

Prozessdiagnostik und kontrollierte Praxis

Günter Krampen und Petra Hank

- 1 Ziele der Prozessdiagnostik und Veränderungsmessung*
 - 1.1 Prozessdiagnostik und wissenschaftlich fundierte Praxis
 - 1.2 Prozessdiagnostik und Indikation
 - 1.3 Prozessdiagnostik und Qualitätssicherung

- 2 Testtheoretische Grundlagen der Veränderungsmessung*
 - 2.1 Veränderungsmessung und Klassische Testtheorie
 - 2.2 Veränderungsmessung und probabilistische Testtheorie

- 3 Methoden und Hilfsmittel der Veränderungsmessung*
 - 3.1 Indirekte Veränderungsmessung
 - 3.2 Direkte Veränderungsmessung
 - 3.3 Kriterien für klinisch bedeutsame Veränderungen

- 4 Varianten der kontrollierten Praxis*
 - 4.1 Kasuistik und Einzelfalldarstellung
 - 4.2 Einzelfallanalyse
 - 4.3 Einzelfallexperiment

- 5 Zusammenfassung*

- 6 Literatur*

Vorbereitet für:

Röhrle, B., Caspar, F. & Schlottke, P. (Hrsg.), Lehrbuch der klinisch-psychologischen Diagnostik. Stuttgart: Kohlhammer (in Vorbereitung).

Autorenangaben:

Prof. Dr. Günter Krampen und Dr. Petra Hank, Universität Trier, Fachbereich I – Psychologie, Abt. Klinische Psychologie, Psychotherapie und Wissenschaftsforschung, D-54295 Trier

1 Ziele der Prozessdiagnostik und Veränderungsmessung

Nach den von Pawlik (1982) unterschiedenen Modell- und Praxisdimensionen der psychologischen Diagnostik kommt der Prozessdiagnostik oder Veränderungsmessung die Aufgabe zu, Veränderungen in der Ausprägung von psychologischen Variablen festzustellen. Zielsetzung ist damit die Abbildung wahrer Merkmalsveränderungen im zeitlichen Verlauf. Dies erfordert die Beschreibung, ggf. auch die Erklärung und Vorhersage endogener und/oder exogen induzierter Merkmalsfluktuationen.

Endogene Merkmalsfluktuationen können das Ergebnis entwicklungsbedingter Reifungsprozesse, im liberalisierten Verständnis einer aktionalen Entwicklungsperspektive (vgl. hierzu etwa Brandtstädter, 2001) aber auch das Resultat entwicklungsregulativer Handlungsbemühungen von Personen sein. Im klinischen Bereich wird dies mit dem *Phänomen der spontanen Remission* bezeichnet. Im Einzelfall ist dies in der Regel nicht alleine durch ein spontanes, d.h., von selbst entstandenes Verschwinden (im Sinne des umgangssprachlichen „Das-wächst-sich-aus“), sondern insbesondere auch durch die Aktivierung individueller personaler und sozialer Ressourcen bedingt. Wird die Ressourcenaktivierung durch professionelle Hilfen angeregt oder verstärkt, handelt es sich bereits um exogen (mit) induzierte Merkmalsfluktuationen. Dies gilt um so mehr für solche Merkmalsveränderungen, die auf umfassendere klinische oder pädagogische Interventionen zurückgehen.

Definition: Prozessdiagnostik (oder Veränderungsmessung) ist die Beschreibung, ggf. auch die Erklärung und die Vorhersage wahrer Veränderungen eines Merkmals (oder von mehreren Merkmalen m) von einer Person (oder mehreren Personen n) über mindestens zwei (oder mehr) Zeitpunkte/Situationen k .

Veränderungsmessung wird nicht als *l'art pour l'art* betrieben, sondern bezieht sich auf mehrere Zielsetzungen, allen voran auf das Bemühen um die wissenschaftliche Fundierung des praktisch psychologischen Handelns. Als ein erstes Kriterium, nach dem sich diese Wissenschaftlichkeit bemisst, führt Perrez (1991) für die Klinische Psychologie Nachweise der beabsichtigten Wirksamkeit des therapeutischen Vorgehens an. Damit sind grundlegende Fragen der Indikation, der Effektivität und der Effizienz von Interventionen angesprochen.

- *Indikation (oder Heilanzeige)* ist die Begründung der Anwendung eines diagnostischen oder therapeutischen Verfahrens bei einem konkreten Fall einer Störung mit Krankheitswert im gesamten Prozess der Diagnostik und der Intervention.

- *Effektivität* betrifft mit Nachweisen der klinischen Wirksamkeit (efficacy) und der klinischen Brauchbarkeit in der Routinepraxis (effectiveness) die Zweckmäßigkeit von Interventionsmaßnahmen.
- *Effizienz* betrifft die Wirtschaftlichkeit von Interventionsmaßnahmen auf dem Hintergrund von Kosten-Nutzen-Analysen unter Einbezug von Alternativbehandlungen und Behandlungsverzichten.

Von zentraler Bedeutung sind bei all dem die Zielsetzungen der differentiellen und der adaptiven Indikationsstellung, mithin die Ausrichtung von Interventionsmaßnahmen an die besonderen Merkmale von Personen im *gesamten* Interventionsprozess. Dies ist zugleich ein wesentliches Element der kontrollierten Praxis, deren Anspruch sich überdies auch auf die Produktevaluation (oder Ergebnisdiagnostik) beim Abschluss einer Interventionsmaßnahme sowie die Prüfung mittel- und längerfristiger Interventionseffekte über Katamnesen bezieht (siehe hierzu etwa Petermann, 1982). Damit liefert die Veränderungsmessung nicht zuletzt einen entscheidenden Beitrag zur Qualitätssicherung in der psychotherapeutischen Versorgung. In der klinischen Anwendungspraxis bildet sie, die kontrollierte Praxis, das Pendant zum Ansatz der evidenzbasierten Psychotherapie (und Medizin), der auf Befunden kontrollierter Studien mit größeren Stichproben aus der klinischen Forschung basiert.

Bevor auf die teststatistischen Grundlagen der Veränderungsmessung eingegangen wird, unterschiedliche Arten der und Hilfsmittel für die Veränderungsmessung erläutert sowie deren Anwendungen in verschiedenen Varianten der kontrollierten Praxis beschrieben werden, werden die benannten Zielsetzungen der Prozessdiagnostik und Veränderungsmessung erläutert.

1.1 Prozessdiagnostik und wissenschaftlich fundierte Praxis

Konstituierende Problemtypen der professionellen psychologischen Praxis sind nach Brocke (1979) Erklärungs-, Prognose- und Gestaltungsprobleme, die sich in ihrem Zusammenspiel als komplexe pragmatische Ausgangsfragen, wie z. B. die Frage nach dem angemessenen *Treatment* zur Beseitigung eines Verhaltensdefizits bei einem Patienten, präsentieren. Der Lösung einer solchen Problemkonstellation kann nach Perrez (1991) eine wissenschaftliche Fundierung attestiert werden, wenn für das mit wissenschaftlichen Erkenntnissen zu vereinbarende therapeutische Vorgehen (zur Erreichung von ethisch legitimen Behandlungszielen mit ethisch vertretbaren und effizienten Methoden unter Befolgung von nomopragmatisch abgeleiteten Handlungsregeln) die beabsichtigte Wirksamkeit unter Ausbleiben von schädigenden Nebenwirkungen nachgewiesen ist.

Diese Forderung nach Wissenschaftlichkeit hat die Fundierung der Anwendung durch die Theorie zum Ziel. Zur Transformation von grundlagentheoretischem Wissen in die klinisch-psychologische Praxis *et vice versa* existieren unterschiedliche wissenschaftstheoretische Modelle (zur Relation zwischen Grundlagenforschung und wissenschaftlich fundierter Anwendung siehe etwa Brocke, 1993; Bunge, 1967; Herrmann, 1979; Perrez, 1991; Scheele, 1982; Westmeyer, 1978, 1979). Bunge (1967) führte in diesem Zusammenhang das Konzept der technologischen Regel ein. Danach ist die praktische Anwendung als eine technologische Regel oder Aussage zu verstehen, die über „den Umweg“ von nomopragmatischen Aussagen aus nomologischen Gesetzesaussagen abgeleitet werden kann. Die Güte dieser technologischen Aussagen bemisst sich nach ihrer Effektivität. Aber nicht nur der Nachweis, *dass* die Regeln in der beabsichtigten Art und Weise wirken, sondern auch *wie* sie wirken, sollte Gegenstand wissenschaftlich kontrollierter Untersuchungen sein: „To decide that a rule is effective it is necessary, though insufficient, to show that it has been successful in a high percentage of cases. (...) Before adopting an empirically effective rule we ought to know why it is effective: we ought to take it apart and reach an understanding of its *modus operandi*. (...) Now, the solid foundation of a rule is a system of law formulas, because these alone can be expected to correctly explain facts – e.g., the fact that a given rule works.” (Bunge, 1967, S. 132-133). Erst die Erklärung der Wirkmechanismen einer technologischen Regel ermöglicht es, technologische Regeln zu optimieren oder sie ggf. gegen effektivere auszutauschen.

Die Erklärung des *modus operandi* strebt Petermann (1982) mit dem *Konzept der kontrollierten Praxis* an. Es sieht vor, Forschungsansätze der Klinischen Psychologie in die therapeutische Praxis zu übertragen und Einzelfallanalysen als kontrollierte Fallexperimente zu betreiben. Als mögliche Erkenntnisziele solcher Einzelfallbetrachtungen benennt Petermann (1982):

- (1.) die Generierung von Hypothesen und wissenschaftliche Illustration,
- (2.) die Prüfung von Hypothesen unter (kontrollierbaren) Laborbedingungen,
- (3.) die Analyse des Einzelfalls mit diagnostischer oder prognostischer Zielsetzung in Verbindung mit der Evaluation von handlungsrelevantem Hintergrundwissen (Hypothesen, Gesetze und Theorien),
- (4.) das Aufstellen von Anwendungsregeln und Effektivitätsbehauptungen,
- (5.) den Vergleich verschiedener Anwendungsregeln und die Bestimmung ihrer relativen Effektivität,
- (6.) die einzelfallanalytische Beurteilung von Anwendungsregeln in der Praxis.

Die zentrale Frage lautet demnach: „Für welche Personen lassen sich durch welche Durchführenden welche Programmziele unter welchen Interventionsbedingungen mit welchen Interventionsmethoden am besten erreichen?“ (Patry & Perrez, 2000, S. 30). Als praktische Aufgaben resultieren neben der fallspezifischen Indikationsstellung, die (formative) Prozessevaluation und die (summative) Produktevaluation. Der wissenschaftlich fundiert arbeitende Praktiker realisiert sie u.a. durch den regelmäßigen bzw. möglichst weitgehenden Einsatz wissenschaftlich fundierter Diagnostik (Brocke, 1993).

Gleichwohl die Beantwortung dieser Fragen primär auf die Fundierung der Praxis durch die Theorie abzielt, ist auch der umgekehrte Weg, die Aufarbeitung theoriespezifischer Defizite oder die theoretische Rekonstruktion psychologischen Anwendungswissens sensu Brocke (1993) sinnvoll. Dies selbst dann, wenn nur ein (loser) heuristischer Zusammenhang zwischen den theoretischen Grundlagen und ihren praktischen Anwendungen besteht. Beispielhaft verweist Bastine (1998) in diesem Zusammenhang auf handlungs- und problemlösetheoretische Ansätze (Kanfer, Reinecker & Schmelzer, 1996) oder schematheoretische Modelle (Grawe, 1988a).

Bedeutsame Schritte einer kontrollierten Anwendungspraxis können in einem einfachen *Ablaufschema des diagnostischen Abklärungsprozesses und des adaptiven Interventionsprozesses* prototypisch beschrieben werden (siehe Abbildung 1). Prozessdiagnostik wird bereits im diagnostischen Abklärungsprozess mit der Vorbereitung der indirekten Veränderungsmessung im Rahmen der allgemeinen Beschwerdendiagnostik und der problembezogenen standardisierten Psychodiagnostik relevant. Während des Interventionsprozesses wird Prozessdiagnostik über die sukzessive Zielkontrolle und andere Maßnahmen der interventionsbegleitenden Diagnostik (etwa über Verlaufsprotokollierungen und Therapiestundenbogen) betrieben. Bei Therapieabschluss und in katamnesticen Untersuchungen manifestiert sich die Prozessdiagnostik in der direkten und indirekten Veränderungsmessung sowie in der Diagnose von Residuen (= bleibender Rückstand einer Störung nach einer Behandlung) und von Rezidiven (= Rückfall, Wiederauftreten einer Störung nach der Abheilung).

