

Robert Percevic

**ENTWICKLUNG UND EVALUATION EINES COMPUTERGESTÜTZTEN
TESTDARBIETUNGSSYSTEMS FÜR THERAPIEBEGLEITENDES
ERGEBNISMONITORING IN DER PSYCHOTHERAPIE**

Inauguraldissertation zur Erlangung des akademischen Grades
eines Doktors der Sozialwissenschaften der Universität

Mannheim

2005

Dekan: Prof. Dr. Peter Drewek
Gutachter: Prof. Dr. Werner W. Wittmann
Prof. Dr. Horst Kächele
Tag der Disputation: 28. April 2005

INHALTSVERZEICHNIS

INHALTSVERZEICHNIS	2
TABELLENVERZEICHNIS	4
ABBILDUNGSVERZEICHNIS	5
ABKÜRZUNGEN UND SYMBOLE	6
EINFÜHRUNG	7
Einsatz von psychometrischen Instrumenten in der Psychotherapie	7
Qualitätssicherung	8
Das Stuttgart-Heidelberger Model und AKQUASI	11
Therapiebegleitendes Ergebnismonitoring	13
Computergestützte Testdarbietung	16
Äquivalenz zwischen computergestützter und Papier/Bleistift-Testdarbietung	19
Computerspezifische Möglichkeiten bei der Testdarbietung	25
Adaptives und entscheidungsorientiertes Testen	26
Erfassen von Itemresponsezeiten	30
FRAGESTELLUNG	35
METHODE	44
Untersuchungsdesign	44
Technik	48
Darstellungsvarianten	49
Entscheidungsorientierte Vorgabe	53
Zusatzbefragung	55
Rückmeldungen an Therapeuten	56
Stichprobengröße und Zeitplan für die Datenerhebung	58
Ablauf	60
Stichprobenbeschreibung	64
AUSWERTUNG UND ERGEBNISSE	67
Psychometrische Äquivalenz	71
Akzeptanz	85
Testökonomie	92
Effekte der Rückmeldungen auf die Behandlung	96

DISKUSSION	100
Psychometrische Äquivalenz	100
Akzeptanz	102
Testökonomie	105
Effekte der Rückmeldungen	107
ZUSAMMENFASSUNG UND AUSBLICK	112
Entwicklung eines computergestützten Monitoringsystems (Web-AKQUASI)	114
Erweiterung der Fragestellung	119
LITERATUR	123

TABELLENVERZEICHNIS

Tabelle 1. Bei der QS eingesetzte Instrumente.....	45
Tabelle 2. Untersuchungsdesign.....	48
Tabelle 3. Reihenfolge der Items und geschätzte Sequenz- Gesamt-Korrelationen bei entscheidungsorientierten Testen	54
Tabelle 4. Ereignisse im Laufe der Studie.....	63
Tabelle 5. Stichprobenumfang.....	64
Tabelle 6. Stichprobenbeschreibung	66
Tabelle 7. Erwartete Effekte.....	70
Tabelle 8. Verteilungsparameter und Reliabilität der computergestützten Vorgabeformen.....	72
Tabelle 9. Verteilungsparameter und Reliabilität der Papier/Bleistift-Vorgabe und der gepoolten computergestützten Vorgabeformen.....	79
Tabelle 10. Durchschnittliche Retestkorrelationen und Interkorrelationen der Zwischenerhebungsinstrumente zu den einzelnen Vorgabeformen.....	80
Tabelle 11. Anpassungsgüte der Messmodelle bei vorgabespezifischer und bei gemeinsamer Schätzung der Faktorladungen	82
Tabelle 12. Verteilungsparameter der Skalenwerte bei entscheidungsorientiertem und handbuchgemäßem Vorgehen.....	83
Tabelle 13. Akzeptanz und Handhabung der einzelnen Vorgabeformen.....	86
Tabelle 14. Zusammenhang zwischen personenbezogenen Variablen und Akzeptanzfragen (τ).....	87
Tabelle 15. Anteil fehlender Werte zur Zwischenerhebung.....	88
Tabelle 16. Häufigkeit und durchschnittliche Länge von Antworten auf offene Fragen.....	89
Tabelle 17. Anteil fehlender Werte zur Entlassung.....	90
Tabelle 18. Häufigkeit einzelner Antwortkategorien zu der Frage „Was hat sie besonders gestört?“.....	91
Tabelle 19. Häufigkeit einzelner Antwortkategorien zu der Frage „Was hat sie besonders gefreut?“.....	91
Tabelle 20. Durchschnittliche Zeiten für das Beantworten des gesamten Fragebogens und durchschnittliche Zeiten zwischen zwei Items.....	93
Tabelle 21. Annahmen zur Schätzung der Testkosten.....	95
Tabelle 22. Dauer der Weiterbehandlung nach der Zwischenerhebung (in Tagen).....	97

ABBILDUNGSVERZEICHNIS

Abbildung 1. Qualitätskreislauf.....	9
Abbildung 2. Seitenweise Darstellung des Fragebogens.....	50
Abbildung 3. Beispiel einer Rückmeldung an Patienten.....	53
Abbildung 4. Zusatzfragebogen.....	56
Abbildung 5. Beispiel einer Rückmeldung an Therapeuten.....	58
Abbildung 6. Beispiel einer Instruktionsseite.....	63
Abbildung 7. Itemschwierigkeiten bei Papier/Bleistift Vorgabe und bei computergestützter Vorgabe.....	76
Abbildung 8. Itemtrennschärfe bei Papier/Bleistift Vorgabe und bei computergestützter Vorgabe.....	77
Abbildung 9. Vergleich der Rohwerte bei Papier/Bleistift- Vorgabe (-) und bei computergestützter Vorgabe (--) [mit Papier/Bleistift-Vorgabe als Referenz].....	78
Abbildung 10. Verteilung der Anzahl vorgegebener Items bei entscheidungsorientierter Testvorgabe.....	94
Abbildung 11. Geschätzte Befragungskosten.....	96
Abbildung 12. Behandlungsdauer in Tagen bis zur Entlassung in klinisch bedeutsam gebesserten Zustand bei Rückmeldung an Therapeuten (-) und ohne Rückmeldung an Therapeuten (--) [Entlassung ohne klinisch bedeutsame Veränderung (+)].....	99

ABKÜRZUNGEN UND SYMBOLE

η	Effektstärke
α	Wahrscheinlichkeit die Null-Hypothese fälschlicherweise abzulehnen
β	Wahrscheinlichkeit die Alternativ-Hypothese fälschlicherweise abzulehnen
χ^2 -Test ...	Chi-Quadrat Test
τ	Kendall-Tau Korrelationskoeffizient
EB45-BEZ .	Ergebnisbogen, Zwischenmenschliche Beziehungen
EB45-SOZ .	Ergebnisbogen, Soziale Integration
EB45-SYB .	Ergebnisbogen, Symptombelastung
EB45-TOT .	Ergebnisbogen, Gesamtskala
ES	Effektstärke
FS	Forschungsstelle für Psychotherapie Stuttgart
GBB	Giesener Beschwerdebogen
GLM	Allgemeines Lineares Modell
KS-Test ..	Kolmogorov-Smirnov Test
M	Mittelwert
Med	Median
n	Stichprobengröße
p	Wahrscheinlichkeit
P/B	Papier/Bleistift Testvorgabe
PAE	Fragebogen zur direkten Veränderungseinschätzung, Patientensicht
PC	Computergestützte Testvorgabe
PDA	Personal Digital Assistent
PTBEZ	Therapeutische Arbeitsbeziehung, Patientensicht
q1	Erstes Quartil
q3	Drittes Quartil
QS	Qualitätssicherung
R	Multiple Korrelation
r	Pearsonscher Korrelationskoeffizient
SD	Standardabweichung
TTBEZ	Therapeutische Arbeitsbeziehung, Therapeutensicht
ZUF8	Fragebogen zu Erfassen der Patientenzufriedenheit

EINFÜHRUNG

Einsatz von psychometrischen Instrumenten in der Psychotherapie

Psychometrische Instrumente, vor allem Selbstbeurteilungsverfahren, werden seit über 80 Jahren in der klinischen Psychologie und Psychotherapie eingesetzt (DuBois, 1970). Versprechen diese doch ein „objektives“ Bild vom Patienten¹, seinen Bedürfnissen, Problemen und Stärken zu geben, welches in die Therapieindikation, Therapieplanung und Therapieevaluation einbezogen werden kann (siehe z.B. Laireiter, 2000; Stieglitz, 2000). Obwohl die Nützlichkeit von psychometrischen Instrumenten für die Therapiegestaltung auch kritisch diskutiert wird (siehe z.B. Houben, 1970; Laireiter, 2000; Evans, 2004), gaben bei einer Umfrage des Testkuratoriums über 67% aller befragten Psychologen an, in ihrer Berufspraxis psychologische Testverfahren einzusetzen, wobei drei der fünf meistbenutzten Verfahren Selbstbeurteilungsverfahren waren (Steck, 1997). Watkins et al. (1995) fanden, dass in den USA sogar 91% aller praktisch tätigen Psychologen Testverfahren einsetzen, wenn auch mit sinkendem Zeitanteil, von 28% ihrer Arbeitszeit 1969 zu 20% 1995. In der Schweiz erklärten 49% der Therapeuten in der

¹Die maskuline Schreibweise wird ausschließlich der besseren Lesbarkeit wegen verwendet. Selbstverständlich sind mit Patienten auch Patientinnen gemeint, mit Therapeuten auch Therapeutinnen usw.

Erwachsenenpsychotherapie Testdiagnostik einzusetzen (Hänsgen & Bernasconi, 2000).

Zahlreiche Untersuchungen über die prognostische Überlegenheit statistischer Entscheidungsverfahren im Vergleich zu klinischen und über die prognostische Überlegenheit von psychometrischen Instrumenten im Vergleich zu Interview und Beobachtung, unterstreichen die (potenzielle) Nützlichkeit dieser Verfahren (siehe Metaanalysen von Meehl, 1954; Sawyer, 1966; Grove et al., 2000; siehe Meyer & Schulte, 2002 speziell zur Psychotherapie).

Qualitätssicherung

Eine besondere Rolle nehmen psychometrische Instrumente bei der Qualitätssicherung in der psychotherapeutischen Versorgung ein. Ansätze einer „Qualitätssicherung“ sind schon seit langem in der psychotherapeutischen Versorgung zu beobachten: so können Supervision und kontinuierliche Weiterbildung als qualitätssichernde Maßnahmen begriffen werden. Systematische Qualitätssicherung verbreite sich in der Gesundheitsversorgung jedoch erst ab Mitte der 70er Jahre.

Die Abfolge systematischer Qualitätssicherung kann durch den Qualitätskreislauf beschrieben werden. Dieser ist von paradigmatischer Bedeutung für alle qualitätssichernden Aktivitäten (Abbildung 1; vgl. AHCPR, 1995). Der Qualitätskreislauf hebt u.a. einen der Unterschiede zwischen Evaluation und Qualitätssicherung hervor: die Datenerhebung

erfolgt bei der Evaluation in der Regel einmalig, bei der Qualitätssicherung dagegen kontinuierlich oder zyklisch (Farin & Bengel, 2003).

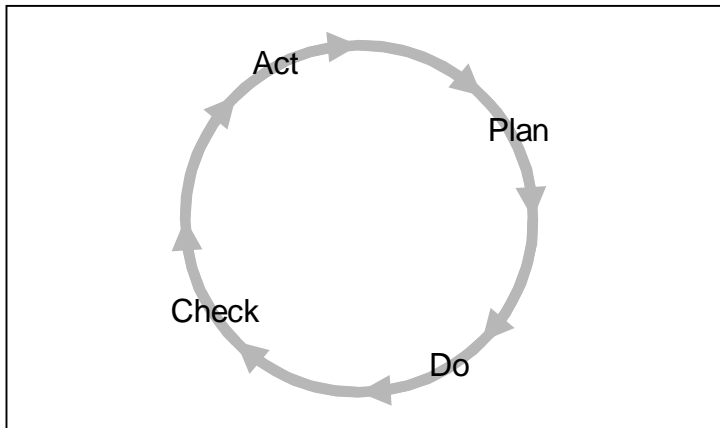


Abbildung 1. Qualitätskreislauf

Bei der Umsetzung einzelner Maßnahmen zur Qualitätssicherung sind im Gesundheitswesen, neben den Entwicklungen in der Industrie (statistische Qualitätskontrolle, DIN ISO-Normen 8402 und 9000-9004 zum Qualitätsmanagement, Konzept des „Total Quality Management“), die Aufteilung in Struktur-, Prozess- und Ergebnisqualität sowie die Unterscheidung von externer und interner Qualitätssicherung von besonderer Bedeutung. Vor dem Hintergrund eines steigenden Effizienz- und Legitimationsdrucks im Gesundheitswesen wurde seit den späten 80ern die Betonung zunehmend auf Ergebnisqualität und externe Qualitätssicherung gelegt.

Die Aufteilung in interne und externe Qualitätssicherung bezieht sich darauf, ob die Qualitätskriterien und gegebenenfalls die Qualitätsfeststellung von einer externen

Institution oder institutionsintern durchgeführt wird. Die Aufteilung in Strukturqualität („physical, organizational, and other characteristics of the system that provides care and of its environment“), Prozessqualität („what is done in caring for patients“) und Ergebnisqualität („what is achieved, an improvement usually in health...“) (Donebian, 1986, pp. 99) unterscheidet Betrachtungsebenen bezüglich der Qualität der medizinischen Versorgung.

In Deutschland wurde Qualitätssicherung im Gesundheitswesen erstmalig 1989 gesetzlich verankert (Sozialgesetzbuch V, Gesundheitsreformgesetz²). Besonderst hervorgehoben wurde dabei ebenfalls die Verpflichtung zu einrichtungsübergreifenden Maßnahmen sowie die Rolle der Ergebnisqualität:

§ 135a Verpflichtung zur Qualitätssicherung

Die Leistungserbringer sind zur Sicherung und Weiterentwicklung der Qualität der von ihnen erbrachten Leistungen verpflichtet. Die Leistungen müssen dem jeweiligen Stand der wissenschaftlichen Erkenntnisse entsprechen und in der fachlich gebotenen Qualität erbracht werden. Vertragsärzte, zugelassene Krankenhäuser sowie Erbringer von Vorsorgeleistungen oder Rehabilitationsmaßnahmen sind verpflichtet, sich an

² Dieses wurde 1993 durch das Gesundheitsstrukturgesetz, 1997 durch das GKV-Neuordnungsgesetz und zuletzt 2000 durch das GKV-Gesundheitsreformgesetz und das Psychotherapeutengesetz erweitert.

einrichtungsübergreifenden Maßnahmen der Qualitätssicherung zu beteiligen, die insbesondere zum Ziel haben, die Ergebnisqualität zu verbessern. Zugelassene Krankenhäuser, stationäre Vorsorgeeinrichtungen und stationäre Rehabilitationseinrichtungen sind verpflichtet, einrichtungsintern ein Qualitätsmanagement einzuführen und weiterzuentwickeln.

(nach Gerlach, 2001, p. 148)

Eine Reihe von Implementierungen qualitätssichernder Maßnahmen, welche diese Vorgaben aufgriffen und für die psychotherapeutische und psychosomatische Versorgung anpassten, folgte (für einen Überblick s. Laireiter & Vogel, 1998; Härter, Linster & Stieglitz, 2003). Der Wert kurzer, symptomorientierter psychometrischer Instrumente kam dabei besonders zum Tragen. Diese wurden verschiedentlich verwendet um Indikatoren, Kriterien und Standards für eine institutionsübergreifende Bewertung der Ergebnisqualität psychotherapeutischer Behandlungen zu entwickeln (z.B. Schmidt et al., 2003; Stieglitz, 2003; Kordy & Hannover, 2000; vgl. Cates, 1999; Piotrowski, 1999).

Das Stuttgart-Heidelberger Model und AKQUASI

Das Stuttgart-Heidelberger Model für Qualitätssicherung entstand 1992/1993 als Kooperationsprojekt zwischen der

Forschungsstelle für Psychotherapie Stuttgart und der Psychiatrischen Universitätsklinik Heidelberg (Kordy & Lutz, 1995). Inzwischen wird das Model an über 25 Kliniken eingesetzt. Dabei werden alle Behandlungsverläufe nach einem konsensuell vereinbarten Bewertungsalgorithmus, welcher auf dem Konzept reliabler, klinisch bedeutsamer Veränderung aufbaut, beurteilt (siehe Jacobson et al., 1999; Kordy & Hannover, 2000; Percevic, Bauer & Kordy, 2004). Dieser Bewertungsalgorithmus ist inzwischen in zentralen Punkten validiert worden (Kordy & Hannover, 1998). Die Grundlage für die Beurteilung bilden Daten, welche bei der Aufnahme und bei der Entlassung erhoben werden. Das dabei eingesetzte Fragebogeninventar basiert auf dem Konzept einer multimodalen Datenerhebung (z.B. Kordy & Kächele, 1996; Schulte, 1993; Wittmann, 1987) und schließt eine Reihe bewährter psychometrischer Instrumente ein, u.a. den SCL-90R (Franke, 1995), den IIP (Horowitz et al. 2000) und den GBB (Brähler & Scheer, 1997). In Qualitätszirkeln werden die Bewertungen an das therapeutische Team rückgemeldet und fallbezogen diskutiert. Schließlich können die Ergebnisse in Form von Jahresberichten zusammengefasst und für den Vergleich zwischen Zeiträumen und Kliniken genutzt werden. Zur Unterstützung der Dateneingabe, Datenverwaltung und Rückmeldung wurde ein spezielles Computerprogramm (AKQUASI) entwickelt (Kordy et al., 2001).

Therapiebegleitendes Ergebnismonitoring

Ungeachtet der Notwendigkeit einer fallbezogenen Qualitätsfeststellung zu Ende der Behandlung, sind damit die Möglichkeiten psychometrischer Instrumente für eine Qualitätsförderung in der Psychotherapie bei weitem nicht ausgeschöpft. Kliniker haben Schwierigkeiten sowohl bei dem Lernen aus klinischer Erfahrung wie auch bei therapiebezogenen Einschätzungen (z.B. Garb, 2000; Schulte & Meyer, 2002). So gibt es seit langem die Forderung nach Verfahren zur Klinischen Differential- und Verlaufsdiagnostik (Steck, 1997) und nach treatmentorientierter Diagnostik (Beutler, 1989; Stein, 1987; Jäger & Krieger, 1994; Groth-Marnat, 2000).

Entsprechende Bemühungen um eine Einbindung der im Rahmen der Qualitätssicherung erfassten Daten in laufende Therapien begannen an der Forschungsstelle Mitte der 90er Jahre. Erste Versuche aus den Aufnahmedaten des Stuttgart-Heidelberger-QS-Inventars sinnvolle Rückmeldungen zu konstruieren scheiterten an der mangelnden Vorhersagekraft dieser Daten für den Therapieverlauf (Hannöver, 2002). Generell stellt sich die Frage welche Art von Information die Therapie sinnvoll unterstützen kann. Nelson-Grey et al. (1989) und Schulte (1991) fanden, dass eine Individualisierung der Therapie anhand von Symptomatik und Problemanalyse sich nicht als förderlich für die Therapieergebnisse erweist. Lambert et al. (2001a; 2001b; 2002a) hingegen konnten zeigen, dass eine Rückmeldung über frühen „Response“ bzw. „Nonresponse“ an

Therapeuten und Patienten die Therapieerfolgsrate erhöht, wobei die Therapiedauer für frühe „Nonresponder“ ansteigt und für frühe „Responder“ sinkt.

Zur Konstruktion eines Rückmeldesignals auf der Basis von Verlaufsdaten wurden von der Forschungsstelle für Psychotherapie zwischen 1999 und 2001 entsprechende Gesundungsverläufe an zwei psychotherapeutischen Kliniken erhoben. Die Ergebnisse dieser Untersuchungen führten zur Konstruktion eines Rückmeldesignals auf Basis der jeweils aktuellsten Messung der Symptombeeinträchtigung, mit dem Ziel einer ergebnisorientierten Steuerung der Therapiedauer (Percevic, Lambert & Kordy, eingereicht; Percevic, Lambert & Kordy, 2004).

Um die Ergebnisse von Zwischenerhebungen rechtzeitig an Therapeuten weiterzugeben, was überhaupt erst die Voraussetzung für einen Einfluss auf die laufende Behandlung schafft, müssen diese Informationen zeitnah zur Verfügung stehen. Mit einer Organisation der Datenerfassung und Datenauswertung wie sie für die Qualitätssicherung mit Prä-Post-Evaluation gut geeignet ist (d. h. Vorlage von Papier/Bleistift-Fragebögen in den Kliniken, Versand an die Forschungsstelle, wo diese gescannt und ausgewertet werden), kann dies nicht oder nur unter erheblichen Schwierigkeiten gewährleistet werden. Zeitnahe Rückmeldungen lassen sich in der Routineanwendung nur bei direkter Beantwortung der entsprechenden Fragebögen am Rechner realisieren. Daher begann

die Forschungsstelle 1999 die Weiterentwicklung von AKQUASI in Richtung eines Systems zur Unterstützung von therapiebegleitendem Ergebnismonitoring, wobei besonderer Wert auf die Möglichkeit zeitnaher Datenerfassung und Rückmeldung gelegt wurde. Die Software wurde als Internetapplikation entwickelt, so dass von jedem Arbeitsplatz aus Daten eingegeben und Rückmeldungen eingesehen werden können (Percevic et al., 2000; Percevic, 2003; vgl. Döge, 1997 zu den Vorteilen von Internettechnologie gegenüber klassischer Client/Server-Technologie).

Vor einem Routineeinsatz der computergestützten Vorgabe des QS-Inventars stellte sich jedoch die Frage nach der Praktikabilität und Legitimität dieses Vorgehens. Entsprechende Vorgaben dazu setzen die Richtlinien für den Einsatz elektronischer Datenverarbeitung in der psychologischen Diagnostik der Föderation Deutscher Psychologenvereinigungen (1986) und die Standards für Testkonstruktion der American Psychological Association (1999) (vgl. auch die DIN 33430, Hornke & Winterfeld, 2004). Im Folgenden ein Ausschnitt aus den Standards für Testkonstruktion:

Standard 6.11

If a test is designed so that more than one method can be used for administration or for recording responses - such as marking responses in a test booklet, on a separate

answer sheet, or on a computer keyboard - then the manual should clearly document the extent to which scores arising from these methods are interchangeable. If the results are not interchangeable, this fact should be reported, and guidance should be given for the interpretation of scores obtained under the various conditions or methods of administration.

(APA, 1999, p. 70)

Computergestützte Testdarbietung

Der erste routinemäßige Einsatz von Computern bei der Testanwendung begann an der Mayo-Klinik in den frühen 60ern mit dem MMPI (Rome et al., 1962) und dem 16PF (Eber, 1964 nach Fowler, 1985). Dabei handelte es sich noch nicht um Testdarbietung am Rechner, sondern um Software für die computerisierte Auswertung und Interpretation dieser Testverfahren. Eine der ersten mechanischen Testdarbietungen erfolgte für den WAIS über Rückwandprojektion (Elwood, 1969) und schnell darauf folgten die ersten Testdarbietungen über Computer (Hedl et al., 1973 für den Slossen Intelligence Test; Lushene et al., 1974 und Butcher & Pancheri, 1976 für den MMPI). Schon 1984 gab es über 112 computerisierte Testverfahren für 72 Instrumente (Fowler, 1985) und inzwischen wird von großen Testanbietern wie National Computer Systems, Psychological Assessment Resources und Psychological Corporation im englischsprachigen und Hogrefe und Schuhfried

im deutschsprachigen Raum wohl jedes gängige Testverfahren in computerisierter Form angeboten. Zunehmend wird auch internetbasiertes Testen angeboten (z.B. bei Hogrefe seit 2004).

Mit dem Einsatz von Rechnern bei der Testanwendung sind zahlreiche Vorteile verbunden. Cohen & Moreland (1992) listen Kostenreduktion, Zeitersparnisse, höhere Konsistenz, Flexibilität und Genauigkeit der Skalenberechnung und sofortige Verfügbarkeit der Ergebnisse auf. Butcher et al. (1985) gehen näher auf die Chancen adaptiver Testdarbietung ein. Eine ausführliche Übersicht zu möglichen Verbesserungen durch den Computereinsatz geben Maiwald & Conrad (1993). Darunter sind die Reduktion von Versuchsleitereffekten, maßgeschneidertes Instruieren, Absicherung gegen Täuschungsversuche, multimediale Darstellung von Situationen, vielfältige Eingabemodalitäten, individuelle Rückmeldung, Erfassung von Bearbeitungszeiten und komplexen Reaktionen, sowie die automatische Testinterpretation. Applikationen, welche auf komplexe Fragebögen, wiederholte Testdarbietung oder zeitnahe Rückmeldungen angewiesen sind, lassen sich einfacher oder nur mit computergestützter Testdarbietung realisieren.

