungen zur verbesserten Problem- und Symptombewältigung nach Grawe (1995, 1998), zu günstigen Erfolgserwartungen und -erlebnissen nach Frank (1982), zu konstruktiven Verhaltensänderungen nach Karasu (1977, 1986) sowie zu positiven Erfolgserwartungen und der Problemmkontrolle nach Weinberger (1995). Damit weist der VFE-PT als Verfahren zur direkten Veränderungsmessung Relationen zur Veränderungstheorie der allgemeinen und differentiellen Psychotherapie auf und fokussiert daraus abgeleitete Veränderungskriterien im Bereich des übergreifenden Wirk- und Effektfaktors der (mehr oder weniger gelungen) aktiven Hilfe zur Problembewältigung und Aspekte der Steigerung von Selbstwirksamkeit.

In unterschiedlichen Teilstichproben wurden zur konvergenten Validierung des VFE-PT anhand von Vortest-Nachtest-Differenzwerten aus der indirekten Veränderungsmessung das *Beck-Depressions-Inventar* (BDI; Hautzinger, Bailer, Worall & Keller, 1995), die *Skala zur Hoffnungslosigkeit* (H-Skalen; Krampen, 1994), die *Änderungssensitive Symptomliste* (ASS-SYM; Krampen, 2006) und die Subskala Neurotizismus aus dem *NEO-Fünf-Faktoren-Inventar* (NEO-FFI; Borkenau & Ostendorf, 1993) eingesetzt. Für die Prüfung der diskriminanten Validität wurden – neben den anderen NEO-FFI-Subskalen – in Teilstichproben Skalen zu Antworttendenzen (*Skala zur sozialen Erwünschtheit*, MCSD nach Crowne & Marlowe, 1964; *Lügen-Skala* aus dem *Eysenck-Persönlichkeits-In-*

ventar, EPI-L nach Eggert, 1974; *Offenheits-Skala* aus dem *Freiburger Persönlichkeitsinventar*, O-Skala des FPI-R nach Fahrenberg, Hampel & Selg, 2001) und zwei Tests zu basalen kognitiven Leistungen (*Zahlen-Verbindungs-Test*, ZVT nach Oswald & Roth, 1987; Subtest „Zahlennachsprechen“ aus dem *Hamburg-Wechsler-Intelligenztest für Erwachsene*, HAWIE-R nach Tewes, 1991) verwendet.

Stichprobe und experimentelles Design von Studie I

An der randomisierten und kontrollierten Studie I waren 150 ambulante Psychotherapiepatienten von 14 niedergelassenen, berufserfahrenen Psychotherapeuten mit Fachkundenachweisen in Verhaltenstherapie ($n = 9$) und/oder tiefenpsychologisch fundierter Therapie ($n = 7$) sowie weiteren Therapieausbildungen bei Grundorientierung an der allgemeinen und differentiellen Psychotherapie im Rahmen der psychotherapeutischen Grundversorgung beteiligt. Bei Kontrolle der Primärdiagnose und des Geschlechts wurden die Patienten nach der Anmeldung und der von den Therapeuten durchgeführten Differentialdiagnostik [*Strukturiertes Klinisches Interview für DSM-IV*, SKID; Wittchen, Zaudig & Fydrich (1997) sowie Vortesterhebungen mit BDI, H-Skala und ASS-SYM] nach dem Zufall einer Therapiegruppe ($n = 75$) oder einer Wartelistekontrollgruppe ($n = 75$) zugewiesen. Alle Patienten stimmten nach der Aufklärung über die Studie der Untersuchungsteilnahme zu (*informed consent*).

Durch die bei der randomisierten Zuweisung der Patienten zur Therapie- oder Wartegruppe durchgeführte *Matching*-Prozedur wurde eine für die Geschlechtszugehörigkeit (jeweils $n = 42$ Frauen und $n = 33$ Männer) und die Primärdiagnose vollkommen ausgewogene Verteilung in der Therapie- und Wartegruppe erreicht, die inferenzstatistische Prüfungen obsolet macht. In beiden Gruppen waren Störungen aus dem Formenkreis der Depression ($n = 32$), Angststörungen ($n = 19$), Anpassungs- und Belastungsstörungen ($n = 15$) sowie somatoformen Störungen ($n = 9$) gleich häufig vertreten. Unterschiedlichste Komorbiditäten (sowohl nach Achse I als auch nach Achse II des DSM-IV) lagen bei 67 der Patienten vor (45 %). Dies entspricht ebenso wie die Verteilung der Primärdiagnosen dem in der Untersuchungsregion typischen Patientenauftreten in der ambulanten psychotherapeutischen Grundversorgung und ist Resultat der naturalistischen Rekrutierung der Untersuchungsteilnehmer.

Ergänzend vorgenommene *post hoc*-Vergleiche bestätigen die Vergleichbarkeit von Therapie- und Wartegruppe nach dem Alter [$M = 33.5$ mit $SD = 10.9$ in der Therapie- und $M = 34.1$ mit $SD = 12.0$ in der Wartegruppe; $t(148) = 0.78$; *Range*: 18 bis 72 Jahre] sowie den Eingangswerten für die Depressivität sowie für die allgemeine Symptom- und Beschwerdenbelastung. Die BDI-Mittelwerte überschreiten in beiden Gruppen das Kriterium für klinisch relevante depressive Störungen [$M = 23.4$ mit $SD = 9.9$ in der Therapie- und $M = 24.2$ mit $SD = 10.2$ in der Wartegruppe; $t(148) = 1.21$]. Im Vergleich zur Normierungsstichprobe liegen die mittleren ASS-SYM-Werte für die allgemeine

Symptom- und Beschwerdenbelastung in beiden Untersuchungsgruppen mit einem Prozentrangwert von $PR = 90$ ($T = 62.5$) ebenfalls im sehr hohen Bereich. Dies gilt auch für die sechs ASS-SYM-Subskalen, wobei sich auch in ihnen keine statistisch bedeutsamen Differenzen in den Eingangswerten zwischen der Therapie- und der Wartegruppe zeigen.

