

DIREKTE VERÄNDERUNGSMESSUNG DER EFFEKTE VON ENTSPANNUNGSTHERAPIE/-TRAINING UND PSYCHOTHERAPIE: IHRE VOR- UND NACHTEILE SOWIE PSYCHOMETRISCHE HILFSMITTEL

GÜNTER KRAMPEN

Auf dem Hintergrund der methodischen Probleme der indirekten Veränderungsmessung und der damit verbundenen Interpretationsbeschränkungen wird in den Ansatz der direkten Veränderungsmessung der Effekte psychologischer Interventionen eingeführt. Erläutert werden die Vor- und Nachteile beider Methoden am Beispiel der Ergebnisevaluation für Entspannungstrainings und -therapien sowie für Psychotherapie anhand der Selbstauskünfte von Patienten im Rahmen einer multimethodalen psychologischen Diagnostik und Evaluation. Es wird hervorgehoben, dass die indirekte und die direkte Veränderungsmessung nicht als alternative, sondern als komplementäre Evaluationsstrategien zu bewerten und einzusetzen sind, da sich ihre unterschiedlichen methodischen Nachteile in der Annäherung an die Bestimmung wahrer Interventionseffekte bei Patienten wechselseitig kompensieren und sich ihre unterschiedlichen methodischen Vorteile kombinieren. In einer Übersicht wird über die im deutschsprachigen Bereich vorliegenden Instrumente zur direkten Veränderungsmessung informiert, von denen die Mehrheit ad hoc entwickelte Verfahren ohne Normen und andere Vergleichswerte sowie breitere Anwendungserfahrungen sind. Auf die drei zur Verfügung stehenden normierten Instrumente zur direkten Evaluation der Effekte von Entspannungstherapien und -trainings sowie von Psychotherapien wird ausführlicher eingegangen.

Schlüsselwörter: direkte Veränderungsmessung, Evaluation, Autogenes Training, Progressive Relaxation, Psychotherapie, Therapieergebnisse, Therapieerfolgskontrolle, Änderungssensitivität

1. Einleitung

Die Methoden und Methodologie der Veränderungsmessung sind für die Prozess- und Ergebnisevaluation psychologischer Interventionen jedweder Art von zentraler Bedeutung. Anforderungen an

die moderne psychologische Diagnostik und Evaluation beziehen sich etwa auf – die Notwendigkeit eines *multimethodalen Vorgehens*, in dem durch Kombinationen unterschiedlicher Datenarten [wie etwa objektive bzw. objektivierbare biographische Daten (L-Daten nach

- Cattell, 1965), Fremdeinschätzungen durch Diagnostiker und/oder Bezugspersonen (L'-Daten), Selbstauskünfte von Patienten (Q-Daten), objektive Testdaten (T-Daten) und projektive Daten] die spezifischen Fehlerquellen jeder Datenart in der diagnostischen Urteilsbildung kompensiert und die unterschiedlichen Vorzüge der jeweiligen Datenarten genutzt werden können;
- die auf dem Hintergrund biopsychosozialer Störungs- und Entwicklungsmodelle gegebene Sinnhaftigkeit eines *multimodalen Vorgehens*, bei dem verstärkt biopsychologische und psychophysiologische (objektive) Ableitungen organischer Korrelate (oder gar Substrate) als Subgruppe der T-Daten (nach Cattell, 1965), berücksichtigt werden;
 - die Abwendung von einer ausschließlich eigenschaftsorientierten Diagnostik und Evaluation zugunsten eines *klinisch relevanten störungsbezogenen, auch Lebensumstände, Befindlichkeiten und soziale (systemische) Aspekte* einbeziehenden Vorgehens;
 - die Umsetzung einer *individualisierten (idiographischen) psychologischen Diagnostik und Evaluation* in Ergänzung der interindividuell (sozial) vergleichenden (nomothetischen) Diagnostik und Evaluation;
 - die Ergänzung der interindividuell (sozial) vergleichenden psychologischen Diagnostik und Evaluation um *eine auch kriterienorientierte Diagnostik und Evaluation*, die von Populationsmerkmalen (also etwa Normen) unabhängig ist.

Veränderungsmessung und Evaluation im Kontext psychologischer Interven-

tionen können sich prinzipiell auf alle Datenarten beziehen. Im Folgenden stehen die Möglichkeiten sowie die Vor- und Nachteile der indirekten und direkten Veränderungsmessung anhand der Selbstauskünfte von Patienten oder Gruppenteilnehmern (also Q-Daten) im Vordergrund. Die dargelegten Probleme und Fragen lassen sich aber ohne weiteres auch auf die anderen Datenarten beziehen. Q-Daten sind in aller Regel ökonomisch zu erheben, und ihre Erhebung wird dann ohne Probleme akzeptiert, wenn die Frageinhalte einen engen Bezug zu den Problemen und Wünschen der Personen bzw. zu der in Frage stehenden psychologischen Intervention aufweisen, also situativ angemessen und nicht lebensfern oder gar lebensfremd sind. Gleichwohl besteht, etwa aufgrund der Möglichkeit von Antworttendenzen, bei der Verwendung von Q-Daten in der psychologischen Diagnostik und Evaluation, ebenso wie bei allen anderen Datenarten in der klinischen Urteilsbildung, nach den Prinzipien der multimethodalen Diagnostik und Evaluation die Notwendigkeit ihrer Kombination mit anderen Datenarten.

„Wie geht's?“ – Antworten auf diese im Lebensalltag häufiger gestellte Frage betreffen einen subjektiven Ist-Zustand, im Fachterminus eine Status-Diagnose in Form von Selbstauskünften (Q-Daten nach Cattell, 1965). Wird das Befinden in einigem zeitlichen Abstand – etwa von Tagen, Wochen, Monaten oder gar Jahren – wiederholt erfragt, können Vergleiche zwischen den verschiedenen Zeitpunkten durchgeführt und ggf. auch quantifiziert werden. Dies ist das Alltags-Pendant zur *indirekten Veränderungsmessung*, da der

zeitliche Vergleich der Antworten (der „Daten“) indirekt, also nicht vom Befragten (dem Patienten) selbst, sondern durch den Fragenden (den Diagnostiker) durchgeführt wird.