Trotz solcher potentieller Vorteile computerunterstützten Testens ist dieses nicht weit verbreitet. Nach Steck (1997) nutzen nur 8% der psychologischen Testanwender in Deutschland computergestützte Testsysteme. In den USA beträgt der Anteil

computergestützten Testens 13% (McMinn et al., 1999). Für die Schweiz fand Hänsgen (1999), dass ca. 11% der Psychologen Computer für die Testdurchführung benutzen.

Roid (1986, nach Booth, 1990) weist darauf hin, dass computergestützte Testdarbietung keinen Erfolg haben kann, solange sie keine ökonomischen und methodischen Vorteile vor einer konventionellen Testvorgabe hat. Obwohl computergestützte Testdarbietung im allgemeinen schneller bearbeitet werden Papier/Bleistift-Vorgaben (z.B. Vansickle & Kapes, 1993) und durchaus Kostenersparnisse bringen kann (z.B. Stein, 1987; Bicanich et al., 1997), sind für die üblichen Zwecke Papier/Bleistift-Fragebögen oft die praktischere und ökonomischere Variante, brauchen diese doch keinen „Gerätepark“, keine speziell eingerichteten Räume, keine spezielle Schulung und haben keinen Wartungsbedarf (Erdman et al, 1985; vgl. Steck, 1997 dessen Probanden als entscheidendes Qualitätsmerkmal für die Anwendung eines Tests seine Testökonomie herausstellen). Da es nur wenige speziell für Computerdarbietung konstruierte Verfahren gibt, meist handelt es sich um Adaptationen von Papier/Bleistift-Verfahren, können viele der methodischen Vorteile einer Computerdarbietung gar nicht zum Tragen kommen. Die ungesicherte Legitimität der meisten dieser Adaptationen³ (Butcher et al., 1985; Kubinger & Farkas, 1991; Cohen & Moreland, 1992; Mertens, 1999), mangelnde Erfahrung und unzureichendes Training der Anwender

³ auch der kommerziell angebotenen

mit Computertechnologie (Olson, 2001) und Befürchtungen von Seiten der Diagnostiker, wichtige Informationen zur Testbearbeitung und den menschlichen Kontakt zu verlieren⁴ (z.B. Fahrenberg et al., 2002), sind sicher weitere Gründe für deren mangelnde Verbreitung.

Äquivalenz zwischen computergestützter und Papier/Bleistift-Testdarbietung

Zur Frage der Äquivalenz von Papier/Bleistift-Testdarbietung und computergestützter Testdarbietung liegen zahlreiche Studien, beginnend in den frühen 70ern, vor (Lushene et al., 1974 der eine Computeradaptation des MMPI untersuchte). Aktuelle deutschsprachige Übersichten geben Franke (1998) und Klinck (2002). Als Untersuchungsdesigns wurden sowohl balancierte Test-Retest-Designs, wie auch unabhängige Zufallsstichproben verwendet (siehe Diskussion zu den jeweiligen Vorteilen und Nachteilen in Kubinger & Farkas, 1991 und Klinck, 2002). Zu einzelnen Instrumenten, vor allem dem MMPI, liegen auch Metaanalysen vor (z.B. Finger & Ones, 1999). Aktuell gehen die Studien vermehrt in Richtung „internetbasiertes Testen“ (z.B. Dillman et al., 1998a; Barak & English, 2002; Lumsden et al., 2002; vgl. Butcher et al., 2004).

⁴ Wobei computergestützte Testdarbietung sicher nicht unpersönlicher als Papier/Bleistift-Verfahren ist und auch die Verhaltensbeobachtung nicht ausschließt vorauf z.B. Fowler, 1985 und Hänsgen, 1999 hinweisen.

Oft wird zwischen psychometrischer Äquivalenz, populationsbezogener Äquivalenz und erfahrungsbezogener Äquivalenz unterschieden (z.B. Klinck, 1998; Schwenkmezger & Hank, 1993; Honaker, 1988). Nach Marco (1981) bezieht sich die psychometrische Äquivalenz auf Änderungen der Reihenfolge der Probanden, Änderungen der Metrik und Verschiebungen im Niveau. Bei Änderungen der Metrik oder Verschiebungen im Niveau können die Testverfahren nach einer Neunormung oder der Anwendung entsprechender Gleichungen zum „Equalizing“ am Rechner vorgegeben und mit Papier/Bleistift-Ergebnissen verglichen werden (siehe z.B. Kolen & Brennan, 1995 zu entsprechenden Verfahren). Problematischer sind Änderungen der Reihenfolge der Probanden da davon die Konstruktvalidität betroffen ist und kein einfaches Angleichen mehr möglich ist (Hofer & Green, 1985). Zusammenfassend lässt sich bezüglich der Befunde zur psychometrischen Äquivalenz sagen, dass bei Leistungstests, vor allem solchen mit einer hohen Speed-Komponente, die Äquivalenz häufig nicht gegeben zu sein scheint und am Rechner oft schlechtere Ergebnisse erzielt werden (z.B. Klinck, 1998; Troche et al., 2002) während bei Selbstbeurteilungsverfahren meist kein Einfluss der Darbietungsform gefunden wurde (z.B. Franke, 1998; Rammstedt, 2004; speziell für Internetbefragungen siehe z.B. Smith & Leight, 1997; Stanton, 1998; Buchanan & Smith, 1999; Buchanan, 1999). Wo Unterschiede bei Selbstbeurteilungsverfahren auftreten, deuten diese meist in Richtung einer höheren Datenqualität mit weniger Tendenz zu

sozial erwünschten Antworten (Joinson, 1999) und größerer Offenheit bei sensiblen Fragen (z.B. Robinson & West, 1992). Als mögliche Ursache werden u.a. der Einfluss eines Eindrucks größerer Wichtigkeit und Objektivität bei Computerbefragung (Tourangeau & Smith, 1996) und der Einfluss größerer Selbstaufmerksamkeit (Franke, 1998) diskutiert. Ein testspezifischer Faktor, der möglicherweise zu größerer Nichtäquivalenz führt, ist das Ausmaß an Zustandsorientierung der Skala, wobei computergestützte Darstellung zu höherer Zustandsangst und höheren Zustandsängsten führen können was sich dann auf entsprechend erhöhten Skalenwerten reflektiert (Hedl et al., 1973; Schwenkmezger & Hank, 1993; siehe dazu auch den Abschnitt über populationsbezogene Äquivalenz). Weiterhin scheint das Alter der Studie mit (nicht)Äquivalenz zusammenzuhängen, wobei neuere Studien eher Äquivalenz konstatieren. Wahrscheinlich sind da Fortschritte in der Technik mit ergonomischerer Gestaltung der Testdarbietung und die größere Erfahrung mit Computern im Alltag entscheidend. Zum Einfluss der Gestaltung der Testdarbietung am Rechner gab es erstaunlicherweise bis vor kurzem wenig vergleichende Untersuchungen. Wo diese durchgeführt wurden, zeigten sich durchaus Einflüsse der Layoutgestaltung auf das Antwortverhalten (siehe z.B. Moser et al., 1990; Christian, 2003; Pomerich, 2004). In letzter Zeit werden vermehrt Gestaltungsfragen im Rahmen von Internet-Befragungen angegangen (z.B. Dillman et al., 1998b; Bowker & Dillman 2000;

Vehovar et al., 2000; Bamert & Heidingsfelder, 2001). Theoretisch wird die Übereinstimmung von perzeptiven und motorischen Anforderungen, welche die Darbietungsformen an den Probanden stellen, diskutiert (z.B. Mead & Drasgow, 1993). Entsprechende empirische Untersuchungen fehlen jedoch weitgehend. Methodisch fällt auf, dass der psychometrische Vergleich meist über Kennwerte der klassischen Testtheorie geschieht, bei der zwei oder mehr Tests als parallel gelten, wenn sie dieselben wahren Werte und die gleiche Messfehlervarianz haben. Entsprechend werden die Kennwerte der klassischen Testtheorie, Rohwertverteilungen, Reliabilitätskoeffizienten und Validitätskoeffizienten verglichen. Der Vergleich von Verteilungen wird meist unter Normalverteilungsannahme vorgenommen (d.h. Mittelwerte und Standardabweichungen werden als suffiziente Statistiken angesehen), obwohl diese sicher oft nicht gegeben ist. Reliabilität wird meist über die interne Konsistenz definiert. Zur Validität liegen weniger Studien vor, wobei eher die Konstruktvalidität, seltener die Kriteriumsvalidität überprüft wird. Dazu werden meist Korrelationskoeffizienten verglichen. Konfirmatorische Faktorenanalyse, bei der bei entsprechender Datenlage Reliabilität und Validität gemeinsam überprüft werden können (z.B. Jöreskog, 1978), findet wenig Anwendung. Auch die Frage nach „Transferäquivalenz“ (im Sinne von Mertens, 1999) wird meist nicht gestellt. Ebenfalls wenig Beachtung bei der Klärung der Äquivalenzfrage finden

Weiterentwicklungen der klassischen Testtheorie, wie die Theorie der Generalisierbarkeit (Cronbach, 1963) und die probabilistischen Testtheorien (siehe Kubinger & Farkas, 1991 für eine der seltenen Ausnahmen).

Weniger untersucht aber immer wieder aufgeworfen wird die Frage zur populationsbezogenen Äquivalenz, welche eng mit der Forderung nach „Testfairness“ zusammenhängt. Diskutiert wird, dass nicht alle Benutzer gleiche Computererfahrung haben und dass diese auch zwischen den Schichten, Altersgruppen und Geschlechtern unterschiedlich ist (Hofer & Green, 1985). Da Computererfahrung mit dem Zurechtkommen am Rechner und mit durch den Rechner verursachter Ängstlichkeit zusammenhängen kann, kann dies zu differenziellen Effekten einer computergestützten Testdarbietung führen. Wo dies untersucht wurde zeigte sich jedoch kein konsistenter Zusammenhang (z.B. Klinck, 2002; Dyck et al., 1999; Miles & King, 1998; Schwenkmezger & Hank, 1993). Auch scheint sich eine durch Computer verursachte Ängstlichkeit schnell zu legen und liegt wohl größtenteils an einer nichtadäquaten Gestaltung der Benutzeroberfläche (Klinck, 2002; Hofer & Green, 1985).

Erfahrungsbezogene Äquivalenz wird meist als Frage nach der Akzeptanz verstanden. Als Indikatoren wurden sowohl indirekte Maße, wie die Anzahl fehlender Werte, der Umfang der Antworten bei offenen Fragen und die Abbruchrate, wie auch direkte Befragungen, sowohl mit offenen, wie auch mit geschlossenen Fragen, verwendet. Die Akzeptanz computergestützter

Testdarbietung scheint bei den Probanden durchgehend sehr hoch zu sein (z.B. Harell et al., 1987; Schwenkmetzger & Hank, 1993; Klinck, 2002), was speziell auch für Patienten psychotherapeutischer Einrichtungen, geriatrische Patienten und psychiatrische Patienten (Gitzinger, 1990; Skinner & Allen, 1983; Freudenmann & Spitzer, 2001) gilt. Diskutiert wurde, ob dies mit dem „Neuigkeitswert“ computergestützter Darbietung zusammenhängt und sich demnach „abnutzt“. Dagegen spricht allerdings die Stabilität der hohen Akzeptanz über die letzten 25 Jahre, wie schon Erdman et al. (1985) anmerken. Etwas niedriger ist die Akzeptanz bei den Diagnostikern, welche Vorbehalte gegenüber der neuen Technik zu haben scheinen (siehe z.B. Farell, 1991). Nach Rosen & Weil (1996) bezeichnet sich über die Hälfte aller Psychologen als „technophob“. Dennoch lehnen sie computergestützte Verfahren nicht grundsätzlich ab (siehe z.B. Johnson & Williams, 1980; Stein, 1987). Nach Hänsgen (1999) ist die Bereitschaft computergestützte Diagnoseverfahren einzusetzen unter den Psychologen mit 54% sogar recht hoch.

Trotz der überwiegend beruhigenden Ergebnisse bezüglich des computergestützten Einsatzes von Selbstbeurteilungsverfahren, wurden auch Ausnahmen zur Äquivalenz der Darbietungsformen gefunden. Die Frage welche Faktoren dazu beitragen ist nicht ausreichend geklärt. Daher gelten immer noch die Empfehlungen der American Psychological Association (1999) und des Testkuratoriums der Föderation Deutscher

Psychologenvereinigungen (1986) jedes Verfahren einzeln zu prüfen (vgl. Mertens, 1999 für eine Gegenposition).

Computerspezifische Möglichkeiten bei der Testdarbietung

Im Gegensatz zur computergestützten Darbietung, Analyse und Interpretation von (ursprünglich) Papier/Bleistift-Tests werden speziell für Computervorgabe konstruierte Verfahren zwar oft diskutiert (z.B. Jäger & Krieger, 1994), selten aber wirklich umgesetzt. Eine Reihe interessanter Entwicklungen gab es dennoch, z.B. computerunterstützte Inhaltsanalyse (Tucker & Rosenberg, 1975), Analyse von Antwortmustern (Brown & Burton 1978), Analyse von Verhalten bei Videospiele (Jones et al., 1981), Erfassung psychophysiologischer Daten (Fahrenberg, 1987), Systeme zur Selbstprotokollierung (Perez & Reicherts, 1989), Darbietung maßgeschneiderter Instruktionen (Bukasa et al. 1990) oder die Analyse von Mimik (Vanger et al., 1997). Heute werden in der psychodiagnostischen Praxis Verfahren wie Planungsspiele und psychophysiologische Monitoringsysteme zwar kommerziell angeboten, z.B. die Managementarbeitsprobe von Etzel & Küppers und der Selbstregulations- und Konzentrationstest für Kinder von Kuhl & Kraska bei Hogrefe oder diverse Stress- und Belastungstests im Rahmen des Testsystems „Biofeedback-2000“ bei Schuhfried, die Verbreitung solcher Verfahren ist jedoch gering. Mehr Beachtung fanden Verfahren, welche sich auf konventionelles Itemmaterial anwenden lassen. Zahlreiche Untersuchungen gibt es zur

adaptiven Itemvorgabe und zur Analyse von Itemresponsezeiten, wobei ersteres auch in der Testpraxis, speziell bei Leistungstests und im Bildungsumfeld, im größeren Umfang Verbreitung fand (s. Wildgrube, 1990).

Adaptives und entscheidungsorientiertes Testen

Ein effizienter Test erfragt genau die (Menge an) Informationen, welche nötig sind, um die gestellte diagnostische Frage zu beantworten (Butcher et al., 1985). Ein konventioneller Test muss ein breites Spektrum an „Schwierigkeiten“ abdecken und enthält so viele Items, welche für den einzelnen Probanden wenig informativ sind. Einen Ausweg bietet adaptives Testen bei dem das Ziel ist jedem Probanden genau die Fragen zu stellen, welche für ihm (und für die spezifische Fragestellung) informativ sind. Adaptive Teststrategien lassen sich auf drei Ebenen unterscheiden: (1) der Ebene der Itemauswahl, d.h. der Methode für die Wahl des nächsten Items, (2) der Ebene der Berechnung der Skalenwerte aus den Einzelitems, was bei adaptivem Testen, bei dem jeder Proband ein anderes Itemset bekommt, spezielle Methoden erfordert und (3) der Ebene des Abbruchkriteriums, welches die Zahl vorgegebener Items für jeden Probanden individuell festlegt. Bei der Itemauswahl unterscheidet Vale (1981) zwischen vorgegebener Verzweigung, mit Interitem-Verzweigung und Intersubtest-Verzweigung, und modelbasierter Verzweigung, welche auf der probabilistischen Testtheorie und den

dazugehörigen Iteminformationsfunktionen aufbaut. Die Berechnung der Skalenwerte erfolgt teils über Maximum-Likelihood und Bayes-Strategien, teils über arbiträre Schätzer und Näherungslösungen (siehe Wild, 1989a für einen Vergleich der Effizienz unterschiedlicher Schätzverfahren). Beim Abbruchkriterium wird messorientiertes und entscheidungsorientiertes Vorgehen unterschieden (Hornke, 2001). Messorientiertes Vorgehen ist an das Erreichen eines bestimmten Messfehlers gebunden, beim entscheidungsorientierten Vorgehen spielen Trennwerte und das Entscheidungsrisiko die wesentliche Rolle (siehe Butcher et al., 1985 für Spezialfälle des entscheidungsorientierten Vorgehens, die Countdown-Strategie und adaptive typologische Ansätze; siehe Hornke, 2001 und Eggen & Straetmans, 2000 für eine allgemeine Gegenüberstellung des messorientierten und des entscheidungsorientierten Ansatzes).

Das Prinzip des adaptiven Testens wurde schon von Binet Anfang des 20. Jahrhunderts angewandt. Items aus seinen Intelligenztest waren nach Altersstufen aufgeteilt. Der erste Itemblock wurde abhängig vom Alter vorgegeben, weitere dann abhängig von den vorhergehenden Antworten. Erst in den späten 50ern wurde die Idee wieder aufgegriffen und weiterführende Strategien wie zweistufige adaptive Tests, pyramidale Tests, flexilevel Tests und stradaptive Tests entwickelt (siehe Weiss, 1974 für eine Übersicht). Eine der ersten Studien zur adaptiven Darbietung von Persönlichkeitstest war die 2-Stufige

Darbietung von MMPI-Items von Kleinmuntz & McLean (1968). Dabei wurden zunächst aus jeder der 15 MMPI-Skalen 5 Items abgefragt und danach weitere Items nur so lange vorgegeben, bis eine Klassifikation in „normalen“ bzw. „abnormalen“ Bereich möglich war. Dies führte zu einer deutlichen Verkürzung der Testzeit, allerdings auf Kosten der Messgenauigkeit. Clavelle & Butcher (1977) stellten eine adaptive Strategie vor, bei der sie psychiatrische Patienten in 5 klinische Gruppen anhand des MMPI klassifizierten. In einer Simulationsstudie auf der Basis von Patientendaten berechneten sie in Blöcken von je 10 Items die Rate korrekter Klassifizierungen, wobei die Itemreihenfolge mit Hilfe von schrittweiser multipler Diskriminanzanalyse bestimmt war. Nach 10 Items erreichten sie 65% korrekter Klassifizierungen, nach 20 Items waren es 72% und nach 50 Items 81%.

Einen großen Sprung vorwärts machte die adaptive Testvorgabe durch die Anwendung der probabilistischen Testtheorie, entwickelt von Rasch und Birnbaum zwischen 1957 und 1960. Diese bietet eine elaborierte theoretische Fundierung für adaptive Strategien mit Parametern, welche das Antwortverhalten suffizient charakterisieren, darauf aufbauenden Informationsfunktionen sowie statistischen Methoden zur Bestimmung dieser Parameter. Sowohl in Simulationsstudien, wie auch im praktischen Einsatz, zeigt sich durchgehend eine größere Effizienz und größere Messgenauigkeit adaptiver Testvorgabe auf Basis der

probabilistischen Testtheorie im Vergleich zur nichtadaptiven Testvorgabe (Weiss, 1985 und Bloxom, 1989 geben Übersichten entsprechender Arbeiten). Im Durchschnitt reduziert sich die Testlänge bei gleichbleibender Messgenauigkeit und Validität auf die Hälfte. Dies gilt auch bei Verletzungen der restriktiven Modellvorgaben, die bei konventionell konstruierten Testverfahren kaum je erfüllt sind. Fehlerbehaftete Parameterschätzungen (Mattson, 1983) und Multidimensionalität (Weiss & Suhadolnik, 1985) beeinträchtigen die adaptive Testvorgabe nicht ernsthaft und verlangen nur die Vorgabe von wenigen Items mehr. Itemheterogenität und die daraus folgende schwache Modellanpassung ist häufig durch wenige Items begründet und lässt sich einfach korrigieren (z.B. zeigen das Carter & Wilkinson, 1984 für den MMPI). Dennoch sind adaptive Strategien nicht einfach umzusetzen. Vor allem gilt dies für die Übertragung konventionellen Testmaterials. Als Faustregel für adaptives Testen gelten mindestens 100, besser 150-200 gleichmäßig über die Skala verteilte Items pro Eigenschaft (Weiss, 1985; siehe auch Hornke, 1983, 1993 der das Problem zu kleiner Itempools eingehend untersuchte). Dies bietet kaum ein konventioneller Test. Modelle für polytome Ratingskalen und multidimensionale Items, wie sie bei Persönlichkeitstests und Symptominventaren üblich sind, wurden zwar entwickelt, die für die Parameterschätzungen nötigen Stichprobengrößen sind allerdings schwer zu erreichen. Dies gilt auch für den

alternativen Ansatz der „Wissensräume“, (Falmagne et al., 1990). Nicht zu vernachlässigen sind auch Effekte der Itemreihenfolge, d.h. Verletzungen der Itemunabhängigkeit, die z.B. durch Ermüdungseffekte, Lerneffekte, Erinnerungseffekte, Kontrasteffekte, Primingeffekte, etc. hervorgerufen werden können und die beim adaptiven Testen, bei dem jeder Proband eine andere Itemreihenfolge erhält, relevant werden (siehe Rost & Hoberg, 1997 allgemein zur Problemlage; siehe Schwarz et al., 1991 zu Kontexteffekten; siehe Wild, 1989b speziell zu diesem Problem beim adaptiven Testen). Ferner verringert sich bei adaptiver Testvorgabe zwar die Anzahl der vorgegebenen Items, die beobachteten Itemresponsezeiten sind allerdings länger als bei nichtadaptiver Vorgabe (siehe Wild, 1989b). Dies kann jedoch eine spezifische Eigenschaft von Leistungstests sein, bei denen nicht nur ein erheblicher Lerneffekt die Beantwortung mit fortschreitendem Test immer schneller macht (z.B. Hornke, 1997), sondern Fragen an der Grenze der jeweiligen Leistungsfähigkeit deutlich längere Responsezeiten fordern als sehr einfache oder sehr schwere Fragen.

Erfassen von Itemresponsezeiten

Die Zeiten, welche erforderlich sind, um einzelne Fragen zu beantworten, sind bei computergestützter Testdarbietung einfach mitzuerfassen. Eine ganze Reihe von Verwendungsmöglichkeiten für Itemresponsezeiten wurde

inzwischen untersucht. Die Itemresponsezeit kann über Schwierigkeiten mit einzelnen Items informieren, was z.B. bei der Testkonstruktion helfen kann; Die Testzeit kann als Indikator für Schwierigkeiten bei der Handhabung dienen und ermöglicht so eine Optimierung testökonomischer Aspekte; Sie kann als klassische Forschungsmethode aus der kognitiven Psychologie zur Erforschung des Testbeantwortungsprozesses herangezogen werden und sie kann individualdiagnostische Zusatzinformationen bieten.

Ausgehend von der Annahme einer funktionalen Abhängigkeit zwischen Entscheidungszeit und Schwierigkeitsgrad bei Urteilsprozessen (Thurner, 1965) wird die Itemresponsezeit mit der Schwierigkeit die Frage zu beantworten in Verbindung gesetzt. Diese kann dann als Indikator benutzt werden, um bei der Testkonstruktion schwer verständliche Items zu identifizieren (z.B. Bassili & Scott, 1996). Weiterhin kann sie für diagnostische Zwecke genutzt werden. So fanden schon Postman & Zimmermann (1945), dass Entscheidungszeiten bei den Behauptungen am längsten sind, denen der Proband indifferent gegenübersteht. Thurner (1965) findet bei dem Ipat Anxiety Questionnaire die längste Testzeit bei einem mittleren Testscore (ein Ergebnis, das allerdings nicht repliziert werden konnte, wie Düker & Schneider, 1977 anmerken). Noch älter ist die Überlegung dass lange Reaktionszeiten mit traumatischen Erfahrungen und erhöhtem emotionalen Engagement in Verbindung stehen. Nach Jung können Persönlichkeitskomplexe

durch lange Assoziationszeiten erkannt werden, was auch eine gewisse empirische Bestätigung fand (z.B. Sutherland & Spilka, 1964). Entsprechend wurde responsezeitgewichtete Scorebildung vorgeschlagen, wobei lange Antwortzeiten bei nichtdevianten Antwortwert als „emotionales Blocken“ verstanden wurde (Gilbert, 1986). Schon früh wurde auch der Zusammenhang zwischen Verfälschung im Sinne sozialer Erwünschtheit und Responsezeit untersucht. Entgegen der häufigen Annahme, dass lange Responsezeiten auf Verfälschung hindeutet, finden Sutherland & Spilka (1964) unter Zeitdruckinstruktion weniger pathologische Werte auf den Security-Insecurity Inventory. Dies gilt insbesondere für Items, welche anfällig für die Tendenz zu sozial erwünschten Antworten sind. Ganz in diesem Sinne finden Schneider & Hübner (1980) bei Verfälschungsinstruktion in Richtung eines guten Eindruckes die kürzesten Antwortzeiten, bei Kontrollinstruktionen mittlere Antwortzeiten und bei Instruktion in Richtung eines schlechten Eindruckes die längsten Antwortzeiten. Maschke (1989) findet allerdings, dass die Bearbeitungsgeschwindigkeit ein von Persönlichkeitskonstrukten weitgehend unabhängiges Merkmal darstellt und, dass die Testvalidität weitgehend unabhängig von der Bearbeitungszeit ist.