Nach dreimonatiger Psychotherapie (mit zwölf Sitzungen bei zumeist wöchentlicher Frequentierung) bzw. Wartezeit beantworteten alle Studienteilnehmer den aus zwölf Komparativaussagen bestehenden VFE-PT (siehe die Items in Tabelle 1). Zudem wurden Nachtest-Erhebungen mit dem BDI, der H-Skala und der ASS-SYM, zum psychosozialen Funktionsniveau nach Achse V des SKID für DSM-IV sowie zu Antworttendenzen (MCSD, EPI-L und FPI-R-O) durchgeführt. Um die Vergleichbarkeit des Untersuchungszeitraums zu gewährleisten, waren die Zeitpunkte der Nacherhebungen individuell auf einen Zeitraum von drei Monaten standardisiert. Innerhalb dieses Zeitraums konnte in der Therapiegruppe keine Behandlung abgeschlossen werden (*Range*: 13 bis 23 weitere Sitzungen; $M = 18.3$, $SD = 5.52$).

Stichprobe und naturalistisches Design von Studie II

Weitere Datenerhebungen mit dem VFE-PT erfolgten bei 275 erwachsenen ambulanten Psychotherapiepatienten,

die ein breites Spektrum an Primär Diagnosen und auch Komorbiditäten nach der ICD-10 aufwiesen ($n = 164$ Patientinnen; $M = 36.7$ Jahre, $SD = 13.20$). Studie II folgte einem naturalistischen *Design* im Rahmen der Versorgungsforschung (vierte Phasen des 4-Phasenprüfmodells für klinische Interventionen), bei dem die Erhebungen 12 bis 20 Wochen nach dem Therapiebeginn (unabhängig vom weiteren Verlauf der Behandlung) durchgeführt wurden. In Teilstichproben erfolgten zusätzlich Vortest- und Nachtest-Erhebungen mit der ASS-SYM, dem NEO-FFI, dem ZVT und dem Subtest Zahlennachsprechen aus dem HAWIE-R sowie Nacherhebungen mit Skalen zu Antworttendenzen (MCSD, EPI-L und FPI-R-O).

Ergebnisse

Änderungssensitivität, Item- und Skalenparameter des VFE-PT

Die Itemparameter des VFE-PT für die direkte Veränderungsmessung, die in der experimentellen *Studie I* drei Monate nach Therapiebeginn bzw. nach einer dreimonatigen Wartezeit erfasst wurden, sind in Tabelle 1 aufgeführt. Sowohl im absoluten Vergleich der Itemmittelwerte (unter Bezug auf positive versus negative Vorzeichen, die positive versus keine/negative Veränderungsangaben indizieren) als auch bei ihrer inferenzstatistischen Prüfung zei-

Tabelle 1. Mittelwerte (M), Standardabweichungen (SD) und Mittelwertvergleiche der VFE-PT-Items und -Skala in einer Psychotherapiegruppe ($n = 75$) und einer Wartelistegruppe ($n = 75$) sowie Item- und Skalenparameter in einer Stichprobe von 275 ambulanten Psychotherapiepatienten

Item-Nr.	Item	Therapiepatienten		Wartegruppe		<i>t-Test</i> ($df=148$)	<i>d</i>	275 Psychotherapiepatienten		
		M	SD	M	SD			M	SD	r_{it}
01.	Ich kann mich besser körperlich entspannen.	+1.4	1.14	-0.2	1.18	6.41**	1.38	+1.5	1.21	.61
02.	Ich kann besser abschalten und mich erholen.	+1.3	1.21	-0.4	1.20	5.89**	1.41	+1.2	1.29	.65
03.	Ich fühle mich insgesamt gesünder.	+1.3	1.40	-0.2	1.37	3.14**	1.08	+1.2	1.38	.78
04.	Ich bin weniger besorgt, wenn ich an die Zukunft denke.	+1.4	1.43	-0.8	1.39	7.31**	1.56	+1.5	1.42	.72
05.	Ich fühle mich ruhiger und ausgeglichener.	+1.2	1.37	-1.1	1.41	6.83**	1.65	+1.3	1.39	.82
06.	Ich schlafe besser.	+1.1	1.52	-0.3	1.55	5.58**	0.91	+1.0	1.64	.46
07.	Ich nehme weniger Medikamente.	+0.8	1.22	0.0	1.31	2.32*	0.63	+0.7	1.33	.47
08.	Ich habe mehr Ausdauer und gebe nicht so schnell auf.	+1.4	0.96	-0.5	1.13	4.88**	1.82	+1.4	1.01	.44
09.	Ich kann mich besser konzentrieren.	+1.3	1.28	-0.1	1.37	3.87**	1.05	+1.4	1.30	.45
10.	Unerwartete Ereignisse nehme ich gelassener hin.	+1.1	1.23	-0.2	1.19	2.59**	1.07	+1.3	1.18	.68
11.	Ich fühle mich wohler.	+1.6	1.42	-0.3	1.35	5.01**	1.12	+1.7	1.42	.77
12.	Ich komme mit Stress und Belastungen besser zu recht.	+1.3	1.23	-0.4	1.19	4.93**	1.40	+1.2	1.20	.72
Skalensummenwert		+15.2	11.8	-4.3	12.1	11.53**	1.64	+15.4	12.6	
interne Konsistenz der Skala (Cronbachs Alpha)		.94		.91				.96		

Anmerkungen: ** $p < .01$; * $p < .05$.