„Geht es Ihnen besser als ... gestern, letzte Woche, letzten Monat, letztes Jahr, etc.?“ – Antworten auf diese ebenfalls im Lebensalltag in Form eines Komparativsatzes häufig gestellte Frage betreffen eine subjektive Veränderungsangabe (wie etwa „ja, besser“, „nein, ebenso“ oder „nein, schlechter“), im Fachterminus somit eine Evaluation oder Prozessdiagnostik in Form von Selbstauskünften (Q-Daten). Dies ist das Alltags-Pendant zur *direkten Veränderungsmessung*, da der zeitliche Vergleich für das Befinden direkt, also vom Befragten (dem Patienten) selbst durchgeführt wird und der Fragende (der Diagnostiker) an diesem Zeitvergleich nicht beteiligt ist.

2. Ursprung der direkten Veränderungsmessung: Methodenimmanente Probleme der indirekten Veränderungsmessung

In seinen Ausführungen zu „einigen beharrlichen/anhaltenden Dilemmata der Veränderungsmessung“ legte Bereiter (1963, S. 3; Übersetzung vom Verf.) vor über 40 Jahren methodenimmanente Probleme der indirekten, psychometrisch fundierten Veränderungsmessung pointiert kritisch und zugleich konstruktiv dar. Dieser Beitrag kann als der entscheidende Impuls für die Entwicklung der Methoden der direkten Verände-

rungsmessung in der Psychometrie bezeichnet werden. Die dabei offengelegten *methodenimmanenten Probleme der indirekten Veränderungsmessung* sind:

- *Regressionseffekte*: Es liegt an der prinzipiell nie gegebenen optimalen Reliabilität ($r_{tt} < 1.00$) und der damit stets gegebenen Messfehlerbehaftetheit psychometrischer Erhebungsverfahren, dass die Testwerte einer Person bei Messwiederholungen unter denselben Bedingungen zum Mittel regredieren. Zu diesen messfehlerbedingten Regressionseffekten treten statistische Regressionseffekte hinzu. Sie bestimmen die Richtung der Zufallsänderungen: Zwischen Vortest- und Differenzwerten bestehen negative Zusammenhänge, Nachtestwerte und Differenzwerte korrelieren positiv miteinander. Niedrige Anfangswerte gehen also mit höheren Differenzwerten einher, ohne dass faktisch ein substantieller Zuwachs gegeben ist. Rechen- und Demonstrationsbeispiele dafür finden sich etwa bei Campbell und Stanley (1963) sowie Petermann, Hehl und Schneider (1977).
- *Das Reliabilitäts-Validitäts-Dilemma*: Die Definition der Reliabilität nach der klassischen Testtheorie als Verhältnis von wahrer Varianz zu beobachteter Varianz bringt es mit sich, dass im Differenzwert aus Vor- und Nachtestung die Fehler beider Messungen kumulieren. Dieser Problematik könnte nur dadurch entgangen werden, dass zwischen Vor- und Nachtest kein Zusammenhang angenommen würde. Dies wäre eine insbesondere bei der Veränderungsmessung höchst problematische Prämisse, da Vor- und Nachtestung schließlich dieselbe Variable vali-

de erfassen sollen. Resultat ist das Reliabilitäts-Validitäts-Dilemma bei der Interpretation von Differenzwerten aus der indirekten Veränderungsmessung.

- Das *Physikalismus-Subjektivismus-Problem*: Die Frage ist, ob numerisch identische Messwert-Differenzen aus Vor- und Nachtests in unterschiedlichen Bereichen eines Messwert-Kontinuums psychologisch (subjektiv) das Gleiche bedeuten? Eine durch ein kognitives Training erreichte (metrische) IQ-Steigerung um zehn Punkte ist etwa bei sehr niedriger oder bei sehr hoher Intelligenz psychologisch anders zu interpretieren als bei mittlerer. Identische oder auch nur ähnliche numerische Messwert-Differenzen aus Vor- und Nachtests repräsentieren damit kaum gleiche oder auch nur ähnliche psychologische Veränderungen. Zielke und Kopf-Mehnert (1978) schlussfolgern: Das „Kontinuum der physikalischen Messwertebene (entspricht nicht dem) Kontinuum auf der subjektiv-psychologischen Bedeutungsebene“ (Zielke & Kopf-Mehnert, 1978, S. 5). Es empfiehlt sich daher, gleiche Messwertdifferenzen in Abhängigkeit vom Ausgangsniveau zu interpretieren.

Inzwischen wurden unterschiedliche Vorschläge für die statistische Korrektur von Messwert-Differenzen aus Gruppenstudien vorgelegt, die zu einer validen Interpretation der Veränderungsindikatoren führen sollen (etwa über Residualmaße, Partialkorrelationen unter Einbezug einer Drittvariablen, regressionsangepasste Veränderungswerte, Change-Quotient, Veränderungsindex; siehe im Überblick etwa Fydrich, 2006; Zielke & Kopf-Meh-

ner, 1978). Für die Beurteilung der klinischen Relevanz von Veränderungen im Einzelfall bleiben die methodenimmanenten Probleme der indirekten Veränderungsmessung aber bestehen, sind damit wahrlich auch über 40 Jahre nach Berriters (1963) Beitrag „beharrlich“ und „anhaltend“.

In der experimentalpsychologischen Perspektive wird die Problematik noch deutlicher: Das Vorgehen der indirekten Veränderungsmessung entspricht einem *einfachen vorexperimentellen Design mit einer Vor- und Nachuntersuchung an einer Interventionsgruppe*, das nach der eingängigen Notation von Campbell und Stanley (1963) mit O X O abgekürzt werden kann. Das O steht dabei für *observation* (hier also Vortest und Nachtest) und das X für *treatment* (also die Intervention). Ursachen mangelnder Validität der Differenzwerte können in einem solchen vor-experimentellen Design neben den oben genannten problemimmanenten Mängeln der indirekten Veränderungsmessung potentiell zusätzlich

- das zwischenzeitliche Geschehen (äußere Ereignisse, die neben der Intervention auftreten),
- die Reifung (biologische und psychische Prozesse, die von äußeren Ereignissen unabhängig sind),
- Vortest-Effekte (Wirkungen der Voruntersuchung),
- die Veränderung diagnostischer Hilfsmittel (*instrumental decay* etwa durch Lernvorgänge oder auch durch Ermüdungserscheinungen etc. bei in Fremdbeurteilungen involvierten Diagnostikern oder Bezugspersonen) sein.