Es gibt mehrere interessante Versuche, diese zum Teil widersprüchlichen Einzelbefunde zu integrieren. Dunn et al (1972) fanden bei einer Untersuchung der Responsezeiten beim MMPI, dass diese über eine multiple Regressionsgleichung

relativ gut vorhergesagt werden können ($r \sim 0.75$), wobei die Länge der Frage die meiste Varianz aufklärt. Aber auch Ambiguität (verlängert die Responsezeit), soziale Erwünschtheit (verkürzt die Responsezeit) und die Varianz der sozialen Erwünschtheit (verlängert die Responsezeit) tragen signifikant bei. Zusätzlich verlängert der Pathologiewert der Antwort die Responsezeit. Aktuellere Untersuchungen gehen in Richtung einer rigorosen Kontrolle von Störfaktoren (itemspezifischen wie z.B. der Itemlänge und personenspezifischen wie z.B. der individuellen Lesegeschwindigkeit) und der Integration in Informationsverarbeitungsmodelle. So erweitern Holden et al. (1992) Rogers Prozessmodell der Beantwortung von Persönlichkeitsitems (Rogers, 1974) um Personenfaktoren und folgert daraus, dass schemakongruente Testantworten schneller sein sollten als nichschemakongruente. Ein Fakingschema, z.B. eins bei dem die Person einen positiven Eindruck erwecken möchte, sollte dazu führen, das Items mit sozial erwünschten Inhalten schneller zugestimmt wird als Items mit sozial unerwünschten Inhalten. Nach einer vierstufigen Datenbereinigungsverfahren (erste Extremwertbereinigung, intraindividuelle Standardisierung, interindividuelle Standardisierung, zweite Extremwertbereinigung), kann Holden diese Vorhersage auch bestätigen (Holden et al., 1992). Esser & Schneider (1998) können diesen Befund replizieren, verdeutlichen aber auch die Grenzen für einen Einsatz in der

psychologischen Einzelfalldiagnostik. Damit bleibt der Einwand von Maschke (1989), dass die Bearbeitungszeit zur Diagnose individueller Verfälschungsneigungen ungeeignet ist, weiter aktuell. Da Testverfahren, welche speziell im Hinblick auf die diagnostische Relevanz von Reaktionszeiten konstruiert wurden, ebenfalls die Ausnahme sind (z.B. Kuipers Instrument zur Diagnose depressiver Selbstschemata, Kuiper et al., 1983), ist die Argumentation nachvollziehbar, dass beim jetzigen Wissensstand die Verwendung von Responsezeiten in der Routinediagnostik nicht sinnvoll ist. Nicht nachvollziehbar ist jedoch die Ablehnung Responsezeiten überhaupt zu messen (z.B. Franke, 1998). Erhobene Responsezeiten können durchaus in einer explorativen Weise ausgewertet und für die Hypothesengewinnung nützlich sein.

FRAGESTELLUNG

Ziel dieser Studie war es, die empirische Grundlage für die Realisierung einer von Patienten und vom therapeutischen Team weitgehend akzeptierten, benutzerfreundlichen, ökonomischen und psychometrisch vertretbaren, computergestützten Darbietung von Instrumenten zum Monitoring von Ergebnissen psychotherapeutischer Behandlungen zu erarbeiten. Folgende vier Einzelaspekte sollten untersucht werden:

Zu einer Optimierung der computergestützten Darstellung wurden in dieser Studie mehrere Formen der Testdarbietung am Rechner gegeneinander und in Relation zur Papier/Bleistift-Darbietung gestellt. Untersucht wurde die Frage, inwieweit Steuerungen des Beantwortungsprozesses, welche durch die computergestützte Darbietung möglich werden, psychometrisch sinnvoll sind und/oder zu einer höheren Akzeptanz beitragen und inwieweit sie zu abweichenden Ergebnissen im Vergleich zur standardisierten Papier/Bleistift-Darbietung führen. Die Auswahl der Darbietungsvarianten erfolgte nicht theoriegeleitet⁵, sondern orientierte sich vorwiegend an pragmatischen Überlegungen. Folgende Formen wurden in dieser Studie verglichen:

⁵ Obwohl ein systematischerer Zugang geleitet durch ein psychologisches Modells zur Fragebogenbeantwortung wünschenswert erscheint, sind solche zur Zeit erst in Ansätzen verfügbar und nicht umfassend genug, diese Studie mit ihrer angewandten Ausrichtung zu leiten.

a) Einzelitemform - Formularform. Bei der Einzelitemform wird jedes Item einzeln am Bildschirm dargestellt und erst nach seiner Beantwortung wird das nächste angezeigt. Bei der Formularform werden mehrere Items gleichzeitig am Bildschirm dargestellt. Die Einzelitemform ist die gängigste Variante bei computergestützter Testdarbietung. Sie ist technisch weniger aufwändig, lässt sich einfach auf unterschiedliche Ausgabegeräte übertragen (vor allem auf Geräte mit beschränkter Displaygröße wie PDA's oder Mobiltelefone) und ist unumgänglich bei adaptiven Testen. Die Probanden bei Schwenkmezger & Hank (1993) gaben als Grund für eine Bevorzugung der computergestützten Darbietung vorrangig die Einzelitemdarstellung an, da sie ablenkende Inhalte ausblendete. Es lässt sich eine erhöhte Aufmerksamkeit bei der Beantwortung jedes einzelnen Items, weniger Flüchtigkeitsfehler sowie weniger Einfluss vorhergegangener Items und damit eine erhöhte Reliabilität und Validität vermuten (siehe dazu Gräf & Heidingsfelder, 1999; Zerr, 2001). Die Formularform dagegen hat durch ihre größere Ähnlichkeit mit der Papier/Bleistift-Darbietung die besseren Aussichten, sich als äquivalente Form zu erweisen und erlaubt eine zügigere Testbearbeitung. Innerhalb der Formularform ließe sich weiter zwischen einer seitenweisen Darstellung, unter Berücksichtigung der Grenzen des Bildschirms, und einer Darstellung des kompletten Formulars, mit einem „Scrollbalken“ als Navigationselement, unterscheiden. Dieser Frage wird aber

nicht weiter nachgegangen, da bereits durchgeführte empirische Untersuchungen eindeutig für die seitenweise Darstellung sprechen (Vehovar et al., 2000).

b) Fehlende Werte erlaubt - fehlende Werte nicht erlaubt.

Während bei der Papier/Bleistift-Vorgabe nur schwer eine sofortige Kontrolle der Vollständigkeit der Antworten möglich ist, ist dies bei computergestützter Darbietung einfach zu realisieren. Gängige Testsysteme benutzen verschiedene Kontrollvarianten, die von keiner Kontrolle, über Hinweise bei fehlenden Werten und der Wiederholung fehlender Fragen, bis zum Zwang, eine Antwort zu geben, reichen. Systematisch untersucht wurden die Effekte dieser Varianten jedoch noch nicht. Erfahrungsgemäß werden bei der Stuttgart-Heidelberger-QS ca. 5% aller Items nicht beantwortet. Dies führt zu Problemen bei der Datenauswertung und einer ad hoc Ersetzung fehlender Werte durch den Mittelwert der gegebenen Antworten. Um zu prüfen ob die Forderung, alle Fragen beantworten zu müssen, zu valideren Skalenwerten führt, als dies durch die Mittelwert-Ersetzung geschieht, wurde die Form „fehlende Werte nicht erlaubt“ untersucht. Die Erwartung dahinter war, dass Patienten in diesem Fall die am ehesten zutreffende Antwort geben. Möglich ist allerdings, dass dieser Zwang zu einer niedrigeren Akzeptanz bis hin zu reaktivem Antwortverhalten führt und Patienten vollständig unsinnige Antworten auf Fragen geben, die sie gezwungenermaßen beantworten müssen, was die Validität eher beeinträchtigen würde.

c) nachträgliche Änderungen möglich - nachträgliche Änderungen nicht möglich. Bestehende Systeme zur computergestützten Testdarbietung handhaben nachträgliche Änderungsmöglichkeiten sehr unterschiedlich: von keiner Möglichkeit für Korrekturen über die Möglichkeit, nur das letzte Item zu korrigieren, bis zur vollständigen Korrigierbarkeit. Die Frage, ob die Möglichkeit zu nachträglichen Änderungen wichtig ist, wurde aus folgenden Gründen untersucht: Zum einen besteht von Testentwicklern oft die mehr oder weniger explizite (wenn auch empirisch weitgehend unbegründete) Anforderung zu „first natural answers“ (siehe Oseas, 1966). Somit könnten nachträgliche Änderungen die Validität beeinträchtigen, wenn sie z.B. gemacht werden, um die Konsistenz mit aktuellen Antworten zu sichern. Weiter ist die Realisation von Änderungsmöglichkeiten mit zusätzlichem technischen Aufwand verbunden und kann bei adaptiven Testen auch zu methodischen Problemen führen. Auf der anderen Seite führt die Möglichkeit von nachträglichen Änderungen zu größerer Ähnlichkeit mit der Papier/Bleistift-Darbietung, kann für die Akzeptanz wichtig sein (so war bei Schwenkmezger & Hank, 1993 der meistgenannte Grund für eine Bevorzugung der Papier/Bleistift-Darbietung die Unmöglichkeit, vorherige Antworten zu kontrollieren) und ermöglicht die Korrektur von Flüchtigkeitsfehlern (was der Validität zugute kommen würde).

d) Rückmeldung an Patienten - keine Rückmeldung. Unter ethischen Gesichtspunkten hat jeder Proband ein Recht darauf, seine Testergebnisse in klarer, direkter, richtiger und verständlicher Form zu erfahren (Pope, 1992; siehe auch American Psychological Association, 1999, Standard 16.2). Zusätzlich kann so eine Rückmeldung im Sinne eines „incentives“ auch motivierend wirken und zu einer erhöhten Akzeptanz und weniger fehlenden Werten führen (siehe z.B. Theobald, 2001). Auch therapeutische Effekte von Rückmeldungen an Patienten wurden beobachtet (Finn & Martin, 1997; Gass & Brown, 1992; Hawkins et al., 2004), wozu möglicherweise größere Selbsterkenntnis und ein gesteigertes Problembewusstsein beitragen. Allerdings ist die automatische Rückmeldung von Testergebnissen, vor allem in psychosomatischen Einrichtungen mit entsprechend sensibler Klientel, nicht unproblematisch und muss vorsichtig angegangen werden. Schädliche Einflüsse auf den Therapieverlauf können bei derzeitiger Kenntnislage nicht ausgeschlossen werden.

Wie Franke (1998) darlegt, ist der Zusammenhang zwischen psychometrischer Äquivalenz, populationsspezifischen Merkmalen und Akzeptanz weitgehend ungeklärt. Annahmen lassen sich Einflüsse von Computererfahrung, Geschlecht und Alter (Zusammenfassend s. Klinck, 2002). Diese Zusammenhänge sollten hier mit dem Ziel untersucht werden, Populationen von vornherein identifizieren zu können, bei welchen besondere

Probleme bei einer computergestützten Testdarstellung zu erwarten sind. In der Routineanwendung sollen die Patienten, bei welchen ein Einfluss der Computerdarbietung auf das Antwortverhalten zu erwarten ist, oder welche eine Computerbefragung als besonders belastend empfinden würden, von vornherein die Papier/Bleistift-Version bekommen. Dazu sollten sowohl offene und geschlossene Fragen als direkte Maße, als auch die Zahl fehlender Angaben und die Testzeit als indirekte Maße erhoben werden.

Wie für das QS-Inventar des Stuttgart-Heidelberger-Modells sollen auch für das Ergebnismonitoring ausreichend normierte, validierte und verbreitete Instrumente zum Einsatz kommen. Dies schloss zumindest zum jetzigen Zeitpunkt die Entwicklung computerspezifischer Verfahren aus. Möglichkeiten, welche die computergestützte Testdarbietung von (ursprünglich) Papier/Bleistift-Fragebögen bietet, sollten jedoch genutzt werden. Insbesondere sollte geprüft werden, inwieweit entscheidungsorientierte Testvorgabe sinnvoll verwendet werden kann. Dies wurde in einer Studie vorbereitet, bei der adaptives Testen mit einer auf jeden Patienten individuell zugeschnittenen Itemreihenfolge und ein entscheidungsorientiertes Abbruchkriterium, bei dem weitere Items nur vorgegeben werden, solange keine Therapieevaluation möglich ist, simuliert wurden. Die Studie basierte auf Daten von 1197 Patienten, welche von 1995 bis 2001 in der Panorama

Fachklinik Scheidegg die kompletten Aufnahme- und Entlassungsbögen beantwortet hatten. Untersucht wurden die SCL-90-R mit 90 Items, das IIP mit 64 Items und der GBB mit 24 Items. Kriterium war die Zahl von Items, welche nötig gewesen wäre, um mit einer Irrtumswahrscheinlichkeit $p < 0.05$ klinisch bedeutsame Veränderungen zu erkennen. Verglichen wurde (a) die Itemreihenfolge, wie sie die Testhandbücher vorsehen, (b) die adaptiven Itemreihenfolgen nach dem "maximum-information"-Ansatz und (c) der Wissensraumtheorie sowie (d) eine nichtadaptive Reihenfolge, gegeben durch die Aufnahmereihenfolge bei schrittweiser multipler Regression. Die Testwerte wurden durchgehend über eine kleinste-Quadrate Schätzung bestimmt. Die beste Lösung ergab sich bei (d). Bei entscheidungsorientierten Abbruchkriterium waren für die Hälfte aller Patienten weniger als 22 Fragen bei der SCL-90-R, weniger als 21 Fragen beim IIP und weniger als 11 Fragen beim GBB nötig (Percevic & Wesiak, 2001). Offen blieb die Frage, ob und inwieweit dieses Vorgehen auch bei tatsächlicher entscheidungsorientierter Vorgabe sinnvoll ist. Roper et al. (1991) und Handel et al. (1999) fanden für den MMPI bei Einsatz der „Countdown Strategie“ vergleichbare Ergebnisse wie anhand von Simulationen erwartet. Bei adaptiver Testvorgabe wurden jedoch auch längere Responsezeiten beobachtet, welche die Einsparungen in der Anzahl präsentierter Items wieder zunichte machen können (Wild, 1989b). Auch kann sich die veränderte Itemreihenfolge auf das Antwortverhalten auswirken.

Insbesondere skalenhomogene Itemblockung kann durchaus zu (geringfügig) veränderten Antwortverhalten führen (siehe Krampen, 1993 für den FPI und Franke, 1999 für den SCL90-R). Die Frage, inwieweit ein entscheidungsorientiertes Vorgehen mit optimierter Itemreihenfolge im Vergleich zu handbuchgemäßem Vorgehen beim EB-45 und beim GBB zu veränderten Antwortverhalten führt, sollte hier untersucht werden.

Da die Studie als Vorbereitung für den Routineeinsatz diente, waren auch organisatorische Fragen zu klären. Fragen nach dem Wartungsaufwand, Schulungsaufwand für Klinikpersonal, Zeitaufwand des Klinikpersonals und allgemein den Betriebskosten der computergestützten Testdarbietung haben eine hohe praktische Relevanz. Die bei der Umsetzung gemachten Erfahrungen sollten hier dargestellt werden. Ferner war es wichtig, einen praktikablen Modus zur Einbettung der Testdarbietung in die klinische Routine zu finden und die Zustimmung von Seiten des therapeutischen Teams zu gewinnen. Durch die computergestützte Darbietung ergab sich die Möglichkeit, den Therapeuten sofortige Rückmeldungen über den Zustand des Patienten zu geben. Insbesondere wurde dies bei dieser Studie als eine Möglichkeit gesehen, die Motivation und Zustimmung des therapeutischen Teams zu erhöhen. Effekte dieser Rückmeldung und allgemein des Computereinsatzes auf den

therapeutischen Prozess und die Therapieergebnisse sollten dennoch als Nebenfrage exploriert werden.

METHODE

Untersuchungsdesign

Die Studie wurde in die routinemäßige Qualitätssicherung der Panorama Fachklinik für Psychosomatik und Psychotherapie Scheidegg eingebunden. Diese Klinik verwendet das Stuttgart-Heidelberger Modell seit 1995. Im Rahmen der Qualitätssicherung werden alle Patienten zur Aufnahme und zur Entlassung schriftlich befragt. Etwa die Hälfte der Patienten nimmt formal an der Qualitätssicherung teil, wobei zu diesen Patienten auch ihre Therapeuten bei der Aufnahme und der Entlassung, sowie in Einzelfällen bei der Falldiskussion schriftlich befragt werden. Diese Patienten nehmen zusätzlich an einer 6-Monats-Katamnese und einer 12-Monats-Katamnese teil. Die Teilnahme bzw. Nichtteilnahme an der Qualitätssicherung wird auf Wochenbasis reguliert, wobei die Patienten, die an zwei von vier zufällig gewählten Wochen aufgenommen werden teilnehmen und die Patienten, welche an den anderen zwei Wochen aufgenommen werden nicht teilnehmen. Seit 1999 werden Patienten zusätzlich nach ca. zwei Wochen Klinikaufenthalt befragt (dies entspricht etwa der halben durchschnittlichen Verweildauer). Klinikintern ist die Fragebogenvergabe und das Einsammeln der Fragebögen Aufgabe einer Arzhelferin. Die Fragebögen beinhalten einen Dokumentationsteil in Anlehnung an die ÜbaDo (Broda et al.,

1993) und einen umfangreichen psychometrischen Teil. Eine Übersicht zu den eingesetzten Instrumenten gibt Tabelle 1.

Tabelle 1. Bei der QS eingesetzte Instrumente

Aufnahme	Zwischenerhebung	Entlassung
<u>Patienten:</u> <ul style="list-style-type: none"> • Symptom Check List 90R (SCL90R; Franke, 1995) • Inventar zur Erfassung Interpersonaler Probleme (IIP; Horowitz et al., 2000) • Therapeutische Arbeitsbeziehung (PTBEZ; Alexander & Luborsky, 1986) • Ergebnisbogen (EB45; Lambert et al., 2002a) • Giesener Beschwerdebogen (GBB; Brähler & Scheer, 1997) <u>Therapeuten:</u> <ul style="list-style-type: none"> • Beschwerdescore (BSS, Schepank, 1995) • Therapeutische Arbeitsbeziehung (PTBEZ; Alexander & Luborsky, 1986) 	<u>Patienten:</u> <ul style="list-style-type: none"> • Direkte Veränderungs-einschätzung (PAE*) • Patienten-zufriedenheit (ZUF8; Schmidt et al., 1989) • EB45 • PTBEZ • GBB 	<u>Patienten:</u> <ul style="list-style-type: none"> • SCL90R • IIP • PTBEZ • Lebenszufriedenheit (Fahrenberg et al., 1986) • EB45 • GBB • PAE <u>Therapeuten:</u> <ul style="list-style-type: none"> • Direkte Veränderungs-einschätzung (THE*) • BSS • Therapeutische Arbeitsbeziehung

*Summenscore der Fragen zur Verbesserung des Allgemeinbefindens, des seelischen Befindens und der Leistungsfähigkeit aus der PsyBaDo (Heuft & Senf, 1998).

Die computergestützte Fragebogenvorgabe wurde für die Zwischenerhebung vorbereitet. Dies stellt einen günstigen Zeitpunkt für therapeutisch relevante Rückmeldung dar. Dass

die Zwischenerhebung nicht in die QS-Evaluation eingeht, gibt Spielräume für die Erprobung. Schließlich sind gerade diejenigen Instrumente enthalten, für welche noch keine Äquivalenzuntersuchungen vorliegen: der GBB, der EB45 und der PTBEZ (der SCL90R und das IIP als weitere Hauptinstrumente der QS wurden im deutschsprachigen Raum schon von Franke, 1998 untersucht).

Bereits vorhandene Papier/Bleistift-Zwischenerhebungen wurden nicht für die Bildung einer Kontrollgruppe verwendet, da Trends und Saisoneffekte, deren Einfluss schwer zu kontrollieren ist, anzunehmen sind (siehe Jahresberichte Panorama Fachklinik, 1996 bis 2000). Stattdessen erhielt zur Zwischenerhebung eine Hälfte der Patienten den Papier/Bleistift-Fragebogen, die andere Hälfte die Computerversion. Zur Aufnahme und zur Entlassung erhielten weiterhin alle Patienten den Papier/Bleistift-Fragebogen. Zwischen Papier/Bleistift-Vorgabe und Computervorgabe wurde auf Basis der durch einen Zufallsmechanismus (siehe vorhergehender Absatz) entschiedenen Teilnahme bzw. Nichtteilnahme an der Qualitätssicherung gewechselt, wobei bei Teilnahme die Computervorgabe und bei Nichtteilnahme die Papier/Bleistift-Vorgabe zum Einsatz kam. Um die Angemessenheit dieses Vorgehens zu überprüfen, wurden schon vorhandene Befragungen (erhoben 1995 und 1996) verglichen, bei welchen ein Teil der befragten Patienten an der QS teilgenommen hat und ein Teil nicht (n=800). Der Vergleich

ergab keinen statistisch signifikanten Einfluss der QS-Teilnahme auf den Behandlungserfolg (χ^2 -Test, $p=0.159$). Da der Behandlungserfolg über die Skalenwerte definiert ist, spricht nichts gegen dieses Vorgehen zur Gruppenbildung⁶.

Innerhalb der computergestützten Fragebogendarbietung wurden im ersten Studienabschnitt die vier Gestaltungsmerkmale computergestützter Vorgabe variiert (siehe Kapitel „Fragestellung“). Im zweiten Studienabschnitt wurde die komplette Vorgabe und die entscheidungsorientierte Vorgabe variiert. Untersucht werden sollten jeweils die Haupteffekte, wobei im ersten Studienabschnitt neben dem Vergleich der computergestützten Vorgabevarianten untereinander der Vergleich mit der Papier/Bleistift-Vorgabe durchgeführt werden sollte. Damit ergab sich ein Untersuchungsdesign wie auf Tabelle 2 dargestellt, wobei p die Wahrscheinlichkeit bezeichnet, zu der jeweiligen Erhebung, in die entsprechende Gruppe zu fallen.

⁶ Die Alternative, eine von der QS-Teilnahme unabhängige Randomisierung, wäre für das Betreuungspersonal organisatorisch zu aufwendig, alle Patienten an der QS teilnehmen zu lassen würde für die Therapeuten zu Mehraufwand führen, und eine Randomisierung innerhalb der QS-Teilnehmer würde den Erhebungszeitraum verdoppeln.

Tabelle 2. Untersuchungsdesign

Aufnahme	Zwischenerhebung				Entlassung
P/B p=1	Abschnitt 1	P/B	p=0.5		P/B p=1
			A	B	
		Form*	p=0.25	p=0.25	
		PC Fehlende Werte**	p=0.25	p=0.25	
		Ändern**	p=0.25	p=0.25	
	Rückmeldung**	p=0.25	p=0.25		
Abschnitt 2	PC	Entscheidungs-orientiert***	p=0.25	p=0.25	

* A - seitenweise, B - itemweise

** A - nein, B - ja

*** A - komplett, B - entscheidungsorientiert

Technik

Die computergestützte Fragebogendarbietung wurde als Internetapplikation realisiert. Dabei wurde serverseitig Windows als Betriebssystem, Apache als Webserver, FoxPro als Datenbank und PHP4 als Applikationssprache verwendet. Clientseitig wurde die Applikation für den InternetExplorer ab Version 4 bei einer Darstellung von 800*600 Bildschirmpunkten und einer Zeichengröße von 12 Punkten optimiert und macht Gebrauch von JavaScript, Cookies und HTML4-Features. Die einzelnen Funktionen waren passwortgeschützt. Die Benutzeroberfläche war so gestaltet worden, dass ein Zurechtkommen ohne große Einarbeitung möglich war (siehe Parshall et al., 2002 für entsprechende Richtlinien und Testverfahren). Dazu wurden standardisierte HTML-Elemente verwendet. Die Auswahl der jeweiligen computergestützten Darbietungsform wurde durch den Zufallsgenerator von

JavaScript gesteuert. Itemresponsezeiten wurden ebenfalls über JavaScript registriert. Ein klinikeigener Rechner wurde so vorkonfiguriert, dass er zugleich als Server und als Client arbeiten konnte und die Applikation beim Hochfahren sofort gestartet wurde. Die Applikation lief anfangs auf einem 386er mit 64MB RAM unter Windows 98, später auf einem Pentium mit 256MB RAM unter Windows 2000. Die Eingabe erfolgte über Maus und Tastatur. Als Bildschirm wurde ein 17'-Farbbildschirm verwendet. Als Hintergrundfarbe war helles Gelb gewählt worden, da Weiß auf einem Bildschirm oft zu hell erscheint und bei dunkleren Farben die Darstellung nicht ausreichend kontrastreich ist.