Tabelle 2. Korrelationen zwischen der direkten Veränderungsmessung mit dem VFE-PT und unterschiedlichen Indikatoren der indirekten Veränderungsmessung (umgepolte Vortest-Nachtest-Differenzen)

Variable der indirekten Veränderungsmessung	VFE-PT-Skala
75 ambulante Psychotherapiepatienten (Behandlungszeit: 12 Wochen)	
ASS-SYM-ERSCHÖ: körperliche und psychische Erschöpfung	.45**
ASS-SYM-ANSPAN: Nervosität und innere Anspannung	.38**
ASS-SYM-DYSREG: psychophysiologische Dysregulationen	.32**
ASS-SYM-LEISTU: Leistungs- und Verhaltensschwierigkeiten	.41**
ASS-SYM-SCHMER: Schmerzbelastungen	.27*
ASS-SYM-SELBST: Probleme in der Selbstbestimmung und -kontrolle	.43**
ASS-SYM-G: allgemeine Symptom- und Problembelastung	.48**
BDI: Depressivität	.42**
H-Skala: Hoffnungslosigkeit	.49**
SKID-I: psychosoziales Funktionsniveau (DSM-Achse V)	.52**
138 ambulante Psychotherapiepatienten (Behandlungszeit: 14 bis 20 Wochen)	
ASS-SYM-ERSCHÖ: körperliche und psychische Erschöpfung	.40**
ASS-SYM-ANSPAN: Nervosität und innere Anspannung	.32**
ASS-SYM-DYSREG: psychophysiologische Dysregulationen	.36**
ASS-SYM-LEISTU: Leistungs- und Verhaltensschwierigkeiten	.44**
ASS-SYM-SCHMER: Schmerzbelastungen	.19*
ASS-SYM-SELBST: Probleme in der Selbstbestimmung und -kontrolle	.37**
ASS-SYM-G: allgemeine Symptom- und Problembelastung	.46**
NEO-FFI: Neurotizismus	.35**
NEO-FFI: Extraversion	.16
NEO-FFI: Offenheit für Erfahrungen	.21*
NEO-FFI: Verträglichkeit	.26**
NEO-FFI: Gewissenhaftigkeit	.05
ZVT: kognitive Leistungsgeschwindigkeit	.19*
HAWIE-Subtest Zahlennachsprechen: Kurzzeitgedächtnis	.23**

Anmerkungen: ** $p < .01$; * $p < .05$.

gen sich bedeutsame Unterschiede zwischen der Therapie- und der Wartegruppe. Für alle zwölf VFE-PT-Items liegen in der Therapiegruppe positive Veränderungsangaben vor, die sich statistisch bedeutsam von den Werten in der Wartegruppe unterscheiden. Die Effektstärke der Gruppenunterschiede liegt nach den Kriterien von Cohen (1988) für alle VFE-PT-Items im großen Bereich (siehe Tabelle 1). Damit ist die spezifische Änderungssensitivität der VFE-PT-Items für Psychotherapieeffekte experimentell nachgewiesen.

Die Änderungssensitivität des VFE-PT für Psychotherapieeffekte bildet sich auch im VFE-PT-Skalenrohwert ab, der in der Therapiegruppe nach drei Monaten Be-

handlung bei durchschnittlich $M = +15.2$ und in der Wartegruppe nach dreimonatiger Wartezeit bei durchschnittlich $M = -4.3$ liegt (siehe vorletzte Zeile in Tabelle 1). Dieser Gruppenunterschied weist eine Effektstärke von $d = 1.64$ auf, was nach den Kriterien von Cohen (1988) ebenfalls einem großen Effekt entspricht. Die Kurzskaala weist überdies in beiden experimentellen Gruppen eine vergleichbare hohe interne Konsistenz (nach Cronbachs Alpha) auf (siehe letzte Zeile von Tabelle 1).

Die Mittelwerte und Standardabweichungen der VFE-PT-Items aus der naturalistischen Studie II sind in den rechten Spalten von Tabelle 1 aufgeführt. Sie stimmen gut mit den Werten überein, die in der Therapiegruppe nach

Tabelle 3. Korrelationen der VFE-PT-Skala mit Antworttendenzen

Korrelat	VFE-PT-Skala
Soziale Erwünschtheit (MCSD)	
75 ambulante Psychotherapiepatienten (Behandlungsdauer: 12 Wochen)	.24*
138 ambulante Psychotherapiepatienten (Behandlungsdauer: 14 bis 20 Wochen)	.31**
112 stationäre Psychotherapiepatienten (Behandlungsdauer: 6 bis 10 Wochen) ^a	.24**
Lügen-Skala (EPI-L)	
75 ambulante Psychotherapiepatienten (Behandlungsdauer: 12 Wochen)	.18
48 stationäre Psychotherapiepatienten (Behandlungsdauer: 6 bis 10 Wochen) ^a	.09
Offenheit (FPI-R)	
75 ambulante Psychotherapiepatienten (Behandlungsdauer: 12 Wochen)	-.08
56 stationäre Psychotherapiepatienten (Behandlungsdauer: 6 bis 10 Wochen) ^a	.11

Anmerkungen: ** $p < .01$; * $p < .05$. ^aWeitere unabhängige Stichproben stationärer Psychotherapiepatienten in Psychosomatischen Fachkliniken, deren spezifische Charakteristika in der vorliegenden Arbeit aus Platzgründen nicht dargestellt werden.

dem experimentellen Design in Studie I gewonnen wurden. Dies gilt auch für den mittleren VFE-PT-Skalenrohwert, der statistisch signifikant, aber numerisch eher niedrig mit der Therapiedauer korreliert ist ($r = .28$; $p < .01$) und seine Standardabweichung.