Diesen Mängeln kann nur dadurch abgeholfen werden, dass bereits in der Testkonstruktion mit echten *experimentellen Designs*, also mit Randomisierung (R) der Ausgangsbedingungen gearbeitet wird. Im einfachsten Fall ist dies eine Interventionsgruppe (R O X O) und eine Kontrollgruppe ohne Intervention (R O O), die im Idealfall eine Wartelistengruppe ist. Die Gruppenzuweisung der Personen erfolgt nach dem Zufall. Nach diesem einfachen Zwei-Gruppen-Design mit randomisierter Zuordnung von Patienten zu einer Interventions- und einer Wartegruppe können in der Testkonstruktion die Validität gefährdende Faktoren des zwischenzeitlichen Geschehens, der Reifung, der Vortest-Effekte und des *instrumental decay* kontrolliert werden. Durch den Vergleich der Messwert-Differenzen der Therapiegruppe (X) mit denen der Wartegruppe (ohne X) kann die *Änderungssensitivität* von Items (etwa zu einzelnen Symptombelastungen) und die von psychometrischen Skalenwerten (etwa zur gesamten Symptombelastung) im Hinblick auf die eingesetzte Interventionsmethode X bestimmt werden (siehe etwa Krampen, 1991, 2006; Zielke, 1979). Dies entspricht dem Grundgedanken von Bereiter (1963) zur psychometrischen Entwicklung *änderungssensitiver Variablen* und Skalen und überwindet zumindest die versuchsplanerischen Schwächen der indirekten Veränderungsmessung. Behoben werden jedoch nicht ihre oben genannten methodenimmanenten Probleme der Regression zur Mitte, des Reliabilitäts-Validitäts-Dilemmas und des Physikalismus-Subjektivismus-Dilemmas.

Es ist zu vermuten, dass es nicht zuletzt wegen des erheblichen Konstruktionsaufwands bislang nur wenige *änderungssensitive Instrumente* für die indirekte Veränderungsmessung gibt, die nach einem experimentellen Design mit Randomisierung entwickelt wurden. In deutscher Sprache existieren lediglich die *Kieler Änderungssensitive Symptomliste* (KASSL; Zielke, 1979) mit empirisch nachgewiesener *Änderungssensitivität* für Effekte der klientenzentrierten Psychotherapie und die *Änderungssensitive Symptomliste zu Entspannungserleben, Wohlbefinden, Beschwerden- und Problembelastungen* (ASS-SYM; Krampen, 2006) mit experimentell nachgewiesener *Änderungssensitivität* für Effekte des Autogenen Trainings und der Progressiven Relaxation.

3. Vorteile und Grenzen der direkten Veränderungsmessung

Bei der direkten Veränderungsmessung werden retrospektive Beurteilungen subjektiv erlebter Veränderungen gefordert, die diese „unabhängig von dem Kontinuum einer physikalischen Messwertebene (als) wahrgenommene Änderungen auf einem Kontinuum der subjektiven Bedeutungsebene abbilden“ (Zielke & Kopf-Mehnert, 1978, S. 8). Diese Vorgehensweise ist ökonomisch, da nur ein Erhebungszeitpunkt benötigt wird. Zudem können Regressionseffekte und das Reliabilitäts-Validitäts-Dilemma hier keine Rolle spielen, da nicht mit Messwert-Differenzen, sondern mit Antworten auf Komparativfragen gearbeitet wird.

Auch dem Physikalismus-Subjektivismus-Dilemma kann zumindest dann eine geringere Bedeutung zugewiesen werden, wenn intraindividuelle Vergleichsnormen herangezogen werden, die experimentell mit Hilfe der Daten aus einer Wartegruppe abgesichert sind.

- Die einfachste Konstruktion entsprechender Instrumente folgt einem *vor-experimentellen Design* mit einer einmaligen (Post-)Erhebung (O) im Anschluss an die Intervention (X) – formal nach Campbell und Stanley (1963): X O. Ursachen mangelnder Validität solcher reiner, direkt über Komparativaussagen erhobener Nachtestbefunde können bei einem solchen einfachen Design potentiell – das zwischenzeitliche Geschehen (äußere Ereignisse, die neben der Intervention auftreten),
- die Reifung (biologische und psychische Prozesse, die von äußeren Ereignissen unabhängig sind),
 - eine selektive Stichprobe und
 - ein selektiver Probandenschwund sein.

Diesen Mängeln kann dadurch abgeholfen werden, dass in der Testkonstruktion echte *experimentelle Designs* mit Randomisierung (R) verwendet werden, also etwa eine zufällige Kontrollgruppenanordnung mit einem Nachtest: Formal ist dies eine Interventionsgruppe (R X O) und eine Kontrollgruppe ohne Intervention (R O), die im Idealfall eine Wartelistegruppe ist. Die Gruppenzuweisung der Personen muss nach dem Zufall erfolgen. Diese experimentelle Anordnung erlaubt es, die Validität gefährdenden Faktoren in der Testkonstruktion zu kontrollieren: Durch den Vergleich der Nachtest-Ergebnisse der Therapiegruppe (X)

mit denen der Wartegruppe (ohne X) kann die *Änderungssensitivität* auf Item- und Skalenebene im Hinblick auf die eingesetzte Interventionsmethode X bestimmt werden.

Diese Vorgehensweise entspricht dem subjektivistischen Konzept von Bereiter (1963), psychometrische Instrumente zu entwickeln, in denen retrospektiv direkte Beurteilungen von Veränderungen für einen definierten Zeitraum erhoben werden. Zielke und Kopf-Mehnert (1978, S. 7-8) fassen zusammen: „Man umgeht also die Offenlegung eines Anfangs- und Endwertes einer zu messenden Eigenschaft, indem man den Vergleichsprozess entlang einer subjektiven Gewichtungsskala durchführen lässt“. Anders gesagt, der methodische „Schwarze Peter“, d.h., die Probleme und Dilemmata der indirekten Veränderungsmessung, wird bei der direkten Veränderungsmessung dem Probanden „zugeschustert“. Dass dies nicht ohne Folgen bleibt, ist trivial. Spezifische potentielle Fehlerquellen und damit Nachteile der direkten Veränderungsmessung beziehen sich auf:

- die größere Bedeutung von Gedächtnisprozessen und Orientierungen in der Zeit: Es ist zu vermuten, dass Erinnerungsprobleme insbesondere bei längeren Therapien und/oder subjektiv unscharf definierten Beurteilungszeiträumen die Antworten verzerren können und so der Verlauf sowie das Ergebnis einer Behandlung nur unzuverlässig abgebildet werden;
- Effekte der sozialen Erwünschtheit, denen „bei einer Ein-Punkt-Erhebung am Ende einer Behandlung“ (Fydrich, 2006, S. 528) eine größere Bedeutung zukommen kann als bei Wiederho-

lungsmessungen: Resultat sind Bedenken, nach denen es zu einer Überschätzung der Veränderungen kommen könnte.