Darstellungsvarianten

Die Darstellung war vom Layout her möglichst ähnlich der Papiervorlage gehalten. Sobald alle Items einer Seite (bzw. das eine Item bei itemweiser Darstellung) beantwortet waren, wurde die nächste Seite angezeigt. Zusätzlich ermöglichten Buttons für „vor“ und „zurück“ eine Navigation durch den Fragebogen. Als Navigationshilfe war die jeweilige Seite und die Anzahl der noch verbleibenden Seiten (bzw. Items) eingeblendet. Abbildung 2 zeigt als Beispiel den Screenshot einer Fragebogenseite.

1. Wenn Sie Ihren jetzigen Zustand ganz kritisch mit dem zu Beginn Ihrer Behandlung vergleichen, wie würden Sie dann folgendes beurteilen?

	viel besser	etwas besser	unver- ändert	etwas schlechter	viel schlechter
Meine körperliche Verfassung ist	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
Meine seelische Verfassung ist	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Mein Allgemeinbefinden ist	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
Meine Leistungsfähigkeit ist	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>

2. Fühlen Sie sich derzeit arbeitsfähig?

- ja, Vollzeit ja, Teilzeit nein fraglich

3. Haben Sie heute andere Beschwerden als zu Beginn der Behandlung?

nein

ja, neue, andere körperliche Beschwerden

wenn ja, welche

ja, neue, andere seelische Beschwerden

wenn ja, welche

4. Wünschen Sie eine Weiterbehandlung?

- ja nein

5. Haben Sie die Beziehung zu Ihrem/r Therapeut/-in als hilfreich erlebt?

- ja eher ja eher nein nein

WEITER

Abbildung 2. Seitenweise Darstellung des Fragebogens

In der Darstellungsvariante, bei welcher nachträgliche Änderungen nicht möglich waren, konnten durch Zurückblättern vorhergehende Fragen zwar angesehen, nicht jedoch verändert werden. Bei der Variante mit der Möglichkeit nachträglicher Änderungen, wurde die Nähe zu der Papier/Bleistift-Version aufrechterhalten, indem das Löschen von schon gemachten Antworten über nochmaliges anklicken des schon angekreuzten Feldes erfolgte (ein gesondertes Feld für „Antwort löschen“ kann zu einer höheren Rate fehlender Werte führen kann, wie Biskin & Kolotkin, 1977 zeigen). Beim Vorblättern wurde geprüft, ob alle Fragen (bis auf die offenen Fragen)

beantwortet waren. Bei unvollständig ausgefüllten Seiten bekamen die Patienten einen entsprechenden Hinweis. Bei der Variante, bei welcher fehlende Werte erlaubt waren, stand es den Patienten frei, die aktuelle Seite mit den fehlenden Antworten weiter zu bearbeiten oder den Hinweis zu ignorieren. Bei der Variante, bei welcher fehlende Werte nicht erlaubt waren, konnte der Hinweis nicht ignoriert werden und die Patienten mussten auf alle, bis auf die offenen Fragen, eine Antwort geben, um weiterzukommen⁷. Bei Rückfragen oder Beschwerden von Seiten der Patienten bekamen diese von der betreuenden Arzthelferin die Instruktion, die angebrachteste Antwort zu geben, bzw. „nie“ anzukreuzen, falls diese Frage überhaupt nicht zutreffen sollte.

Die Rückmeldung der Testergebnisse an Patienten wurde vorrangig als Motivationselement eingeführt. Ein Therapieeffekt war nicht beabsichtigt. Durch den Hinweis, dass Unklarheiten mit dem Therapeuten zu besprechen sind, sind aber Einflüsse, zumindest auf die therapeutische Arbeitsbeziehung, durchaus denkbar. Die Patienten, welche Rückmeldungen bekommen, wurden schon zu Beginn, d.h. bei den Instruktionen, darauf hingewiesen, dass sie am Ende eine kurze Rückmeldung erwartet. Patienten mussten die Rückmeldung nicht einsehen, sondern konnten diese auch überspringen. Eine patientengerechte Form der Rückmeldung wurde über eine am

⁷ Eine vollständige Entsprechung zur Papier/Bleistift-Erhebung, bei der überhaupt keine Kontrolle der Antworten stattfindet, wurde nicht implementiert, um die Chance auf eine Optimierung nicht zu vergeben.

Trennwert orientierte Verbalisierung der Ergebnisse und einen tendenziell positiven Ton (siehe Kluger & DeNisi, 1996; Hawkins & Lambert, 2002) realisiert. Falls bestimmte Skalen aufgrund fehlender Werte nicht berechnet werden konnten, wurde dies ebenfalls angegeben, in der Erwartung die Patienten bei der Entlassungsbefragung zu kompletteren Antworten zu motivieren. Abbildung 3 zeigt eine Beispielsrückmeldung. Im Anhang ist eine Aufstellung aller verwendeten Textbausteine und Regeln. Mehr noch als bei den anderen Testvorgabevarianten war es hier wichtig durch kontinuierliche Auswertung der Ergebnisse und der Kommentare von Seiten des therapeutischen Teams und der Patienten, die Unbedenklichkeit sicherzustellen, und gegebenenfalls diese Form von Rückmeldungen einzustellen oder zu modifizieren. Tatsächlich wurde, angeregt durch Proteste von Seiten einiger Patienten mit schlechter Zustandsbeurteilung, beim erstem Softwareupdate (siehe Tabelle 4) die Rückmeldung dahingehend verändert, dass vor dem Symptomzustand zunächst die (unseres Wissens nach immer) guten Aussichten des Patienten für die Weiterbehandlung thematisiert wurden.

Patientenauswertung ✓

Hier die Ergebnisse aus der Auswertung Ihres Fragebogens. Diese Auswertung gibt nur grobe Hinweise. Besprechen Sie deshalb Unklarheiten mit Ihrem Therapeuten.

Eins vorweg: Sie haben sehr gute Chancen, vom weiteren Verlauf der Behandlung zu profitieren.

Über Ihre Gedanken, sich das Leben zu nehmen, sollten Sie unbedingt mit Ihrem Therapeuten reden.

Die Auswertung deutet darauf hin, dass Sie sich stark durch Symptome belastet fühlen. Sie geben viele psychische Beschwerden an. Zwischenmenschliche Beziehungen könnten einen Problempunkt darstellen. Ihren Angaben nach könnten Sie im sozialen und beruflichen Umfeld Probleme haben. Sie leiden nicht unter außergewöhnlich vielen körperlichen Beschwerden. Im Vergleich zum Behandlungsbeginn fühlen Sie sich nicht besser.

ZURÜCK

*Panorama Fachklinik für Psychosomatik, Psychotherapeutische Medizin und Naturheilverfahren
Scheidegg/Allgäu*

HILFE **INFO**

Abbildung 3. Beispiel einer Rückmeldung an Patienten

Entscheidungsorientierte Vorgabe

Zur Vorbereitung der entscheidungsorientierten Vorgabe wurden Patientendaten zum GBB und EB45 von 1999 bis 2001 herangezogen (N=295). Die Aufnahmereihenfolge bei schrittweiser multipler Regression mit dem jeweiligen Skalenwert als Kriterium und den einzelnen Items als Prädiktoren ergab die Itemreihenfolge. Die Verrechnung der Items erfolgte ohne differenzielle Gewichtung, d.h. die einzelnen Antworten wurden einfach zusammengezählt⁸. Tabelle 3 zeigt die Reihenfolge der Items sowie die Korrelation der

⁸ Dieses Vorgehen ist einfacher, weniger anfällig für Überanpassung und gab in der Simulation fast so gute Ergebnisse wie die optimal gewichtete Summe.

jeweiligen Zwischensumme mit der Gesamtsumme. Aus diesen Korrelationen wurden die Standardfehler geschätzt. Die Items wurden nacheinander vorgegeben bis sich der geschätzte Skalenwert ± 1.96 Standardfehler mit dem Trennwert nicht mehr überlappte. Der Trennwert war, wie in der Stuttgart-Heidelberger-QS üblich, bei einen Wert angesetzt, der dem 68. Perzentil bei der Allgemeinbevölkerung entspricht.

Tabelle 3. Reihenfolge der Items und geschätzte Sequenz-Gesamt-Korrelationen bei entscheidungsorientierten Testen

EB45-SYB		EB45-BEZ		EB45-SOZ		GBB	
Reihe	R	Reihe	R	Reihe	R	Reihe	R
42	.81	18	.7	4	.6	15	.76
15	.85	37	3	38	7	21	.82
9	.90	16	.8	39	.8	23	.87
25	.92	43	2	21	3	5	.89
45	.93	7	.8	28	.8	2	.93
31	.94	30	8	12	8	17	.94
36	.95	17	.9	14	.9	12	.95
33	.96	19	2	44	1	8	.95
22	.97	1	.9	32	.9	4	.96
34	.97	26	5		4	20	.97
41	.97	20	.9		.9	10	.98
3	.98		7		5	18	.98
5	.98		.9		.9	1	.98
35	.98		8		8	22	.99
27	.98		.9		.9	7	.99
24	.99		8		9	3	.99
23	.99		.9		1	14	.99
6	.99		9			19	.99
29	.99		.9			6	.99
8	.99		9			11	.99
2	.99		1			13	.99
40	.99					24	.99
10	.99					16	1
13	1					9	1
11	1						

Die Vorgabe des Fragebogens erfolgte itemweise. Nur Items des GBB und des EB45 wurden nach der entscheidungsorientierten Teststrategie vorgegeben. Aus Gründen der einfacheren technischen Realisierbarkeit war keine Möglichkeit zum Zurückblättern oder für nachträgliche Korrekturen gegeben. Da die psychometrische Qualität der so gewonnenen Daten fraglich war, wurde auf die Patientenrückmeldung verzichtet. Um die Ähnlichkeit zur kompletten Vorgabe zu erhöhen, wurde auch diese entsprechend angepasst. Der einzige Unterschied bestand schließlich darin, dass bei kompletter Vorgabe alle Items des GBB und des EB45 beantwortet werden mussten und diese Items in der im Testhandbuch vorgesehenen Reihenfolge dargestellt wurden.

Zusatzbefragung

Zusätzlich zu den Items, welche der Papier/Bleistift-Fragebogen erhält, wurde bei der computergestützten Vorgabe ein kurzer Fragebogen zur Akzeptanz und Handhabung vorgelegt. Dieser wurde an das Ende der Befragung gehängt (d.h. nach der Patientenrückmeldung, falls diese gegeben wurde, sonst nach der letzten Frage des QS-Inventars). Es handelt sich um einen selbstentwickelten Fragebogen, der eine grobe Einschätzung der Akzeptanz, Schwierigkeiten bei der Benutzung und Computererfahrung ermöglichen sollte. Dieser Fragebogen wurde ohne eine formale psychometrische Überprüfung, jedoch unter Berücksichtigung von allgemeinen Empfehlungen zur

Fragebogenkonstruktion entwickelt (Krauth, 1995; Lienert & Raatz, 1994). Abbildung 4 zeigt die einzelnen Fragen.

Die letzten 7 Fragen!

1. Wie anstrengend fanden Sie die Befragung?
 nicht kaum einigermaßen erheblich stark

2. Wie sind Sie mit der Benutzung dieses Programmes zurechtgekommen?
 ohne Schwierigkeiten
 ohne nennenswerte Schwierigkeiten
 einigermaßen
 mit erheblichen Schwierigkeiten

3. Wie verständlich waren die Fragen?
 verständlich
 größtenteils verständlich
 kaum verständlich
 nicht verständlich

4. Wieviel Erfahrung haben Sie im Umgang mit Computern?
 keine Erfahrung wenig Erfahrung benutzte Computer häufig

5. Würden Sie bei einer weiteren Erhebung einen computerbasierten Fragebogen wie diesen bevorzugen, oder lieber einen Papierfragebogen wie bei der Aufnahme ausfüllen?
 Computerbasierten Fragebogen Papierfragebogen

6. Hat Sie etwas besonders gestört?

7. Hat Ihnen etwas besonders gefallen?

Vielen Dank für Ihre Mitarbeit!

Abbildung 4. Zusatzfragebogen

Rückmeldungen an Therapeuten

Um die Zusammenarbeit und Akzeptanz des therapeutischen Teams zu gewinnen, wurde eine Rückmeldung der Ergebnisse der Zwischenerhebung an die Therapeuten vorbereitet. Die Therapeuten erhielten graphisch Informationen über den Zustand des Patienten, wie er sich im GBB und den Subskalen des EB45 (im Vergleich zur Bevölkerung) sowie dem PTBEZ und ZUF8 (im Vergleich zu anderen Patienten dieser Klinik) niederschlägt. Daneben erhielten die Therapeuten Informationen zu häufigen

Suizidgedanken. Die zusammenfassende Beurteilung des Zustands des Patienten wurde durch farbliche Kodierungen weitergegeben: „rot“ bedeutete, dass der Patient häufige Suizidgedanken angegeben hat oder seinen Zustand als verschlechtert bezeichnet hat, und „grün“ bedeutete, dass die Messwerte des Patienten auf allen Symptomskalen unterhalb des Trennwertes lagen (68. Perzentil bei der Allgemeinbevölkerung). Andernfalls wurde „schwarz“ als Farbcode vergeben. Falls aufgrund fehlender Werte keine Farbcodierung erfolgen konnte, wurde dies ebenfalls angegeben. Abbildung 5 zeigt eine Beispielsrückmeldung. Beim erstem Softwareupdate (siehe Tabelle 4) wurden auf Wunsch vieler Therapeuten die verbalen Rückmeldungen, wie sie die Patienten erhielten, mit aufgenommen. Die Überprüfung möglicher Effekte dieser Rückmeldung auf die Therapie ist kein Hauptgegenstand dieser Studie und wird demnach nur als Nebenfragestellung verfolgt.

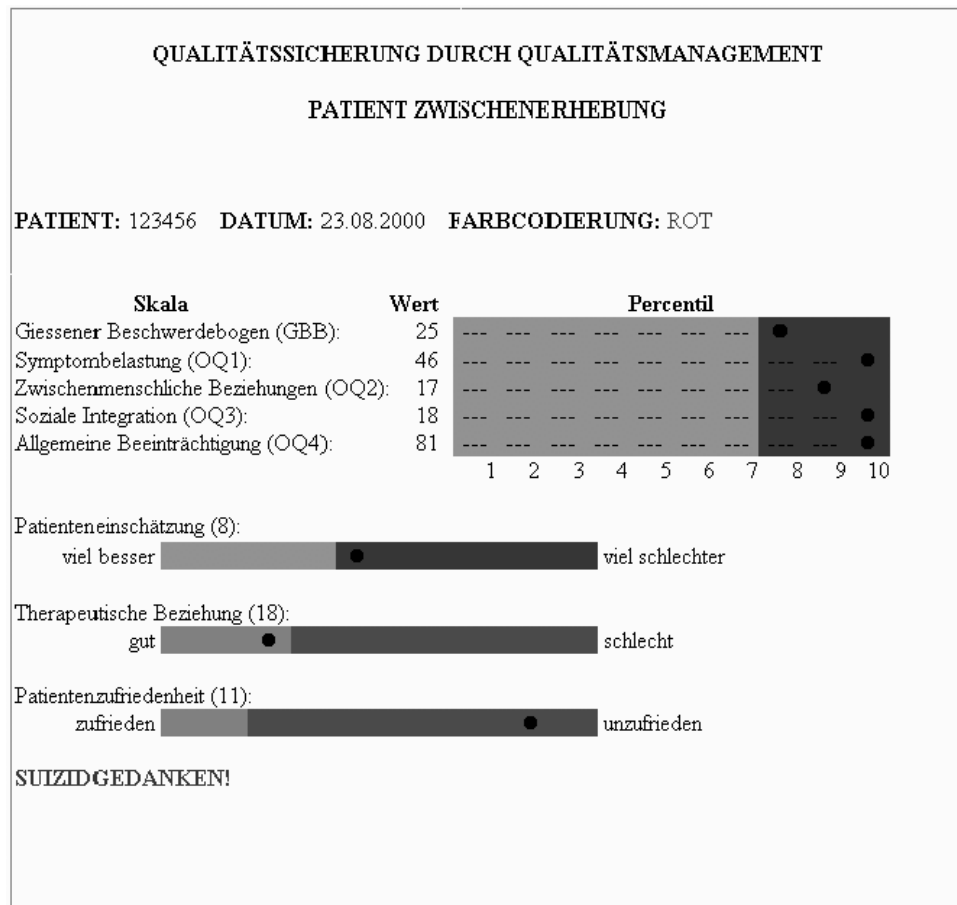


Abbildung 5. Beispiel einer Rückmeldung an Therapeuten

Stichprobengröße und Zeitplan für die Datenerhebung

Ein potenzieller Schwachpunkt zahlreicher Äquivalenzuntersuchungen, vor allem solcher aus dem klinischen Bereich, sind kleine, häufig unbegründete Stichprobengrößen. Viele der Vorbehalte gegenüber statistischen Hypothesentests sind in diesem Fall zutreffend (siehe dazu z.B. Wilkinson & TSFI, 1999 und Hager, 2000). Um diesen Vorbehalten zu begegnen, wurden für diese Studie vorab Power-Überlegungen für die Einschätzung der nötigen Stichprobengröße durchgeführt. Der Befund „keinen statistisch signifikanten Unterschied gefunden“ sollte im Sinne „praktischer Irrelevanz“

interpretiert werden können⁹. Dazu sollte die Wahrscheinlichkeit für die Entdeckung eines praktisch relevanten Unterschiedes, wenn er gegeben ist, möglichst groß sein.

Die Schätzung der dazu nötigen Stichprobengröße wurde ausgehend von der Frage nach Verteilungsäquivalenz zwischen den Vorgabeformen vorgenommen. Ausgehend von einer Reliabilität der eingesetzten Instrumente von ca. 0.90, d.h. einem Standardmessfehler von 0.31, bot sich eine mittlere Effektstärke ($\eta^2 \sim 0.06$; Cohen, 1988) als Grenzwert für den praktisch relevanten Unterschied an. Für die gängigen Fehlklassifikationsrisikos $\alpha = .05$ und $1 - \beta = .80$ und den Mittelwertsvergleich von zwei gleichgroßen Gruppen ergibt sich dann eine minimale Stichprobengröße von $n = 2 \cdot 64$. Da bei den eingesetzten Skalen nicht von Normalverteilungen auszugehen und entsprechend der Einsatz nichtparametrischer Verfahren geplant war sowie multiple Vergleiche durchgeführt werden sollten, stellt diese Rechnung nur einen groben Anhaltspunkt dar. Zum Erreichen der nominalen Fehlklassifikationsrisikos sind eher größere Stichproben nötig (vgl. Bortz & Lienert, 2003). Als minimale Stichprobengröße für den Vergleich der einzelnen computergestützten Varianten in ihren jeweiligen dichotomen Ausprägungen wurde daher pro Ausprägung $n = 100$

⁹ Umgekehrt brauchte der Befund eines statistisch signifikanten Unterschiedes nicht automatisch als praktisch relevanter Unterschied interpretiert werden sondern sollte nur die Aufmerksamkeit und die weitere Datenanalyse lenken.

angestrebt, d.h. $n_{\text{computergestützt}}=2 \cdot 100$. Weiter bestimmt das Design, dass 50% der Patienten eine der computergestützten Vorgabevarianten bekommen und 50% die Papier/Bleistift Vorgabe. Damit ergibt sich für den gesamten ersten Studienabschnitt eine Stichprobe von $n_{\text{gesamt}}=200+2 \cdot 100$. Selbst im ungünstigsten Fall, falls sich die computergestützten Vorgabevarianten unterscheiden und getrennt mit der Papier/Bleistift Vorgabe verglichen werden müssen, ergeben sich so Powerreserven für den Vergleich der einzelnen computergestützten Darstellungsvarianten mit der Papier/Bleistift Vorgabe. Bei jährlich ca. 550 Patienten ergibt sich ein akzeptabler Zeitrahmen von 9 Monaten für den ersten Studienabschnitt. Für den zweiten Studienabschnitt, den Vergleich entscheidungsorientierter und handbuchgemäßer Vorgabe, sollten pro Vorgabevariante ebenfalls $n=100$ erhoben werden, d.h. $n_{\text{computergestützt}}=2 \cdot 100$. Obwohl für diesen Vergleich keine Papier/Bleistift-Gruppe benötigt wurde, konnte aus organisatorischen Gründen weiterhin nur die Hälfte der Patienten am Rechner arbeiten. Daher ergaben sich auch für diesen Studienabschnitt geplante 9 Monate Erhebungszeit, also 18 Monate für die gesamte Erhebung.

Ablauf

Ein Behandlungszimmer, in dem die Patienten ungestört arbeiten konnten, wurde seitens der Klinikleitung für die

Computerbefragung bereitgestellt. Die Arzthelferin, welche für die QS verantwortlich ist und üblicherweise die Papier/Bleistift-Fragebögen austeilt, kümmerte sich auch um die computergestützte Patientenbefragung. Sowohl sie wie auch ihre Vertretung erhielten eine ausführliche Einführung in das System sowie ein schriftliches Manual (siehe Anhang). Das gesamte Therapeutenteam bekam eine kurze Demonstration und schriftliche Unterlagen zur Interpretation der Rückmeldungen (siehe Anhang). Bei Fragen oder Störungen konnte jederzeit die Forschungsstelle kontaktiert werden.

Mit den Patienten, welche computergestützte Fragebogendarbietung zur Zwischenerhebung erhielten, machte die Arzthelferin einen Termin aus. Sie führte den Patienten dann zum Rechner. Am Rechner gab die Arzthelferin zuerst den Patientencode ein. Falls dieser schon abgespeichert war, d.h. der Patient bereits einen Teilabschnitt bearbeitet hatte, konnte der Patient die Bearbeitung fortsetzen (was für den Fall von ungewollten Abbrüchen gedacht war). Im Normalfall wurde ein neuer Patient angelegt und automatisch einer der computergestützten Vorgabeformen zugeordnet, worauf die vorgabespezifischen Instruktionen mit Übungsfeldern erschienen (siehe Abbildung 6). Diese ging die Arzthelferin mit dem Patienten durch. Danach konnte sie sich wieder anderen Aufgaben widmen, blieb aber jederzeit für den Patienten ansprechbar (siehe Styles, 1991 welche die Wichtigkeit einer persönlichen Betreuung hervorhebt und Effekte auf Reliabilität

und Validität findet). Nach Beendigung der Befragung erhielt der Patient die Meldung, dass er fertig sei und sich an das Klinikpersonal wenden solle. Nun konnte die Arzthelferin die Rückmeldung für die Therapeuten zu diesem Patienten drucken und dem jeweiligen Therapeuten in sein Fach legen. Falls ein Patient sich strikt weigern sollte, am Computer zu arbeiten, oder schon im Vorfeld erkennbar war, dass ihm dies nicht möglich sein würde, konnte die betreuende Arzthelferin einen Papier/Beistift-Bogen austeilen, wobei dies vermerkt wurde (siehe American Psychological Association, 1999, Standard 15.1, welcher fordert, dass immer eine der Testperson angemessene Befragungsform benutzt wird). Bei Papier/Bleistift-Vorgabe wurde die Prozedur aus der QS-Routine beibehalten, d.h. die Arzthelferin teilte den Fragebogen aus, der Patient sollte den Fragebogen innerhalb eines Zeitrahmens von 2-3 Tagen bearbeiten, und dann wurde er wieder von der Arzthelferin eingesammelt.