Die Trennschärfekoeffizienten der zwölf VFE-PT-Items bestätigen die gute psychometrische Qualität aller Items (siehe rechte Spalte von Tabelle 1) und münden in eine interne Konsistenz (nach Cronbachs Alpha) von $r_{tt} = .96$, die übereinstimmend zu den Befunden aus Studie I die hohe instrumentelle Reliabilität des VFE-PT bestätigt.

Validität der direkten Veränderungsmessung mit dem VFE-PT

Die Interkorrelationen des VFE-PT-Skalenwerts mit konstruktnahen Indikatoren aus der indirekten Veränderungsmessung sind durchgängig statistisch bedeutsam und belegen die *konvergente Validität* der direkten zur indirekten Veränderungsmessung: In beiden Studien gilt dies sowohl für die mit der *Änderungssensitiven Symptomliste* erfasste Reduktion der allgemeinen Symptom- und Problembelastung der Patienten als auch für die sechs ASS-SYM-Subskalen zur Verringerung spezifischer Symptombelastungen (siehe Tabelle 2). In Studie I ist der VFE-PT-Skalenwert ferner mit Vortest-Nachtest-Differenzen im *Beck-Depressions-Inventar* und in den *Skalen zur Hoffnungslosigkeit* statistisch bedeutsam korreliert. Das Gleiche gilt für Vortest-Nachtest-Differenzen in dem mit dem *NEO-Fünf-Faktoren Inventar* erfassten Neurotizismus in einer Teilstichprobe von Studie II (siehe Tabelle 2). Diese Befunde bestätigen die konvergente Validität des VFE-PT zur direkten Veränderungsmessung psychotherapeutischer Effekte anhand von klinisch relevanten Indikatoren aus der indirekten Veränderungsmessung. Die gemeinsame Varianz beider Ansätze der Veränderungsmessung va-

riert zwischen 12 % (Neurotizismus) und 24 % (Hoffnungslosigkeit).

In Studie I wurde zudem der Zusammenhang zwischen dem VFE-PT und Vortest-Nachtest-Differenzen in den diagnostischen Fremdeinschätzungen des psychosozialen Funktionsniveaus (L'-Daten) nach der Achse V des *Diagnostischen und Statistischen Manuals Psychischer Störungen* (DSM-IV-TR; nach dem SKID) überprüft. Die Korrelation von $r = .52$ ($p < .01$) belegt die konvergente Validität des VFE-PT für Vortest-Nachtest-Differenzen dieser klinischen Fremdbeurteilungen im therapeutischen Prozess.

Hinweise auf die *diskriminante Validität* des VFE-PT ergeben sich aus seinen schwachen Zusammenhängen mit konstruktfernen Indikatoren aus der indirekten Veränderungsmessung für eine Teilstichprobe aus Studie II: Zwar partiell statistisch bedeutsame, jedoch von der Varianzbindung (maximal 7 %) her niedrige Korrelationen der VFE-PT-Skala liegen aus der zweiten Studie mit Psychotherapiepatienten vor (siehe Tabelle 2). Vortest-Nachtest-Differenzen in den Skalen zur Extraversion, Offenheit für Erfahrungen, Verträglichkeit und Gewissenhaftigkeit – erfasst mit dem *NEO-Fünf-Faktoren Inventar* (NEO-FFI) – sind ebenso nur schwach mit dem VFE-PT korreliert wie Vortest-Nachtest-Differenzen in zwei Indikatoren basaler kognitiver Funktionen (Leistungsgeschwindigkeit und Kurzzeitgedächtnisleistung) nach dem ZVT und dem Subtest „Zahlennachsprechen“ aus dem HAWIE-R.

Faktorielle Struktur des VFE-PT

Da es sich bei dem VFE-PT nicht um ein faktorenanalytisch konstruiertes psychometrisches Verfahren handelt, wurden dimensionsanalytische Auswertungen ausschließlich unter exploratorischen Zielsetzungen (im Sinne *a posteriori* realisierter exploratorisch-faktorieller Validierungsanalysen) durchgeführt. Ähnlich wie bei den An-

Tabelle 4. Absolute (f) und relative (%) Veränderungshäufigkeiten im VFE-PT-Skalenwert (nach den kritischen Grenzwerten) in der Gruppe der Psychotherapiepatienten ($N = 75$) und der Wartelistekontrollgruppe ($N = 75$)

Veränderungskriterium	kritisches Intervall des VFE-PT-Skalenwertes	Therapiegruppe		Wartelistegruppe	
		f	%	f	%
Verbesserung bei $p < .001$	+ 17 bis + 36	39	52 %	8	11 %
Verbesserung bei $p < .01$	+ 9 bis +16	13	17 %	0	0 %
Verbesserung bei $p < .05$	+ 7 und +8	6	8 %	2	3 %
Verbesserung bei $p < .10$	+6	2	3 %	0	0 %
keine Veränderung ($p > 10$)	+5 bis -5	15	20 %	27	36 %
Verschlechterung bei $p < .10$	-6	0	0 %	11	15 %
Verschlechterung bei $p < .05$	-7 und -8	0	0 %	5	7 %
Verschlechterung bei $p < .01$	-9 bis -11	0	0 %	3	4 %
Verschlechterung bei $p < .001$	-12 bis -36	0	0 %	19	25 %
Chi-Quadrat-Wert ($df = 8$)		78.88**			
Veränderungskriterium (aggregiert nach + / 0 / -)	kritisches Intervall des VFE-PT-Skalenwertes	Therapiegruppe		Wartelistegruppe	
		f	%	f	%
Verbesserung (+)	+ 6 bis + 36	60	80 %	10	13 %
keine Veränderung (0)	+ 5 bis -5	15	20 %	27	36 %
Verschlechterung (-)	-6 bis -36	0	0 %	38	51 %
Chi-Quadrat-Wert ($df = 2$)		77.14**			