4. Kombination beider Methoden: Multimethodale Diagnostik und Evaluation

Damit ist deutlich, dass die direkte Veränderungsmessung einige der bedeutenden Probleme und Dilemmata indirekter Veränderungsmessungen vermeidet, andererseits eigene, jedoch andere methodische Probleme und Interpretationsbeschränkungen aufweist. Beide Strategien der Veränderungsmessung stellen daher keine Alternativen (die sich womöglich wechselseitig ausschließen) dar, sondern sind als jeweils eigenständige Methoden der Veränderungsmessung zueinander komplementär (vgl. hierzu auch Baumann, Sodemann & Tobien, 1980; Fydrich, 2006; Krampen, 1991; Michalak, Kosfelder, Meyer & Schulte, 2003; Zielke, 1982). Sie ergänzen sich nicht zuletzt deswegen sinnvoll, weil jede Methode ihre eigenen Beschränkungen und Mängel aufweist, diese jedoch voneinander unabhängig sind und sich somit in der Zusammenschau der klinischen Bewertung von Veränderungsergebnissen die Fehlerquellen wechselseitig aufheben.

Im konkreten Einzelfall ist die *Konsistenz der Befunde aus beiden Strategien der Ergebnisevaluation* von besonderem Interesse. Stimmen die Befunde aus indirekter und direkter Veränderungsmessung überein, so kann – nach dem Prinzip der multimethodalen und der multimoda-

len Messung selbstverständlich bei Berücksichtigung weiterer evaluativ relevanter Informationen (wie Fremdeinschätzungen durch Diagnostiker und Bezugspersonen, L'-Daten, objektive Veränderungen in Lebensverhältnissen und Lebensstil, L-Daten nach Cattell, 1965) – von einem psychometrisch gut abgesicherten Therapieergebnis ausgegangen werden.

Stimmen die Befunde aus den beiden Strategien der Veränderungsmessung nicht überein, so wird es psychologisch im eigentlichen Sinne interessant. Ebenso wie bei *Inkonsistenzen* zwischen unterschiedlichen Datenmodalitäten muss dann für den jeweiligen Einzelfall die potentielle Bedeutsamkeit der unterschiedlichen Fehlerquellen im diagnostisch-evaluativen Prozess der klinischen Urteilsbildung abgewogen werden. So ist etwa bei Patienten/Personen mit einer Persönlichkeitsausprägung im Sinne der ursprünglichen Theorie zur sozialen Erwünschtheit von Crowne und Marlowe (1964) bei Inkonsistenzen zu den Befunden aus der indirekten Veränderungsmessung Vorsicht geboten. Dies gilt auch bei Personen 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.

Andererseits sollte gerade im Einzelfall bedacht werden, dass auch ein umgekehrter Effekt denkbar ist, dass also die in der Datenerhebung verwendeten direkten Komparativaussagen von Patienten im Sinne einer subjektiven selbst- und therapiekritischen Bestandsaufnahme (=

Bilanz) beantwortet werden könnten, somit also eher in Richtung einer *Unterschätzung* wahrer Veränderungen gehen könnten. Ähnliches kann dann gegeben sein, wenn von Patienten etwa auf der Basis von Selbsterfahrungsmotivation oder aufgrund von sekundärem Krankheitsgewinn auf eine Verlängerung der Behandlung gedrängt wird.

Bei kürzen zeitlichen Abständen zwischen Vor- und Nacherhebungen sind bei der indirekten Veränderungsmessung entsprechende Verzerrungen wahrscheinlicher als bei längeren Zeitintervallen zwischen den Datenerhebungen, da zum Nacherhebungszeitpunkt die bei der Vorerhebung gemachten Angaben besser erinnert werden können. Andererseits werden bei der direkten Veränderungsmessung bei kürzeren Beurteilungsintervallen Gedächtnisverzerrungen und Probleme mit der Orientierung in der Zeit eine geringere Bedeutung haben als bei längeren Beurteilungsintervallen.

5. Überblick zu deutschsprachigen Instrumenten der direkten Veränderungsmessung

Die Darstellung der Grundprinzipien der direkten Veränderungsmessung gehört inzwischen zum Standard von Übersichtsbeiträgen und Einführungen in die klinisch-psychologische Diagnostik. Gleichzeitig ist häufiger neben einer geringen Empirie eine gewisse Scheu vor und Skepsis gegenüber den Verfahren der direkten Veränderungsmessung – ganz im Unterschied zu denen der indirekten –

zu konstatieren (siehe etwa Fydrich, 2006; Stieglitz, 2003; Wittchen & Hoyer, 2006).

Fydrich (2006, S. 529) spricht etwa nach der Beschreibung des *Veränderungsfragebogens des Erlebens und Verhaltens* (VEV; Zielke & Kopf-Mehnert, 1978) die Empfehlung aus, den VEV „zur Überprüfung des Therapieverlaufs und Therapieerfolgs nur in Kombination mit anderen Verfahren“ oder „als ‚kleine Lösung‘“ nur dann einzusetzen, „wenn keine symptom- oder prozessbezogenen Daten zu Beginn oder im Verlauf der Therapie erhoben wurden“. Trifft der zweite Teil dieser Empfehlung mit Sicherheit zu (wenngleich dabei implizit die Nachlässigkeit moniert wird, keine Vortestdaten erhoben zu haben), so scheint aufgrund der oben beschriebenen Methodenprobleme der indirekten Messung deren implizite Bewertung als „große Lösung“ gegenüber der direkten Veränderungsmessung als „kleine Lösung“ übertrieben. Ähnliches gilt für den ersten Teil der Empfehlung Fydrichs, da ebenso für die Methoden der indirekten Veränderungsmessung die Empfehlung gilt, dass diese zur Prüfung von Therapieverlauf und -ergebnis nur „in Kombination mit anderen Verfahren“ eingesetzt werden sollte. Auf diese Empfehlung wird aber verzichtet.