Etwa alle 4 Wochen schickte die Arzthelferin eine Diskette mit den computererhobenen Zwischenerhebungen und den Papier/Bleistift-QS-Daten an die Forschungsstelle. Ab Dezember 2001 wurden die Daten wöchentlich über das Internet verschickt. Die Datenerhebung begann im Dezember 2000 und endete im Juli 2002. Im Juli 2001 wurde die Darstellungsvariante abgeschafft, bei der keine nachträglichen Änderungen möglich waren, da sich Patienten häufig dagegen beschwerten. Im Dezember 2001 begann der zweite

Studienabschnitt. Tabelle 4. gibt einen Überblick aller relevanter Ereignisse.

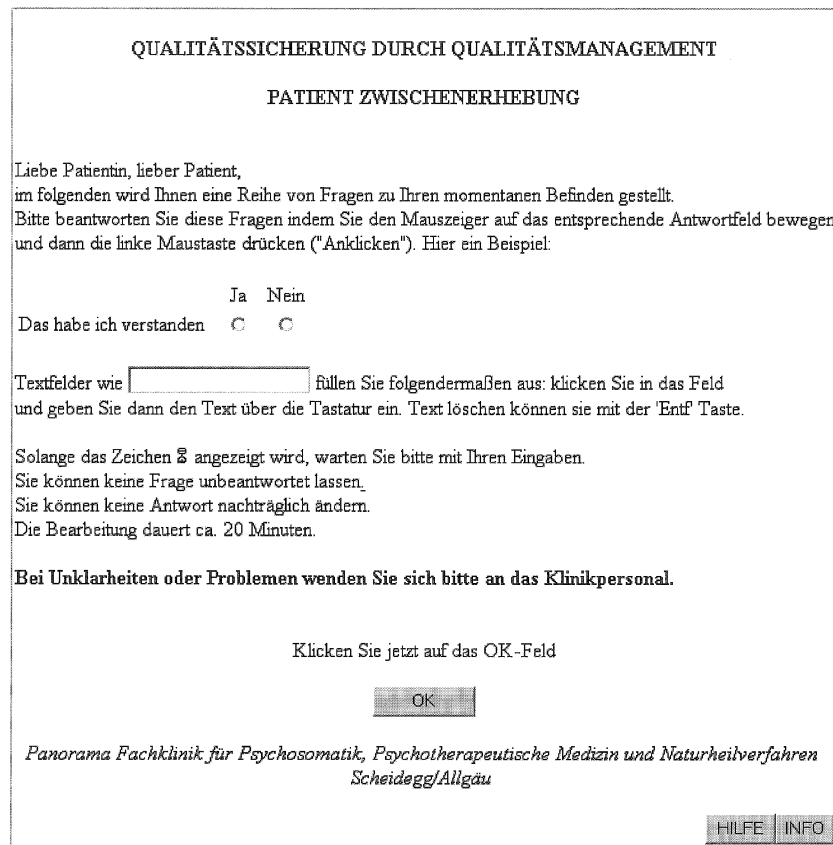


Abbildung 6. Beispiel einer Instruktionsseite

Tabelle 4. Ereignisse im Laufe der Studie

Datum	Ereignis
10.10.2000	Beginn der Softwareentwicklung
30.10.2000	Installation in der Klinik; Schulung; Beginn der Datenerhebung
14.12.2000	Update mit veränderter Rückmeldung an Patienten, erweiterter Rückmeldung an Therapeuten und veränderten Farben für S/W-Druck
24.03.2001	Rechner defekt
26.07.2001	Neuer Rechner wird Aufgebaut; Darstellungsvariante „keine Änderungen möglich“ wird abgesetzt.
30.08.2001	Fehler beim Speichern von Responsezeiten, da WIN2000 auf neuen Rechner alte Treiber nicht unterstützt, wird korrigiert

...

07.12.2001	Diskettenlaufwerk defekt; Internetanschluss wird installiert; Entscheidungsorientierte Vorgabe wird installiert; Abschnitt 2 beginnt
20.12.2001	Programmierfehler bei entscheidungsorientierter EB45-Vorgabe wird korrigiert
26.02.2001	Sporadisch auftretende Abstürze bei entscheidungsorientierter Vorgabe aufgrund von Bug im IE/Apache/PHP werden umgangen
01.07.2002	Ende der Datenerhebung

Stichprobenbeschreibung

Insgesamt wurden 1160 Patienten während der Erhebungszeit aufgenommen. Tabelle 5 gibt einen Überblick über die Anzahl der eingegangenen Fragebögen zu den einzelnen Stichprobengruppen. Die computergestützte Zwischenerhebung wurde durchschnittlich 14 Tage (SD=3.8) nach Therapiebeginn durchgeführt. Für die Papier/Bleistift-Vorgabe liegen keine so verlässlichen Angaben vor, da der Fragebogen mehrere Tage bei den Patienten bleiben konnte. Im Durchschnitt wurde er 18 Tage (SD=4.8) nach Therapiebeginn eingesammelt.

Tabelle 5. Stichprobenumfang

Aufnahme	Zwischenerhebung				Entlassung
	P/B N=372				
P/B N=504	Abschnitt 1	A B		P/B N=485	
		Form*	N=116		N=112
		PC Fehlende Werte**	N=119		N=109
		Ändern**	N= 51		N=177
		Rückmeldung**	N=113		N=115
	Abschnitt 2	PC	Entscheidungs- orientiert***	N=112	N=114

* A - seitenweise, B - itemweise

** A - nein, B - ja

*** A - komplett, B - entscheidungsorientiert

Fehlende Fragebögen wurden als „missing at random“ angesehen, die Berechnungen erfolgten bei paarweisen Fallausschluss. Die Annahme, dass fehlende Papier/Bleistift-Fragebögen am ehesten ihre Ursache in organisatorischen Problemen der Klinik haben, welche unabhängig von den Patientenbeschwerden sind, scheint plausibel. So wurden die Fragebögen, die nicht zur eigentlichen QS gehören, nicht so sorgfältig eingesammelt, es fehlten Patientencodes, so dass einzelne Bögen nicht zugeordnet werden konnten, oder Patientencodes waren mehrfach vergeben. Skalenwerte zur Aufnahme ließen keine Rückschlüsse auf fehlende Bögen zur Zwischenerhebung zu ($p=0.46$, GLM). Bei fünf Patienten wurden trotz Zuweisung an den Rechner Papier/Bleistift-Fragebögen verteilt. In all diesen Fällen waren die Patienten älter und würden nach Einschätzung der betreuenden Arzthelferin am Rechner nicht zurechtkommen. Wegen der kleinen Anzahl dieser Fälle wurden sie bei der Auswertung weggelassen. Alle Analysen bis auf die, bei denen entscheidungsorientierte Vorgabe ausdrücklich miteinbezogen war, wurden mit der im Laufe des ersten Studienabschnitts gesammelten Stichprobe durchgeführt.

Tabelle 6 enthält eine Beschreibung der Stichprobe anhand von ausgewählten demographischen und anamnestischen Variablen sowie Skalenwerten zur Aufnahme¹⁰. Es zeigt sich kein statistisch signifikanter Zusammenhang zwischen der

¹⁰ Eine ausführliche Darstellung der Grundgesamtheit ist im Jahresbericht der Panorama Fachklinik (2000) gegeben

Vorgabeform und den in Tabelle 6 aufgeführten Aufnahmevariablen ($p > 0.28$, GLM). Die Randomisierung wird als gelungen angesehen und es werden keine statistischen Angleichungen durchgeführt.

Tabelle 6. Stichprobenbeschreibung

Alter	48 Jahre (SD=14.7)
Geschlecht	Männlich: 19% Weiblich: 81%
Familienstand	Ledig: 20% Verheiratet: 53% Verwitwet: 9% Geschieden: 11% Getrennt lebend: 6%
Berufsabschluss	Noch in Berufsausbildung: 4% Lehre: 33% Meister/Fachschule: 15% Fachhochschule/Universität: 27% Ohne Abschluss: 9% Sonstiger Abschluss: 13%
Beschwerdendauer	Bis 1 Jahr: 23% Bis 2 Jahre: 19% Bis 5 Jahre: 26% Bis 10 Jahre: 14% Bis 15 Jahre: 8% Bis 20 Jahre: 10%
Therapiemotivation	3.36 (SD=0.85) (0=nicht motiviert...5=sehr motiviert)
GBB	32.5 (SD=15.75)
EB45-TOT	73.9 (SD=22.87)
EB45-SYB	44.6 (SD=14.35)
EB45-BEZ	15.7 (SD=7.24)
EB45-SOZ	13.6 (SD=4.95)
Frühere Behandlung	Keine: 86% 1-2 Aufnahmen: 10% 3-5 Aufnahmen: 3% Mehr als 5 Aufnahmen: 1%
Erstdiagnose	Affektive Störungen (F3): 31% Neurotische, Belastungs & Somatoforme (F4): 14% Verhaltensauffälligkeiten (F5): 5% Persönlichkeits- & Verhaltensstörungen (F6): 9% Alle anderen und fehlende: 42%
Computererfahrung*	Keine: 26% Wenig: 29% Häufige Benutzung: 45%

*Nur erfasst bei computergestützter Testvorgabe

AUSWERTUNG UND ERGEBNISSE

Ausgehend von den im Kapitel „Fragestellung“ dargestellten Zielen der Studie, wurde die Auswertung vorgenommen. Als Hauptfragen wurde untersucht: (1) können computergestützt erfasste Werte mit Papier/Bleistift-Werten verglichen werden und unter welchen Bedingungen, (2) können und wollen Patienten die Fragebögen am Rechner ausfüllen und unter welchen Bedingungen und (3) welche testökonomischen Folgen (Testzeit, Testkosten) hat computergestützte Testdarbietung. Als Nebenfrage wurde untersucht (4) welchen Effekt haben Rückmeldungen der Testergebnisse auf die Behandlung.

Zu (1) wurden die Rohwertverteilungen als Maß für Änderungen der Metrik und Verschiebungen im Niveau, die internen Konsistenzen als Maß der Reliabilität, die Retestkorrelationen als Maß für Änderungen der Patientenreihenfolge, die Skaleninterkorrelationen als Maß der Konvergenzvalidität und die Faktorenstrukturen als weiteres Maß der Konstruktvalidität zwischen den einzelnen computergestützten Vorgabevarianten untereinander und mit der Papier/Bleistift-Vorgabe verglichen. Um die Frage zu klären, ob die Bevorzugung von computergestützter oder Papier/Bleistift-Vorgabe die Antworten bei Computervorgabe beeinflusst, wurden die entsprechenden Patientengruppen verglichen. Um den Einfluss der Itemreihenfolge bei entscheidungsorientierter Vorgabe auf die Antworten zu klären, wurden die Lageparameter bei

entscheidungsorientierter und handbuchgemäßer Testvorgabe sowie die Verteilungen und Korrelationen bei tatsächlicher entscheidungsorientierter und simulierter entscheidungsorientierter Testvorgabe verglichen.

Zu (2) wurden die Verteilungen der Antworten auf die Fragen nach Anstrengung, Zurechtkommen, Verständlichkeit und Bevorzugung zwischen den einzelnen computergestützten Vorgabevarianten verglichen. Der Zusammenhang dieser Fragen mit Alter, Geschlecht, Berufsabschluss und Computererfahrung wurde untersucht. Unter der Annahme, dass die Bevorzugung von Computerdarbietung von Alter, Computererfahrung und der bearbeiteten Darstellungsvariante abhängt, wurde versucht diese Präferenz vorherzusagen. Die Antworten auf offene Fragen danach, was besonders gestört oder gefreut hat, wurden im Hinblick auf Hinweise zu Akzeptanz und Handhabung kategorisiert und ausgezählt. Als indirekte Indikatoren für die Akzeptanz und Handhabung wurde die Rate fehlender Werte sowie die Beantwortung offener Fragen zwischen den einzelnen computergestützten Vorgabevarianten untereinander und mit der Papier/Bleistift-Vorgabe verglichen. Zusätzlich wurden als Indikatoren für Akzeptanz und Handhabung die Bearbeitungszeiten zwischen den computergestützten Vorgabevarianten verglichen, und der Zusammenhang zwischen Bearbeitungszeiten und Alter, Geschlecht, Berufsabschluss und Computererfahrung untersucht.

Zu (3) wurden neben den Bearbeitungszeiten die geschätzten Befragungskosten herangezogen. Diese Schätzung basiert auf den in dieser Studie gemachten Erfahrungen mit computergestützter Fragebogenvorgabe.

Zu (4) wurde der Einfluss der Rückmeldungen an Patienten und Therapeuten auf die therapeutische Arbeitsbeziehung zur Entlassung und das Erreichen klinisch bedeutsamer Veränderung untersucht. Um zu überprüfen, inwieweit die Empfehlung, die Therapiedauer an dem erreichten Symptomzustand zu orientieren, von den Therapeuten befolgt wurde, und inwieweit die Rückmeldung dieses Zustandes den Therapeuten dabei hilft, wurde die Interaktion zwischen der Therapiedauer nach der Zwischenerhebung, dem Symptomzustand zur Zwischenerhebung und der Rückmeldung über diesen Symptomzustand untersucht.

Es fand kein hypothesenprüfendes Vorgehen im engeren Sinne statt¹¹. Bestimmte Erwartungen lagen jedoch aufgrund theoretischer Überlegungen oder empirischer Befunde nahe und diese sind in Tabelle 7 zusammengefasst. Diese Erwartungen sind bei der Einführung, bei der Fragestellung oder bei der Darstellung der Methode begründet.

¹¹ In dem Sinne, dass Hypothesen aus einer Theorie abgeleitet und getestet wurden, um diese Theorie zu falsifizieren oder zu stützen.

Tabelle 7. Erwartete Effekte

Rohwertverteilung
<p>Wenig Unterschied zwischen PC und P/B Vorgabe Größere Unterschiede zwischen PC und P/B bei itemweiser Vorgabe als bei seitenweiser Größere Unterschiede zwischen PC und P/B wenn keine nachträglichen Änderungen möglich sind Größere Unterschiede zwischen PC und P/B bei Rückmeldungen an Patienten Größere Unterschiede zwischen PC und P/B wenn keine fehlenden Werte erlaubt sind Varianz bei entscheidungsorientierter Vorgabe kleiner als bei kompletter Testvorgabe Unterschiede in der Vergleichbarkeit von PC und P/B in Abhängigkeit von der Präferenz</p>
Validität & Reliabilität
<p>Größer bei PC als bei P/B Vorgabe Größer bei itemweiser Vorgabe als bei seitenweiser Unterschiede in Abhängigkeit davon ob nachträgliche Änderungen möglich sind oder nicht Größer bei Rückmeldung an Patienten Unterschiede in Abhängigkeit davon ob fehlende Werte erlaubt sind oder nicht Größer bei kompletter Testvorgabe als bei entscheidungsorientierter Vorgabe</p>
Präferenz & Handhabung
<p>Präferenz für PC Vorgabe Präferenz für entscheidungsorientierte Vorgabe Präferenz für Vorgabe bei der nachträgliche Änderungen möglich sind Präferenz für Vorgabe mit Rückmeldung an Patienten Präferenz für Vorgabe bei der fehlende Werte erlaubt sind Unterschiedliche Präferenz für itemweise und seitenweise Vorgabe Je älter der Patient, desto größer die Präferenz für P/B Je weniger Computererfahrung, desto größer die Präferenz für P/B Je älter der Patient, desto schwerer die Handhabung von PC Je weniger Computererfahrung, desto schwerer die Handhabung von PC Einfachere Handhabung bei itemweiser Vorgabe Entscheidungsorientierte Vorgabe weniger anstrengend als die komplette Vorgabe</p>
Effekte auf die Behandlung
<p>Mehr Anpassung der Therapiedauer an den Symptomzustand bei Rückmeldung an Therapeuten Bessere Therapieergebnisse bei Rückmeldung an Therapeuten Bessere Therapieergebnisse bei Rückmeldung an Patienten Bessere Arbeitsbeziehung bei Rückmeldung an Therapeuten Bessere Arbeitsbeziehung bei Rückmeldung an Patienten</p>

...

Fehlende Werte
Weniger fehlende Werte bei PC als bei P/B-Vorgabe
Weniger fehlende Werte bei itemweiser Vorgabe als bei seitenweiser
Weniger fehlende Werte in der aktuellen Erhebung bei Rückmeldung an Patienten
Weniger fehlende Werte zur Entlassung bei Rückmeldung an Patienten
Testzeit
Je älter der Patient, desto länger die Testzeit
Je weniger Computererfahrung, desto länger die Testzeit
Längere Testzeit bei kompletter als bei entscheidungsorientierter Vorgabe
Kürzere Itemresponsezeit bei kompletter als bei entscheidungsorientierter Vorgabe
Längere Testzeit bei itemweiser Vorgabe als bei seitenweiser
Kürzere Testzeit wenn keine nachträglichen Änderungen möglich sind
Kürzere Testzeit bei PC als bei P/B-Vorgabe
Kosten
Geringere Kosten bei PC als bei P/B-Vorgabe
Geringere Kosten bei entscheidungsorientierter Vorgabe als bei kompletter Vorgabe

Psychometrische Äquivalenz

Tabelle 8 zeigt die Verteilungsparameter der Skalenwerte und die Ergebnisse der Tests auf Verteilungsunterschiede für die einzelnen computergestützten Vorgabeformen. Zur Überprüfung der Verteilungsäquivalenz wurde der KS-Test eingesetzt, da nicht von Normalverteilung ausgegangen werden konnte. Die Rohwertverteilungen zwischen den einzelnen computergestützten Vorgabevarianten sind vergleichbar. Es zeigten sich keine statistisch signifikanten Unterschiede (p zwischen 0.07 und 0.99). Dennoch wurden, um den Befund weiter absichern und in seiner Bedeutung besser einschätzen zu können, die multivariaten Effektstärken der beobachteten Mittelwertsunterschiede über GLM geschätzt. Die geschätzten

Effektstärken sind mit $\eta^2 < 0.04$ klein und die Effekte nicht signifikant ($p > 0.05$).

Tabelle 8. Verteilungsparameter und Reliabilität der computergestützten Vorgabeformen

EB45-SYB	min	q1	med	q3	max	p	M	SD	r _{xx}	p
Seitenweise	9	33	40	48	82		41.1	13.91	.89	
Itemweise	7	27	41	51	80	.52	40.5	14.97	.92	.95
Fehl. erlaubt	8	28.5	41	50	82		40.5	14.70	.91	
Fehl. n. erlaubt	7	33.75	41	49.25	80	.88	41.1	14.24	.90	.29
Änderungen möglich	8	32	41	50	82		40.8	14.21	.90	
Änderungen n. möglich	7	32	40	47	80	.90	40.8	15.33	.92	.85
Rückmeldung	7	34	41	50	82		42	14.19	.89	
Keine Rückmeldung	8	27	41	50	75	.27	39.6	14.63	.91	.85
EB45-BEZ	min	q1	med	q3	max	p	M	SD	r _{xx}	p
Seitenweise	3	11	16	20.25	31		16.0	6.62	.77	
Itemweise	1	11	15	19.5	35	.90	15.3	6.47	.80	.76
Fehl. erlaubt	1	9	15	19	35		15.2	6.94	.81	
Fehl. n. erlaubt	3	11	16	20	33	.68	16.1	6.17	.75	.08
Änderungen möglich	1	11	16	20	31		15.7	6.55	.79	
Änderungen n. möglich	3	12	15	18	35	.45	15.6	6.58	.79	.51
Rückmeldung	2	10	15	20.25	35		15.4	7.06	.82	
Keine Rückmeldung	1	12	17	19	31	.51	15.9	5.97	.74	.03
EB45-SOZ	min	q1	med	q3	max	p	M	SD	r _{xx}	p
Seitenweise	0	10	14	16	22		12.9	4.33	.57	
Itemweise	3	9	13	16	26	.69	13.0	4.70	.64	.81
Fehl. erlaubt	0	9	13	16	26		12.8	4.89	.65	
Fehl. n. erlaubt	4	10	14	16	23	.89	13.2	4.16	.55	.10
Änderungen möglich	0	9.25	13	16	26		13.0	4.60	.62	
Änderungen n. möglich	4	9	13	16	23	.99	12.9	4.22	.52	.18
Rückmeldung	3	10	14	16.5	23		13.2	4.49	.62	
Keine Rückmeldung	0	9	13	16	26	.89	12.7	4.54	.58	.31
EB45-TOT	min	q1	med	q3	max	p	M	SD	r _{xx}	p
Seitenweise	15	55	70	83	126		70.2	21.28	.91	
Itemweise	18	51	69	85	132	.61	68.7	23.32	.93	.91
Fehl. erlaubt	15	51.5	66.5	84.75	126		68.4	23.22	.93	
Fehl. n. erlaubt	20	54	70.5	82	132	.41	70.4	21.47	.92	.24
Änderungen möglich	15	53	70	83	126		69.5	22.20	.92	
Änderungen n. möglich	21	54	68	85	132	.99	69.4	22.86	.92	.52
Rückmeldung	20	54	69	84	132		70.7	22.73	.92	
Keine Rückmeldung	15	50	70	82	124	.79	68.1	21.86	.92	.50
GBB	min	q1	med	q3	max	p	M	SD	r _{xx}	p
Seitenweise	0	12	22	32	61		23.6	13.31	.89	
Itemweise	2	12	24	35	65	.36	25.1	14.62	.91	.85
Fehl. erlaubt	2	12.25	24	34	65		25.5	14.44	.91	
Fehl. n. erlaubt	0	12	21	33	58	.26	23.3	13.53	.89	.15
Änderungen möglich	2	12	22	33	65		24.3	14.09	.90	
Änderungen n. möglich	0	14	23	35	56	.99	24.4	13.75	.91	.69
Rückmeldung	4	16	24	34	61		25.7	12.63	.87	
Keine Rückmeldung	0	11	20	33	65	.07	22.9	15.20	.92	.99

...

PTBEZ	min	q1	med	q3	max	p	M	SD	r _{xx}	p
Seitenweise	-15	14	19	24	33		18.3	8.41	.84	
Itemweise	0	17	21	25	33	.11	20.3	6.60	.79	.08
Fehl. erlaubt	-15	14.25	21	24	33		18.7	8.46	.84	
Fehl. n. erlaubt	-3	16	20	25	32	.32	19.9	6.68	.79	.08
Änderungen möglich	-15	16	20	25	33		19.1	7.88	.83	
Änderungen n. möglich	4	16	21	24	33	.99	20.0	6.51	.78	.15
Rückmeldung	-15	15	19	23	33		18.4	7.88	.81	
Keine Rückmeldung	-5	16	21	25	33	.23	20.3	7.17	.83	.71
ZUF8	min	q1	med	q3	max	p	M	SD	r _{xx}	p
Seitenweise	16	25	27	30	32		27.5	3.37	.88	
Itemweise	18	26	28	30	32	.98	27.8	3.08	.82	.02
Fehl. erlaubt	16	26	28	30	32		27.3	3.34	.85	
Fehl. n. erlaubt	17	26	28	31	32	.57	27.9	3.10	.86	.63
Änderungen möglich	16	26	28	30	32		27.6	3.27	.85	
Änderungen n. möglich	18	25	28	30	32	.99	27.8	3.06	.88	.83
Rückmeldung	18	25	27	30	32		27.5	3.03	.84	
Keine Rückmeldung	16	26	28	30	32	.18	27.7	3.41	.87	.85
PAE	min	q1	med	q3	max	p	M	SD	r _{xx}	p
Seitenweise	3	5	7	8	14		6.9	2.51	.83	
Itemweise	3	5	6	8	15	.86	6.6	2.58	.85	.71
Fehl. erlaubt	3	5	7	8	15		6.8	2.70	.85	
Fehl. n. erlaubt	3	5	6	8	14	.99	6.8	2.42	.83	.29
Änderungen möglich	3	5	6	8	14		6.8	2.63	.86	
Änderungen n. möglich	3	6	6	8	15	.84	6.7	2.27	.76	.03
Rückmeldung	3	6	7	8	15		7.1	2.46	.84	
Keine Rückmeldung	3	4	6	8	14	.17	6.5	2.62	.84	.50

Tabelle 8 zeigt weiter die internen Konsistenzen (Chronbachs Alpha Koeffizient) für alle computergestützten Vorgabevarianten und die Ergebnisse der entsprechenden Vergleichstests. Bei dem Vergleich der internen Konsistenzen unterschieden sich 3 Gruppen statistisch signifikant ($p < 0.05$, Feldt-Test). Um die weitere Datenanalyse zu begründen, wird dieser Befund hier diskutiert. Die signifikanten Unterschiede sind bei einer beobachteten Differenz in der Reliabilität zwischen 0.06 und 0.10 praktisch relevant. Jedoch findet sich jede dieser drei signifikanten Differenzen auf einer anderen Skala und bei einer anderen Vorgabevariante. Bei den 32 durchgeführten Vergleichen ist die Wahrscheinlichkeit, unter der Annahme, dass es keine tatsächlichen Unterschiede gibt,

mit ca. 0.22 recht hoch, drei oder mehr signifikante Unterschiede zu finden. Da sich in den beobachteten Differenzen insgesamt kein Trend oder Muster zeigt, scheint dies die plausiblere Erklärung für die gefundenen statistischen Signifikanzen zu sein als die Annahme tatsächlicher Unterschiede.