Anmerkung: ** $p < .01$.

wendungen der VFE-Items im Kontext der Evaluation von Veränderungen aufgrund der Teilnahme an Einführungen in das Autogene Training (siehe Krampen, 1991) führten Hauptkomponentenanalysen (mit Varimax-Rotation) der VFE-PT-Daten aus verschiedenen Stichproben erwachsener Psychotherapiepatienten zu einem inkonsistenten Befundmuster. Durchgängig ergibt sich zwar für die Datensätze aus unterschiedlichen Stichproben in der Regel ein varianzstarker erster Faktor (mit einer Varianzbindung zwischen 54 % und 64 %), auf dem die deutliche Mehrheit der VFE-PT-Items (zwischen neun und elf Items) bedeutsame Ladungen aufweist. Explorativ zusätzlich berechnete zwei-, drei- und vierfaktorielle Lösungen führten zu varianzschwachen zweiten, dritten und vierten Faktoren (< 7 % der Varianzbindung), die in jedem Fall nur durch jeweils ein Item, maximal zwei Items markiert werden und sich zwischen den Faktorenanalysen der Daten aus verschiedenen Stichproben unterscheiden. Der konsistent in Faktorenanalysen der Daten aus unterschiedlichen Stichproben ermittelte varianzstarke erste Faktor weist einerseits auf eine relativ hohe faktorielle Homogenität des VFE-PT, wodurch die itemanalytischen Befunde bestätigt werden: In Einklang mit den in Tabelle 1 aufgeführten Trennschärfekoeffizienten der VFE-PT-Items weisen die Faktorladungen darauf, dass die mit dem VFE erfasste Dimension der subjektiven Veränderung im Entspannungserleben und Befinden vor allem durch die Items „Ich fühle mich ruhiger und ausgeglichener“, „Ich fühle mich wohler“, „Ich fühle mich insgesamt gesunder“ und „Ich bin weniger besorgt, wenn ich an die Zukunft denke“ markiert werden, die Indikatoren von Veränderungen im Erleben sind. Andererseits finden sich – in Abhängigkeit von der jeweiligen Stichpro-

be – unterschiedliche „Ausreißer“-Items, die weitere, allerdings nicht stabil interpretierbare Faktoren markieren. Auf die ausführliche Darstellung entsprechender faktorenanalytischer Lösungen wird hier verzichtet. Nach der vorliegenden Befundlage würden durch sie ohnehin nur für die jeweiligen Datensätze höchst spezifische Faktormatrizen wiedergegeben, die in hohem Maße stichprobenabhängig sind.

Bezüge der direkten Veränderungsmessung zu Antworttendenzen

Die Zusammenhänge des VFE-PT-Skalenwertes mit Antworttendenzen, für die bei der direkten Veränderungsmessung höher ausgeprägte Kovariationen vermutet werden, wurden in mehreren unabhängigen Stichproben ambulanter (aus Studie I und II) und – in weiteren unabhängigen Stichproben – auch stationärer Psychotherapiepatienten untersucht (siehe Tabelle 3). Für die *Skala zur sozialen Erwünschtheit* (MCSD) zeigen sich übereinstimmend in drei Stichproben statistisch signifikante Korrelationen zum VFE-PT-Skalenwert. Die gemeinsame Varianz beider Variablen bleibt mit 6 % bis 10 % im unteren Bereich. Nicht statistisch bedeutsam sind dagegen die Bezüge des VFE-PT-Skalenwertes zur „Lügen-Skala“ aus dem EPI und zur „Offenheits-Skala“ aus dem FPI-R. In Studie I ergaben sich keine signifikanten ($p > .20$) Unterschiede zwischen den Korrelationen des VFE-PT-Skalenwertes und den Antworttendenzen in der Therapie- und der Warteliste-Kontrollgruppe.

Veränderungsquoten in Therapie- und Wartelistekontrollgruppe

Für die Anwendungspraxis sind auf der Basis von Gruppenvergleichen Nachweise von Behandlungserfolgen im Einzelfall erforderlich. Für den exakten Nachweis intraindividuelle Veränderungen im VFE-PT-Skalengesamtwert wurden daher kritische Veränderungswerte bestimmt, die über- bzw. unterschritten werden müssen, um eine Änderung als reliable Verbesserung versus Verschlechterung qualifizieren zu können. Ebenso wie beim VEV von Zielke und Kopf-Mehnert (1978) wird dabei davon ausgegangen, dass therapiebedingte Veränderungen nur dann vorliegen, wenn sie über Veränderungen in einer unbehandelten Vergleichsgruppe hinausgehen. Die kritischen Grenzwerte bzw. Intervalle (siehe Tabelle 4) wurden daher anhand der statistischen Parameter Standardabweichung und Reliabilität sowie den daraus resultierenden Bestimmungen des Standardmessfehlers und der Konfidenzintervalle für unterschiedliche Signifikanzniveaus (siehe hierzu etwa Lienert, 1967) des VFE-PT in der unbehandelten Wartelistekontrollgruppe (aus Studie I) bestimmt.