In der folgenden Übersicht (siehe auch Tabelle 1) werden die wenigen in deutscher Sprache vorliegenden Instrumente zur direkten Veränderungsmessung benannt und kurz erläutert. Lediglich drei dieser Instrumente (VEV, VFE-AT und VFE-PT) sind normiert, verfügen also über Möglichkeiten des Vergleichs des bei einem Patienten erhaltenen Mess-

Tab. 1: Übersicht zu deutschsprachigen Verfahren der direkten Veränderungsmessung

Name	Autor(en)	Altersbereich	Skala (Konstrukt)	Itemzahl	Konstruktion
VEV	Zielke & Kopf-Mehnert (1978)	Erwachsene	Veränderungen des Erlebens und Verhaltens auf der Dimension „Spannung, Unsicherheit, Pessimismus“ vs. „Entspannung, Gelassenheit, Optimismus“	42	<u>quasi-experimentell</u> : klientenzentrierte Therapie- und Wartegruppe mit Parallelisierung, aber ohne Randomisierung
VEV-R-2001	Zielke & Kopf-Mehnert (2001)	Erwachsene	wie VEV	42	<u>vor-experimentell</u> : Patienten in Psychosomatischen Reha-Kliniken
VEV-K-25	Kriebel et al. (2001)	Erwachsene	wie VEV	25	<u>vor-experimentell</u> : Patienten in Psychosomatischer Reha-Klinik
VEV-K-12	Stapel et al. (2005)	9- bis 12-Jährige	wie VEV	56	<u>vor-experimentell</u> : Kinder in stationärer medizinischer Reha
VEV-K-17	Stapel et al. (2005)	13- bis 17-Jährige	wie VEV	50	<u>vor-experimentell</u> : Jugendliche in stationärer medizinischer Reha
VFE-AT VFE-PR VFE-ET	Krampen (1991, 2006)	Erwachsene und ältere Jugendliche	Veränderungen im Entspannungserleben und Befinden nach AT- bzw. PR-Kursen.	12	<u>experimentell</u> : Trainings- vs. Wartegruppen mit Randomisierung
VFE-PT	Krampen (2008)	Erwachsene	Veränderungen im Entspannungserleben und Befinden während und nach einer ambulanten Psychotherapie	12	<u>experimentell</u> : Therapie- vs. Wartegruppe mit Randomisierung
FGV	Krampen & von Delius (1981)	Erwachsene	erlebte gesundheitliche Veränderungen von Krankenhauspatienten	12	<u>vor-experimentell</u> : Patienten in Allgemeinkrankenhäusern
FGV-II	Krampen & Ohm (1985)	Erwachsene	erlebte allgemeine und herzbezogene gesundheitliche Veränderungen von Infarktpat.	20	<u>vor-experimentell</u> : Infarktpat. in Psychosomatischen Reha-Kliniken

werts mit denen aus systematisch erhobenen Normierungsstichproben. Alle anderen im Folgenden skizzierten Verfahren wurden als Forschungsinstrumente nach dem oben skizzierten vor-experimentellen Vorgehen (Ein-Gruppen-Design mit Nachtest) entwickelt und sind somit nur in sehr begrenztem Maße psychometrisch evaluiert.

5.1 Veränderungsfragebogen des Erlebens und Verhaltens (VEV)

Der VEV (Zielke & Kopf-Mehnert, 1978) wurde spezifisch für die direkte Veränderungsmessung der Effekte klientenzentrierter Einzel-Psychotherapien entwickelt und normiert. Seine 42 Items wurden „aus der Veränderungstheorie der Gesprächspsychotherapie und dem Erfahrungskonzept von Gesprächspsychotherapeuten“ (Zielke & Kopf-Mehnert, 1978, S. 9) abgeleitet und aus einem Ausgangspool von 114 Items nach itemstatistischen Parametern und den Ergebnissen einer Faktorenanalyse selektiert.

Die 42 VEV-Items beziehen sich auf Fragen, die in Komparativform mögliche Veränderungen des Erlebens und Verhaltens für einen definierten Zeitraum (etwa seit Beginn der Therapie vor neun Wochen in der Konstruktionsstichprobe) beschreiben. Verwendet werden bipolare Antwortskalen (von +3 für starke Änderung in der Itemrichtung über 0 für keine Änderung bis zu -3 für starke Änderung in entgegengesetzter Richtung). 28 Items sind positiv (in Richtung subjektiver Besserung), 14 Items negativ (in Richtung subjektiver Verschlechterung) formuliert.

Durch Summierung der Punkte über alle 42 Items wird (unter Beachtung der Itempolungen) ein Gesamtskalenwert gebildet, der zwischen 42 und 294 variieren kann und das Ausmaß der berichteten Veränderungen des Erlebens und Verhaltens auf einer Dimension mit den Polen „Spannung, Unsicherheit, Pessimismus“ versus „Entspannung, Gelassenheit, Optimismus“ abbildet.

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 knapp neun Wochen (mit wöchentlichen Therapiesitzungen) und 45 nach einer Wartezeit von knapp neun Wochen den VEV beantwortet haben. Therapie- und Wartegruppe sind nicht randomisiert, aber nach dem Alter sowie den Persönlichkeitsmerkmalen Neurotizismus, Extraversion und Intelligenz parallelisiert worden. Die Konstruktion des VEV erfolgte also nach einem quasi-experimentellen Design.

Der VEV differenziert sowohl auf Item- als auch auf Skalenniveau statistisch und praktisch signifikant zwischen der Therapie- und der Wartegruppe, was seine Brauchbarkeit für die Veränderungsmessung in der klientenzentrierten 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. Faktorenanalysen führten zu einem varianzstarken bipolaren Faktor, wodurch die relative Ho-

mogenität der VEV-Items auf der Dimension „Spannung, Unsicherheit, Pessimismus“ versus „Entspannung, Gelassenheit, Optimismus“ empirisch belegt wird.

5.2 Varianten des VEV

Neben der Urform des VEV, der inzwischen sehr breite, weit über den Einsatzbereich der klientenzentrierten Psychotherapie hinausgehende Anwendung gefunden hat (siehe hierzu etwa Fydrich, 2006; Zielke & Kopf-Mehnert, 2001a), existieren inzwischen mehrere Varianten und Kurzformen, unter ihnen der VEV-R-2001 (Zielke & Kopf-Mehnert, 2001b). Im VEV-R-2001 wurde durchgängig auf negativ gepolte Items verzichtet, da diese bei der Bearbeitung des VEV häufig zu Missverständnissen geführt haben. Untersuchungen an 300 Patienten in Psychosomatischen Fachkliniken bestätigen die Faktorenstruktur sowie günstige teststatistische Parameter und geben erste Hinweise auf die Validität des VEV-R-2001. Auf eine Aktualisierung und Anpassung der Veränderungsnormen an das stationäre Patientengut und die stationäre Interventionspraxis wurde mit dem Hinweis auf nahezu identische Testwertstatistiken von VEV und VEV-R-2001 verzichtet. Die Konstruktion und Validierung des VEV-R-2001 erfolgte somit nach einem vor-experimentellen Design, was auch für die anderen VEV-Varianten gilt.