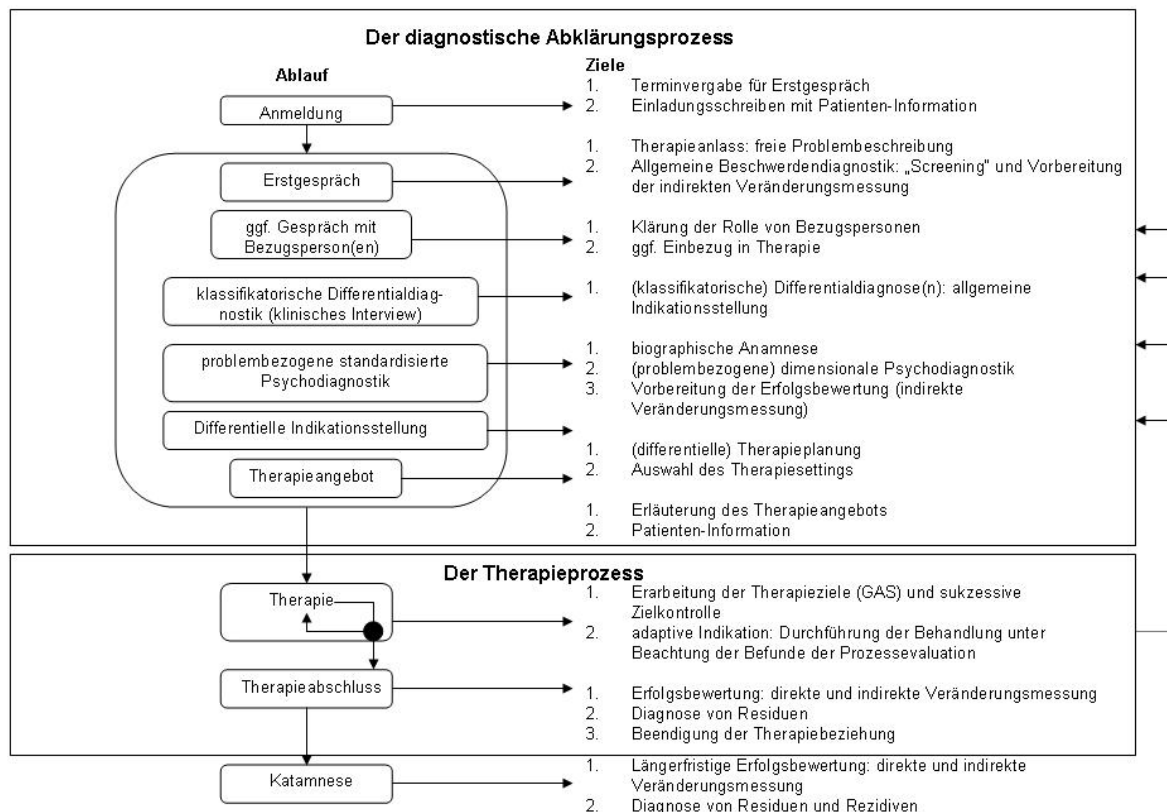


Abbildung 1: Ablaufschema des diagnostischen Abklärungsprozesses und des adaptiven Therapieprozesses in der kontrollierten klinisch-psychologischen Praxis (modifiziert nach Grawe, 1998, S. 632 und Krampen, 2002a, S. 22)

1.2 Prozessdiagnostik und Indikation

Die Indikation (*lat. indicare: anzeigen*), die sogenannte „Heilanzeige“, ist die Begründung der Anwendung eines bestimmten diagnostischen oder therapeutischen Verfahrens bei einem konkreten Fall einer Störung mit Krankheitswert im *gesamten* Prozess der Diagnostik und der Intervention. Die Anwendung der Methode ist mit der Indikationsstellung hinreichend nach den *Kriterien der Notwendigkeit, Zweckmäßigkeit und Wirtschaftlichkeit* (etwa nach dem deutschen Sozialgesetzbuch V; SGB V) zu rechtfertigen. Dabei besteht grundsätzlich Aufklärungspflicht gegenüber Patienten, die auch die Information über alternative Behandlungszugänge und Risiken umfassen muss. In all diesen Bestimmungstücken unterscheidet sich eine *professionelle Indikation* von nicht-professionellen „Indikationen“, die im Lebensalltag zum Teil stigmatisierend und diskriminierend (etwa nach dem Motto „Du brauchst ´ne Therapie“), zum Teil aber auch aus Besorgnis aufgrund fehlerhafter Zuschreibungen erfolgen.

Professionelle Indikationsstellungen implizieren im Rahmen der kontrollierten Anwendungspraxis minimal drei Aspekte (siehe auch Abbildung 1):

(1.) Die *allgemeine Indikationsstellung* bezieht sich auf die klassifikatorische Differentialdiagnose einer (ggf. komorbiden) Störung nach ICD-10 oder DSM-IV-TR, der Krankheitswert nach SGB V und damit Behandlungsnotwendigkeit zukommt.

- (2.) Aus dieser allgemeinen Indikationsstellung wird eine *differentielle Indikationsstellung*, wenn Personmerkmale und Interventionsmethoden systematisch unter Rückgriff auf explikative, normative und prognostische Hypothesen aufeinander bezogen werden. Voraussetzung dafür ist eine über die klassifikatorische Diagnostik hinausgehende ätiologisch und prognostisch ausgerichtete (Entwicklungs-)Diagnostik.
- (3.) Die differentielle Indikationsstellung wird zu einer *adaptiven Indikation*, wenn die Diagnostik interventionsbegleitend mit dem Ziel durchgeführt wird, Personmerkmale und Interventionsmethoden während der Behandlung systematisch unter Rückgriff auf explikative, normative und prognostische Hypothesen aufeinander zu beziehen, um die Behandlungsprognose positiv zu gestalten, ggf. zu verbessern.

Allgemeine und differentielle Indikationsstellung werden zusammenfassend auch häufig als *selektive Indikation* bezeichnet, da sie sich auf das Ergebnis des diagnostischen Abklärungsprozesses vor der Behandlung und damit die (selektive) Zuweisung von Patienten zu bestimmten Behandlungsmaßnahmen beziehen (siehe Abbildung 1). Prozessdiagnostik wird hier systematisch vorbereitet, aber nicht realisiert. Prozessdiagnostik ist vielmehr die Basis der *adaptiven, verlaufsorientierten Indikation* und – darüber hinaus – auch der mittel- und längerfristigen Erfolgskontrolle von Interventionsmaßnahmen. In Tabelle 1 sind diese unterscheidbaren Aspekte der Indikation in eine allgemeine Systematik der klinisch-psychologischen Diagnostik und Evaluation eingeordnet.

Auf dem Hintergrund der weitgehend konsensuellen Aufgabenbestimmung der Psychologie, die sich auf die Beschreibung, Erklärung und Vorhersage von Verhalten und Erleben, die auf Verhalten und Erleben bezogene Zielanalyse sowie die Intervention zur korrektiven oder präventiven Veränderung von Verhalten und Erleben und deren Evaluation bezieht, werden in Tabelle 1 die folgenden *diagnostischen Termini technici* unterschieden:

- (1.) deskriptive Statusdiagnostik (als Basis der allgemeinen Indikation),
- (2.) entwicklungsbezogene Diagnostik (als Basis der differentiellen Indikation),
- (3.) (formative) Prozessdiagnostik (als Basis der adaptiven Indikation),
- (4.) (summative) Produkt- oder Ergebnisdiagnostik (als Basis der Behandlungsbewertung).

Die diesen diagnostischen Schritten spezifisch zugeordneten *diagnostischen Strategien* beziehen sich auf:

- (1.1) das *Screening* anhand freier Problembeschreibungen von Patienten, die etwa durch den Einsatz standardisierter Beschwerdenlisten sinnvoll ergänzt werden können,

- (1.2) die deskriptive klassifikatorische Diagnostik nach der ICD-10 oder dem DSM-IV-TR anhand teilstrukturierter klinischer Interviews,
- (2.1) die ätiologisch orientierte Diagnostik mit Hilfe biographischer Anamnesen, psychodynamischer Explorationen, Verhaltens- und Problemanalysen sowie problembezogener standardisierter psychometrischer Instrumente,
- (2.2) die prognostisch orientierte Diagnostik von Therapiemotivation und anderen Therapievoraussetzungen, Ressourcen und Beziehungsmustern der Patienten sowie ihren Konfliktmustern und Bewältigungsstilen,
- (3.0) die interventionsbegleitende Diagnostik anhand von Sitzungs- und Verlaufsprotokollierungen sowie Therapiestundenbogen,
- (4.1) die Veränderungsdiagnostik über Vorher-Nachher-Vergleiche (indirekte Veränderungsmessung) und direkte Veränderungsmessungen sowie
- (4.2) die Katamnese über Nachuntersuchungen mit Hilfe von Katamnesebogen und/oder durch Einsatz von Methoden der direkten und indirekten Veränderungsmessung.

Tabelle 1: Systematik der klinisch-psychologischen Diagnostik und Evaluation mit zugeordneten Aufgabenstellungen der Psychologie und diagnostischen Strategien (modifiziert nach Krampen, 1998, S. 46-47)

Aufgabenstellung der Psychologie	Beschreibung von Verhalten und Erleben		Bedingungs- und Zielanalyse sowie Prognose von Verhalten und Erleben		Psychologische Intervention	Evaluation	
Terminus technicus	(deskriptive) Statusdiagnostik		(dynamische) Entwicklungsdiagnostik		(formative) Prozessevaluation	(summative) Produktevaluation	
diagnostische Strategie	Screening	deskriptive Differentialdiagnostik	ätiologisch orientierte Diagnostik	prognostisch orientierte Diagnostik	interventionsbegleitende Diagnostik	Veränderungsdiagnostik	Katamnese
Methoden (exemplarisch)	freie Problem- beschreibung Beschwerdenlisten - SCL-90-R - KASSL - GBB - AT-EVA: SYM	ICD-10 DSM-IV-TR klinische Interviews - SKID - IDCL, IDCL-P - SCAN - SIDAM - IPDE - DIPS, MINI-DIPS	Biographische Anamnese psychodynamische Exploration; motivationale Schemata Verhaltens- und Problemanalysen: Störungskomponenten problembezogene standardisierte (dimensionale) Psychodiagnostik - BDI, H-Skalen - IAF, STAI - TAI, MALT - PSSI, FPI-R	Therapiemotivation und Therapie- voraussetzungen - OPD - FMP, TEB, FLL - Intelligenztests Ressourcen- und Beziehungsdagnostik - OPD - SONET - PFU Konfliktmuster und Bewältigungsstile - OPD - SVF, FKV - ISE, ISE-Z Lebenssituation Therapiezielklärung - Goal Attainment Scaling (GAS)	Verlaufsprotokollierung - Aktenführung - Protokollbogen Therapiestundenbogen - SB-K, SB-T - BIKEB - STEP Verlaufsevaluation anhand - von Lebensveränderungen - von Bezugspersonen-Info. - wiederholter standardisierter Psychodiagnostik - Goal Attainment Scaling (GAS)	Indirekte Veränderungs- messung Direkte Veränderungs- messung	Nachunter- suchung Wie Veränderungs- diagnostik Katamnesebogen
Resultat	Hypothese über Krankheitswertigkeit	Klassifikatorische ICD-10- bzw. DSM- IV-Diagnose(n)	Hypothese(n) zur Ätiologie/Genese	- Hypothese(n) zu zielführenden Maßnahmen - Therapieplanung	- Evaluation therap. Ansprechbarkeit - Veränderung der ICD- 10/DSM-IV-Diagnosen - Adaptive Therapiegestaltung	Residuen nach ICD-10/DSM-IV-TR?	
	⇔ allgemeine Indikationsstellung		⇔ differentielle Indikationsstellung		⇔ adaptive Indikation	⇔ Behandlungsbewertung?	

Neben dieser auf den Diagnose- und Interventionsprozess bezogenen Systematisierung unterschiedlicher Konzepte der Indikation können mit Mans (1998) drei *klassische Indikationsmodelle* nach den Theorien unterschieden werden, die bei der Indikationsstellung für die Prognose von Behandlungserfolgen verwendet werden:

- Im störungs- und symptomorientierten *Diagnose-Modell der Indikation* werden aufgrund der klassifikatorischen Diagnostik nach ICD-10 oder DSM-IV-TR und aufgrund der vorliegenden Effektivitätsnachweise aus kontrollierten empirischen Studien für bestimmte Störungsbilder weitgehend theoriefrei aufgrund der empirischen Befundlage Therapieplanungen vorgenommen (vgl. hierzu etwa Fiedler, 1996, 2000; Schneider & Margraf, 2005).
- Nach dem *Persönlichkeitsmodell der Indikation* erfolgt die Vorhersage anhand von Persönlichkeitsmerkmalen unter engem Bezug auf eine Persönlichkeitstheorie (vgl. etwa Becker, 1995; Beutler & Clarkin, 1990).
- Im *Beziehungsmodell der Indikation* stehen Erwartungen, Motivationen und ähnliche Behandlungsvoraussetzungen beim Patienten im Vordergrund der Therapieplanung (vgl. etwa Schneider, 1990).

Während der Schwerpunkt dieser drei von Mans (1998) aufgeführten „klassischen“ Indikationsmodelle auf der selektiven Indikationsstellung liegt, umfassen *Wirkfaktoren-Modelle der Indikation* explizit auch die Aufgabenstellungen der adaptiven Indikation. Dies gilt etwa für das von Grawe, Donati und Bernauer (1994) vorgelegte Dreikomponentenmodell der Wirkungsweise von Psychotherapie, da angenommen wird, dass allgemeine (schulübergreifende) Wirkfaktoren wie (1) die aktive Hilfe zur Problembewältigung, (2) die motivationale Klärung sowie (3) die therapeutische Unterstützung und Beziehung am gesamten Interventionsprozess mit unterschiedlichem Gewicht beteiligt sind, dabei stetig zusammenwirken und gemeinsam für Behandlungserfolge verantwortlich sind. Daher sind diese Wirkfaktoren nicht nur zu Behandlungsbeginn für die selektive Indikation und Therapieplanung direkt relevant, sondern insbesondere auch in allen Behandlungsphasen für die adaptive Indikation. Nur so kann im Interventionsprozess adaptiv und flexibel der jeweils indizierte Wirkfaktor schwerpunktmäßig verstärkt bzw. der aktuell weniger indizierte Wirkfaktor reduziert werden.

- Das Wirkfaktoren-Modell der Indikation entspricht dem Ansatz einer *schulenspezifischen oder schulübergreifenden Indikation* und repräsentiert gut die dritte Phase des Drei-Phasen-Modells zur Geschichte der Psychotherapie im 20. Jahrhundert (Krampen, 2002b), die durch den Ansatz einer Allgemeinen und Differentiellen Psychotherapie gekennzeichnet ist (siehe hierzu Grawe et al., 1994; Grawe, 1998; Krampen, 2002a).

- Der zweiten Phase der Psychotherapiegeschichte, die mit dem Schlagwort des Eklektizismus (vgl. hierzu Krampen, 2002b) beschrieben werden kann, entspricht das Konzept der *schulenvergleichende Indikation*.
- Der ersten Phase der Psychotherapiegeschichte, die sich auf die Begründung und Neuentwicklung eigenständiger psychotherapeutischer Behandlungsansätze (Therapieschulen) bezieht, entspricht das Konzept der *schulenspezifischen Indikation* mit ihrem Universalitätsanspruch.

1.3 Prozessdiagnostik und Qualitätssicherung

Ein weiterer Grund, Prozessdiagnostik zu betreiben, ergibt sich vor dem Hintergrund der Forderung nach Qualitätssicherung und Qualitätsmanagement in der ambulanten und stationären Psychotherapie sowie der psychosozialen Versorgung. Dieser Gedanke an sich ist nicht neu. Neu ist in diesem Zusammenhang nach Laireiter und Vogel (1998, S. 19) vielmehr „die systematische Implementierung der Qualitätsphilosophie und von Methoden der ständigen Leistungsverbesserung“, begründet in gesundheitspolitischen und wirtschaftlichen Überlegungen, Regelungen zum Patientenschutz und der Patientenrechte sowie in versorgungspolitischen und wissenschaftlichen Aspekten.

In Deutschland nahm die Entwicklung der Qualitätssicherung im Gesundheitssystem ihren Anfang im Globalziel 31 der Weltgesundheitsorganisation (WHO) und den daraus resultierenden Veränderungen der Gesetzgebung. Die §§ 135 – 138 des SGB V sowie die ihnen nachgeordneten Psychotherapierichtlinien regeln die Qualitätssicherung in der Gesundheitsversorgung und damit auch im Hinblick auf die psychotherapeutische Versorgung mit dem Ziel, einen für alle Patienten gleichmäßig hohen Leistungsstandard im ambulanten und stationären Versorgungsbereich zu gewährleisten.