In Tabelle 4 sind die absoluten und relativen Veränderungshäufigkeiten, die mit dem VFE-PT in einer Therapiegruppe und in einer Wartelistegruppe (Studie I) nach diesem Kriterium der kritischen Grenzen für Einzelfallentscheidungen erreicht wurden, zusammengefasst. Danach können 80 % der behandelten Patienten (im Vergleich zu 13 % der nicht behandelten Patienten, die hypothetisch als spontane Remissionen eingeordnet werden können) nach zwölf Sitzungen als gebessert angesehen werden. Für 52 % kann dies bei $p < .001$, für 17 % bei $p < .01$, für 8 % bei $p < .05$ und für 3 % bei $p < .10$ inferenzstatistisch abgesichert werden. Statistisch bedeutsame Verschlechterungen werden im Beurteilungszeitraum von drei Monaten in der Therapiegruppe nicht, in der Wartelistekontrollgruppe dagegen bei 51 % beobachtet. Bei 20 % der Therapiegruppe können mit dem VFE-PT nach dreimonatiger Psychotherapie (noch?) keine statistisch bedeutsamen Veränderungen nachgewiesen werden. In der Wartelistekontrollgruppe beträgt dieser Anteil 36 %. Anhand signifikanter Befunde in Chi-Quadrat-Tests werden die unterschiedlichen Verteilungen der Veränderungsquoten in der Therapie- und Wartelistekontrollgruppe aus Studie I inferenzstatistisch bestätigt (siehe Tabelle 4).

Diskussion

Sowohl in der experimentellen Testkonstruktion (Studie I) als auch in seiner breiteren Anwendung im Rahmen der psychotherapeutischen Versorgungsforschung (Studie II) hat sich der *Veränderungsfragebogen zum Entspannungserleben und Befinden – Psychotherapie* (VFE-PT) empirisch bewähren können. Seine Änderungssensitivität für Effekte der ambulanten Psychotherapie bei Erwachsenen in der ambulanten Grundversorgung konnte sowohl auf Item- als auch auf Skalenebene empirisch bestätigt werden. In der Stichprobe sind die häufigsten

psychischen Störungsbilder (z. T. mit Komorbiditäten) der ambulanten psychotherapeutischen Versorgung vertreten. Hinweise auf eine Störungsspezifität des VFE-PT ergaben sich – nicht zuletzt aufgrund der kleinen Substichprobengrößen – bislang nicht. Die Trennschärfekoeffizienten seiner Items als auch die interne Konsistenz des VFE-PT sind psychometrisch zufriedenstellend. Hervorzuheben ist auch seine mit nur zwölf Items gegebene Ökonomie in der (ggf. wiederholten) Anwendung, die – neben den für Psychotherapiepatienten plausiblen Iteminhalten – zu seiner guten Akzeptanz bei Patienten und Psychotherapeuten beiträgt.

Günstig ist die Befundlage zur konvergenten Validität des VFE-PT gegenüber klinisch-relevanten Vortest-Nachtest-Differenzwerten aus der indirekten Veränderungsmessung. Die konvergenten Validitätskoeffizienten erreichen zwar nicht die zum Teil von anderen Autoren für alternative Methoden der direkten Veränderungsmessung berichteten numerischen Werte (mit gemeinsamen Varianzen bis zu 40 %; vgl. etwa Stieglitz & Baumann, 2001; Zielke & Kopf-Mehnert, 1978), die für den VFE-PT ermittelten Kovariationen mit Vortest-Nachtest-Differenzwerten zur Depressivität, Hoffnungslosigkeit, emotionalen Labilität sowie allgemeinen Symptom- und Beschwerdenbelastung von Psychotherapiepatienten sind aber durchgängig statistisch bedeutsam und weisen im Mittel eine gemeinsame Varianz von 20 % auf. Bedeutsamer ist noch, dass die mit dem VFE-PT bei Patienten direkt erfassten subjektiven Veränderungen des Erlebens und Verhaltens (Q-Daten) essentiell mit Vortest-Nachtest-Differenzwerten aus klinischen Fremdbeurteilungen (L'-Daten) des psychosozialen Funktionsniveaus (nach Achse V des DSM) korreliert sind und eine gemeinsame Varianz von 27 % aufweisen. Mit diesen Befunden wird als Fazit bestätigt, dass beide Ansätze der Veränderungsmessung zwar zum Teil zu unterschiedlichen Ergebnissen führen, damit nicht äquivalent sind, aber als komplementäre Methoden methodologisch einen gleichberechtigten Status aufweisen. Sinnvoll ist ihr kombinierter Einsatz, um – auch unter Einbezug anderer Datenquellen (wie klinischen Fremdbeurteilungen, objektivierbaren Veränderungen in Lebensstil und -umständen etc.) – wahre Veränderungen bei Psychotherapiepatienten nach den Strategien der multimethodalen und -modalen Diagnostik und Evaluation möglichst valide abschätzen zu können.

Die diskriminante Validität der direkten Veränderungsmessung psychotherapeutischer Effekte mit dem VFE-PT konnte nicht nur für konstruktfernere Variablen (wie Extraversion, Offenheit für Erfahrungen, Verträglichkeit, Gewissenhaftigkeit und zwei Indikatoren der basalen kognitiven Leistungsfähigkeit), sondern auch gegenüber Antworttendenzen empirisch belegt werden. Während sich für zwei gängige Indikatoren der „Lügen-Tendenz“ und der „Offenheit“ in jeweils zwei unabhängigen Stichproben von Psychotherapiepatienten keine statistisch bedeutsamen Bezüge zum VFE-PT zeigten, liegen aus drei unabhängigen Patientenstichproben Hinweise auf statistisch signifikante, jedoch numerisch nicht stark ausgeprägte Zusammenhänge mit der Tendenz vor, sich in sozi-