Für den Vergleich mit der Papier/Bleistift-Vorgabe wurden die einzelnen computergestützten Vorgabeformen daher gepoolt. Ansonsten erfolgte die Auswertung wie bei dem Vergleich der computergestützten Vorgabeformen, mit der Ausnahme, dass zusätzlich eine Auswertung auf Itemebene und eine detailliertere Betrachtung der Verteilungsunterschiede erfolgte. Möglich wurde dies durch die größere Stichprobe nach Zusammenführen der einzelnen computergestützten Vorgabevarianten.

Die Itemkennwerte sind im Anhang angegeben. Die Itemschwierigkeiten (Itemmittelwerte; Abbildung 7), die Itemtrennschärfen (korrigierte Item-Skala Korrelationen; Abbildung 8) und die Itemdispersionen (Itemstandardabweichungen) korrelieren hoch zwischen den beiden Vorgabemodalitäten ($r^2_{\text{schwierigkeit}}=0.98$; $r^2_{\text{trennschärfe}}=0.75$; $r^2_{\text{dispersion}}=0.93$). Es wurden keine statistisch signifikanten Unterschiede im Mittelwert der Itemdispersion gefunden ($M_{\text{pb}}=0.98$; $M_{\text{pc}}=0.97$; t-Test: $p=0.17$). Ebenfalls wurden keine statistische signifikanten Unterschiede im Mittelwert der Itemtrennschärfe gefunden ($M_{\text{pb}}=0.50$; $M_{\text{pc}}=0.49$; t-Test: $p=0.34$).

Die durchschnittlichen Itemschwierigkeiten unterscheiden sich jedoch signifikant in ihrem Ausmaß ($M_{pb}=2.66$; $M_{pc}=2.59$; t-Test: $p<0.01$). Bei computergestützter Testvorgabe wurden tendenziell mehr Beschwerden, weniger Zufriedenheit, eine schlechtere therapeutische Arbeitsbeziehung und weniger Veränderung berichtet. Allerdings ist die Effektstärke mit 0.1 klein. Die Signifikanz ist mit der Verwendung des t-Tests für gepaarte Stichproben und der hohen Korrelation zwischen den Vorgabemodalitäten zu erklären.

Bei Betrachtung Item für Item finden sich unter den 88 untersuchten Items vier, deren Mittelwerte sich signifikant unterscheiden ($p<0.05$; t-Test), sieben, deren Varianzen sich signifikant unterscheiden ($p<0.05$; Levene-Test), und ein Item, wo mit dem KS-Test signifikante Verteilungsunterschiede ($p<0.05$) erkannt wurden¹². Ein Muster findet sich nur bei den Itemschwierigkeiten: die größten Effektstärken treten bei Items der Skalen PTBEZ und ZUF8 auf. Der Unterschied geht systematisch in Richtung weniger Zufriedenheit mit der therapeutischen Arbeitsbeziehung und weniger Zufriedenheit mit der Behandlung bei computergestützter Testvorgabe. Jedoch ist auch Größe des Effektes klein (ES bis max. 0.29 für das zweite Item des ZUF8). Keine systematischen Unterschiede sind bei den Itemdispersionen und Itemtrennschärfen zu erkennen. Auffallend sind jedoch die niedrigen und zum Teil negativen

¹² Bei Adjustierung der Signifikanzgrenze bezüglich der Anzahl der Items, lassen sich keine signifikanten Unterschiede mehr feststellen.

Trennschärfeindizes für einzelne Items (vgl. Anpassungsgüte der Messmodelle weiter im Text).

Abbildung 7. Itemschwierigkeiten bei Papier/Bleistift Vorgabe und bei computergestützter Vorgabe

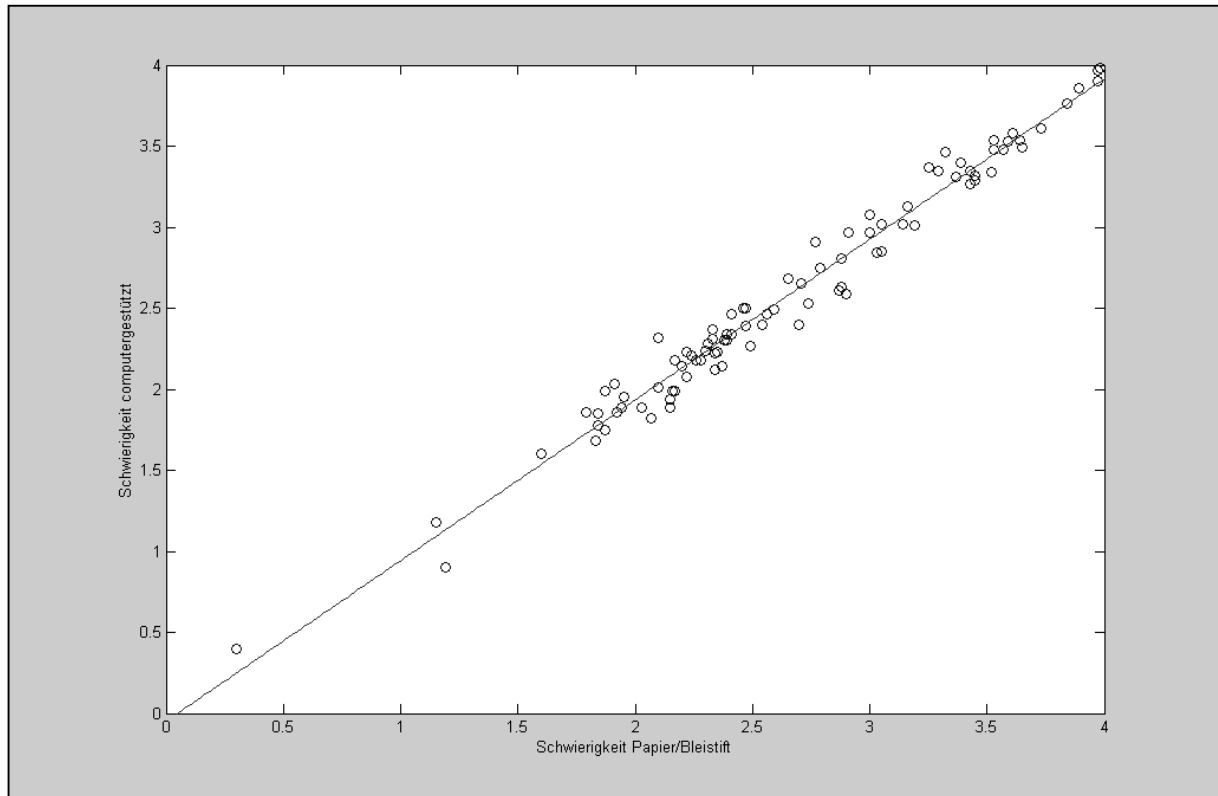
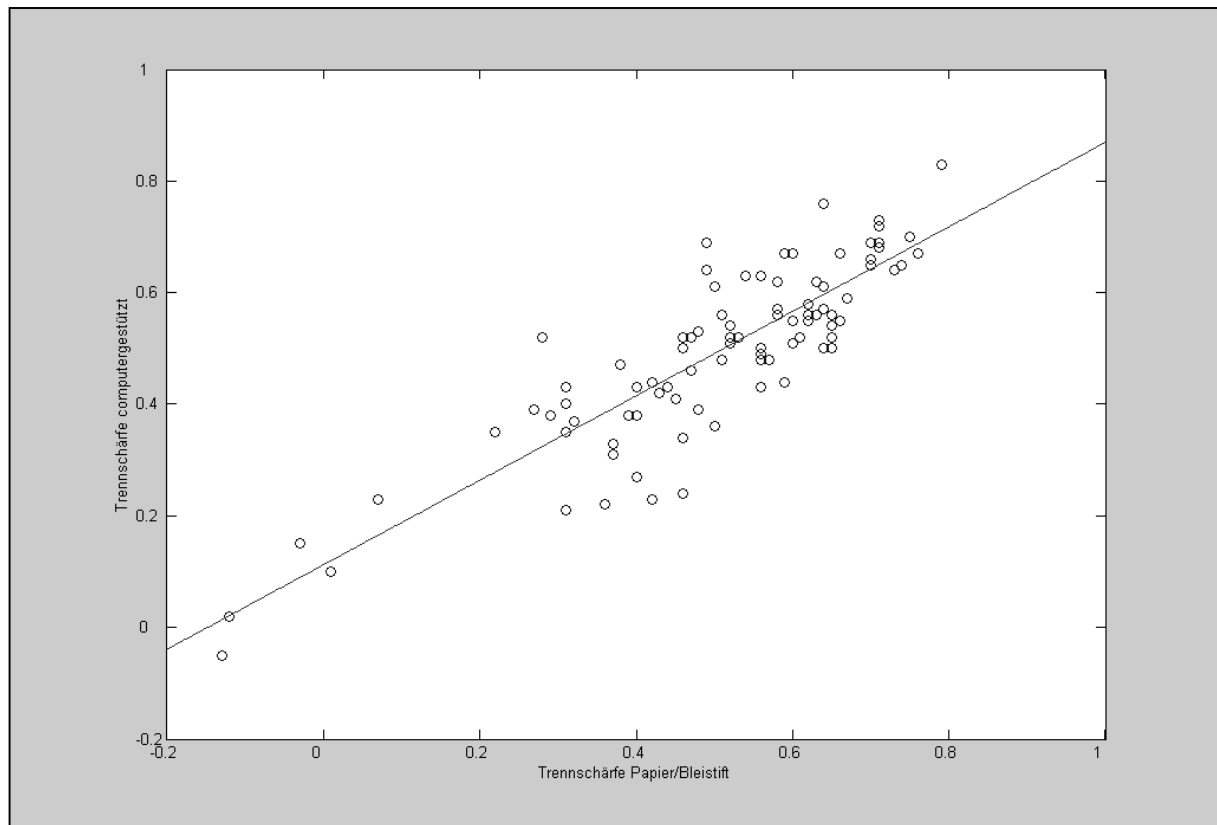


Abbildung 8. Itemtrennschärfe bei Papier/Bleistift Vorgabe und bei computergestützter Vorgabe



Die beobachteten Verteilungsunterschiede zwischen computergestützter und Papier/Bleistift-Vorgabe auf Skalenebene sind auf Abbildung 9 dargestellt. Es zeigt sich das gleiche Bild wie schon bei der Betrachtung auf Itemebene: ein leichter Trend zu etwas negativeren Einschätzungen bei computergestützter Vorgabe. Tabelle 9 zeigt die entsprechenden deskriptiven Verteilungsparameter sowie die Ergebnisse der dazugehörigen Vergleichstests. Keiner der Vergleichstests zeigt statistisch signifikante Unterschiede (p zwischen 0.06 und 0.89). Die geschätzte Effektstärke der Mittelwertsunterschiede fiel ebenfalls klein und nicht

signifikant aus ($\eta^2=0.01$; $p>0.05$). Die Ergebnisse auf Itemebene konnten auf Skalenebene somit nicht bestätigt werden. Ebenso konnte der visuelle Eindruck von den Verteilungsdarstellungen nicht statistisch abgesichert werden. Unter Berücksichtigung der gegebenen Teststärke und der kleinen beobachteten Effekte ist dieser Befund jedoch nachvollziehbar. Bezüglich der internen Konsistenzen sind die Befunde im Einklang mit der Analyse auf Itemebene (vgl. Tabelle 9): Keine statistisch signifikanten Unterschiede zwischen den Reliabilitätskoeffizienten wurden gefunden (p zwischen 0.19 und 0.76).

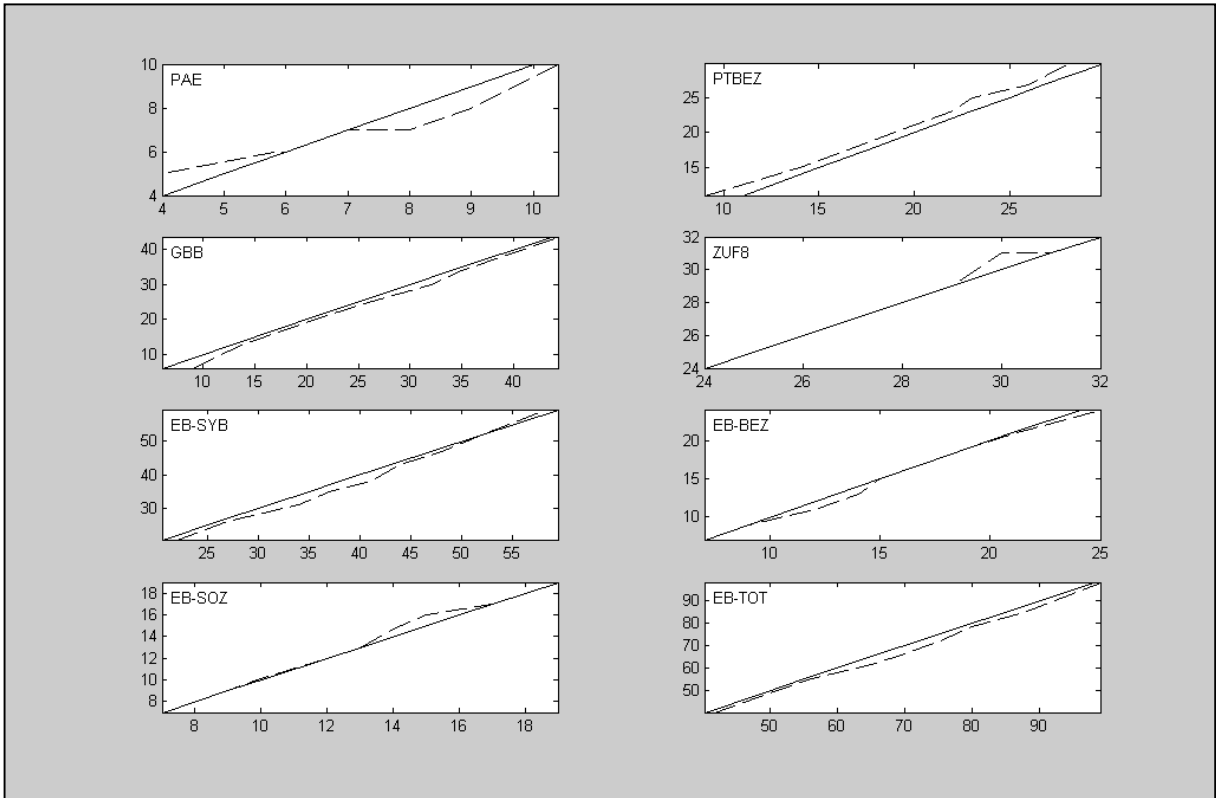


Abbildung 9. Vergleich der Rohwerte bei Papier/Bleistift-Vorgabe (-) und bei computergestützter Vorgabe (--) [mit Papier/Bleistift-Vorgabe als Referenz]

Tabelle 9. Verteilungsparameter und Reliabilität der Papier/Bleistift-Vorgabe und der gepoolten computergestützten Vorgabeformen

EB45-SYB	min	q1	med	q3	max	p	M	SD	r _{xx}	p
Papier/Bleistift	9	29	38	50	86		39.6	14.64	.92	
Computergestützt	7	32	41	50	82	.37	40.8	14.43	.91	.20
EB45-BEZ	min	q1	med	q3	max	p	M	SD	r _{xx}	p
Papier/Bleistift	0	10	15	20	33		15.1	6.68	.79	
Computergestützt	1	11	15	20	35	.79	15.7	6.54	.79	.50
EB45-SOZ	min	q1	med	q3	max	p	M	SD	r _{xx}	p
Papier/Bleistift	3	10	13	17	30		13.1	4.71	.63	
Computergestützt	0	9	13	16	26	.88	12.9	4.51	.60	.30
EB45-TOT	min	q1	med	q3	max	p	M	SD	r _{xx}	p
Papier/Bleistift	16	52	66	81	131		67.6	22.26	.93	
Computergestützt	15	53.75	70	83.25	132	.38	69.4	22.30	.92	.19
GBB	min	q1	med	q3	max	p	M	SD	r _{xx}	p
Papier/Bleistift	0	12	21	32	71		22.7	14.13	.91	
Computergestützt	0	12	22	33	65	.53	24.3	13.99	.90	.22
PTBEZ	min	q1	med	q3	max	p	M	SD	r _{xx}	p
Papier/Bleistift	-16	17	21	25	40		20.4	7.80	.82	
Computergestützt	-15	16	20	24	33	.06	19.3	7.59	.82	.50
ZUF8	min	q1	med	q3	max	p	M	SD	r _{xx}	p
Papier/Bleistift	17	26	28	31	32		28.0	3.14	.85	
Computergestützt	16	26	28	30	32	.13	27.6	3.22	.86	.69
PAE	min	q1	med	q3	max	p	M	SD	r _{xx}	p
Papier/Bleistift	3	5	6	8	15		6.6	2.18	.82	
Computergestützt	3	5	6	8	15	.89	6.8	2.55	.84	.76

Als Hinweis auch die Konstruktinvarianz wurden zunächst die Retestkorrelationen sowie die Interkorrelationen der Zwischenerhebungsinstrumente zwischen den einzelnen Vorgabemodalitäten verglichen. Die zugrundeliegende Korrelationstabelle kann im Anhang eingesehen werden. Da keine spezifischen Hypothesen bezüglich der Instrumente vorlagen, wurde nicht jeder einzelne Korrelationskoeffizient verglichen, sondern es wurden die durchschnittlichen absoluten Korrelationen herangezogen. Damit wurde die Teststärke erhöht und die Anzahl der notwendigen Vergleiche verkleinert. Tabelle

10 zeigt die durchschnittlichen Ausprägungen der Interkorrelationen und der Retestkorrelationen für die einzelnen Vorgabemodalitäten sowie die Ergebnisse der entsprechenden Vergleichstests. Die Unterschiede in den durchschnittlichen Korrelationen wurden über ein Bootstrap-Verfahren auf statistisch signifikante Differenz geprüft (siehe Shao & Tu, 1996 für die Verfahrensklasse), da der üblicherweise eingesetzte Box-Test nicht die mittlere Ausprägung, sondern paarweise alle Unterschiede zweier Korrelationsmatrixen testet. Es wurden keine statistisch signifikanten Unterschiede gefunden (p zwischen 0.09 und 0.91). Sowohl die durchschnittliche Ausprägung der Korrelationen der Instrumente untereinander, als Indikator für die Konvergenzvalidität, wie auch die durchschnittliche Ausprägung der Retestkorrelationen, als Maß für Änderungen in der Patientenreihenfolge, erwiesen sich als vergleichbar.

Tabelle 10. Durchschnittliche Retestkorrelationen und Interkorrelationen der Zwischenerhebungsinstrumente zu den einzelnen Vorgabeformen

	Zwischen Zeitpunkten	p	Zwischen Instrumenten	p
Seitenweise Vorgabe	.60		.47	
Itemweise Vorgabe	.68	.17	.42	.48
Fehlende Werte erlaubt	.59		.47	
Fehlende Werte nicht erlaubt	.69	.15	.42	.58
Änderungen möglich	.69		.44	
Änderungen nicht möglich	.63	.32	.44	.91
Rückmeldung an Patienten	.62		.44	
Keine Rückmeldung an Pat.	.67	.48	.46	.41
Papier/Bleistift-Vorgabe	.69		.45	
Computergestützte Vorgabe	.64	.25	.52	.09

Ebenfalls zwischen allen Vorgabemodalitäten vergleichbar ist die Faktorstruktur der Zwischenerhebungsinstrumente auf Skalenebene. Die Komponentenladungen für die einzelnen Vorgabemodalitäten sind im Anhang aufgeführt. Unabhängig von der Vorgabemodalität wurden je zwei Komponenten extrahiert (Hauptkomponentenanalyse; Extraktionskriterium $\text{Eigenwerte} > 1$; erklärte Gesamtvarianz zwischen 66% und 72%), welche „Beschwerden“ (EB45-SYB, EB45-SOZ, EB45-BEZ und GBB) sowie „Behandlungszufriedenheit“ (ZUF8, PTBEZ) abzubilden scheinen. Die direkte Veränderungseinschätzung PAE lädt auf beiden Komponenten signifikant. Die Kongruenz der Komponentenladungen lag für die einzelnen Vorgabevarianten zwischen 0.91 und 0.99 (wobei ab einer Kongruenz von 0.80 von Faktorenäquivalenz ausgegangen wird; s. Fulgosi, 1988).

Für die gepoolten computergestützten Vorgabevarianten und die Papier/Bleistift Vorgabe wurde zusätzlich für jedes Instrument getrennt ein Vergleich der Faktorladungen auf Itemebene durchgeführt. Dazu wurde die Anpassungsgüte der skalenkonformen Messmodelle (siehe Anhang) bei vorgabespezifischer und bei gemeinsamer Parameterschätzung verglichen (LISREL 8.54, Maximum Likelihood Schätzung). Die Ergebnisse sind in Tabelle 11 zusammengefasst. Insgesamt ist die Anpassungsgüte der Messmodelle bis auf PAE eher schlecht. Es wurden jedoch keine signifikanten Unterschiede in der Anpassungsgüte in Abhängigkeit davon ob die Parameterschätzung gemeinsam oder getrennt erfolgte gefunden (p zwischen 0.05 und

0.99). Bei EB-SOZ ist der Unterschied auf dem Chi-Quadrat an der Signifikanzgrenze ($p=0.05$), der RMSEA bestätigt dies jedoch nicht. Entsprechend den Empfehlungen, verschiedene Indices für die Anpassungsgüte zu nutzen, um sich ein Gesamtbild zu verschaffen (z.B. Byrne, 1989), widerlegen die Ergebnisse auch für den EB-SOZ nicht die Annahme von Konstruktinvarianz.

Tabelle 11. Anpassungsgüte der Messmodelle bei vorgabespezifischer und bei gemeinsamer Schätzung der Faktorladungen

		Chi-Quadrat	d.f.	p	RMSEA
GBB	Vorgabeunspezifisch	2815.00	552		0.16
	Vorgabespezifisch	2802.50	528	0.97	0.16
EB-SYB	Vorgabeunspezifisch	2239.25	600		0.11
	Vorgabespezifisch	2216.25	575	0.58	0.11
EB-BEZ	Vorgabeunspezifisch	499.24	90		0.14
	Vorgabespezifisch	493.69	80	0.85	0.15
EB-SOZ	Vorgabeunspezifisch	205.54	72		0.11
	Vorgabespezifisch	188.72	63	0.05	0.11
EB-TOT	Vorgabeunspezifisch	6855.02	1980		0.11
	Vorgabespezifisch	6828.76	1935	0.99	0.11
PAE	Vorgabeunspezifisch	11.44	6		0.06
	Vorgabespezifisch	5.51	3	0.12	0.06
ZUF8	Vorgabeunspezifisch	159.70	56		0.09
	Vorgabespezifisch	147.52	48	0.14	0.10
PTBEZ	Vorgabeunspezifisch	803.12	110		0.17
	Vorgabespezifisch	788.94	99	0.22	0.18

Tabelle 12 zeigt die Verteilungsparameter der Skalenwerte bei entscheidungsorientierter und bei handbuchgemäßer Vorgabe. Der Vergleich erfolgte mit Hilfe des Mann-Whitney-Tests, da Unterschiede in der Verteilung bei entscheidungsorientierter Vorgabe von vornherein erwartet waren, und nur Unterschiede in der Lage von Interesse waren. Kein statistisch signifikanter

Unterschied der Lageparameter der Skalenwerte wurde gefunden (p zwischen 0.17 und 0.33). Die durchschnittliche Ausprägung der Korrelationen aller Zwischenerhebungsinstrumente mit dem GBB und den EB45-Skalen ist bei entscheidungsorientiertem Abbruchkriterium allerdings deutlich niedriger als bei Vorgabe aller Items ($r_{entsch}=0.36$, $r_{kompl}=0.46$, $p<0.01$).