Was ist gemeint, wenn von Qualitätssicherung resp. Qualitätsmanagement die Rede ist? Nübling und Schmidt (1998a) subsumieren unter Qualitätssicherung „eine Reihe verschiedener Ansätze und unterschiedlicher Maßnahmen (...), deren globales Ziel darin besteht, die Qualität gesundheitsbezogener Dienstleistungen, d.h. die konkrete Versorgungspraxis systematisch und kontinuierlich zu hinterfragen, zu bewerten, zu fördern und – im Falle von Mängeln oder Schwachstellen – zu verbessern“ (S. 51). So verstanden ist Qualitätsmanagement eine Facette der Qualitätssicherungsmaßnahmen, die sich mit den organisatorischen bzw. strategischen Komponenten von qualitätsorientierten Systemen beschäftigt. Demgegenüber betonen Rückert und Linster (1998), dass Qualitätssicherung

primär ergebnisorientiert ist, während Qualitätsmanagement darauf abziele, Qualität durch eine kontinuierliche Qualitätsförderung prospektiv zu sichern.

Es geht also bei der Herstellung und Sicherung von Qualität um die Feststellung und Bewertung eines vorhandenen Ist-Zustandes vor dem Hintergrund eines gesetzten Soll-Zustandes (siehe hierzu auch das Paradigma der Qualitätssicherung von Viethen, 1994) und dies im Hinblick auf Struktur, Ergebnis und Prozess (Donabedian, 1966). Letztgenannte Dimension führt geradewegs zur Veränderungsmessung. Angewandt auf den Kontext der Psychotherapie stehen bei der Prozessorientierung alle für die Behandlung und Beratung der Patientinnen und Patienten erforderlichen Tätigkeiten auf dem Prüfstand. Mit anderen Worten: Die Patientenbehandlung ist das Herzstück, um das sich alle anderen Arbeitsprozesse gruppieren (Janßen, 1997). Um diesen Anspruch zu realisieren, bedarf es einerseits empirisch bewährter oder konsensuell begründeter diagnostischer und therapeutischer Maßnahmen, andererseits qualitätsdefinierender Standards (Kriterien), anhand derer die erreichte Qualität kontinuierlich bewertet werden kann.

2 *Testtheoretische Grundlagen der Veränderungsmessung*

2.1 Veränderungsmessung und Klassische Testtheorie

Mit den genannten Zielsetzungen der Veränderungsmessung verbinden sich bestimmte (methodologische) Ansprüche an die zugrunde liegende Testtheorie. Erfasst werden sollen wahre Veränderungen der Merkmalsausprägung im zeitlichen Verlauf, eine mit den Axiomen der Klassischen Testtheorie (Gulliksen, 1950; Lord & Novick, 1968) nicht zu vereinbarende Zielsetzung, da die Definition des *True Scores*, also des „wahren“ Merkmalswertes, ausschließlich zeit- und bedingungs*invariante* Merkmale zulässt und intraindividuelle Veränderungen in der Merkmalsausprägung als Messfehler interpretiert werden (siehe etwa Lienert, 1989). Differenzwerte zwischen einem ersten und zweiten Messzeitpunkt (z. B. die Differenz der Messwerte vor bzw. nach Implementation einer Trainingsmaßnahme) vereinen demzufolge die Messfehler beider Messungen auf sich. Diesem Dilemma entgeht man nur, wenn zwischen den in Differenz gesetzten Messwerten keine positive Korrelation besteht, also etwas inhaltlich Verschiedenes gemessen wird. Eine inhaltliche Interpretation der Differenzwerte ist dann aber nicht mehr möglich (vgl. hierzu Zielke, 1982). Nach Pawlik (1982) „sollte es auch gar nicht überraschen, dass ein Messmodell, das in den wahren Werten Veränderungen ausschließt, testkonstruktiv nichts mehr bringt, wenn es just unter der Fragestellung verwandt wird, die im Rahmen dieses Modells nicht erklärt ist: nämlich zur

Messung von wahren Veränderungen in diesen wahren Werten“ (S. 26). Als Ausweg im Rahmen der Klassischen Testtheorie schlägt er die Annahme von zwei „wahren“ Quellen interindividueller Unterschiede vor: *intraindividuell stabile* Werte und *intraindividuell variable* Werte, um die Frage nach dem auf wahre bedingungs*invariante* und dem auf wahre bedingungs*variable* Merkmalsvarianz zurück gehenden Prozentanteil beantworten zu können (Pawlik, 1978). Diese Überlegungen implizieren neben einer Theorie paralleler Tests eine Theorie paralleler Situationen, d.h. von Situationen, die im Hinblick auf die wahren Merkmalschwankungen äquivalent sind. Neben den Varianzquellen der n Personen und der m parallelen Testverfahren ergibt sich als weitere Varianzquelle die der k parallelen Situationen, die die zweidimensionale Datenstruktur auf eine dreidimensionale erweitert. Abbildung 2 zeigt die resultierende Datenstruktur.

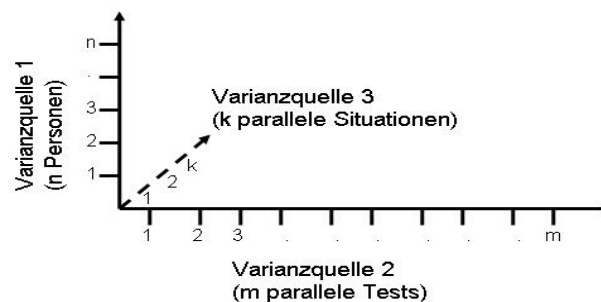


Abb. 2. Datenstruktur der Veränderungsmessung

Betrachtet man diesen Datenkubus (Abb. 2) als dreifaktoriellen Versuchsplan mit den unabhängigen Faktoren Personen, Items und Zeitpunkte, lassen sich in dem Haupteffektmodell die Einflüsse der drei Faktoren auf das Antwortverhalten separat voneinander analysieren. Dabei wird angenommen, dass Veränderungen für alle Personen und Items in gleichem Ausmaß stattfinden. In Analogie zur Varianzanalyse bildet die Interaktion zwischen den Faktoren Person und Items personspezifische Veränderungen ab. Vergegenwärtigt man sich die zugehörige (zweidimensionale) Datenorganisation, werden allerdings die genannten klassischen Probleme der Veränderungsmessung deutlich. Darüber hinaus fehlt nach wie vor der für die (herstellende) Realisation paralleler Situationen von Pawlik (1982) geforderte ökopsychologische Bezugsrahmen.

2.2 Veränderungsmessung und probabilistische Testtheorien

Der Versuch, die Probleme der Veränderungsmessung durch Erweiterung von zwei auf drei Faktoren des Antwortverhaltens zu umgehen, ist auch im Rahmen probabilistischer Testtheorien durch die Einführung des Parameters „Zeitpunkteffekt“ unternommen worden. Rost und Spada (1983) führen die Verallgemeinerung für quantitative Testmodelle auf die dreifaktorielle Datenstruktur Personenfähigkeit, Itemschwierigkeit und Zeitpunkteffekt aus. In Abhängigkeit der getroffenen Annahmen ergibt sich ein System von acht Veränderungsmodellen, die sich darin unterscheiden, ob sie Veränderungen itemspezifisch, personenspezifisch und/oder global modellieren. Neben dem (restriktiven) dreifaktoriellen Raschmodell, das unter der Annahme, Veränderung finde für alle Personen und Items in gleichem Ausmaß statt, die separate Analyse der drei Faktoren erlaubt, ergeben sich durch Berücksichtigung der verschiedenen Wechselwirkungseffekte zwischen je zwei der drei Einflussfaktoren drei Modelle, die personenspezifische Veränderungen, itemspezifische Veränderungen sowie eine Analyse des globalen (Therapie-)Effektes ermöglichen. Gemeinsam ist diesen Modellen, das sie sich auf das (normale) zweifaktorielle Raschmodell reduzieren lassen – mit den bekannten klassischen Problemen der Veränderungsmessung: Die Messgenauigkeit der Differenzen der geschätzten Wechselwirkungsparameter zwischen den Zeitpunkten ist gering, die Differenzen korrelieren negativ mit dem ersten Zeitpunktparameter und sind nur dann valide interpretierbar, wenn tatsächlich der Faktor Personen resp. Items für alle realen und virtuellen Item- bzw. Personengruppen konstant ist. Auf der nächst höheren Verallgemeinerungsstufe existieren drei Modelle mit je zwei Wechselwirkungsparametern. Will man auf dieser Ebene item- und personenspezifische Veränderungen untersuchen, stellt sich das Validitätsproblem, da alle Personen zu jedem neuen Zeitpunkt andere Fähigkeiten und alle Items andere Schwierigkeiten haben, die ursprüngliche Invarianzvoraussetzung von Rasch (1960) also aufgegeben wird.

Ein zweiter Ansatz, die mit der Messung von Veränderungen verbundenen Probleme in den Griff zu bekommen, stellen so genannte *dynamische Testmodelle* dar (zum Konzept des dynamischen Testens siehe Guthke, Böttcher & Sprung, 1991). Hier wird Veränderung so verstanden, dass sich die Fähigkeit der Person *während* der Testvorgabe verändert, die Person lernt mit jedem Item, das sie bearbeitet dazu. Damit wird das Stabilitätsprinzip der Personenfähigkeit aufgegeben. Diese Lernprozesse können wiederum personenspezifisch, itemspezifisch oder global konzipiert werden (vgl. Rost, 1996). Durch die Einführung des Parameters „Lernfähigkeit“ als zweiten Personenparameter wird beispielsweise die individuelle Lernfähigkeit parametrisiert. Um dabei das allen Modellen mit

personenspezifischem Lernen inhärente Problem der Veränderungsmessung, die messfehlerbedingte Korreliertheit der beiden Parameter zu umgehen, schätzen Klauer und Sydow (1992) die Korrelation Personfähigkeit und Lernfähigkeit unter der Annahme ihrer bivariaten Normalverteilung. Itemspezifische Lerneffekte und reaktionskontingente Lernprozesse lassen sich mittels Varianten des linear-logistischen Testmodells (*LLTM*; Fischer, 1974), einem Spezialfall des normalen Raschmodells, untersuchen (siehe hierzu zusammenfassend Rost, 1996).

Ein dritter Ansatz, die Testtheorie für die Veränderungsmessung tauglich zu machen, besteht in der Bestimmung der Ursachen einer Veränderung. Hier geht es darum, die Wirksamkeit verschiedener Veränderungsmaßnahmen zu quantifizieren und zwischen Personengruppen vergleichend darzustellen. Da einfache Mittelwertvergleiche von Personenparametern voraussetzen, (1.) dass allen Personen - auch wenn sie unterschiedlichen Maßnahmen ausgesetzt wurden - dieselben Items vorgegeben werden, (2.) die Personenparameter mit einem mehr oder weniger großen Schätzfehler behaftet sind und (3.) keine Untersuchungen ausgewertet werden können, bei denen dieselben Personen unterschiedlichen *Treatment*-Kombinationen ausgesetzt sind, empfiehlt Rost (1996) auch hier die Anwendung des linear-logistischen Testmodells. Insbesondere das linear-logistische Modell mit abgeschwächten Annahmen (linear logistic model with relaxed assumptions, *LLRA*), das auf die Annahme der Itemhomogenität zu allen Messzeitpunkten verzichtet, bietet einen eleganten Ausweg im Hinblick auf die Validitätsproblematik. Das Modell bildet eine globale Veränderung ab und dies sowohl im Hinblick auf die Items als auch bezüglich der Personen. Anders ausgedrückt. Es wird angenommen, dass der Effekt der Maßnahmen für alle Items und alle Personen gleich groß ist, unabhängig davon, ob die Items homogen sind bzw. dasselbe messen.

3 *Methoden und Hilfsmittel der Veränderungsmessung*

Veränderungen in Merkmalsausprägungen von Personen zwischen verschiedenen Zeitpunkten können indirekt oder direkt erfasst werden. Bei der indirekten Veränderungsmessung werden die Differenzen zwischen unabhängig erfassten Vortestwerten und Nachtestwerten verwendet. Aus diesen Differenzwerten wird das Ausmaß der Merkmalsveränderung vom Diagnostiker indirekt erschlossen. Demgegenüber wird bei der direkten Veränderungsmessung eine Änderung anhand von Veränderungsaussagen im Zeitvergleich (z.B. „Ich fühle mich weniger gehetzt.“) direkt abgelesen (vgl. hierzu etwa Baumann, Sodemann & Tobien, 1980; Krampen

& von Delius, 1981). Diese Veränderungsaussagen können Selbstaussagen von Patienten oder Fremdbeurteilungen von Diagnostikern sein.

3.1 Indirekte Veränderungsmessung

Bereits 1963 hat Bereiter auf die methodischen Probleme der *indirekten Veränderungsmessung* über Differenzwerte von Vor- und Nachtesthebungen aufmerksam gemacht: (1.) Regressionseffekte, (2.) das schon beschriebene Reliabilitäts-Validitäts-Dilemma und (3.) das Physikalismus-Subjektivismus-Dilemma (Bedeutsamkeitsproblem) schränken die Aussagekraft dieser Daten ein. Bei der statistischen Regression sind die Richtungen der Zufallsänderungen eindeutig determiniert, da Vortestwerte und Differenzwerte negativ korreliert sind, zwischen den Nachtestwerten und den Differenzwerten dagegen eine positive Korrelation besteht. Diese statistische Abhängigkeit führt dazu, dass z.B. bei Patienten mit höheren Ausgangswerten höhere Änderungswerte erwartet werden als bei Patienten mit mittleren oder niedrigeren Werten im Vortest. Dieses Artefakt verleitet dazu, die Veränderung bei Patienten mit hohen Vortestwerten zu überschätzen. Was das Physikalismus-Subjektivismus-Dilemma betrifft, so stellt sich die Frage, ob dem Kontinuum auf der physikalischen Messwertebene ein Kontinuum auf der subjektiv-psychologischen Ebene entspricht. Haben gleiche Differenzwerte zwischen verschiedenen Punkten der Messwertskala die gleiche subjektive Bedeutung für Patienten?