Weitere Varianten des VEV sind die in der Psychosomatischen Rehabilitation von Erwachsenen entwickelte Kurzform VEV-K-25 mit 25 durchgängig positiv gepolten Items (Kriebel, Paar, Schmitz-

Buhl & Raatz, 2001) sowie die von Stapel, Zielke und Hoff-Emden (2005) vorgelegten Varianten für den Kontext der stationären medizinischen Rehabilitation von neun- bis zwölfjährigen Kindern (VEV-K-12 mit 56 Items) und 13- bis 17-jährigen Jugendlichen (VEV-K-17 mit 50 Items). Varianten des VEV bilden nahezu einen Schwerpunkt der Forschung zur direkten Veränderungsmessung im Kontext der Evaluation stationärer psychosomatischer Rehabilitationsmaßnahmen in Deutschland (vgl. im Überblick Schmid, 2008).

5.3 Veränderungsfragebogen zum Entspannungserleben und Befinden (VFE-AT, VFE-PR bzw. VFE-ET)

Der VFE-AT wurde ursprünglich als Bestandteil des *Diagnostischen und Evaluativen Instrumentariums zum Autogenen Training* (AT-EVA; Krampen, 1991) spezifisch für die Evaluation von Einführungen in die Grundstufe des Autogenen Trainings nach einem randomisierten Kontrollgruppen-Design konstruiert und normiert. Zu zwölf Komparativaussagen werden Kursteilnehmer am Ende eines entsprechenden Einführungskurses (und auch in Katamnesen) darum gebeten, im Vergleich zum Beginn des Kurses retrospektiv dazu Angaben zu machen, in welchem Maße sich ihr Befinden und Entspannungserleben positiv (+3, +2, +1), nicht (0) oder negativ (-1, -2, -3) verändert hat.

Jeweils zwei der zwölf Items betreffen die sechs Indikations- und Effektbe-

reiche des Autogenen Trainings (siehe hierzu Krampen, 1998). Berechnet wird ein Summenwert, der kriterial (nach positiven versus keinen versus negativen Veränderungsangaben) anhand kritischer intraindividuelle Veränderungsnormen (anhand der statistischen Parameter einer Wartekontrollgruppe bestimmt) und zusätzlich anhand sozialer Vergleichsnormen (Prozentränge und T-Werte, berechnet nach den Daten aus einer Stichprobe von 342 AT-Kursteilnehmern) interpretiert werden kann.

Alle Vergleichswerte erwiesen sich inzwischen auch für die Evaluation von Einführungskursen in die Progressive Relaxation als tauglich. Daher wird der VFE-AT inzwischen auch als VFE-PR oder – allgemeiner – als VFE-ET bezeichnet (ET steht für Entspannungstraining und -therapie). Das Verfahren ist ökonomisch und weist sehr gute Item- sowie Reliabilitätskennwerte, nur schwache Bezüge zu Indikatoren der sozialen Erwünschtheit und stimmige Befunde zur konvergenten Validität gegenüber Indikatoren aus der indirekten Veränderungsmessung auf (siehe hierzu Krampen, 1991, 2006).

5.4 Veränderungsfragebogen zum Entspannungserleben und Befinden – Psychotherapie (VFE-PT)

Der VFE-PT (Krampen, 2008) ist eine Weiterentwicklung des VFE-AT bzw. VFE-ET für den Anwendungsbereich der direkten Veränderungsmessung während und nach der Durchführung einer Psy-

chotherapie bei Erwachsenen. Die Fragebogenkonstruktion bezieht überdies die Grundlagen und die Veränderungstheorie der allgemeinen und differentiellen Psychotherapie (in Anlehnung an Grawe, 1998) ein. Erfasst werden mit dem VFE-PT Veränderungen des Erlebens und Verhaltens im Bereich des Befindens und Entspannungserlebens, die Beziehungen zum allgemeinen psychotherapeutischen Wirkfaktor der aktiven Hilfe zur Problembewältigung aufweisen und die mit Verbesserungen der Selbstwirksamkeitsüberzeugungen sowie der Bewältigung von Handlungs- und Lebensproblemen von Patienten verbunden sind.

Die Testkonstruktion folgte einem echten experimentellen Design mit Randomisierung und bezieht sich auf Daten von 150 ambulanten Psychotherapiepatienten von 14 niedergelassenen, berufserfahrenen Psychotherapeuten. Die Patienten wurden nach dem Zufall einer Therapiegruppe ($n = 75$) oder einer Wartelistekontrollgruppe ($n = 75$) zugewiesen. Nach dreimonatiger Psychotherapie bzw. Wartezeit beantworteten alle Studienteilnehmer den aus zwölf Komparativaussagen bestehenden VFE-PT. Die Änderungssensitivität dieses Kurzfragebogens für Psychotherapieeffekte wird im Vergleich zur unbehandelten Kontrollgruppe auf Item- und Skalenniveau empirisch bestätigt. In der Therapiegruppe liegt die Quote signifikant positiver Therapieeffekte bei 80 % und signifikant negativer Effekte bei 0 %, in der Wartelistekontrollgruppe für positive Veränderungen bei 13 % und für negative bei 51 %.

Eine kreuzanalytische Validierungsstudie zum VFE-PT folgte einem natura-

listischen 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 statistisch bedeutsame Korrelationen zu Indikatoren klinisch relevanter Veränderungen aus der indirekten Veränderungsmessung (etwa Vortest-Nachtest-Differenzwerte aus der Änderungssensitiven Symptomliste, ASS-SYM, dem Beck-Depressions-Inventar, BDI, der Skala zur Hoffnungslosigkeit, H-Skalen und zum Psychosozialen Funktionsniveau nach Achse V, DSM-IV) bestätigt.

Der VFE-PT weist in beiden Stichproben ambulanter Psychotherapiepatienten weder zur Skala „Offenheit“ aus dem Freiburger-Persönlichkeits-Inventar (FPI-R) noch zur „Lügen-Skala“ aus dem Eysenck-Persönlichkeits-Inventar (EPI) statistisch bedeutsame Korrelationen auf. Signifikante, aber mit einer gemeinsamen Varianz von maximal 9% moderate Beziehungen bestehen allerdings zur sozialen Erwünschtheit (erfasst mit der Marlowe-Crowne-Skala).