al erwünschter Weise zu verhalten. Die gemeinsame Varianz beider Variablen überschreitet aber nicht 10%. Damit werden Bedenken, die sich auf eine stärkere, erhebliche Abhängigkeit der Befunde aus der direkten Veränderungsmessung von Antworttendenzen beziehen und nach denen Überschätzungen der wahren Veränderungen durch Patienten befürchtet werden (siehe etwa Reinecker-Hecht & Baumann, 2005; Fydrich, 2006), relativiert. Diese Befunde sprechen auf der Stichprobenebene gegen stärkere Effekte von Antworttendenzen auf die Befunde der direkten Veränderungsmessung mit dem VFE-PT. Selbstverständlich bleibt eine kritische Prüfung im jeweiligen Einzelfall notwendig. Dies gilt insbesondere für Patienten mit einer entsprechenden Persönlichkeitsausprägung (etwa im Sinne der ursprünglichen Theorie zur sozialen Erwünschtheit von Crowne & Marlowe, 1964) und auch für solche mit einer abhängigen, asthenischen Persönlichkeitsausprägung oder gar Persönlichkeitsstörung (ICD-10: F60.7), also mit einer erhöhten Tendenz, sich – etwa auch im therapeutischen Kontext – in sozial erwünschter Weise darzustellen und zu verhalten. Bedeutsam werden hier vor allem Konsistenzen zu den Befunden aus der indirekten Veränderungsmessung und aus der klinischen Urteilsbildung sowie weitere fremd-evaluative Daten. Andererseits sollte gerade im Einzelfall bedacht werden, dass auch eine umgekehrte Argumentation denkbar ist, dass also die beim VFE-PT verwendeten direkten Komparativaussagen von Patienten im Sinne einer subjektiven selbst- und therapie-kritischen Bestandsaufnahme (= Bilanz) beantwortet werden könnten, somit also eher in Richtung einer *Unterschätzung* wahrer Veränderungen gehen könnten.

Hervorzuheben ist, dass mit dem VFE-PT schon aufgrund seines Ursprungs (Evaluation systematischer Entspannungsverfahren) nur ein begrenzter Ausschnitt der Gesamtmenge möglicher Zielbereiche psychotherapeutischer Interventionen erfasst werden kann. Insbesondere werden in ihm keine störungsspezifischen Therapieziele thematisiert. Es kann vermutet werden, dass daher keine bedeutsamen Unterschiede im VFE-PT-Skalenwert zwischen Patienten mit unterschiedlichen Störungsbildern ermittelt werden konnten. Der Evaluationsbereich des VFE-PT ist damit globaler auf Aspekte der Befindlichkeit und des Entspannungserlebens von Patienten begrenzt, der gleichwohl bei vielen psychischen (und auch physischen) Störungen mit Krankheitswert von allgemeiner Bedeutung ist.

Literatur

- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84, 191–215.
- Bereiter, C. (1962). Using tests to measure change. *Personnel and Guidance Journal*, 41, 6–11.
- Bereiter, C. (1963). Some persisting dilemmas in the measurement of change. In C. W. Harris (Ed.), *Problems in measuring change* (pp. 3–20). Madison, WI: University of Wisconsin Press.
- Beßmann, A. (2004). *Krankheitsverläufe bei chronischen Erkrankungen: Welche Methode der Veränderungsmessung eignet sich zu ihrer Beschreibung und Prognose?* Unveröffentlichte Dissertation, Universität Bielefeld.
- Borkenau, P. & Ostendorf, F. (1993). *NEO-Fünf-Faktoren Inventar (NEO-FFI)*. Göttingen: Hogrefe.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.) New York, NY: Academic Press.
- Crowne, D. P. & Marlowe, D. (1964). *The approval motive: Studies in evaluative dependence*. New York, NY: Wiley.
- Eggert, D. (1974). *Eysenck-Persönlichkeits-Inventar (EPI)*. Göttingen: Hogrefe.
- Fahrenberg, J., Hampel, R. & Selg, H. (2001). *Freiburger Persönlichkeitsinventar (FPI-R)*. Göttingen: Hogrefe.
- Frank, J. D. (1982). Therapeutic components shared by all psychotherapies. In J. H. Harvey & M. M. Parks (Eds.), *Psychotherapy research and behavior change* (= The master lecture series, Vol. 1, pp. 5–38). Washington, DC: American Psychological Association.
- Fydrich, T. (2006). Diagnostik und Intervention in der Klinischen Psychologie. In M. Amelang & L. Schmidt-Atzert (Hrsg.), *Psychologische Diagnostik und Intervention* (4. Aufl., S. 495–532). Heidelberg: Springer.
- Grawe, K. (1995). Grundriß einer Allgemeinen Psychotherapie. *Psychotherapeut*, 40, 130–145.
- Grawe, K. (1998). *Psychologische Therapie*. Göttingen: Hogrefe.
- Hautzinger, M., Bailer, M., Worall, H. & Keller, F. (1995). *Beck-Depressions-Inventar (BDI; 2. Aufl.)*. Göttingen: Hogrefe.
- Karasu, T. B. (1977). Psychotherapies: An overview. *American Journal of Psychiatry*, 134, 851–863.
- Karasu, T. B. (1986). The psychotherapies: Benefits and Limitations. *American Journal of Psychotherapy*, 40, 324–342.
- Kohlmann, T. & Raspe, H. (1998). Zur Messung patientennaher Erfolgskriterien in der medizinischen Rehabilitation. *Rehabilitation*, 37, 30–37.
- Krampen, G. (1991). *Diagnostisches und Evaluatives Instrumentarium zum Autogenen Training (AT-EVA)*. Göttingen: Hogrefe.
- Krampen, G. (1994). *Skalen zur Erfassung von Hoffnungslosigkeit (H-Skalen)*. Göttingen: Hogrefe.
- Krampen, G. (2006). *Änderungssensitive Symptomliste zu Entspannungserleben, Wohlbefinden, Beschwerden- und Problembelastungen (ASS-SYM)*. Göttingen: Hogrefe.
- Krampen, G., Schui, G. & Wiesenhütter, J. (2008). Evidenzbasierte Psychotherapie und Therapie-Ressourcen. *Zeitschrift für Klinische Psychologie und Psychotherapie*, 37, 43–51.
- Kriebel, R., Paar, G. H., Schmitz-Buhl, S. M. & Raatz, U. (2001). Veränderungsmessung mit dem Veränderungsfragebogen (VEV): Entwicklung einer Kurzform und deren Anwendung in der Psychosomatischen Rehabilitation. *Praxis Klinische Verhaltensmedizin und Rehabilitation*, 53, 20–32.
- Lienert, G.A. (1967). *Testaufbau und Testanalyse* (2. Aufl.). Weinheim: Beltz.
- Michalak, J., Kosfelder, J., Meyer, F. & Schulte, D. (2003). Messung des Therapieerfolges: Veränderungsmaße oder retrospektive Erfolgsbeurteilung. *Zeitschrift für Klinische Psychologie und Psychotherapie*, 32, 94–103.
- Nübling, R., Steffanowski, A., Wittmann, W. W. & Schmidt, J. (2004). Strategien der Ergebnismessung am Beispiel der psychosomatischen Rehabilitation. *Praxis Klinische Verhaltensmedizin und Rehabilitation*, 65, 35–44.
- Oswald, W.-D. & Roth, E. (1987). *Der Zahlen-Verbindungs-Test (ZVT; 2. Aufl.)*. Göttingen: Hogrefe.
- Raspe, H., Weber, U., Voigt, S., Kosinski, A. & Petras, H. (1997). Qualitätssicherung durch Patientenbefragung in der medizinischen Rehabilitation. *Rehabilitation*, 36, 31–42.
- Reinecker-Hecht, C. & Baumann, U. (2005). Klinisch-psychologische Diagnostik. In M. Perrez & U. Baumann (Hrsg.), *Lehrbuch Klinische Psychologie – Psychotherapie* (3. Aufl., S. 128–146). Bern: Huber.
- Schmid, F. (2008). *Stabilität und Prädiktion von Behandlungsergebnissen stationärer psychosomatischer Rehabilitation*. Unveröffentlichte Dissertation, Universität Trier.