Vortest-Nachtest-Differenzen können in der Prozessdiagnostik anhand einer Vielzahl standardisierter psychometrischer Instrumente bestimmt werden. Minimal sind dabei drei Gruppen diagnostischer Hilfsmittel zu unterscheiden (zu einer allgemeinen Systematik klinisch-psychologischer Erhebungsmethoden siehe Stieglitz in diesem Band):

1. *Störungsübergreifende Verfahren* wie etwa Beschwerden- und Symptomlisten sowie Breitband-Persönlichkeitsinventare, mit denen über Selbstauskünfte und/oder Fremdeinschätzungen von Patienten ein größeres Spektrum von Personmerkmalen erfasst wird.
2. *Störungsspezifische Verfahren*, mit denen über Selbstauskünfte und/oder Fremdeinschätzungen gezielt die Personmerkmale erhoben werden, denen nach den Befunden aus der klassifikatorischen Diagnostik für die Indikation und die Bewertung des Behandlungserfolges eine besondere Bedeutung zukommt.
3. *Therapiestundenbogen*, mit deren Hilfe Aspekte des Erlebens und der Wahrnehmung einzelner Therapiesitzungen über Selbstauskünfte und/oder Fremdeinschätzungen von Patienten wiederholt erfasst werden.

ad 1: Beispiele für Störungsübergreifende Verfahren

- Die 90 fünffach gestuften Items der *Symptom Check Liste-90-R (SCL-90-R*; deutsche Adaptation von Franke, 2002) erfassen die subjektiv erlebte Befindlichkeit einer Person in den vorangegangenen sieben Tagen in den Bereichen Somatisierung, Zwanghaftigkeit, Unsicherheit im Sozialkontakt, Depressivität, Ängstlichkeit, Aggressivität/Feindseligkeit, Phobische Angst, Paranoides Denken und Psychotizismus. Drei zusätzliche globale Kennwerte lassen Schlussfolgerungen auf die grundsätzliche psychische Belastung zu und geben Auskunft über Intensität und Anzahl belastender Symptome. Nach Franke (2002) schließt die SCL-90-R „die Lücke zwischen zeitlich extrem variabler Befindlichkeit und zeitlich überdauernder Persönlichkeitsstruktur“ und eignet sich damit zur Messwiederholung. Die Messgenauigkeit der Skalen ist mit internen Konsistenzen zwischen $.51 \leq \alpha \leq .89$ als mäßig bis gut zu bewerten. Die Reliabilität des Testprofils bleibt aufgrund hoher Skaleninterkorrelationen unbefriedigend, was Hessel, Schumacher, Geyer und Braehler (2001) dazu Anlass gab, ihre bevölkerungsrepräsentative Normierung der SCL-90-R auf die globalen Kennwerte (summiert über alle 90 Items) zu beschränken. Bei Franke (2002) werden dagegen auch subskalenspezifische Normen, getrennt nach Geschlecht und Alter vorgelegt. Die durchschnittliche Bearbeitungsdauer liegt bei zehn bis fünfzehn Minuten, mithin im zumutbaren und ökonomischen Bereich.
- Die *Änderungssensitive Symptomliste (AT-SYM*; Krampen, 1991a) ist Bestandteil des *Diagnostischen und Evaluativen Instrumentariums zum Autogenen Training (AT-EVA*; Krampen, 1991b). Sie dient der mit der Eingangsdiagnostik verbundenen Indikationsstellung und der Erfolgskontrolle durch die indirekte Veränderungsmessung. Mit der AT-SYM werden sechs Indikationsbereiche von Entspannungstrainings und –therapien (wie Autogenes Training und Progressive Relaxation) auf der Ebene der subjektiven Beschwerdenbelastung erhoben, die direkt aus den störungsunspezifischen (allgemeinen) Ziel- und Effektbereichen von Entspannungsverfahren abgeleitet sind. Die Subskalen der änderungssensitiven Symptomliste sind inhaltlich aber auch von allgemeinerer klinischer Relevanz und beziehen sich mit jeweils acht Items auf (1) körperliche und psychische Erschöpfung, (2.) Nervosität und innere Anspannung, (3.) psychophysiologische Dysregulationen, (4.) Leistungs- und Verhaltensschwierigkeiten, (5.) Schmerzbelastungen sowie (6.) Probleme in der Selbstbestimmung und –kontrolle.

Überprüfungen der internen Konsistenz ($.77 \leq \alpha \leq .93$) und Testhalbierungsreliabilität ($.76 \leq r_{tt} \leq .91$) ergaben befriedigende Werte. Ähnliches gilt für die Werte der Testwiederholungs-Reliabilität ($.40 \leq r_{tt} \leq .92$) mit deutlichen Unterschieden zwischen Personen, die am Autogenen Training teilgenommen versus nicht teilgenommen haben, was gleichzeitig ein Hinweis auf die Validität der Skala ist und die interventionsspezifische Änderungssensitivität von AT-SYM auf Skalenebene empirisch belegt. Weitere Validitätshinweise liefern Vergleiche mit den Kennwerten der Änderungssensitivität der Items der „Kieler Änderungssensitiven Symptomliste“ (KASSL; Zielke, 1979a, 1979b). Sie ergeben erwartungskonform interventionsspezifische (Gesprächspsychotherapie/Einzeltherapie versus Autogenes Training/Gruppenangebot) Unterschiede in der Änderungssensitivität beider Instrumente. Darüber hinaus bestätigen Expertenbeurteilungen die inhaltliche Validität von AT-SYM auf Itemniveau.

- Das *Freiburger Persönlichkeitsinventar (FPI-R)*; Fahrenberg, Hampel & Selg, 2001) ist ein faktorenanalytisch und itemmetrisch begründetes Persönlichkeitsverfahren zu den Konstruktbereichen der Lebenszufriedenheit, Sozialen Orientierung, Leistungsorientierung, Gehemmtheit, Erregbarkeit, Aggressivität, Beanspruchung, Körperlichen Beschwerden, Gesundheitssorgen und Offenheit sowie zu den Sekundärskalen Extraversion und Emotionalität im Sinne Eysencks. Die insgesamt 138 Items sind nach dem Zutreffen auf die eigene Person zu beantworten. Die interne Konsistenz (Cronbachs Alpha) der Skalen liegt zwischen $.73 \leq \alpha \leq .83$. Die interne Validität ist durch die Konvergenz von faktorenanalytischen und itemmetrischen Analysen sowie durch konfirmatorische Clusteranalysen gesichert. Der FPI-R ist für Jugendliche ab 16 Jahren und Erwachsene einsetzbar. Es liegen nach Geschlecht und Altersgruppen differenzierte Normen vor. Die Bearbeitungszeit liegt zwischen 20 und 30 Minuten.

ad 2: Beispiele für störungsspezifische Verfahren

- Das *Beck-Depressions-Inventar (BDI)*; Hautzinger, Bailer, Worall, & Keller, 2000), ein ökonomisches Selbstbeurteilungsinstrument zur Erfassung der Schwere einer Depression, enthält 21 Gruppen von vierfach gestuften Aussagen zu depressiven Symptomen, wie Taurige Stimmung, Pessimismus, Versagen, Unzufriedenheit, Schuldgefühle u. a., die hinsichtlich Auftreten und Intensität zu beurteilen sind. Das

Zeitfenster beträgt laut Instruktion eine Woche. Analysen zur psychometrischen Güte weisen den BDI als ein zuverlässiges, konsistentes und veränderungssensitives Instrument aus. Das Instrument kann bei depressiven Patienten bzw. bei Verdacht auf Vorliegen einer Depression im Altersbereich von 18 bis 80 Jahren eingesetzt werden. Seine durchschnittliche Bearbeitungszeit beträgt je nach Antriebshemmung zehn bis fünfzehn Minuten. Für eine ausführliche Darstellung der Diagnostik depressiver Störungen siehe den Beitrag von Hautzinger in diesem Band.

- Die *Skalen zur Erfassung von Hoffnungslosigkeit (H-Skalen)*; Krampen, 1994) stellen die deutsche Bearbeitung und Weiterentwicklung der H-Skala von Beck und Mitarbeitern dar. Basierend u. a. auf dem handlungstheoretischen Partialmodell der Persönlichkeit erfassen ihre Items negative Erwartungen der Person über sich selbst, die personspezifische Umwelt und ihr künftiges Leben. Die H-Skalen liegen in zwei Versionen à 20 Items und in zwei parallelen Halbformen à zehn Items vor. Die für klinische Gruppen indizierte Standardversion der H-Skala (H-S-Skala) umfasst 20 dichotome Aussagen, die nach ihrem Zutreffen auf die eigene Person zu beantworten sind. Überprüfungen der Messgenauigkeiten der Skalen via Konsistenzschätzungen, Testhalbierung und –wiederholung belegen die Zuverlässigkeit aller H-Skalen-Versionen und –Formen und weisen sie als eine änderungssensible Methode zur Beurteilung therapeutischer Effekte aus. Ihre inhaltliche, konvergente sowie diskriminante und differenzielle Validität konnte empirisch nachgewiesen werden. Die H-Skalen können ab dem Alter von 13 Jahren bis ins hohe Alter als Einzel- oder Gruppentest vorgegeben werden. Bevölkerungsrepräsentative Normen liegen ab dem 18. Lebensjahr vor. Die Bearbeitungsdauer beträgt bei den H-Skalen 10 bis 15 Minuten.
- Intensität und Verlauf der Zustandsangst können durch die (wiederholte) Vorlage der *State-Angst Skala (STAI-S)* des State-Trait-Angstinventars von Laux, Glanzmann, Schaffner und Spielberger (1981) beschrieben werden. Die 20 Items messen die aktuell erlebte Angst als einen emotionalen Zustand, charakterisiert durch Anspannung, Besorgtheit, innere Unruhe und Furcht vor zukünftigen Ereignissen sowie eine erhöhte Aktivität des autonomen Nervensystems. Die innere Konsistenz der Skala beträgt $\alpha = .90$, die Restestrelabilität fällt erwartungsgemäß niedrig aus. Ihre Sensitivität gegenüber situativen Veränderungen konnte experimentell nachgewiesen werden. Die Skala kann ab dem Alter von 15 Jahren im klinischen Bereich und in der experimentellen

Angstforschung benutzt werden. Aufgrund der multiplen Abhängigkeit des State-Angstwertes sollte die Interpretation eines individuellen Wertes nur auf dem Hintergrund eines Vergleiches erfolgen, bei dem die verschiedenen Einflussgrößen ihrer jeweiligen Bedeutung entsprechend berücksichtigt werden. Die Diagnostik von Ängsten und Phobien wird ausführlich von Schneider, Ertl und Margraf in diesem Band behandelt.

- Der *Interaktions-Angst-Fragebogen (IAF)*; Becker, 1997) wurde zur Erfassung der bereichsspezifischen Angstneigungen konstruiert. Beschreibungen von insgesamt 55 Situationen zu den Bereichen Angst vor physischer Verletzung, Auftreten, Normüberschreitung, Erkrankungen und ärztlichen Behandlungen, Selbstbehauptung sowie Abwertung und Unterlegenheit sind auf einer siebenstufigen Skala danach zu beurteilen, wie angenehm oder unangenehm es wäre, sich in der betreffenden Lage zu befinden. Zusätzlich kann mit einer 10. Skala die globale Angstneigung gemessen werden. Die Messgenauigkeit des IAF ist mit Konsistenzschätzungen $.72 \leq \alpha \leq .91$ als zufriedenstellend bis gut zu bewerten. Die Reliabilitätskoeffizienten des Testprofils genügen den Anforderungen. In einer Reihe von Validierungsstudien wurden die Zusammenhänge zwischen den Skalen des IAF und anderen Konstrukten mit plausiblen Ergebnissen überprüft. Der als Einzel- oder Gruppentest im Altersbereich von 18 bis 65 Jahren u. a. im klinisch-psychologischen Bereich zur Differenzialdiagnostik sowie für die Evaluation therapeutischer Maßnahmen einsetzbare IAF wurde an einer repräsentativen Stichprobe normiert. T- und Stanine-Werte sowie Prozentränge liegen getrennt für Männer und Frauen aus verschiedenen Altersstufen vor. Die Bearbeitung des IAF dauert 15 Minuten.
- Das *Trierer Alkoholismusinventar (TAI)*; Funke, Funke, Klein & Scheller, 1987) ist ein standardisierter Test für alkoholabhängige Männer und Frauen, der behandlungsrelevante, differenzialdiagnostische Informationen liefert. Die 90 Fragen des Tests fokussieren spezifische Erlebens- und Verhaltensweisen, wie sie von alkoholabhängigen Personen selbst als Antezedenzen und Konsequenzen des exzessiven Alkoholkonsums dargestellt werden. Die sieben Skalen des TAI betreffen (1) Verlust der Verhaltenskontrolle und negative Gefühle nach dem Trinken, (2) Soziales Trinken, (3) Süchtiges Trinken, (4) Trinkmotive, (5) Folgen, Schädigung und Versuche der Selbstbehandlung von physiologischen Begleiterscheinungen, (6) Partnerprobleme

wegen des Trinkens und (7) Trinken wegen Partnerproblemen. Die interne Konsistenz der Skalen ist zufriedenstellend. In einer Reihe von Studien wiesen die Zusammenhänge zwischen den TAI-Skalen und verschiedenen Variablen die Validität des Verfahrens aus. Retest-Befunde belegen eine ausreichende Stabilität der Skalen. Für das Instrument liegen geschlechtsspezifische Stanine-, T- und Prozentrang-Werte vor. Seine Bearbeitung dauert 30 Minuten.

ad 3: Beispiele für Therapiestundenbogen

Stunden- oder Sitzungsbogen in der Prozessdiagnostik und -evaluation von Psychotherapien hatten ihren Ursprung im klient-zentrierten Ansatz (vgl. etwa Eckert, 1976; Zielke, 1980; im Überblick Frohburg, 1999), waren daher zunächst inhaltlich eher schulenspezifisch aufgebaut, wurden aber bald auf dem Hintergrund von Plausibilitätserwägungen auf allgemeinere Variablen erweitert.