Für die Anwendungspraxis liegen für den exakten Nachweis signifikanter Veränderungen im VFE-PT-Skalengesamtwert kritische Veränderungswerte vor, 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 Verände-

rungen in einer unbehandelten Vergleichsgruppe hinausgehen. Die kritischen Grenzwerte bzw. Intervalle wurden daher anhand der statistischen Parameter Standardabweichung und Reliabilität sowie den daraus resultierenden Bestimmungen des Standardmessfehlers und der Konfidenzintervalle für unterschiedliche Signifikanzniveaus des VFE-PT in der unbehandelten Wartelistenkontrollgruppe bestimmt. 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 Wartelistenkontrollgruppe 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 Wartelistenkontrollgruppe beträgt dieser Anteil 36 %. Anhand signifikanter Befunde in Chi-Quadrat-Tests werden die unterschiedlichen Verteilungen der Veränderungsquoten in der Therapie- und Wartelistenkontrollgruppe inferenzstatistisch bestätigt.

Ergänzend zu diesen intraindividuellen Veränderungsnormen, die aus dem Vergleich mit 75 Patienten einer randomisierten Wartelistenkontrollgruppe als kritische intraindividuelle Signifikanzgrenzen resultieren, liegen (soziale) Ver-

gleichswerte (Prozent-Rang- und T-Wert-Normen) aus der Stichprobe von 275 ambulanten Psychotherapiepatienten vor.

5.5 Fragebogen zu erlebten gesundheitlichen Veränderungen (FGV)

Der FGV (Krampen & von Delius, 1981) ist ein Forschungsinstrument für die direkte Veränderungsmessung der subjektiv erlebten gesundheitlichen Veränderungen nach stationären medizinischen und psychologischen Behandlungen in Allgemeinkrankenhäusern und Fachkliniken. Das Kurzinstrument besteht aus zwölf in bipolarer Form gegenübergestellten Komparativaussagen, für die retrospektiv im Vergleich zur Klinikaufnahme positive versus keine versus negative Veränderungen angegeben werden sollen (z.B.: „Mein Appetit hat zugenommen“ versus „... abgenommen“).

Mit einer erweiterten Version des FGV (FGV-II; Krampen & Ohm, 1985) können ergänzend über acht Zusatz-Items die erlebten herzbezogenen gesundheitlichen Veränderungen speziell bei Herzinfarktpatienten in der Psychosomatischen Rehabilitation erfasst werden. Die Befunde zur Reliabilität und Validität des FGV und FGV-II genügen den Anforderungen, die an Forschungsinstrumente zu stellen sind. Angaben zu Mittelwerten und Standardabweichungen der FGV-Skalen in unterschiedlichen Patientengruppen liegen vor, Normierungen dagegen nicht. Beide Verfahren wurden ad hoc nach vor-experimentellen Designs entwickelt.

5.6 Veränderungsprozessbogen (VPB) und Veränderungsfragebogen für Lebensbereiche (VLB)

Der VPB (Grawe, 1982) und seine Weiterentwicklung zum VLB (Itten & Grawe, 2002) dienen der direkten Erfassung psychotherapeutisch induzierter Veränderungen in der deskriptiven Einzelfallanalyse. Zwei Mal wöchentlich sollen Patienten jeweils für die letzten drei bis vier Tage Veränderungen in der allgemeinen Stimmung, im Medikamentenverbrauch, in den Zukunftserwartungen sowie in sieben Lebensbereichen (Beziehungen zu Familienangehörigen, andere enge Beziehungen, allgemeiner Kontakt und Umgang mit anderen, Beruf, Freizeit, Therapie und Symptome/Probleme) auf neunstufigen Skalen angeben.

Die sehr große Menge anfallender Daten wird im Einzelfall zeitreihenanalytisch und anhand dreier aggregierter Maße (Veränderungsindex, -spektrum und -muster) ausgewertet. Eine Normierung wurde bislang nicht vorgelegt. VPB und VLB sind *ad hoc* konstruierte Erhebungsverfahren, deren Entwicklung auf Daten aus vor-experimentellen Designs fußt.

5.7 Goal Attainment Scaling (GAS)

Das GAS wurde ursprünglich im Rahmen der Qualitätssicherung gemeindenaher gesundheitspsychologischer Programme entwickelt (vgl. Kiresuk & Sherman, 1968; Kiresuk, Smith & Cardillo, 1968) und wird seit geraumer Zeit

auch für die Evaluation der Psychotherapie (vor allem mit kognitiv-behavioralem Schwerpunkt) eingesetzt.

- Bei der GAS, der „Zielerreichungsanalyse“ nach Wittchen und Hoyer (2006, S. 380), werden im ersten psychotherapeutischen Schritt die Lebens- und Verhaltensbereiche – im Sinne einer Ziel- und Problemanalyse – bestimmt und möglichst konkret sowie verhaltensnah beschrieben, bei denen Veränderungsnotwendigkeiten und -wünsche bestehen;
- diese Zielbereiche werden im zweiten Schritt auf einer mehrstufigen Skala nach der Erwünschtheit therapeutischer Ergebnisse konkret anhand von Verhaltensindikatoren und dem Bewältigungsverhalten für kritische Lebenssituationen operationalisiert;
- im dritten Schritt wird wiederholt im psychotherapeutischen Prozess und zusätzlich am Ende der Therapie das Ausmaß der Annäherung an diese Ziele (bezogenen auf die ihnen zugeordneten konkreten Verhaltensweisen und kritischen Lebenssituationen) direkt quantitativ beurteilt.

Für dieses rein idiographische Vorgehen liegen weder Befunde zu den Gütekriterien noch zur Normierung vor. Schon die Bestimmung von Vergleichswerten für das GAS ist schwierig, da sowohl die Anzahl der berücksichtigten Lebens-, Verhaltens- und Zielbereiche als auch die Anzahl der diesen jeweils zugeordneten Verhaltensindikatoren interindividuell sehr stark variiert.

5.8 Varia

Weitere Vorschläge für Forschungsinstrumente zur direkten Veränderungsmessung – durchweg ohne Normierung – finden sich im deutschen Sprachraum für die Evaluation der Effekte von Kommunikationstrainings für Paargruppen (Henss & Böning, 1984) und von Gesundheitsmanagement-Programmen in Industrieunternehmen (Pfaff & Bentz, 2001), für subjektiv erlebte Veränderungen durch Alterungs- und Demenzsymptome bei älteren Menschen (Oswald & Fleischmann, 1995) sowie für subjektiv erlebte Veränderungen nach kritischen Lebensereignissen (Klauer, 1994). Alle Instrumente sind *ad hoc* konstruierte Erhebungsverfahren, und ihre Entwicklung basiert auf Daten aus vor-experimentellen Designs.