Tabelle 12. Verteilungsparameter der Skalenwerte bei entscheidungsorientiertem und handbuchgemäßem Vorgehen

	M (SD)		p
	Entscheidungsorientiert	Handbuchgemäß	
EB45-SYB	38.7 (14.49)	42.4 (15.76)	.17
EB45-BEZ	16.2 (5.95)	15.5 (6.69)	.33
EB45-SOZ	12.6 (4.69)	13.4 (5.36)	.28
GBB	25.9 (13.55)	24.4 (14.48)	.23

Um zu überprüfen, inwieweit neben den erwarteten statistischen Effekten¹³ auch ein verändertes Antwortverhalten dabei eine Rolle spielt, wurde an der Teilstichprobe mit handbuchgemäßer Vorgabe entscheidungsorientierte Vorgabe simuliert und mit den Werten bei tatsächlicher entscheidungsorientierter Vorgabe verglichen. Auf Itemebene zeigten sich weder signifikante Unterschiede zwischen den Itemschwierigkeiten ($M_{entsch}=1.25$, $M_{simu}=1.27$, $ttest$: $p=0.87$) noch in deren Dispersion ($M_{entsch}=0.88$, $M_{simu}=0.86$, $ttest$: $p=0.19$). Itemschwierigkeiten und Dispersion korrelieren zwar nur moderat ($r^2_{schwierigkeit}=0.64$; $r^2_{dispersion}=0.41$), kein Item

¹³ Aufgrund der Regressionsschätzung und der am Trennwert orientierten Itemanzahl

weicht jedoch derart ab, dass von einem ausgeprägten Itembias ausgegangen werden muss. Ebenfalls vergleichbar ist die durchschnittliche Ausprägung der Korrelationen der Items untereinander ($r_{entsch}=0.36$, $r_{simu}=0.38$). Auf Skalenebene unterscheiden sich die Verteilungen bei simulierter und bei tatsächlicher entscheidungsorientierter Vorgabe ebenfalls nicht statistisch signifikant (KS-Test: $p>0.57$). Ebenso sind die vorgabespezifischen Unterschiede in den Korrelationen der entscheidungsorientiert vorgegebenen Instrumenten zu den nicht entscheidungsorientiert vorgegebenen Instrumenten (PAE, PTBEZ und ZUF8) gering ($r_{entsch}=0.24$, $r_{simu}=0.22$). Größere Unterschiede wurden jedoch bei den durchschnittlichen Interkorrelationen der entscheidungsorientiert vorgegebenen Instrumenten beobachtet ($r_{entsch}=0.30$, $r_{simu}=0.42$).

Die simulierte Vorgabe wurde zuletzt noch zum Replizieren der Untersuchung zur Klassifikationsgüte bei entscheidungsorientierter Testvorgabe genutzt (Percevic & Wesiak, 2001; s. Methodenteil). Dazu wurde die Klassifikation „Zielbereich erreicht“ vs. „Zielbereich nicht erreicht“ anhand des kompletten Fragebogens mit der anhand des simuliert entscheidungsorientiert vorgegebenen Fragebogens verglichen. Demnach würden bei entscheidungsorientierter Vorgabe 5 Skalenwerte fehlerhaft als „im Zielbereich“ klassifiziert werden und 20 Skalenwerte fehlerhaft als „außerhalb des Zielbereiches“. Bei einer Stichprobengröße von 112 Patienten und vier Skalen, d.h. insgesamt 448 Skalenwerten, entspricht

dies einer Rate korrekter Klassifikationen von 94.4%. Dabei lag die Entfernung zum Zielbereich auf Grundlage des kompletten Fragebogens bei den fehlklassifizierten Fällen unterhalb eines Standardmessfehlers: es wurde ein maximaler Abstand vom Zielbereich von 3 Punkten auf dem GBB beobachtet.

Akzeptanz

Tabelle 13 zeigt die Verteilung der Antworten auf die Fragen nach Anstrengung, Zurechtkommen, Verständlichkeit und Bevorzugung (Fragen 1, 2, 3 und 5 der Zusatzbefragung) für die einzelnen Vorgabevarianten. Mehr als 2/3 der Patienten geben an, dass sie computergestützte Darbietung bevorzugen, die Befragung nicht oder kaum anstrengend war, sie mit der Software ohne Schwierigkeiten zurechtgekommen und die Fragen verständlich waren. Es fand sich kein statistisch signifikanter Unterschied zwischen den Vorgabeformen (p zwischen 0.12 und 0.91; χ^2 -Tests). Bei den Vorgabeformen wo keine nachträglichen Änderungen möglich waren, zeigte sich jedoch tendenziell eine niedrigere Bevorzugung für die computergestützte Vorgabe (70% bis 77% im Vergleich zu den sonst beobachteten 78% bis 84%). Die niedrigste beobachtete Präferenz war mit 70% bei entscheidungsorientierter Vorgabe.

Tabelle 13. Akzeptanz und Handhabung der einzelnen Vorgabeformen

	Itemweise Vorgabe		Fehlende Werte nicht erlaubt		Fehlende Werte erlaubt		Änderungen nicht möglich		Änderungen möglich		keine Rückmeldung an Pat. Rückmeldung an Patienten		Standardvorgabe		entscheidungsorientierte Vorgabe	
	%	p	%	p	%	p	%	p	%	p	%	p	%	p	%	p
Anstrengung																
nicht	38	.48	43	.43	45	.37	46	.41	38	.42			38	.42		
kaum	33	.36	32	.36	32	.41	33	.35	35	.31			35	.31		
einigermaßen	24	.11	21	.14	18	.16	15	.20	21	.22			21	.22		
erheblich	4	.04	3	.05	4	.06	4	.04	6	.04			6	.04		
stark	1	.12	1	.64	1	.61	2	.59	1	.91			1	.91		
Zurechtkommen (Software)																
ohne Schwierigkeiten	72	.79	81	.71	76	.75	73	.79	71	.73			71	.73		
ohne nennenswerte Schwierigkeiten	21	.11	11	.20	16	.16	16	.16	16	.19			16	.19		
einigermaßen	5	.06	4	.07	5	.08	8	.04	10	.07			10	.07		
mit erheblichen Schwierigkeiten	3	.30	5	.09	4	.80	2	.48	3	.67			3	.67		
Verständlichkeit (Fragen)																
verständlich	71	.77	76	.72	73	.77	75	.72	76	.76			76	.76		
größtenteils verständlich	28	.24	23	.28	26	.24	24	.28	22	.23			22	.23		
kaum verständlich	1	.00	1	.00	1	.00	1	.00	1	.01			1	.01		
nicht verständlich	0	.39	0	.39	0	.79	0	.77	1	.97			1	.97		
Bevorzugung																
Computergestützt	84	.79	80	.82	83	.77	78	.84	70	.77			70	.77		
Papier/Bleistift	16	.39	20	.73	17	.42	22	.30	30	.25			30	.25		

Tabelle 14 zeigt alle untersuchten Zusammenhänge zwischen den demographischen Variablen und den Antworten auf die Fragen 1 bis 5 der Zusatzbefragung. Es ergaben sich statistisch signifikante Zusammenhänge zwischen Alter und Zurechtkommen an Rechner (je älter der Patient desto schwerer fällt ihm die Handhabung des Computers; $\tau=0.31$, $p<0.05$) sowie zwischen der Erfahrung im Umgang mit Computern und der Anstrengung der

Befragung (je unerfahrener desto anstrengender; $\tau=0.17$, $p<0.05$) und der Erfahrung im Umgang mit Computern und dem Zurechtkommen an Rechner (je unerfahrener desto schwerer; $\tau=0.49$, $p<0.05$). Die Zeit, welche ein Patient braucht, um den Fragebogen zu beantworten, hängt vom Alter, von der Computererfahrung und vom Berufsabschlusses ab. Ältere Patienten ($\tau=0.19$; $p<0.05$), computerunerfahrenere Patienten ($\tau=0.29$; $p<0.05$) und Patienten mit niedrigerem Berufsabschluss ($\tau=0.12$; $p<0.05$) brauchten länger. Ebenfalls zeigte sich ein Zusammenhang zwischen Alter und Computererfahrung, je älter der Patient desto weniger Computererfahrung hat er ($\tau=0.24$; $p<0.05$).

Tabelle 14. Zusammenhang zwischen personenbezogenen Variablen und Akzeptanzfragen (τ)

	Alter	Geschlecht	Berufsabschluss	Computererfahrung
Anstrengung	.01	.06	.09	.17*
Zurechtkommen	.31*	.06	.04	.49*
Verständlichkeit	.01	.03	.03	.01
Bevorzugung	.04	.05	.12	.09
Testzeit	.19*	.02	.12*	.29*

* $p<0.05$

Die Präferenz für computergestützte bzw. Papier/Bleistift-Vorgabe ließ sich nicht über das Alter, die Computererfahrung und die gerade bearbeitete Computervariante bestimmen. Die entsprechende logistische Regressionsrechnung führte zu keinem statistisch signifikanten Modell ($R^2=0.11$, $p=0.29$). Ebenso ließen sich bei Kontrolle des Alters die Skalenwerte zur

Zwischenerhebung nicht über die Bevorzugung vorhersagen ($\eta^2=0.11$, $p=0.07$; GLM).

Tabelle 15 zeigt den Anteil fehlender Werte für die einzelnen Vorgabeformen. Der Gesamtanteil fehlender Werte, bei Ausnahme der offenen Fragen, war bei allen computergestützten Vorgabevarianten sehr niedrig und lag im Mittel bei 0.53%. Deshalb wurde auf eine skalen- oder itembasierte Auswertung verzichtet. Es wurden keine Unterschiede zwischen den computergestützten Vorgabevarianten gefunden (p zwischen 0.12 und 0.99; χ^2 -Tests). Im Vergleich dazu gab es bei Papier/Bleistift-Vorgabe deutlich mehr fehlende Angaben (4.38%, $p<0.05$; χ^2 -Test) und komplett fehlende Fragebögen (34%).

Tabelle 15. Anteil fehlender Werte zur Zwischenerhebung

	Fehlend	p
Seitenweise Vorgabe	0.44%	.99
Itemweise Vorgabe	0.60%	
Änderungen möglich	0.61%	.93
Änderungen nicht möglich	0.17%	
Rückmeldung an Patienten	0.64%	.12
Keine Rückmeldung an Pat.	0.41%	
Computergestützt gesamt	0.53%	.01
Papier/Bleistift	4.38%	
	+ 34%*	

*komplett fehlende Fragebögen

Tabelle 16 zeigt für die einzelnen Vorgabeformen den Prozentsatz an Patienten welche eine Antwort auf offene Fragen (Textfragen) gaben, sowie die durchschnittliche Länge der

Antwort falls eine gegeben wurde. Offene Fragen wurden bei itemweiser Vorgabe öfter beantwortet als bei seitenweiser (86% zu 66%; $p < 0.05$; χ^2 -Test) und ebenso öfter, wenn keine Änderungen möglich als wenn Änderungen möglich waren (89% zu 72%; $p < 0.05$; χ^2 -Tests).

Tabelle 16. Häufigkeit und durchschnittliche Länge von Antworten auf offene Fragen

	Antwort	p	Zeichen	p
Seitenweise Vorgabe	66%		64	
Itemweise Vorgabe	86%	<.01	72	.18
Fehlende Werte Erlaubt	74%		66	
Fehlende Werte nicht Erlaubt	78%	.55	71	.68
Änderungen möglich	72%		68	
Änderungen nicht möglich	89%	.01	71	.69
Rückmeldung an Patienten	72%		73	
Keine Rückmeldung an Pat.	80%	.13	64	.18

Um die Annahme zu testen, dass die Rückmeldung an Patienten die Kooperationsbereitschaft in der Testsituation verbessert, wurde der entsprechende Anteil fehlender Fragebögen und die Anzahl von Skalen zur Entlassung verglichen, welche aufgrund fehlender Werte nicht berechnet werden konnten. In Tabelle 17 ist der Anteil fehlender Bögen und nicht berechenbarer Skalen zur Entlassung angegeben. Beim Anteil fehlender Bögen zeigte sich kein Unterschied abhängig vom Vorhandensein der Rückmeldung, es fehlten 23% bzw. 24% der Bögen. Bei dem Anteil berechenbarer Skalen wurde ein Unterschied gefunden: bei Rückmeldung an Patienten konnten zur Entlassung weniger Skalen berechnet werden. Bei 51% aller Patienten, die zur

Zwischenerhebung eine Rückmeldung erhalten haben, und bei 37% aller Patienten, welche keine Rückmeldung erhalten haben, konnte zur Entlassung mindestens eine Skala nicht berechnet werden ($p < 0.05$, χ^2 -Test).

Tabelle 17. Anteil fehlender Werte zur Entlassung

	Bogen fehlend	p	Skalen nicht berechenbar	p
Rückmeldung an Patienten	24.3%		Alle berechenbar: 49% Eine: 37% Mehr als eine: 14%	
Keine Rückmeldung an Pat.	22.5%	.76	Alle berechenbar: 63% Eine: 30% Mehr als eine: 7%	.04

Die Tabellen 18 und 19 geben eine Übersicht der Häufigkeiten, mit welcher einzelne Themen in den Fragen danach, was besonders gestört oder gefreut hatte (Fragen 6 und 7 des Zusatzfragebogens), angesprochen wurden. Die Kategorisierung der Antworten erfolgte nach inhaltlichen Überlegungen. Die Auswertung zeigt, dass insgesamt deutlich mehr positive Kommentare als negative gegeben wurden. Dies gilt insbesondere auch für die Kommentare, welche sich auf die Computerbefragung beziehen¹⁴. Häufig wurden die Betreuung bei der computergestützten Fragebogenvorgabe, die Arbeit am Computer an sich, die einfache Handhabung und die schnelle Bearbeitung lobend hervorgehoben. Einige Patienten saßen zum

¹⁴ Anfänglich war die Überlegung, dass sich diese Fragen ausschließlich auf die Befragungssituation beziehen sollen, wofür sie aber offensichtlich ungeschickt formuliert waren. Da viele der Bemerkungen auf die so formulierten Fragen jedoch für die Qualitätssicherung relevant waren, wurden diese so beibehalten.

ersten Mal vor einem Computer und drückten ihre Zufriedenheit damit aus, dass sie daran zurechtgekommen sind. Die häufigsten Kritikpunkte betrafen die Unmöglichkeit vorhergehende Antworten zu Korrigieren und unzutreffende Fragen.

Tabelle 18. Häufigkeit einzelner Antwortkategorien zu der Frage „Was hat sie besonders gestört?“

		Anzahl
	Keine Angabe	102
	Nichts	267
Unabhängig vom Computer	Fragen	51
	Umgebung, Therapie	26
	Sonstiges	10
Computerbefragung allgemein	Software, Hardware	11
	Arbeitsplatz, Organisation	8
	Computer	4
	Keine freien Kommentare möglich	4
		$\Sigma=27$
Darstellungsvarianten	Itemweise Darstellung	1
	Keine fehlenden Werte	1
	Keine Korrekturen	13

Tabelle 19. Häufigkeit einzelner Antwortkategorien zu der Frage „Was hat sie besonders gefreut?“

		Anzahl
	Keine Angabe	149
	Nichts	97
Unabhängig vom Computer	Fragen	17
	Umgebung, Therapie	98
	Qualitätssicherung	17
	Sonstiges	34
Computerbefragung allgemein	Handhabung Software	21
	Schnell / Kurz	23
	Spaß	6
	Ruhe	5
	Computer	24
		$\Sigma=79$
Darstellungsvarianten	Auswertung	7
	Itemweise Darstellung	2

Testökonomie

Tabelle 20 zeigt die durchschnittlichen Ausfüllzeiten und die durchschnittlichen Itemresponsezeiten sowie die Spannweite und die Ergebnisse der entsprechenden Vergleichstests für alle Vorgabemodalitäten. Obwohl die Verteilungen eindeutig linksschief sind, sind Mittelwerte angegeben, da diese für die Frage der Testkosten das geeignete Maß darstellen. Mit Ausnahme der entscheidungsorientierten Vorgabe waren die Bearbeitungszeiten zwischen den computergestützten Vorgabevarianten vergleichbar (p zwischen 0.11 und 0.47; KS-Test) und lagen im Mittel bei 23-26 Minuten für die Gesamtbearbeitungszeit mit 13-14 Sekunden pro Item. Relativ große, wenn auch nicht statistisch signifikante Unterschiede, gab es bei der Gesamtzeit zwischen den Vorgabevarianten „Änderungen möglich“ und „Änderungen nicht möglich“ (24 Minuten zu 26 Minuten) und bei der Itemresponsezeit zwischen seitenweiser und itemweiser Vorgabe (13 Sekunden zu 15 Sekunden). Die kürzeste Testzeit betrug 8 Minuten, alle Patienten scheinen sich also ausreichend Zeit für den Fragebogen genommen zu haben¹⁵. Die längste Testzeit betrug 95 Minuten, wobei dies deutlich von den übrigen Testzeiten abweicht, 99% der Patienten brauchten weniger als 60 Minuten.

¹⁵ die Zeit, welche nötig ist, um den Fragebogen einfach durchzuklicken beträgt ca. 2 Minuten, alle Fragen und Instruktionstexte zu lesen dauert ca. 7 Minuten

Tabelle 20. Durchschnittliche Zeiten für das Beantworten des gesamten Fragebogens und durchschnittliche Zeiten zwischen zwei Items

	Gesamtzeit (Min.)	p	Itemresponsezeit (Sec.)	p
Seitenweise Vorgabe	24 (10...95)		13 (4...58)	
Itemweise Vorgabe	24 (9...73)	0.19	15 (4...39)	0.13
Fehlende Werte erlaubt	24 (12...95)		14 (4...58)	
Fehlende Werte nicht erlaubt	24 (9...73)	0.23	14 (4...42)	0.25
Änderungen möglich	24 (9...95)		14 (4...58)	
Änderungen nicht möglich	26 (10...52)	0.11	14 (6...29)	0.11
Rückmeldung an Patienten	25 (9...95)		14 (4...58)	
Keine Rückmeldung an Pat.	23 (10...57)	0.47	13 (4...43)	0.39
Entscheidungsorientiert	16 (5...55) [3*]		11* (4...36)	
Handbuchgemäß	23 (8...84) [8*]	>0.01	7* (3...20)	>0.01

*Zeiten für EB45 & GBB allein

Bei entscheidungsorientierter Vorgabe wurden trotz längerer Bearbeitungszeiten pro Item (11 statt 7 Sekunden, $p < 0.05$; KS-Test) deutlich kürzere Gesamtbearbeitungszeiten gefunden (im Mittel 16 statt 23 Minuten, $p < 0.05$; KS-Test), da die Anzahl vorgegebener Items bei entscheidungsorientierter Vorgabe von 69 Items auf im Mittel 18 Items sank. Die Verteilung der Anzahl vorgegebener Items bei entscheidungsorientierter Testvorgabe ist auf Abbildung 10 dargestellt.

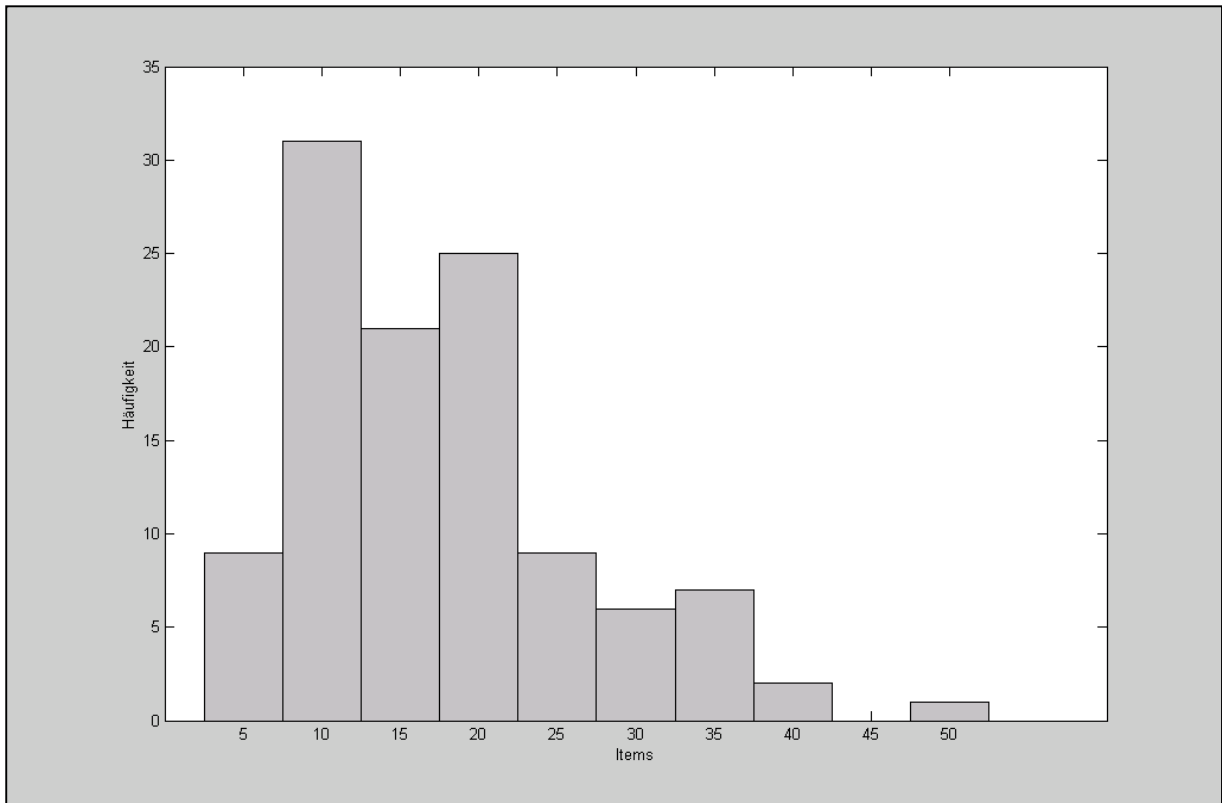


Abbildung 10. Verteilung der Anzahl vorgegebener Items bei entscheidungsorientierter Testvorgabe

Zum Vergleich der Erhebungskosten von Papier/Bleistift-Fragebogenvorgabe (mit nachträglicher Dateneingabe von Hand oder über Scanner) und von computergestützter Fragebogenvorgabe, wurde auf die Erfahrungen aus dieser Studie und die bisherigen Erfahrungen der FS mit Dateneingabe zurückgegriffen. Die entsprechenden Schätzungen sind in Tabelle 21 dargestellt. Sie gelten für eine QS-Erhebung nach dem Stuttgart-Heidelberger Modell. Nicht mit aufgenommen sind Kosten, die in allen Fällen in gleicher Höhe anfallen (z.B. Lizenzgebühren für die Testverfahren). Nicht verglichen wurde die Papier/Bleistift Anwendung ohne Auswertung über Rechner, da der Personalaufwand für die Skalenberechnung, Berechnung

von Qualitätsindikatoren und die Darstellung der Ergebnisse weit über der für die Dateneingabe liegt. Ferner wurde nicht zwischen entscheidungsorientierter Vorgabe und kompletter Vorgabe sowie zwischen Online-Befragung und Offline-Befragung unterschieden, da die erwarteten finanziellen Auswirkungen nur gering sind.

Tabelle 21. Annahmen zur Schätzung der Testkosten

	Scanner	Dateneingabe von Hand	Vorgabe am PC
Hardware	7000€ / 4 Jahre / 1000 Patienten	1000€ / 4 Jahre / 600 Patienten	1000€ / 4 Jahre / 250 Patienten
Software	8000€ / 4 Jahre / 1000 Patienten	2000€ / 4 Jahre	3000€ / 4 Jahre
Betriebskosten	20% (Hardware + Software) / Jahr	20% (Hardware + Software) / Jahr	20% (Hardware + Software) / Jahr
Personalkosten (Betreuung, Dateneingabe)	20 min. = 3€ / Patient	40 min. = 6€ / Patient	10 min. = 1.5€ / Patient
Sonstiges (Druckkosten)	40 Seiten * 0.05€ / Patient	40 Seiten * 0.05€ / Patient	

Computergestützte Testdarbietung erwies sich schon bei kleinen Patientenzahlen als das kostengünstigste Verfahren, die Eingabe über Scanner durchgehend als das teuerste (siehe Abbildung 11).