- Dies gilt etwa für die „*Stundenbeurteilung (Klient) – Selbstbeurteilungsverfahren*“ (*SB-K*) von Schindler, Hohenberger-Sieber und Hahlweg (1990a), mit der über 20 Items Erwartungen von Patienten, die von ihnen erlebte Transparenz, Aktivität und Zusammenarbeit, ihre Zuversicht und ihr Vertrauen sowie der erlebte Aufbau von Selbstkontrolle erfasst werden.
- Informationen aus der Therapeutensicht zu Erwartungen an die Sitzung, Struktur und Stringenz, Motivierung, Anleitung, Information, neue Erkenntnisse sowie Fortschritt und Bestätigung können mit der „*Stundenbeurteilung (Therapeut)*“ (*SB-T*; Schindler et al., 1990b) erhoben werden.
- Der „*Bielefelder Klienten-Erfahrungsbogen*“ (*BIKEB*) von Höger und Eckert (1997) ist faktorenanalytisch fundiert und erfasst mit jeweils vier Items die Aspekte des Zurechtkommens mit dem Therapeuten/der Therapeutin, des Zurechtkommens mit sich selbst, des Veränderungserlebens, der persönlichen Sicherheit und Zuversicht, der Beruhigung sowie der körperlichen Entspannung versus Erschöpfung (6 Subskalen x 4 Items = 24 Items).
- Entsprechende allgemeine prozessdiagnostische Verfahrensvorschläge liegen inzwischen auch für die *Gruppenpsychotherapie* (vgl. etwa Enke & Lermer, 1978; Strauss & Eckert, 1994) vor.
- Da die vorliegenden Stunden- oder Sitzungsbogen entweder durch ihre Therapieschulen-spezifische oder sehr allgemeine inhaltliche Ausrichtung keine Bezüge

zu den allgemeinen Wirkfaktoren aufweisen, wurde von Grawe (1998, S. 669ff; siehe etwa auch Grawe & Braun, 1994) der „*Klienten/Klientinnen-Stundenbogen*“ entwickelt, der in identischen Literaturquellen manchmal auch als „*Patientenstundenbogen*“ bezeichnet wird. Über 29 Items wird das Erleben der Therapiesitzung durch Patienten auf siebenstufigen bipolaren Antwortskalen erfasst. Die Auswertung erfolgt nach fünf faktorenanalytisch begründeten Skalen zu (1) der Zufriedenheit mit der Sitzung, (2) den vom Patienten erlebten Fortschritten innerhalb der Therapie, (3) den erlebten Fortschritten außerhalb der Therapie, (4) der erlebten Qualität der Therapiebeziehung und (5) der erlebten emotionalen Beteiligung in der Therapie. Anhand der Skalenwerte werden idiographische Prozessfigurationen und Musteranalysen (Feature Pattern Analysis) therapeutischer Verläufe realisiert, deren Befunde zu Kombinations-Mustertypen therapeutischer Verläufe in größeren Stichproben verdichtet und mit Therapieergebnisvariablen in Beziehung gesetzt werden (Grawe, 1998; Grawe & Braun, 1994; Grawe, Caspar & Ambühl, 1990a, 1990b; Smith, 2001). Kritisch anzumerken bleibt, dass der Bezug der verwendeten Skalen dieses Stundenbogens zum Dreikomponentenmodell der Wirkungsweise von Psychotherapie gering bleibt und dass das Verfahren mit 29 Items für eine routinemäßig durchgeführte, ökonomische Prozessdiagnostik am Ende jeder Therapiesitzung relativ lang ist. Hinzu kommt, dass Grawe und Braun (1994) den Messfehler vernachlässigen. Dies hat zur Konsequenz, dass eine Post-Prä-Differenz, die ausschließlich auf Messfehlern beruht, genauso groß aussehen kann, wie eine Post-Prä-Differenz, die auf perfekt reliablen Messungen basiert (siehe Steyer, Hannover, Telser & Kriebel, 1997).

- Die „*Stundenbogen für die Allgemeine und Differentielle Einzel-Psychotherapie*“ (STEP; Krampen, 2002a) wurden als ökonomische, standardisierte Instrumente zur Erfassung der von Grawe et al. (1994) unterschiedenen allgemeinen Wirkfaktoren psychotherapeutischer Prozesse und Effekte aus der Sicht von Patienten/innen und der von Therapeuten/innen entwickelt. Ihre Items beziehen sich direkt auf das Erleben einer Therapiesitzung im Einzelsetting durch die Patienten/innen (STEPP) sowie die darauf bezogenen Fremdwahrnehmungen und -beurteilungen ihrer Therapeuten/innen (STEPT). Diese beiden Versionen des STEP sind inhaltlich parallel aufgebaut und umfassen jeweils 12 komplementär formulierte Items. Beim STEPP ist von Patienten/innen nach einer Therapiesitzung auf siebenstufigen Antwortskalen anzugeben, ob die jeweilige Aussage für die eigene Person zutrifft (mit "1" für "stimmt ganz genau") oder nicht (mit "7" für "stimmt überhaupt nicht"). Beim STEPT beurteilen

die Therapeuten/innen inhaltlich analoge Fragen unter Bezug auf ihre Wahrnehmungen des Erlebens und Verhaltens ihres Patienten bzw. ihrer Patientin auf identischen Antwortskalen. Für die Patienten/innen- und die Therapeuten/innen-Perspektive existieren jeweils drei inhaltlich komplementäre Subskalen zu der in einer Sitzung erlebten bzw. wahrgenommenen

- motivationalen Klärung bei dem Patienten/der Patientin (Klärungsperspektive; STEPP-K bzw. STEPT-K) mit fünf Items,
- aktiven Hilfe zur Problembewältigung (Problembewältigungsperspektive; STEPP-P bzw. STEPT-P) mit vier Items und
- therapeutischen Beziehung (Beziehungsperspektive; STEPP-B bzw. STEPT-B) mit drei Items.

Zusammen mit einer auf Therapieziele bezogenen, etwa über das „*Goal Attainment Scaling*“ (GAS; Kirusek, Smith & Cardillo, 1994) realisierten sukzessiven Prozess- und Ergebnisevaluation dienen Stundenbogen der adaptiven Indikation therapeutischer Maßnahmen. Zugleich bilden sie eine wertvolle, strukturierte Hilfe bei der *Dokumentation von Psychotherapiesitzungen und psychotherapeutischer Fallverläufe* (vgl. hierzu etwa Laireiter, 1994; Stangier, Hoyer, Kosfelder & Meyer, 1998), für die nicht nur eine formale Verpflichtung besteht, sondern die auch als ein Instrument der Qualitätssicherung von Psychotherapie unentbehrlich ist. Zu betonen ist, dass strukturierte Verfahren wie die STEP stets nur eine Ergänzung (freilich eine hilfreiche und wichtige) freier und teilstrukturierter Protokollierungen sein können und diese keinesfalls ersetzen dürfen.

3.2 Direkte Veränderungsmessung

Bei der *direkten Veränderungsmessung* werden die oben genannten methodischen Probleme der indirekten Veränderungsmessung dadurch umschifft, dass als Vergleichsprozess die subjektive Messwertskala der Person selbst herangezogen wird und in nur einer Messung subjektiv erlebte Veränderungen in ihrer Richtung für einen vorgegebenen Zeitraum direkt erhoben werden. Unabhängig von dem physikalischen Messwertkontinuum werden die wahrgenommenen Veränderungen durch den impliziten Vergleich zwischen einem intern wahrgenommenen Anfangszustand mit einem Endzustand auf einem Kontinuum der subjektiven Bedeutsamkeit des Probanden quantifiziert. Regressionseffekte können ebenso wenig auftreten wie die mit dem Reliabilitäts-Validitäts-Dilemma verbundenen Interpretationsprobleme. Die direkte Veränderungsmessung ist auch ökonomischer als die indirekte, wirft dafür aber Fragen nach dem Ausmaß von Verzerrungen durch die retrospektive Einschätzung der stattgefundenen Veränderung, nach einer möglichen

Überschätzung der wahrgenommenen Veränderung (etwa bedingt durch sozial erwünschtes Antwortverhalten) sowie nach der numerischen Äquivalenz zwischen direkt versus indirekt gewonnenen Änderungswerten auf. Daher darf die direkte Veränderungsmessung nicht als eine Alternative zur indirekten Messung betrachtet werden. Sie stellt vielmehr eine eigenständige Methode der Veränderungsmessung dar, die ergänzend zur indirekten Veränderungsmessung herangezogen werden sollte, wobei die Übereinstimmung der Befunde aus beiden Strategien der Veränderungsmessung von besonderem Interesse ist (vgl. hierzu auch Baumann et al., 1980; Krampen & von Delius, 1981; Zielke, 1982).

Dies wird durch die empirischen Untersuchungsbefunde von Michalak, Kosfelder, Meyer und Schulte (2003) bestätigt. Die Studie bezieht sich auf Daten einer großen Stichprobe von 775 ambulanten Psychotherapie-Patienten (davon $n = 583$ mit Angststörungen und $n = 95$ mit affektiven Störungen) mit einer durchschnittlichen Behandlungsdauer von 26 Sitzungen. Neben einer Vielzahl indirekter Veränderungsmaße (störungsspezifische und störungsübergreifende Erfolgsmaße und Fragebogen zu Störungsfolgen im Vortest-Nachtest-Vergleich) wurden direkte Veränderungsmaße über retrospektive Einschätzungen des Globalerfolges der Behandlung und der Therapiezielerreichung in der Patienten- und der Therapeutesicht sowie über den „Veränderungsfragebogen des Erlebens und Verhaltens“ (VEV; Zielke & Kopf-Mehnert, 1978) erfasst. Bei statistisch durchgängig signifikanten Korrelationen der direkt und der indirekt erhobenen Erfolgsmaße (in moderater Höhe im Bereich von $.18 \leq r \leq .54$; $p \leq .01$) ergaben sich für verschiedene Subgruppen faktorenanalytisch zwei eindeutig bestimmte Faktoren. Ein Faktor ist jeweils durch essentielle Ladungen der indirekten Erfolgsmaße, der zweite durch essentielle Faktorladungen der direkten Erfolgsmaße markiert.

Zu den Instrumenten der direkten Veränderungsmessung gehören zum Beispiel:

- Die große Gruppe der *Zufriedenheitsmaße*, bei denen am Ende der Behandlung der Behandlungserfolg bzw. die Zufriedenheit mit der Behandlung durch Patienten und/oder Therapeuten retrospektiv eingeschätzt wird (siehe etwa Michalak et al., 2003). Prominentes Beispiel ist die *Consumer Reports Study* (Seligman, 1995), bei der ehemalige Patienten rückblickend nach dem Ende der Behandlung nach ihrer Zufriedenheit mit einer Psychotherapie befragt wurden. Angesichts der vielfältigen empirischen Belege zum Rückschaufehler (siehe zusammenfassend hierzu Hank, Schwenkmezger & Schumann, 2001), erscheinen Reliabilität und Validität dieser Erfolgsmaße und damit ihre isolierte Verwendung fraglich.

- Das *Goal Attainment Scaling* (GAS; Kiresuk, Smith & Cardillo, 1994) ist eine einzelfallbezogene Methode, mit der sich aktuelle Interventionsziele präzisieren und das subjektive Ausmaß ihrer Erreichung quantifizieren lässt. Die Konstruktion des GAS sollte möglichst frühzeitig im therapeutischen Prozess erfolgen. Sie umfasst mehrere Schritte. Im ersten Schritt werden bis zu fünf für den Patienten wichtige Zielbereiche benannt, die entsprechend ihrer subjektiven Bedeutsamkeit in eine Rangreihe zu bringen und entsprechend zu gewichten sind. Anschließend werden für jedes Problem fünf objektiv zu beschreibende, möglichst verhaltensnahe Indikatoren (= Kriterien) der Zielerreichung definiert und auf einer fünffach gestuften Skala mit den sprachlichen Bezeichnungen „viel weniger als die erwarteten Resultate“ (-2), „etwas weniger als die erwarteten Resultate“ (-1), „am meisten erwartete oder angenehme Resultate“ (0), „etwas mehr als die erwarteten Resultate“ (+1), „viel mehr als die erwarteten Resultate“ (+2) angeordnet. In regelmäßigen Abständen kann der Patient dann während und am Ende der Therapie seine Zielerreichung in einem Wertebereich von -2 bis +2 einschätzen. Damit trägt das GAS zur Evaluation des Therapieprozesses mit den individuellen Therapiezielen als Erfolgskriterien bei. Ihre Validität bemisst sich nach (1) der Bedeutsamkeit der Zielbereiche für den jeweiligen Patienten und des Indikatorwertes der Kriterien für die Zielbereiche, (2.) deren Repräsentativität für die zu evaluierende Behandlung und (3.) der Akzeptanz für externe Urteiler (Kordy & Hannover, 1999). Da Patient und Therapeut die Therapieziele gemeinsam erarbeiten, stärkt das Instrument darüber hinaus Motivation und Selbstverantwortung des Patienten. Ein weiterer Vorteil des GAS liegt darin, dass es schulen- und methodenunabhängig eingesetzt werden kann.
- Der *Veränderungsfragebogen des Erlebens und Verhaltens-R-2001* (VEV-R-2001; Zielke & Kopf-Mehnert, 2001; Vorversion: VEV, Zielke & Kopf-Mehnert, 1978) quantifiziert die Einschätzung der subjektiv wahrgenommenen Veränderungen im Erleben und Verhalten von Psychotherapiepatienten/innen. Dazu werden 42 Items nach der eingeschätzten Veränderung, die im Verlauf der Therapie erfolgt ist, direkt abgefragt. Die Items bestehen aus siebenfach gestuften Aussagen, (z. B. „Ich bin ruhiger geworden“), basieren inhaltlich vor allem auf dem klient-zentrierten psychotherapeutischen Ansatz und bilden einen bipolaren Faktor mit den beiden Polen „Entspannung, Gelassenheit und Optimismus“ vs. „Spannung, Unsicherheit und Pessimismus“. Der VEV-R-2001 ist ein objektives und konsistentes Verfahren mit einer erwartungsgemäß geringen zeitlichen Stabilität. Signifikante Korrelationen mit den

Veränderungsbeurteilungen VERA des „Freiburger Fragebogens zur Krankheitsverarbeitung“ (Muthny, 1989) bestätigen seine kriterienbezogene Validität. Die Autoren teilen kritische Veränderungswerte mit, vor deren Hintergrund die individuelle Veränderung ggf. als signifikant beurteilt werden kann. Die Bearbeitungszeit beträgt 10-15 Minuten.

- Analog zum VEV und zum „*Fragebogen zu erlebten gesundheitlichen Veränderungen*“ (FGV; Krampen & von Delius, 1981), der für die direkte Veränderungsmessung bei stationären medizinischen und psychiatrischen Patienten entwickelt wurde und bislang lediglich als Forschungsinstrument vorliegt, aufgebaut, ist der „*Veränderungsfragebogen zum Entspannungserleben und Wohlbefinden*“ (VFE; Krampen, 1991c). Als Separatum des „*Diagnostischen und Evaluativen Instrumentariums zum Autogenen Training*“ (AT-EVA; Krampen, 1991b) fokussieren die nur 12 Items des VFE das direkte Veränderungserleben in den bereits oben genannten sechs Indikationsbereichen von Entspannungstrainings und –therapien. Wegen der weit über den engeren Bereich der Entspannungstherapien hinausgehenden Bedeutung der Frageninhalte (z.B. „Ich bin weniger besorgt, wenn ich an die Zukunft denke“; „Ich fühle mich wohler“; „Ich habe mehr Ausdauer und gebe nicht so schnell auf“) und seiner Kürze ist der VFE auch für Anwendungen im Rahmen der Psychotherapie und häufigere Wiederholungsmessungen besonders gut geeignet. Für den bei Jugendlichen und Erwachsenen einsetzbaren VFE liegen sowohl soziale Vergleichsnormen (Prozentrang- und T-Werte) als auch intraindividuelle Veränderungsnormen vor.