Literatur

- Baumann, U., Sodemann, U. & Tobien, H. (1980). Direkte versus indirekte Veränderungsdiagnostik. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 1, 201-216.
- 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.
- Campbell, D.T. & Stanley, J.C. (1963). Experimental and quasi-experimental designs for research on teaching. In N.L. Gage (Ed.), *Handbook of research on teaching* (pp. 171-246). Chicago, IL: Rand McNally.

- Cattell, R.B. (1965). The scientific analysis of personality. Harmondsworth (GB): Penguin [deutsche Übersetzung (1973). Die wissenschaftliche Erforschung der Persönlichkeit. Weinheim: Beltz].
- Crowne, D.P. & Marlowe, D. (1964). The approval motive: Studies in evaluative dependence. New York, NY: Wiley.
- 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. (1982). Der Veränderungsprozessbogen. In M. Zielke (Hrsg.), Diagnostik in der Psychotherapie (S. 231-252). Stuttgart: Kohlhammer.
- Grawe, K. (1998). Psychologische Therapie. Göttingen: Hogrefe.
- Henss, R. & Böning, U. (1984). Auswirkungen eines Kommunikationstrainings für Paargruppen aus der Sicht unterschiedlicher Beurteiler. Psychologische Beiträge, 26, 224-238.
- Itten, S. & Grawe, K. (2002). VLB: Veränderungsfragebogen für Lebensbereiche. In E. Brähler, J. Schumacher & B. Strauß (Hrsg.), Diagnostische Verfahren in der Psychotherapie (S. 382-384). Göttingen: Hogrefe.
- Kiresuk, T.J. & Sherman, R.E. (1968). Goal attainment scaling: A method for evaluating comprehensive community mental health programs. Community Mental Health Journal, 4, 443-453.
- Kiresuk, T.J., Smith, A. & Cardillo, J.E. (Eds.) (1968). Goal attainment scaling: Applications, theory, and measure. Hillsdale, NJ: Erlbaum.
- Klauer, T. (1994). Veränderungserleben und Krankheitsbewältigung. Berlin: Verlag für Wissenschaft und Forschung.
- Krampen, G. (1991). Diagnostisches und Evaluatives Instrumentarium zum Autogenen Training (AT-EVA). Göttingen: Hogrefe.
- Krampen, G. (1998). Einführungskurse zum Autogenen Training: Ein Lehr- und Übungsbuch für die psychosoziale Praxis (2. Aufl.). Göttingen: Hogrefe Verlag für Angewandte Psychologie.
- Krampen, G. (2006). Änderungssensitive Symptomliste zu Entspannungserleben, Wohlbefinden, Beschwerden- und Problembelastungen (ASS-SYM). Göttingen: Hogrefe.
- Krampen, G. (2008). Veränderungsfragebogen zum Entspannungserleben und Befinden – Psychotherapie (VFE-PT). Göttingen: Hogrefe.
- Krampen, G. & Ohm, D. (1985). Klinikwahrnehmung und Genesung von Herzinfarktpatienten in einer Rehabilitationsklinik. Rehabilitation, 24, 64-68.
- Krampen, G. & von Delius, A. (1981). Zur direkten Messung subjektiv erlebter gesundheitlicher Veränderungen. Medizinische Psychologie, 7, 166-174.
- 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.
- Michalak, J., Kosfelder, J., Meyer, F. & Schulte, D. (2003). Messung des Therapieerfolges: Veränderungsmaße oder retrospektive Erfolgsbeurteilung. Zeit-

- schrift für Klinische Psychologie und Psychotherapie, 32, 94-103.
- Oswald, W.-D. & Fleischmann, U.M. (1995). Nürnberger-Alters-Inventar (NAI; 3. Aufl.). Göttingen: Hogrefe.
- Petermann, F., Hehl, F.J. & Schneider, W. (1977). Veränderungsmessung im Rahmen der klassischen Testtheorie. In F. Petermann (Hrsg.), *Methodische Grundlagen klinischer Psychologie* (S. 129-154). Weinheim: Beltz.
- Pfaff, H. & Bentz, J. (2001). Intervention und Evaluation im DaimlerChrysler-Werk Berlin: Das Change Assessment Inventar (CAI) als Evaluationsinstrument des Gesundheitsmanagements. In B. Badura, M. Litsch & C. Vetter (Hrsg.), *Fehlzeiten-Report: Zukünftige Arbeitswelten* (S. 176-190). Berlin: Springer.
- Schmid, F. (2008). Stabilität und Prädiktion von Behandlungsergebnissen stationärer psychosomatischer Rehabilitation. Unveröffentlichte Dissertation, Universität Trier.
- 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. In M. Härter, H.W. Linster, R.-D. Stieglitz & A. Beiroth (Hrsg.), *Qualitätsmanagement in der Psychotherapie* (S. 97-117). Göttingen: Hogrefe.
- 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.
- Zielke, M. (1978). Validierung eines therapiebezogenen Veränderungsfragebogens für die Gesprächspsychotherapie an einer klinischen Klientenstichprobe. *Diagnostica*, 14, 89-102.
- Zielke, M. (1979). Kieler Änderungssensitive Symptomliste (KASSL). Weinheim: Beltz.
- Zielke, M. (1982). Probleme und Ergebnisse der Veränderungsmessung. In M. Zielke (Hrsg.), *Diagnostik in der Psychotherapie* (S. 41-59). Stuttgart: Kohlhammer.
- Zielke, M. & Kopf-Mehnert, C. (1978). Veränderungsfragebogen des Erlebens und Verhaltens (VEV). Weinheim: Beltz.
- Zielke, M. & Kopf-Mehnert, C. (2001a). 22 Jahre wissenschaftliche und klinische Erfahrungen mit dem Veränderungsfragebogen des Erlebens und Verhaltens VEV. *Praxis Klinische Verhaltensmedizin und Rehabilitation*, 53, 3-6.
- Zielke, M. & Kopf-Mehnert, C. (2001b). 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.

Günter Krampen
Fachbereich I – Psychologie und ZPID
Leibniz-Institut
Universität Trier
54286 Trier
krampen@uni-trier.de