- Schmidt, J., Steffanowski, A., Nübling, R., Lichtenberg, S. & Wittmann, W. W. (2003). *Ergebnisqualität stationärer psychosomatischer Rehabilitation: Vergleich unterschiedlicher Evaluationsstrategien*. Regensburg: Roderer.
- Stapel, M., Zielke, M. & Hoff-Emden, H. (2005). VEV-K: Veränderungsfragebogen des Erlebens und Verhaltens für Kinder und Jugendliche für die Altersgruppen bis 12 Jahre (VEV-K-12) und bis 17 Jahre (VEV-K-17). *Praxis Klinische Verhaltensmedizin und Rehabilitation*, 68, 51–60.
- Stieglitz, R.-D. (2003). Psychodiagnostische Verfahren für Kinder. In M. Härter, H. W. Linster, R.-D. Stieglitz & A. Beiroth (Hrsg.), *Qualitätsmanagement in der Psychotherapie* (S. 97–117). Göttingen: Hogrefe.
- Stieglitz, R.-D. & Baumann, U. (2001). Veränderungsmessung. In R.-D. Stieglitz, U. Baumann & H. J. Freyberger (Hrsg.), *Psychodiagnostik in Klinischer Psychologie, Psychiatrie, Psychotherapie* (2. Aufl., S. 21–38). Stuttgart: Thieme.
- Tewes, M. (1991). *Hamburg-Wechsler-Intelligenztest für Erwachsene – Revision (HAWIE-R)*. Göttingen: Hogrefe.
- Weinberger, J. (1995). Common factors aren't so common: The common factors dilemma. *Clinical Psychology – Science and Practice*, 2 (1), 45–69.
- Wittchen, H.-U. & Hoyer, J. (2006). Diagnostische Prozesse in der Klinischen Psychologie und Psychotherapie. In H.-U. Wittchen & J. Hoyer (Hrsg.), *Klinische Psychologie & Psychotherapie* (S. 349–382). Heidelberg: Springer.
- Wittchen, H.-U., Zaudig, M. & Fydrich, T. (1997). *Strukturiertes Klinisches Interview für DSM-IV (SKID)*. Göttingen: Hogrefe.
- Zielke, M. (1978). Validierung eines therapiebezogenen Veränderungsfragebogens für die Gesprächspsychotherapie an einer klinischen Klientenstichprobe. *Diagnostica*, 14, 89–102.
- Zielke, M. & Kopf-Mehnert, C. (1978). *Veränderungsfragebogen des Erlebens und Verhaltens (VEV)*. Weinheim: Beltz.
- Zielke, M. & Kopf-Mehnert, C. (2001a). Der VEV-R-2001: Entwicklung und testtheoretische Reanalyse der revidierten Form des Veränderungsfragebogens des Erlebens und Verhaltens (VEV). *Praxis Klinische Verhaltensmedizin und Rehabilitation*, 53, 7–19.
- Zielke, M. & Kopf-Mehnert, C. (2001b). 22 Jahre wissenschaftliche und klinische Erfahrungen mit dem Veränderungsfragebogen des Erlebens und Verhaltens VEV. *Praxis Klinische Verhaltensmedizin und Rehabilitation*, 53, 3–6.

Prof. Dr. Günter Krampen

Universität Trier
Fachbereich I – Psychologie und ZPID
54286 Trier
E-Mail: krampen@uni-trier.de