3.3 Kriterien für klinisch bedeutsame Veränderungen

Bleibt die Frage, was die Veränderungen – gleich ob direkt oder indirekt gemessen – klinisch bedeutsam macht. Neben der Verwendung inhaltlicher, vor-psychometrischer Kriterien (wie etwa Symptombefreiheit nach ICD-10 oder DSM-IV-TR, objektive Veränderungen in bedeutsamen Lebensbezügen, direkt beobachtbare Zunahmen von Handlungsfreiheiten etc.) sind psychometrisch fundierte Aussagen notwendig. Klassisch bei Versuchen der Beantwortung dieser Frage ist der Rückgriff auf psychometrische Parameter wie etwa Vertrauensintervalle, Standardabweichungen, mittlere bzw. günstigere Normwerte etc., die das Ausmaß von direkt oder indirekt erfassten Veränderungen und deren klinische Bedeutsamkeit abschätzen helfen sollen. Insbesondere die Arbeitsgruppe um Jacobson (Jacobson, Follette & Revenstorf, 1984; Jacobson & Revenstorf, 1988; Jacobson & Truax,

1991) hat sich damit intensiv beschäftigt und einen systematischen Operationalisierungsvorschlag vorgelegt, mit dem sowohl in Wirksamkeits- und Effektivitätsstudien als auch in der Qualitätssicherung gearbeitet wird.

Jacobson und Truax (1991) gehen dabei von der Annahme aus, dass klinische und nicht-klinische Stichproben unterschiedlichen Populationen angehören, die durch einen geeigneten Trennwert unterschieden werden können. Eine klinisch bedeutsame Verbesserung impliziert demnach den reliablen Wechsel von der „kranken“ zur „gesunden“ Population oder – auf der Grundlage von Skalenwerten – den reliablen Wechsel aus einem für Patienten typischen Skalenabschnitt in einen für die Normalbevölkerung typischen Skalenabschnitt.

Da sich die Messwerte von klinischen und nicht-klinischen Populationen auf psychometrischen Instrumenten allerdings zumeist überlappen, kommt es beim Einsatz dieser Instrumente zu Fehlklassifikationen. Tatsächlich Kranke werden nicht entdeckt (= F1), und Gesunde werden fälschlicherweise als krank klassifiziert (= F2). Jacobson und Truax (1991) schlagen drei Varianten zur Bestimmung des Zielbereichs vor:

- (1.) Im günstigsten Fall kennt man die Verteilungsparameter der klinischen und nicht-klinischen Stichprobe und legt den Kriteriumswert so fest, dass die Fehlklassifikationsraten für F1 (1-Sensitivität) und F2 (1-Spezifität) gleich groß ausfallen. Sind beide Populationen normal verteilt, bestimmt sich der Trennwert nach $C_1 = (\delta_0 M_1 + \delta_1 M_0) / (\delta_0 + \delta_1)$. Die Symptomatik sollte nach einer Behandlung so weit reduziert sein, dass die behandelte Person näher am Mittelwert der nicht gestörten Population als an dem der gestörten Population liegt.
- (2.) Sind die Parameter δ_0 und M_0 nicht bekannt und es stehen nur Patientendaten zur Verfügung, entspricht der Trennwert dem Wendepunkt der Normalverteilung bei $C_2 = M_1 - 2\delta_1$. Das Ausmaß der Symptomatik sollte nach der Behandlung mindestens zwei Standardabweichungen unterhalb des Mittelwertes der zugehörigen „kranken“ Population liegen.
- (3.) Sind die Verteilungsparameter der klinischen Stichprobe nicht bekannt, soll der Trennwert nach $C_3 = M_0 + 2\delta_0$ berechnet werden. Die Stärke der Symptomatik sollte also nach einer Behandlung innerhalb von zwei Standardabweichungen vom Mittelwert einer gesunden Population liegen.

Wie Percevic, Bauer und Kordy (2004) eindrucksvoll zeigen, führen diese drei Verfahren zu unterschiedlichen Trennwerten und damit zu sehr unterschiedlichen Einschätzungen der klinischen Bedeutsamkeit und Fehlklassifikationsraten. Sie plädieren daher für eine inhaltliche Validierung des Kriteriumswertes unter Berücksichtigung von

Fehlklassifikationsraten und Kriteriumsvaliditäten – eine Forderung, der man sich nur anschließen kann und die der Bearbeitung bedarf.

4 Varianten der kontrollierten Praxis

Die empirische Evaluation klinisch-psychotherapeutischer Interventionen kann durch verschiedene Arten von Einzelfalldarstellungen erfolgen. Gemeinsam ist ihnen der Fokus auf die intraindividuelle Merkmalsfluktuation im zeitlichen Verlauf mit dem Bemühen, diese Verlaufsscharakteristik als Funktion anderer intra- oder interindividuell variierender Variablen zu erklären. In diesem Sinne postuliert Hilliard (1993) „all single research is intrasubject research“ (S. 374). Sie kann nach inhaltlichen und methodischen Gesichtspunkten systematisiert werden. Inhaltlich betrachtet, lassen sich Einzelfalldarstellungen danach unterscheiden, ob sie sich auf das gesamte Leben eines Individuums beziehen oder nur bestimmte, z. B. krankheitsbezogene Ausschnitte thematisieren. Im erstgenannten Fall sprechen Selg, Klapproth und Kamenz (1992) von einer Biographie, im letztgenannten Fall handelt es sich um eine Kasuistik oder Einzelfalldarstellung. Je nach angewandter Methodik lassen sich Einzelfalldarstellungen weiter differenzieren in (beschreibende) Falldarstellungen oder Kasuistiken, (quantitative) Einzelfallanalysen und (kontrollierte) Einzelfallexperimente (Hilliard, 1993; Kern, 1997).

4.1 Kasuistik und Einzelfalldarstellung

Die narrative Falldarstellung ist die älteste Form der Einzelfallanalyse und gilt als Vorläufer der experimentellen Einzelfallanalyse (Schroeder & Wildman, 1988). Historisch gesehen wurzelt dieser Ansatz in der Psychoanalyse. Bekanntes Beispiel ist Freuds Studie über die Patientin Anna O.. Falldarstellungen sind qualitativ ausgerichtete retrospektive Fallbeschreibungen ohne Spezifikation der Intervention und systematische Messungen der Interventionseffekte. Sie sind weder intern noch extern valide. Gleichwohl die Fallstudie auch zur Bestätigung von Hypothesen durchgeführt werden könnte, finden sich in der psychotherapeutischen Forschung und Praxis nur Darstellungen im Sinne einer Hypothesengenerierung (Hilliard, 1993). In Übereinstimmung dazu ist diese hermeneutisch-epistemologische Vorgehensweise nach Grawe (1988b) dann indiziert, wenn das Ziel ist, besser zu verstehen, was überhaupt in der Psychotherapie geschieht. Ähnlich argumentiert Thorngate (1986):

„To find out what people do in general, we must first discover what each person does in particular,

then determine what, if anything, these particulars have in common. ... Nomothetic laws lie at the intersection of idiographic laws; the former can be discovered only after we find the latter" (pp. 75-76).

Dieses Ziel strebt Jüttemann (1981, 1990) mittels *Komparativer Kasuistik (KK)* an, deren diagnostische Modellvariante sich nach seinem Verständnis aufgrund ihrer engen Verbindung zur einzelfalldiagnostischen Vorgehensweise besonders als eine Form der Praxiskontrolle anbietet. Die Vorgehensweise der KK besteht in einer qualitativ vergleichenden Betrachtung von hochgradig ähnlichen Einzelfällen und dient der Generierung von potentiellen funktional relevanten Hypothesen, die als Übereinstimmungen zwischen den Fällen sichtbar werden. Diese „theorievorbereitende“ Strategie kann als Umsetzungsversuch einer kombiniert idiographisch-nomothetischen Forschungsstrategie verstanden werden.

Untersuchungsmethodologisch handelt es sich dabei um einen Kleingruppen-Ansatz, der über „den Umweg“ von qualitativen Einzelfallbetrachtungen realisiert wird. Im Anwendungskontext der klinisch-psychotherapeutischen Praxiskontrolle soll so der Wirkungszusammenhang bestimmter therapeutischer Interventionsmaßnahmen und die Entstehung differentieller Therapieeffekte *post hoc* durch die theoretische Beschreibung der beobachteten Datenmuster erklärbar werden. Dabei schreibt Jüttemann (1990) der Erhebung retrospektiver Daten und der qualitativen Aufbereitung biographischen Materials eine besondere Bedeutung zu. Damit macht diese Strategie gleichsam die Nahtstelle zwischen qualitativer Falldarstellung und quantitativer Fallanalyse aus.

In neueren Fallstudien flankieren quantitative Daten die qualitative Fallbeschreibung und umgekehrt. So werden qualitativ zu beschreibende Fälle nach quantitativen Kriterien (z. B. Testbefunde) ausgewählt. Beispiele für diese *quantitatively informed qualitative case studies* (Hilliard, 1993) bieten etwa die Arbeiten von Strupp (1980a, 1980b, 1980c, 1980d; Strupp, Schacht, Henry & Binder, 1992; vgl. auch die Beispiele bei Krampen, 2001; Meinschmidt & Krampen, 2003). Umgekehrt basieren *qualitatively informed single case quantitative analyses* auf qualitativen Auswahlentscheidungen, wie beispielsweise in der Berner Therapievergleichsstudie geschehen (Grawe, 1992). Systematische Überlegungen zur Kombination der qualitativen Technik mit quantitativen Daten stehen allerdings noch aus (Hilliard, 1993).

4.2 Einzelfallanalyse

Die quantitative Einzelfallanalyse beruht auf der systematischen Beobachtung einer spezifizierten Intervention. Die Interventionseffekte werden jedoch nicht durch einen

systematischen Versuchsplan evaluiert. Das Vorgehen kann sowohl konfirmatorisch als auch explorativ angelegt sein (Hilliard, 1993). Die dieser Kategorie zuzuordnenden Einzelfalldarstellungen machen eine große Bandbreite aus. Sie reichen von Beiträgen in überwiegend beschreibender Form mit nur einigen wenigen quantitativen Detailangaben (meist prä/post erhobene Testwerte) bis hin zu viele Messzeitpunkte umfassenden Falldarstellungen mittels Zeitreihenanalysen, sequentiellen Analysen und Wachstumskurven

Im Rahmen der Psychotherapieforschung sei beispielhaft auf die Arbeit von Lutz (2003) hingewiesen. Mit Hilfe von Wachstumsanalysen modelliert er individuelle psychotherapeutische Verläufe basierend auf Patientenausgangscharakteristika (z. B. Chronifizierung der Probleme, therapeutische Vorerfahrung, Vertrauen des Patienten in den Erfolg der Psychotherapie, Beeinträchtigungsschwere aus Sicht des Therapeuten) und Prozessinformationen mit dem Ziel, selektive und adaptive Indikationsentscheidungen zu ermitteln sowie diese als Entscheidungs- und Handlungsregeln für die klinische Praxis nutzbar zu machen.

In der psychotherapeutischen Praxis dominieren Dokumentationen psychotherapeutischer Fallverläufe die einzelfallanalytischen Darstellungen. Sie sind aus rechtlichen, ethischen, berufsständischen und wissenschaftlichen Gründen erforderlich (Baumann & Ühlein, 1994) und stellen ein wesentliches qualitätssicherndes Element in der psychotherapeutischen Praxis dar. Allgemein verbindliche Richtlinien zu ihrer Struktur und ihrem Inhalt gibt es allerdings (noch) nicht. Einen Überblick über die vorhandenen *Dokumentationssysteme in der ambulanten und stationären Psychotherapie* sowie in der psychiatrischen und psychosozialen Versorgung geben Laireiter und Vogel (1998). Gemeinsam sind diesen Dokumentationssystemen (z. B. Braun, 1998; Kordy, Hannover & Bauer, 2002; Krischker, Fleischmann & Cording, 1998; Nübling & Schmidt, 1998b; Palm, 1998; Stein & Herzog, 2002) die strukturellen Elemente Behandlungsbeginn, Behandlungsverlauf und Therapieende, die je nach Behandlungssetting eine andere inhaltliche Ausgestaltung aufweisen. Während bei den älteren Verfahren die Basisdokumentation im Vordergrund stand, fokussieren neuere Verfahren stärker die Verlaufsdokumentation und kommen damit Forderungen nach einer angemessenen Prozessevaluation (Grawe, 1988b, Richter, 1994) nach.

Stellvertretend für diese neue Generation sei das *Allgemeine Dokumentationssystem für Psychotherapie (PSYCHODOK)*; Laireiter, Lettner & Baumann, 1998) genannt. Dieses teilstandardisierte Papier-Bleistiftverfahren zielt auf eine schulen-, tätigkeitsfeld- und altersunabhängige Basis- und Verlaufsdokumentation ab. Die Basisdokumentation umfasst

Angaben zu Personendaten, zur Anmeldung, zur sozialen, biographischen und störungsbezogenen Anamnese, zur Problemanalyse, zum Therapieverlauf, zum Behandlungsverlauf sowie zur Katamnese. Elemente der Verlaufsdocumentation sind Patienten-Stundenbogen, Prozess-(Stunden-)Evaluationsbogen und ein Verlaufsevaluationsbogen. Die Instrumente sind sowohl durch den Psychotherapeuten selbst als auch durch den Patienten zu bearbeiten. Hervorzuheben ist der modulare Aufbau der Basisdocumentation, der ihre fallspezifische und damit ökonomische Durchführung gestattet.

Erwähnenswert ist in diesem Zusammenhang ferner die *Figurationsanalyse* nach Grawe und Baltensperger (1998), eine computerunterstützte Methode zur Prozess- und Ergebnisevaluation von Psychotherapie, die den Vergleich individueller Therapieeffekte vor dem Hintergrund relevanter Vergleichsgruppen zu bewerten ermöglicht. Damit verbindet dieses Verfahren gruppenstatistische Analysen mit der Einzelfallanalyse. Das Dokumentationssystem kann in ambulanten und stationären Einrichtungen der Psychotherapie und psychosozialen Versorgung verwendet werden.

4.3 Einzelfallexperiment

Ein kontrolliertes Einzelfallexperiment ist „ein planmäßiges und replizierbares Verfahren, bei dem durch systematische Variation der unabhängigen Variable (Intervention) und Konstanthaltung anderer Bedingungen die Veränderung der abhängigen Variable im zeitlichen Verlauf über zahlreiche Messzeitpunkte registriert werden kann“ (Fichter, 1989, S. 61) – eine Rarität in der Vorgehensweise der deutschsprachigen Therapieforschung (Kern, 1997; Perst & Baumann, 1999). Da es beim Einzelfallexperiment darum geht, Veränderungen im zeitlichen Verlauf abzubilden, steht die Definition der verschiedenen Zeitabschnitte in einem definierten Zeitkontinuum im Vordergrund.

Eckpunkte bei der Wahl des Untersuchungsdesigns (Petermann, 1996) sind

- (1) die Anzahl der verschiedenen Phasen im Untersuchungszeitraum,
- (2) die Sequenz der Wiederholungen dieser Phasen,
- (3) die Anzahl der Messungen an verschiedenen Personen zum gleichen Zeitpunkt,
- (4) die Anzahl der pro Phase zu erfassenden Variablen,
- (5) die Auswahl der Bezugsgrößen - Verhalten, Personen oder Situationen – und
- (6) bei Interventionsstudien die Veränderungen der Anforderungen in der Zeit.

Je nach Kombination dieser Aspekte ergeben sich verschiedene Untersuchungsdesigns, die sich in die Gruppen AB-Versuchspläne, Versuchspläne mit multiplen Grundraten über Verhaltensweisen, Personen und/oder Situationen, Versuchspläne mit verändertem Kriterium und alternierende Versuchspläne einteilen lassen (siehe hierzu z. B. Fichter, 1989; Kern, 1997; Petermann, 1996).

Voraussetzung für den Nachweis der internen Validität von Einzelfallexperimenten ist, dass die Behandlung unmittelbar wirksam wird, erzeugte Interventionseffekte (z. B. bei ABAB-Versuchsplänen) umkehrbar sind, Transfereffekte bei Designs mit multiplen Grundraten ausbleiben, Interventions- und Reihenfolgeeffekte nicht konfundiert sind und eine mögliche wechselseitige Beeinflussung von Merkmalen der Person und solchen der Intervention beachtet werden. Eine Maximierung der experimentellen Kontrolle kann durch die Kombination der verschiedenen Grunddesigns realisiert werden. Diese kombinierten Designs eignen sich darüber hinaus zur Untersuchung von Generalisationseffekten. Beispiele dazu finden sich in Fichter (1989).

Für die Auswertung der seriell anfallenden Daten sind Zeitreihenanalysen geeignet: Dabei wird die Zeitreihe als Funktion der internen (sequentiellen) Abhängigkeit und als Funktion der Designparameter (Intervention, Kovariate) verstanden. Das anzupassende Zeitreihenmodell liefert Hinweise auf die serielle Abhängigkeit und Periodik der untersuchten Prozesse und erhellt damit ihre Struktur. Anhand der Korrelationen über die Zeit (Kreuzkorrelationen) lassen sich mögliche Wirkzusammenhänge erschließen, und Interventionseffekte können statistisch abgesichert werden (Revenstorf & Keeser, 1989). Über die Voraussetzungen zur Durchführung einer Zeitreihenanalyse informiert Noack (1989). Die Möglichkeiten für die Analyse qualitativer Verlaufsdaten werden Revenstorf und Vogel (1989) erläutert.

5 Zusammenfassung

Das vorangegangene Kapitel beschäftigt sich mit dem Beitrag der Prozessdiagnostik zu einer kontrollierten Praxis, die dem Anspruch eines wissenschaftlich fundierten praktisch psychologischen Handelns genügen will. Erkenntnisziel ist es, den *modus operandi* von klinischen oder pädagogischen Interventionen zu erklären. Diese Zielsetzung impliziert, wie weiter gezeigt wurde, sich Fragen nach der selektiven und adaptiven Indikation, der Effektivität und der Effizienz von Interventionen zu stellen. Anhand eines Ablaufschemas zum diagnostischen Abklärungsprozess und dem adaptiven Interventionsprozess wurden die bedeutsamen Schritte einer so verstandenen kontrollierten Anwendungspraxis verdeutlicht. Ausgehend von einer Darstellung der testtheoretischen Grundlagen von Veränderungsmessungen, wurden im nächsten Schritt Methoden und Hilfsmittel der (in)direkten Veränderungsmessung vorgestellt und Kriterien zur Beurteilung von klinisch bedeutsamen Veränderungen diskutiert, bevor abschließend Kasuistik und Einzelfalldarstellung, Einzelfallanalyse und Einzelfallexperiment als Varianten der kontrollierten Praxis dargestellt wurden. Die Überlegungen sollten deutlich gemacht haben, dass die Prozessdiagnostik das *sine qua non* einer kontrollierten Praxis ist, die – wissenschaftlich fundiert – verstehen will, warum bei welchen Personen durch welche Durchführenden welche Interventionsziele unter welchen Bedingungen mit welchen Methoden am besten realisiert werden können. Bleibt zu wünschen, dass die bereits 1988 in diesem Kontext von Grawe gestellte Forderung „zurück zur psychotherapeutischen Einzelfallforschung“ stärker als bisher Gehör findet (Grawe, 1988b).

6 Literatur

- Bastine, R. (1998). Grundlagen der Klinischen Psychologie. In T. Herrmann, W. Tack & R. Bastine (Hrsg.), *Klinische Psychologie* (Bd. 1, S. 47-147). Stuttgart: Kohlhammer.
- Baumann, U., Sodemann, U. & Tobien, H. (1980). Direkte versus indirekte Veränderungsdiagnostik. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 1, 201-216.
- Baumann, U. & Ühlein, H. (1994). *Leitsätze zur Dokumentation klinisch-psychologischer/psychotherapeutischer Interventionen*. Köln: Deutscher Psychologenverlag.
- Becker, P. (1995). *Seelische Gesundheit und Verhaltenskontrolle*. Göttingen: Hogrefe.
- Becker, P. (1997). *Interaktions-Angst-Fragebogen IAF* (3. rev. und erw. Aufl.). Göttingen: Beltz-Test GmbH.
- Bereiter, C. (1963). Some persisting dilemmas in the measurement of change. In C.W. Harris (Ed.), *Problems in measuring change* (pp. 3-20). Madison: The University of Wisconsin Press.
- Beutler, L.E. & Clarkin, J.F. (1990). *Systematic treatment selection*. New York, NY: Brunner/Mazel.
- Brandstädter, J. (2001). *Entwicklung, Intentionalität, Handeln*. Stuttgart: Kohlhammer.
- Braun, U. (1998). Qualitäts-Monitoring: Ein praktikables Ergebnis-Feedback-System für die Psychotherapie. In A.-R. Laireiter & H. Vogel (Hrsg.), *Qualitätssicherung in der Psychotherapie und psychosozialen Versorgung – Ein Werkstattbuch* (S. 225-251). Tübingen: dgvt.
- Brocke, B. (1979). Aspekte einer Methodologie der angewandten Sozial- und Verhaltenswissenschaften. *Zeitschrift für Sozialpsychologie*, 10, 2-29.
- Brocke, B. (1993). Wissenschaftliche Fundierung psychologischer Praxis: Theoretische Psychologie, angewandte Psychologie und professionelle psychologische Praxis. In W. Bungard & T. Herrmann (Hrsg.), *Arbeits- und Organisationspsychologie im Spannungsfeld zwischen Grundlagenorientierung und Anwendung* (S. 15-47). Bern: Huber.
- Bunge, M (1967). *Scientific research. Volume 2: The Search for Truth*. Berlin: Springer.
- Donabedian, A. (1966). Evaluating the quality of medical care. *Milbank Memorial Funds Quarterly*, 44, 166-203.

- Eckert, J. (1976). Zur Prognose von psychotherapeutischen Effekten bei unterschiedlichen Behandlungsmethoden. *Zeitschrift für Klinische Psychologie*, 5, 153-163.
- Enke, H. & Lermer, S.P. (1978). Selbsterleben von Gruppenmitgliedern im Spiegel empirischer Untersuchungsbefunde. *Praxis der Psychotherapie*, 23, 13-25.
- Fahrenberg, J., Hampel, R. & Selg, H. (2001). *Das Freiburger Persönlichkeitsinventar FPI. Revidierte Fassung FPI-R und teilweise geänderte Fassung FPI-A1. Handanweisung* (7. Aufl.). Göttingen: Hogrefe.
- Fichter, M.M. (1989). Versuchsplanung experimenteller Einzelfalluntersuchungen in der Psychotherapieforschung. In F. Petermann (Hrsg.), *Einzelfallanalyse* (2. Aufl., S. 61-79). München: Oldenbourg.
- Fiedler, P. (1996). Psychotherapie. In Psychologie Verlags Union (Hrsg.), *Perspektiven der Psychologie* (S. 169-201). Weinheim: Psychologie Verlags Union.
- Fiedler, P. (2000). *Integrative Psychotherapie bei Persönlichkeitsstörungen*. Göttingen: Hogrefe.
- Fischer, G. (1974). *Einführung in die Theorie psychologischer Tests*. Bern: Huber.
- Franke, G.H. (2002). *Die Symptom-Checkliste von L.R. Derogatis (SCL-90-R): Deutsche Version* (2. Aufl.). Weinheim: Beltz.
- Frohburg, I. (1999). Qualitätssicherung in der ambulanten Gesprächspsychotherapie. *Gesprächspsychotherapie und Personenzentrierte Beratung*, 30(1), 14-27.
- Funke, W., Funke, J., Klein, M. & Scheller, R. (1987). *Trierer Alkoholismus-Inventar (TAI)*. Göttingen: Hogrefe.
- Grawe, K. (1988a). Heuristische Psychotherapie: Eine schematheoretisch fundierte Konzeption des Psychotherapieprozesses. *Integrative Psychotherapie*, 4, 309-324.
- Grawe, K. (1988b). Zurück zur psychotherapeutischen Einzelfallforschung. *Zeitschrift für Klinische Psychologie*, 17, 1-7.
- Grawe, K. (1992). Psychotherapieforschung zu Beginn der neunziger Jahre. *Psychologische Rundschau*, 43, 132-162.
- Grawe, K. (1998). *Psychologische Therapie*. Göttingen: Hogrefe.
- Grawe, K. & Baltensperger, C. (1998). Figurationsanalyse. Ein Konzept und Computerprogramm für die Prozess- und Ergebnisevaluation in der Therapiepraxis. In A.-R. Laireiter & H. Vogel (Hrsg.), *Qualitätssicherung in der Psychotherapie und psychosozialen Versorgung – Ein Werkstattbuch* (S. 179-207). Tübingen: dgvt.
- Grawe, K. & Braun, U. (1994). Qualitätskontrolle in der Psychotherapiepraxis. *Zeitschrift für Klinische Psychologie*, 23, 242-267.

- Grawe, K., Caspar, F. & Ambühl, H. (1990a). Die Berner Therapievergleichsstudie: Prozessvergleich. *Zeitschrift für Klinische Psychologie*, 19, 316-337.
- Grawe, K., Caspar, F. & Ambühl, H. (1990b). Die Berner Therapievergleichsstudie: Wirkungsvergleich und differentielle Indikation. *Zeitschrift für Klinische Psychologie*, 19, 338-361.
- Grawe, K., Donati, R. & Bernauer, F. (1994). *Psychotherapie im Wandel*. Göttingen: Hogrefe.
- Gulliksen, H. (1950). *Theory of mental tests*. New York: Wiley.
- Guthke, J., Böttcher, H. & Sprung, L. (1991). *Psychodiagnostik Bd. 1 und Bd. 2*. Berlin: Deutscher Verlag der Wissenschaften.
- Hank, P., Schwenkmezger, P. & Schumann, J. (2001). Daily Mood reports in hindsight: results of a computer-assisted time sampling study. In J. Fahrenberg & M. Myrtek (Eds.), *Progress in ambulatory assessment* (pp. 143-156). Göttingen: Hogrefe.
- Hautzinger, M., Bailer, M., Worrall, H. & Keller, F. (2000). *Beck-Depressions-Inventar (BDI). Testhandbuch* (3. Aufl.). Bern: Huber.
- Herrmann, T. (1979). *Psychologie als Problem*. Stuttgart: Klett-Cotta.
- Hessel, A., Schumacher, J., Geyer, M. & Braehler, E. (2001). Symptom-Checkliste SCL-90-R: Testtheoretische Überprüfung und Normierung an einer bevölkerungsrepräsentativen Stichprobe. *Diagnostica*, 47, 27-39.
- Hilliard, R. (1993). Single-case methodology in psychotherapy process and outcome research. *Journal of Consulting and Clinical Research*, 61, 373-380.
- Höger, D. & Eckert, J. (1997). Der Bielefelder Klienten-Erfahrungsbogen (BIKEB). Ein Verfahren zur Erfassung von Aspekten des „Post-Session Outcome“ bei Psychotherapien. *Zeitschrift für Klinische Psychologie*, 26, 129-137.
- Jacobson, N., Follette, W. & Revenstorf, D. (1984). Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. *Behavior Therapy*, 15, 336-352.
- Jacobson, N. & Revenstorf, D. (1988). Statistics for assessing the clinical significance of psychotherapy techniques: Issues problems and new developments. *Behavioral Assessment*, 10, 133-145.
- Jacobson, N. & Truax, P. (1991). Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59, 12-19.
- Janßen, H. (1997). Total Quality Management im Gesundheitswesen. In H. Spörkel, B. Ruckriegel, H. Janßen & A. Eichler (Hrsg.), *Total Quality management im*

- Gesundheitswesen. Methoden und Konzepte des Qualitätsmanagement für Gesundheitseinrichtungen* (S. 1-13). Weinheim: Psychologie Verlags Union.
- Jüttemann, S. (1981). Komparative Kasuistik als Strategie psychologischer Forschung. *Zeitschrift für Klinische Psychologie und Psychotherapie*, 29, 109-119.
- Jüttemann, S. (1990). Komparative Kasuistik als klinisch-psychologische Praxiskontrolle. In S. Jüttemann (Hrsg.), *Komparative Kasuistik* (S. 43-48). Heidelberg: Ansanger.
- Kanfer, F.H., Reinecker, H. & Schmelzer, D. (1996). *Selbstmanagement-Therapie: Ein Lehrbuch für die klinische Praxis*. Berlin: Springer.
- Kern, H. (1997). Einzelfallforschung in der Klinischen Psychologie: Multiple-Baseline-Versuchspläne. *Zeitschrift für Klinische Psychologie und Psychotherapie*, 45, 45-66.
- Kiresuk, T., Smith, A. & Cardillo, J. (Eds.). (1994). *Goal attainment scaling: Applications, theory, and measurement*. Hillsdale: Lawrence Erlbaum Associates.
- Klauer, K. & Sydow, H. (1992). Interindividuelle Unterschiede in der Lernfähigkeit: Zur Analyse von Lernprozessen bei Kurzzeitleerntests. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 3, 175-190.
- Kordy, H. & Hannover, W. (1999). Zur Evaluation psychotherapeutischer Behandlungen anhand individueller Therapieziele. In H. Ambühl & B. Strauß (Hrsg.), *Therapieziele* (S. 75-90). Göttingen: Hogrefe.
- Kordy, H., Hannover, W. & Bauer, S. (2002). Das Stuttgart-Heidelberger Modell zur Qualitätssicherung in der stationären Psychotherapie. In M. Härter, H. Linster & R.-D. Stieglitz (Hrsg.), *Qualitätsmanagement in der Psychotherapie* (S. 289-304). Göttingen: Hogrefe.
- Krampen, G. (1991a). Änderungssensitive Symptomliste (AT-SYM). In G. Krampen, *Diagnostisches und Evaluatives Instrumentarium zum Autogenen Training (AT-EVA)* (S. 14-16). Göttingen: Hogrefe.
- Krampen, G. (1991b). *Diagnostisches und Evaluatives Instrumentarium zum Autogenen Training (AT-EVA)*. Göttingen: Hogrefe.
- Krampen, G. (1991c). Veränderungsfragebogen zum Entspannungserleben und Wohlbefinden (VFE). In G. Krampen, *Diagnostisches und Evaluatives Instrumentarium zum Autogenen Training (AT-EVA)* (S. 17-18). Göttingen: Hogrefe.
- Krampen, G. (1994). *Skalen zur Erfassung von Hoffnungslosigkeit (H-Skalen)*. Handanweisung und Verbrauchsmaterialien. Göttingen: Hogrefe.

- Krampen, G. (1998). Diagnostik nach der ICD-10: Ihr Stellenwert im psychodiagnostischen Prozess sowie Hinweise zu einschlägiger Fachliteratur und psychodiagnostischen Hilfsmitteln. *Report Psychologie*, 23, 44-52 und 57-63.
- Krampen, G. (2001). Differenzialdiagnostik sowie allgemeine und differenzielle Psychotherapie pathologischer Aggressivität: Differenzialdiagnostische Einordnung, Fallbeispiele, Spekulationen. *Report Psychologie*, 26, 540-558.
- Krampen, G. (2002a). Stundenbogen für die Allgemeine und Differentielle Einzelpsychotherapie (STEP). Göttingen: Hogrefe.
- Krampen, G. (2002b). Drei-Phasen-Modell zur Geschichte der Psychotherapie im 20. Jahrhundert. *Report Psychologie*, 27, 330-346.
- Krampen, G. & von Delius, A. (1981). Zur direkten Messung subjektiv erlebter gesundheitlicher Veränderungen. *Medizinische Psychologie*, 7, 166-174.
- Krischker, S., Fleischmann, H. & Cording, C. (1998). Basisdokumentation in der Psychiatrie. In A.-R. Laireiter & H. Vogel (Hrsg.), *Qualitätssicherung in der Psychotherapie und psychosozialen Versorgung – Ein Werkstattbuch* (S. 401-417). Tübingen: dgvt.
- Laireiter, A.-R. (1994). Dokumentation psychotherapeutischer Fallverläufe. *Zeitschrift für Klinische Psychologie*, 23, 236-241.
- Laireiter, A.-R., Lettner, K. & Baumann, U. (1998). PSYCHO-DOK. Allgemeines Dokumentationssystem für Psychotherapie: Beschreibung und Funktionen im Rahmen des Qualitätsmanagements von Psychotherapie. In A.-R. Laireiter & H. Vogel (Hrsg.), *Qualitätssicherung in der Psychotherapie und psychosozialen Versorgung – Ein Werkstattbuch* (S. 209-224). Tübingen: dgvt.
- Laireiter, A.-R. & Vogel, H. (1998). Qualitätssicherung in der Psychotherapie und psychosozialen Versorgung – Einblick in die Werkstatt. In A.-R. Laireiter & H. Vogel (Hrsg.), *Qualitätssicherung in der Psychotherapie und psychosozialen Versorgung – Ein Werkstattbuch* (S. 17-46). Tübingen: dgvt.
- Laux, L., Glanzmann, P., Schaffner, P. & Spielberger, C.D. (1981). *Das State-Trait-Angstinventar (STAI). Theoretische Grundlagen und Handanweisung*. Weinheim: Beltz.
- Lienert, G.A. (1989). *Testaufbau und Testanalyse* (4. Aufl.). Weinheim: Beltz.
- Lord, F.M. & Novick, M.R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Lutz, W. (2003). *Die Wiederentdeckung des Individuums in der Psychotherapieforschung*. Tübingen: dgvt.

- Mans, E.J. (1998). Perspektiven der empirischen Erforschung der differentiellen Indikation zwischen heterogenen Therapierichtungen. *Praxis Klinische Verhaltensmedizin und Rehabilitation*, 11, 44-49.
- Meinlschmidt, G. & Krampen, G. (2003). Weltweite psychopathologische Folgen gravierender Terrorakte und ihrer Jahrestage: Ein Fallbericht zur Entwicklung einer depressiven Störung im Kontext des Jahrestages der Anschläge vom 11. September 2001. *Psychotherapie*, 8, 214-216.
- Michalak, J., Kosfelder, J., Meyer, F. & Schulte, D. (2003). Messung des Therapieerfolgs. Veränderungsmaße oder retrospektive Erfolgsbeurteilung. *Zeitschrift für Klinische Psychologie und Psychotherapie*, 32, 94-103.
- Muthny, F. A. (1989). *Fragebogen zur Krankheitsverarbeitung (FKV)*. Göttingen: Hogrefe.
- Noack, H. (1989). Neuere Entwicklungen, Software und Kritik zur Zeitreihenanalyse nach der ARIMA-Methode. In F. Petermann (Hrsg.), *Einzelfallanalyse* (2. Aufl., S. 213-234). München: Oldenbourg.
- Nübling, R. & Schmidt, J. (1998a). Qualitätssicherung in der Psychotherapie: Grundlagen, Realisierungsansätze, künftige Aufgaben. In A.-R. Laireiter & H. Vogel (Hrsg.), *Qualitätssicherung in der Psychotherapie und psychosozialen Versorgung – Ein Werkstattbuch* (S. 49-74). Tübingen: dgvt.
- Nübling, R. & Schmidt, J. (1998b). Interne Qualitätssicherung in der stationären psychosomatischen Rehabilitation: Erfahrungen mit dem „zweigleisigen Modell“. In A.-R. Laireiter & H. Vogel (Hrsg.), *Qualitätssicherung in der Psychotherapie und psychosozialen Versorgung – Ein Werkstattbuch* (S. 335-353). Tübingen: dgvt.
- Palm, W. (1998). Prozeßkontrolle für die ambulante Psychotherapie: ein Beitrag zur Qualitätssicherung aus der Perspektive psychotherapeutischer Praxistätigkeit. In A.-R. Laireiter & H. Vogel (Hrsg.), *Qualitätssicherung in der Psychotherapie und psychosozialen Versorgung – Ein Werkstattbuch* (S. 253-273). Tübingen: dgvt.
- Patry, J.-L. & Perrez, M. (2000). Theorie-Praxis-Probleme und die Evaluation von Interventionsprogrammen. In W. Hager, J.-L. Patry & M. Perrez (Hrsg.), *Handbuch Evaluation psychologischer Interventionsmaßnahmen* (S. 19-40). Bern: Huber.
- Pawlik, K. (1978). Umwelt und Persönlichkeit: Zum Verhältnis von Ökologischer und Differentieller Psychologie. In C.F. Graumann (Hrsg.), *Die ökologische Fragestellung in der Psychologie* (S. 112-134). Bern: Huber.
- Pawlik, K. (1982). Modell- und Praxisdimensionen psychologischer Diagnostik. In K. Pawlik. (Hrsg.), *Diagnose der Diagnostik* (2. Aufl., S. 13-43). Stuttgart: Klett-Cotta.

- Percevic, R., Bauer, S. & Kordy, H. (2004). Was ist klinisch bedeutsam an klinisch bedeutsamer Veränderung? Demonstration einer Validierungsstrategie am Beispiel des SCL-90R GSI. *Zeitschrift für Klinische Psychologie und Psychotherapie*, 33, 165-171.
- Perrez, M. (1991). Wissenschaftstheoretische Grundbegriffe der klinisch-psychologischen Interventionsforschung. In M. Perrez & U. Baumann (Hrsg.), *Lehrbuch Klinische Psychologie. Band 2: Intervention* (S. 51-63). Bern: Huber.
- Petermann, F. (1982). *Einzelfalldiagnose und klinische Praxis*. Stuttgart: Kohlhammer.
- Petermann, F. (1996). Kontrollierte Praxis. In H.S. Reinecker & D. Schmelzer (Hrsg.), *Verhaltenstherapie, Selbstregulation, Selbstmanagement* (S. 75-82). Göttingen: Hogrefe.
- Perst, A. & Baumann, U. (1999). Einzelfallstudien in klinisch-psychologischen, psychotherapeutischen Fachzeitschriften. *Zeitschrift für Klinische Psychologie*, 28, 205-213.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Kopenhagen: Nielsen & Lydiche.
- Revenstorf, D. & Keeser, W. (1989). Zeitreihenanalyse von Therapieverläufen – ein Überblick. In F. Petermann (Hrsg.), *Einzelfallanalyse* (2. Aufl., S. 167-212). München: Oldenbourg.
- Revenstorf, D. & Vogel, B. (1989). Zur Analyse qualitativer Verlaufsdaten. In F. Petermann (Hrsg.), *Einzelfallanalyse* (2. Aufl., S. 235-256). München: Oldenbourg.
- Richter, R. (1994). Qualitätssicherung in der Psychotherapie (Editorial). *Zeitschrift für Klinische Psychologie*, 23, 233-235.
- Rost, J. (1996). *Lehrbuch der Testtheorie und Testkonstruktion*. Bern: Huber.
- Rost, J. & Spada, H. (1983). Die Quantifizierung von Lerneffekten anhand von Testdaten. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 4, 29-49.
- Rückert, D. & Linster, H.W. (1998). Qualitätssicherung und Qualitätsmanagement im Rahmen der ambulanten Psychotherapie mit Kindern, Jugendlichen und ihren Bezugspersonen. In A.-R. Laireiter & H. Vogel (Hrsg.), *Qualitätssicherung in der Psychotherapie und psychosozialen Versorgung – Ein Werkstattbuch* (S. 421-456). Tübingen: dgvt.
- Scheele, B. (1982). Psychotherapie und ihre allgemein-psychologische Grundlegung. In R. Bastine, P.A. Fiedler, K. Grawe, S. Schmidtchen & G. Sommer (Hrsg.), *Grundbegriffe der Psychotherapie* (S. 143-148). Weinheim: Edition Psychologie.

- Schindler, L., Hohenberger-Sieber, E. & Hahlweg, K. (1990a). Stundenbeurteilung (Klient). In G. Hank, K. Hahlweg & N. Klann (Hrsg.), *Diagnostische Verfahren für Berater* (S. 331-335). Weinheim: Beltz-Test.
- Schindler, L., Hohenberger-Sieber, E. & Hahlweg, K. (1990b). Stundenbeurteilung (Therapeut). In G. Hank, K. Hahlweg & N. Klann (Hrsg.), *Diagnostische Verfahren für Berater* (S. 337-339). Weinheim: Beltz-Test.
- Schneider, S. & Margraf, J. (2005). *Diagnostisches Interview bei psychischen Störungen (DIPS für DSM-IV-TR)*. Berlin: Springer (im Druck).
- Schneider, W. (1990). Die Psychotherapiemotivation. In W. Schneider (Hrsg.), *Indikationen zur Psychotherapie* (S. 183-201). Weinheim: Beltz.
- Schroeder, H. & Wildman, B. (1988). Single case designs in clinical settings. *Hospice Journal*, 4, 3-24.
- Selg, H., Klapproth, J. & Kamenz, R. (1992). *Forschungsmethoden der Psychologie*. Stuttgart: Kohlhammer.
- Seligman, M. (1995). The effectiveness of psychotherapy – The consumer reports study. *American Psychologist*, 50, 965-974.
- Smith, E. (2001). *Die Gestaltung produktiver Therapiesitzungen in der psychologischen Psychotherapie*. Vortrag auf dem 19. Symposium der Fachgruppe Klinische Psychologie und Psychotherapie der DGPs in Bern (Schweiz), 24.-26. Mai 2001.
- Stangier, U., Hoyer, J., Kosfelder, J. & Meyer, F. (1998). Dokumentation von Psychotherapiesitzungen. *Zeitschrift für Klinische Psychologie*, 27, 172-176.
- Stein, B. & Herzog, T. (2002). Qualitätsmanagement in der psychosomatisch-psychotherapeutischen Konsiliar- und Liaisonversorgung. In M. Härter, H. Linster & R.-D. Stieglitz (Hrsg.), *Qualitätsmanagement in der Psychotherapie* (S. 305-324). Göttingen: Hogrefe.
- Steyer, R., Hannöver, W., Telser, C. & Kriebel, R. (1997). Zur Evaluation intraindividueneller Veränderung. *Zeitschrift für Klinische Psychologie*, 26, 291-299.
- Strauss, B. & Eckert, J. (1994). Dimensionen des Gruppenerlebens. *Zeitschrift für Klinische Psychologie*, 23, 188-201.
- Strupp, H. (1980a). Success and failure in time-limited psychotherapy: A systematic comparison of two cases (Comparison 1). *Archives of General Psychiatry*, 37, 595-603.
- Strupp, H. (1980b). Success and failure in time-limited psychotherapy: A systematic comparison of two cases (Comparison 2). *Archives of General Psychiatry*, 37, 708-716.

- Strupp, H. (1980c). Success and failure in time-limited psychotherapy: With special reference to the performance of a lay counselor (Comparison 3). *Archives of General Psychiatry*, 37, 831-841.
- Strupp, H. (1980d). Success and failure in time-limited psychotherapy: Further evidence (Comparison 4). *Archives of General Psychiatry*, 37, 947-954.
- Strupp, H., Schacht, T., Henry, W. & Binder, J. (1992). Jack M.: A case of premature termination. *Psychotherapy*, 29, 191-205.
- Thorngate, W. (1986). The production, detection, and explanation of behavioral patterns. In J. Valsiner (Ed.), *The individual subject and scientific psychology* (pp. 71-93). New York: Plenum Press.
- Viethen, G. (1994). Qualitätssicherung in der Medizin. Teil I. *QualiMed*, 2, 9-16.
- Westmeyer, H. (1978). Wissenschaftstheoretische Grundlagen Klinischer Psychologie. In U. Baumann, H. Berbalk & G. Seidenstücker (Hrsg.), *Klinische Psychologie: Trends in Forschung und Praxis* (Bd. 1, S. 108-132). Bern: Huber.
- Westmeyer, H. (1979). Die rationale Rekonstruktion einiger Aspekte psychologischer Praxis. In H. Albert & K.H. Stapf (Hrsg.), *Theorie und Erfahrung* (S. 139-161). Stuttgart: Klett-Cotta.
- Zielke, M. (1979a). *Kieler Änderungssensitive Symptomliste (KASSL)*. Weinheim: Beltz.
- Zielke, M. (1979b). *Indikation zur Gesprächspsychotherapie*. Stuttgart: Kohlhammer.
- Zielke, M. (1980). Untersuchung der Gütekriterien des Klienten-Erfahrungs-Bogens (KEB). *Diagnostica*, 26, 57-73.
- 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. (2001). 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.