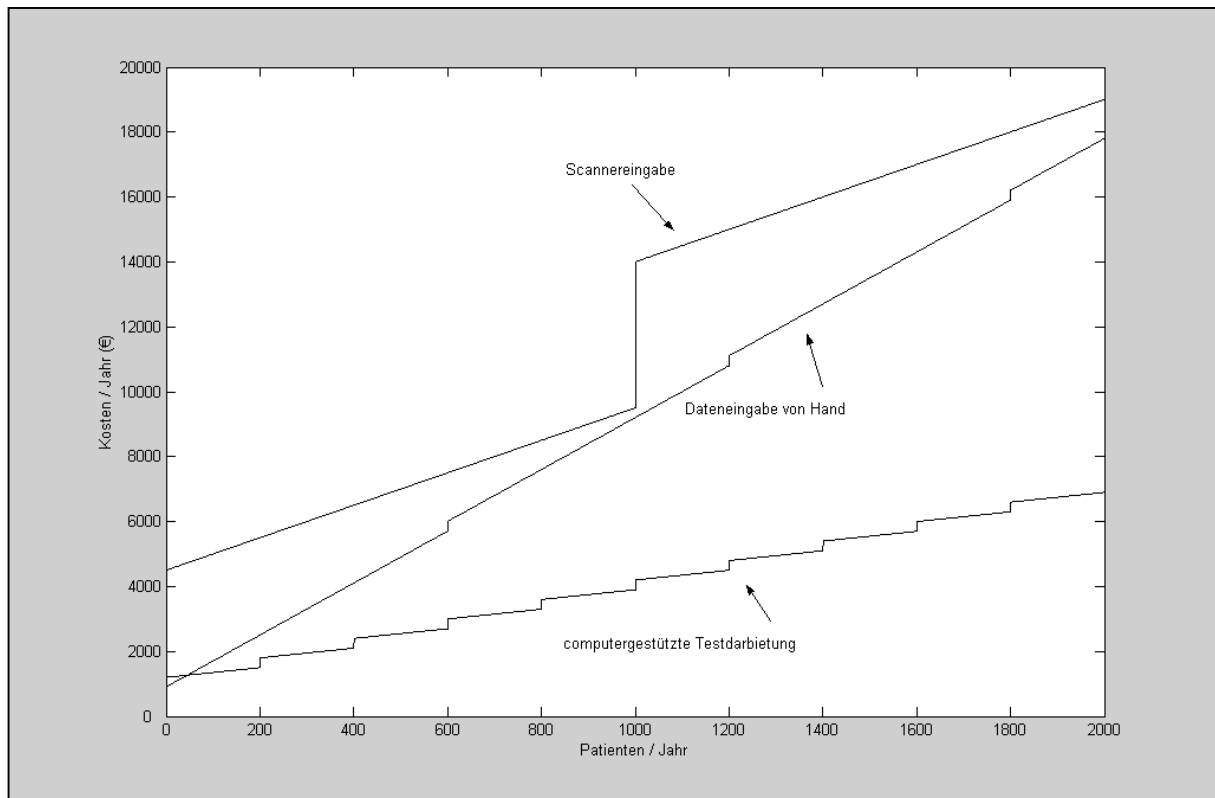


Abbildung 11. Geschätzte Befragungskosten

Effekte der Rückmeldungen auf die Behandlung

Tabelle 22 zeigt die Dauer der Weiterbehandlung nach der Zwischenerhebung aufgeschlüsselt nach dem Ergebnis zur Zwischenerhebung und dem Vorliegen einer Rückmeldung. Es wurde ein Zusammenhang zwischen der Dauer der Weiterbehandlung nach der Zwischenerhebung und dem erreichten Zustand zur Zwischenerhebung gefunden ($p < 0.05$, GLM). Es wurde kein Zusammenhang zwischen der Dauer der Weiterbehandlung und der Rückmeldungen ($p > 0.46$, GLM) sowie keine signifikante Interaktion zwischen dem erreichten Zustand zur Zwischenerhebung, der Dauer der Weiterbehandlung und der Rückmeldungen gefunden ($p > 0.27$; GLM). Im Hinblick auf die therapeutische Arbeitsbeziehung zur Entlassung zeigte sich

ebenfalls kein Einfluss der Rückmeldungen, weder der Therapeutenrückmeldung noch der Patientenrückmeldung. Dies gilt sowohl aus Sicht der Patienten (PTBEZ, Therapeutenrückmeldung: $p=0.42$, Patientenrückmeldung: $p=0.68$; KS-test) wie auch aus Sicht der Therapeuten (TTBEZ, Therapeutenrückmeldung: $p=0.29$, Patientenrückmeldung: $p=0.64$; KS-test).

Tabelle 22. Dauer der Weiterbehandlung nach der Zwischenerhebung (in Tagen)

Ergebnis zur Zwischenerhebung	Rückmeldung an Therapeut	Keine R. an Therapeut	Rückmeldung an Patient	Keine R. an Patient
Grün	14 (SD=5.4)	9 (SD=3.4)	14 (SD=6.8)	15 (SD=3.6)
Schwarz	20 (SD=9.2)	19 (SD=8.9)	20 (SD=8.8)	20 (SD=9.7)
Rot	24 (SD=9.2)	27 (SD=14.4)	24 (SD=7.7)	25 (SD=10.5)
p	Ergebnis: $>.01$ Rückmeldung: $.58$ Interaktion: $.27$		Ergebnis: $>.01$ Rückmeldung: $.46$ Interaktion: $.76$	

Ein positiver Effekt der Rückmeldung von Zwischenerhebungswerten auf den Therapieerfolg deutet sich dennoch an. Für diese Auswertung wurde Überlebensanalyse verwendet, um die Therapiedauer als wichtigen Faktor für den Therapieerfolg in die Schätzung mit einzubeziehen und die Sensitivität der Auswertung zu steigern, indem bei Fällen die vor Erreichen einer klinisch bedeutsamen Besserung entlassen wurden zumindest die beobachtete Therapiedauer zu berücksichtigen. Bei der Schätzmethode erschien bei der gegebenen feinen Gradierung der Zeit von einem Tag Kaplan

Maier das geeignete Verfahren. Da globale Unterschiede in den Überlebensfunktionen von Interesse waren, war der Log Rank Test für deren Vergleich geeignet.

Therapieerfolg wurde als klinisch bedeutsame Besserung auf dem EB45-TOT operationalisiert, mit funktionalen Bereich unter 64 Punkten und reliabler Veränderung bei mehr als 6 Punkten (siehe Lambert & Finch, 1999). Abbildung 12 zeigt die entsprechenden Überlebensfunktionen. Die geschätzte durchschnittliche (Median) Behandlungsdauer bei einer Entlassung in klinisch bedeutsam gebessertem Zustand betrug ohne die Rückmeldung an Therapeuten 57 Tage und mit Rückmeldung 46 Tage ($p=0.01$; Kaplan Maier, Log Rank Test). Die Rückmeldung an Patienten zeigt dagegen keinen signifikanten Einfluss ($p=0.64$; Kaplan Maier, Log Rank Test). Die geschätzte Behandlungsdauer stellt die maximale Zeit bis zum Erreichen klinisch bedeutsamer Besserung dar, da nicht der genaue Zeitpunkt dieses Ereignisses bekannt ist, sondern nur das Ergebnis zur Entlassung. Die problematische Annahme der Unabhängigkeit zwischen Zensierung (d.h. Entlassung ohne klinisch bedeutsame Besserung zu zeigen) und Gruppenzugehörigkeit (d.h. Rückmeldung oder keine Rückmeldung) wird in der Diskussion thematisiert.

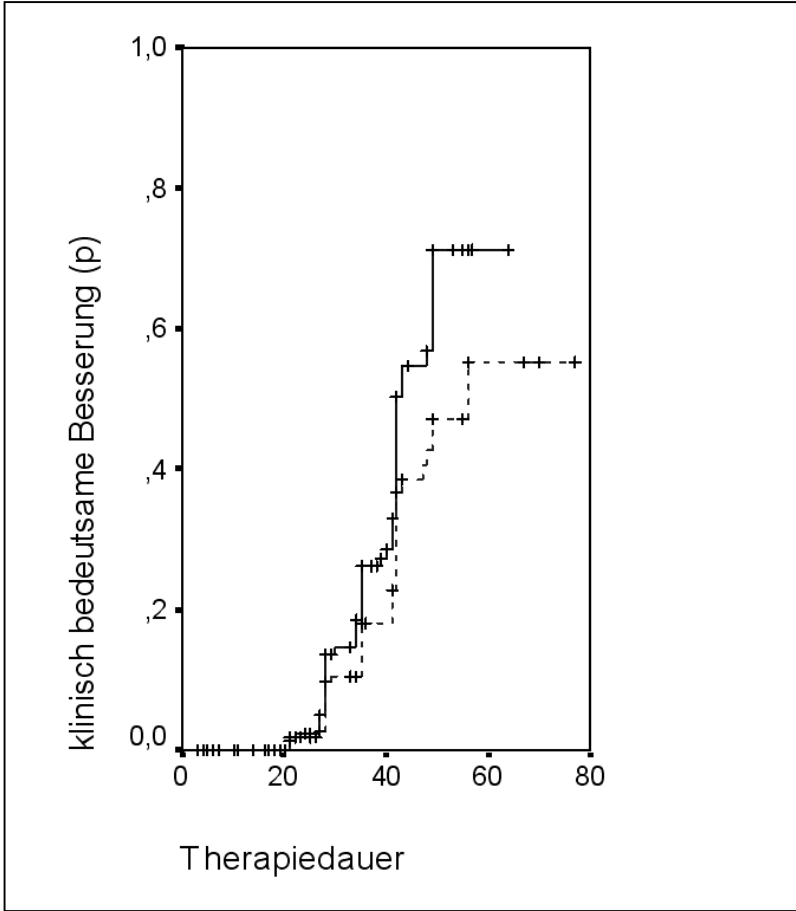


Abbildung 12. Behandlungsdauer in Tagen bis zur Entlassung in klinisch bedeutsam gebesserten Zustand bei Rückmeldung an Therapeuten (-) und ohne Rückmeldung an Therapeuten (--)
 [Entlassung ohne klinisch bedeutsame Veränderung (+)]

DISKUSSION

Psychometrische Äquivalenz

Bezüglich der Rohwertverteilungen und der Kennwerte zur Reliabilität und Validität wurden keine Unterschiede von praktisch relevanter Größenordnung zwischen den einzelnen Vorgabemodalitäten (mit Ausnahme der entscheidungsorientierten Vorgabe) gefunden. Der Befund psychometrischer Vergleichbarkeit von Papier/Bleistift Vorgabe und computergestützter Vorgabe deckt sich mit der Mehrzahl der Befunde zu Selbstbeurteilungsverfahren. Eine Generalisierbarkeit über alle Tests dieser Art ist damit zwar nicht nachgewiesen, liegt aber nahe, zumindest für Items, die sich auf eine Zeitperspektive von mehreren Tagen und länger beziehen. Die computergestützten Vorgabeformen wurden sorgfältig gestaltet, so dass sie ein möglichst komfortables Arbeiten bei größtmöglicher Übereinstimmung mit dem Papier/Bleistift-Layout gewährleisten. Es kann nicht ausgeschlossen werden, dass anders gestaltete Testdarbietungen einen Einfluss auf das Antwortverhalten haben, doch deuten die vorliegenden Ergebnisse auf eine hohe Robustheit des Antwortverhaltens gegenüber der Vorgabeform hin.

Die Grenzen der Robustheit von Selbstbeurteilungsverfahren gegenüber Vorgabemodalitäten und Gestaltungsaspekten deuten sich möglicherweise bei den Ergebnissen zu den Skalen ZUF8 und PTBEZ an. Bei einzelnen Items dieser Skalen wurden zwar

geringe, aber systematische Verteilungsunterschiede zwischen computergestützter Testvorgabe und Papier/Bleistift-Vorgabe gefunden: Bei computergestützter Darbietung wurde weniger Zufriedenheit mit der Behandlung und mit der therapeutischen Arbeitsbeziehung angegeben. Dafür könnte zum einen die häufig berichtete größere Offenheit bei computergestützten Testen verantwortlich sein (vgl. Robinson & West, 1992; Hofer, 1985). Zum anderen wurde bei der Umsetzung der Skala PTBEZ auf computergestützte Vorgabe stärker von der Papier/Bleistift-Vorlage abgewichen als bei den übrigen Skalen. Die Papier/Bleistift-Vorlage sieht zu jedem Item ein Feld vor, in welches die Beurteilung als Zahl zwischen -3 und +3 eingetragen wird. Bei der Umsetzung auf computergestützte Vorgabe wurden, entsprechend dem Antwortformat bei den übrigen Skalen sowie um den Patienten die Zahleneingabe zu ersparen, Ankreuzfelder (Radio-Buttons) verwendet.

Bei entscheidungsorientierter Vorgabe verändert die Schätzung der Skalenwerte über Regressionsgleichungen und die Orientierung der Testlänge am Trennwert wie erwartet die Verteilung und hat Einfluss auf die Reliabilität und damit auch auf die Außenkorrelationen. Zudem deuten etwas geringere Zusammenhänge als erwartet bei den Interkorrelationen der entscheidungsorientiert vorgegebenen Instrumenten auf ein geringfügig verändertes Antwortverhalten. Für den gedachten Zweck, der Bestimmung klinisch signifikanter Veränderung, ergeben sich bei entscheidungsorientierter Vorgabe jedoch

keine schwerwiegenden Einschränkungen. Fehlklassifikationen sind in ca. 5% der Fälle zu erwarten, wobei diese Fälle typischerweise nahe am erwarteten Zielbereich liegen. Bei weitergehender Nutzung entscheidungsorientiert erhobener Daten müssen die Unterschiede in der Verteilung, der Reliabilität und den Korrelationen aber berücksichtigt werden. Die kürzere Befragung ist dann gegen die Einbußen bei der Reliabilität und den Außenkorrelationen sowie die Verteilungsverzerrungen abzuwägen. In der Versorgung könnte sich ein Modell bewähren, bei dem Routinemäßig eine entscheidungsorientierte Testung erfolgt, bei größerer Diskrepanz zwischen klinischen Eindruck und Testwerten jedoch die Möglichkeit gegeben ist, die Testung mit handbuchgemäßer Vorgabe zu wiederholen.

Akzeptanz

Wie schon in zahlreichen Studien belegt (vgl. Klinck, 2002), ist die Akzeptanz des computergestützten Testens auch in der vorliegenden Studie hoch. Die meisten Patienten kommen nach kurzer Einführung gut am Rechner zurecht, und würden bei der nächsten Befragung lieber wieder am Rechner arbeiten, statt Papier/Bleistift Bögen auszufüllen. Bei den offenen Bemerkungen wird die Computereingabe als angenehmer, spaßiger und schneller bezeichnet. Diese Präferenz für computergestützte Testvorgabe im Vergleich zur Papier/Bleistift-Vorgabe ermutigt, sie auch in der Routine einzusetzen. Da die Akzeptanz weitgehend unabhängig von der

Gestaltung der computergestützten Vorgabe ist, steht es frei, diese anhand anderer Kriterien, z.B. technischer Überlegungen, zu wählen. Dass dies nur eingeschränkt gilt, zeigt die Erfahrung mit der Vorgabeform „keine Änderungen möglich“. Diese fiel bei den freien Kommentaren negativ auf. Wo keine nachträglichen Änderungen erlaubt waren, d.h. bei der entsprechenden Vorgabeform im ersten Studienabschnitt und im gesamten zweiten Studienabschnitt, ist auch die beobachtete Präferenz für computergestützte Testdarbietung am niedrigsten (auch wenn die Unterschiede hinsichtlich der Präferenz nicht statistisch signifikant wurden).

Demographische Variablen sind nicht mit der Bevorzugung einer bestimmten Vorgabeform verbunden und lassen sich somit nicht für eine präferenzbezogene Zuordnung zu computergestützter Vorgabe bzw. Papier/Bleistift Vorgabe nutzen. Da auch die psychometrischen Eigenschaften unabhängig von der Bevorzugung sind, scheint es aus psychometrischer Sicht nicht notwendig, Patienten aufgrund ihrer Präferenz alternative Formen anzubieten. Dies bleibt dann eine Entscheidung zwischen Praktikabilität und dem Eingehen auf spezielle Wünsche der Patienten. Höheres Alter und geringere Computererfahrung erschweren allerdings das Zurechtkommen am Rechner, erhöhen die Anstrengung und verlängern die Eingabedauer. Das kann bei der Ausgestaltung des computergestützten Testens berücksichtigt werden, etwa indem älteren und computerunerfahrenen Patienten eine längere

Einarbeitung gegeben wird und für diese mehr Zeit am Rechner reserviert wird.

Die insgesamt niedrige Rate fehlender Werte bei computergestützter Testvorgabe im Vergleich zur Papier/Bleistift Vorgabe ist bemerkenswert. Inwieweit dabei die computergestützte Vorgabe als solche und inwieweit der Hinweis auf unbeantwortete Fragen eine Rolle spielen, bleibt offen. Die Unterschiede bei den offenen Fragen deuten auf geringe Einflüsse der jeweiligen computergestützten Vorgabemodalität. Bei seitenweiser Vorgabe scheint es einfacher, diese Fragen zu übersehen oder zu ignorieren, als bei itemweiser Vorgabe.

Überraschend ist, dass sich bei entscheidungsorientierter Vorgabe trotz deutlich kürzerer Bearbeitungszeit keine höhere Akzeptanz zeigt. Tendenziell ist eher das umgekehrte der Fall. Das kann daran liegen, dass in den ersten Wochen bei entscheidungsorientierter Testvorgabe mehreren Patienten aufgrund eines Softwarefehlers der Rechner während der Bearbeitung abstürzte. Ungeachtet dessen, scheint die Bearbeitungszeit, in dem Umfang wie sie bei der Zwischenerhebung anfällt, keinen kritischen Akzeptanzfaktor darzustellen.

Bei keinem Patienten wurden Bearbeitungszeiten gefunden, die darauf hindeuten, dass der gesamte Fragebogen oder einzelne Fragen beantwortet wurden, ohne diese überhaupt gelesen zu haben. Bei einer Anwendung von Testverfahren in Kontext der

stationären QS scheint eine Kontrolle solchen Antwortverhaltens nicht nötig zu sein.

Erfreulich ist die hohe Akzeptanz des Verfahrens bei den Therapeuten und dem Pflorgeteam. Nachdem die Datenerhebung für diese Studie beendet wurde, votierten diese für eine Fortsetzung der computergestützten Fragebogenvorgabe und das weitere Erhalten der Rückmeldungen.

Testökonomie

Aus Sicht der Patienten und der Ressourcenausnutzung ist die Dauer der Erhebung das relevante Maß für die Testökonomie. Zwar stehen keine Zeiten für Papier/Bleistift-Befragung zur Verfügung, in den freien Kommentaren deutet sich jedoch an, dass die computergestützte Testdarbietung schneller ist. Dies kann zwar daran liegen, dass die Patienten als Vergleich die längere Aufnahmeerhebung hatten, es deckt sich jedoch mit Angaben aus der Literatur und persönlichen Eindrücken.

Dass die Zeiten für alle Vorgabeformen, mit Ausnahme der entscheidungsorientierten Vorgabe, vergleichbar sind, ist überraschend. Speziell zwischen itembasierter und seitenbasierter Darbietung waren Unterschiede schon aus technischen Gründen zu erwarten (ca. 2 Sec. für den Bildschirmaufbau). Ein Unterschied in dieser Größenordnung wurde im Mittel pro Item auch beobachtet, ist jedoch bei den gegebenen Verteilungen nicht groß genug, um statistisch signifikant zu werden. Interessanterweise verlor sich dieser

Zeitunterschied pro Item bei der Gesamtzeit der Befragung komplett. Bei seitenweiser Darstellung ist es einfacher durch den Fragebogen zu navigieren und Patienten scheinen diese Möglichkeit öfter zu nutzen, was entsprechend Zeit kostet. Wo keine Änderungen möglich waren, wurde trotz gleichbleibender Itemresponsezeit eine längere Gesamtzeit beobachtet (wenn der Unterschied auch nicht statistisch signifikant wurde): Einige Patienten haben Korrekturen versucht, wo diese nicht möglich waren und beschwerten sich während der Testbearbeitung bei der betreuenden Arzthelferin; Dies verlängerte deren Testzeiten entsprechend.

Eindeutig sind die Unterschiede bei entscheidungsorientierter Testvorgabe. Die meisten Patienten müssen deutlich weniger Fragen beantworten. Obwohl die durchschnittliche Zeit pro Item zunimmt, ist die durchschnittliche Gesamtzeit damit niedriger. Die längere Responsezeit pro Item liegt wohl daran, dass die ersten Fragen diskriminierender sind und somit eine längere Entscheidungszeit fordern. Dass Items, welche dem EB45 und GBB zugeordnet sind, insgesamt schneller beantwortet werden als die übrigen, liegt wohl daran, dass in diesem Fragebogenteil wenig Instruktionstexte zu lesen und keine offenen Fragen zu beantworten sind.

Aus Betreibersicht gehören zur Frage der Testökonomie nicht nur die Testzeiten, auch Betriebskosten und Datenqualität gehen darin ein. Die Frage der Betriebskosten konnte hier

nicht detailliert geklärt werden, diese hängt zu sehr von dem konkreten Einsatz ab. Ein approximativer Vergleich bot sich aber an und wurde für den Einsatz in der QS nach dem Stuttgart-Heidelberger Modell durchgeführt. Computergestützte Testdarbietung ist schon bei kleinen Patientenzahlen das kostengünstigste Verfahren. Die Eingabe über Scanner scheint durchgehend teurer zu sein als die Eingabe von Hand. Andererseits resultiert sie in weniger Eingabefehlern (Percevic et al., 1999). Bei computergestützter Testdarbietung wird bei deutlich geringeren Kosten diese Fehlerquelle gänzlich umgangen und es werden weniger fehlende Werte provoziert.

Effekte der Rückmeldungen

Zunächst wichtig ist, dass die Rückmeldung an Patienten keine negativen Effekte auf die Behandlungsergebnisse, die therapeutische Arbeitbeziehung oder die Akzeptanz der Befragung zeigt. Nach der anfänglichen Modifikation der Rückmeldung sind auch die offenen Bemerkungen dazu ausnahmslos positiv. Jedoch scheinen entgegen der Annahme, dass die Rückmeldung die Kooperationsbereitschaft erhöht, eher mehr fehlende Werte bei der Entlassungsbefragung die Folge zu sein. Die Bemerkungen einiger Patienten legen die Vermutung nahe, dass dies möglicherweise daran liegt, dass durch die Rückmeldung klarer wird, welche Informationen der Therapeut aus ihren Antworten entnehmen kann und dass er dabei

möglicherweise etwas entdeckt, was sie nicht offenbaren wollen (z.B. negative Gefühle gegenüber dem Therapeuten). Da keine eindeutigen positiven Effekte auf den Therapieverlauf zu erkennen sind, und die automatisierte Rückmeldung immer ein gewisses Risiko der Übersimplifizierung und Fehlinterpretation beinhaltet, scheint es ratsam, diese Form von Patientenrückmeldung nicht weiter anzubieten. Andere Formen der Rückmeldung z.B. solche, bei denen die Therapeuten die Testergebnisse mit ihren Patienten besprechen, könnten sich als sinnvoller erweisen.

Das Design dieser Studie war nicht dahingehend ausgerichtet, einen Effekt der Rückmeldung an Therapeuten zu untersuchen. Dennoch wurde diese Frage in einer explorativen Weise angegangen. Dabei hat sich gezeigt, dass Patienten, zu welchen Therapeuten Rückmeldungen erhalten, zwar eher in einen klinisch bedeutsam gebesserten Zustand entlassen werden, dies jedoch wohl nicht auf den erwarteten Effekt der Rückmeldung, nämlich einer verstärkten Adaptation der Therapiedauer aufgrund der Rückmeldung beruht. Tendenziell ist eher das Umgekehrte der Fall: Patienten welche als Farbcode „grün“ erhielten, wurden bei Rückmeldung länger behandelt und Patienten welche als Farbcode „rot“ erhielten, wurden bei Rückmeldung kürzer behandelt als ohne Rückmeldung. Warum zeigt sich nicht der erwartete Effekt? In dieser Studie wurden die Rückmeldungen in einer suboptimalen Weise gegeben. Mit nur einer Zwischenerhebung konnte das hinter dem

therapiebegleitenden Ergebnismonitoring liegende Konzept, die Behandlung so lange durchzuführen, bis der Patient das Therapieziel nachweisbar erreicht hat, nur teilweise nachgebildet werden. Die Therapeuten erhielten nur eine kurze Einführung in die Interpretation der Rückmeldung. Ob, wie und inwieweit sie diese nutzten, nutzen konnten und nutzen wollten, blieb ihnen überlassen und wurde nicht kontrolliert. Dass Therapeuten an einer weitergehenden Schulung interessiert waren, zeigt sich u.a. an den Anfragen nach Interpretationen der Rückmeldung und dem Wunsch, die graphische Darstellung durch die verbale zu ergänzen (obwohl diese ja keine zusätzlichen Informationen beinhaltet). Das Anbieten von Rückmeldungen ohne entsprechende Korrekturvorschläge hat sich in vielen Bereichen als nicht effektiv erwiesen (siehe z.B. Fiedler, 1983; Yates, 1990) und inwieweit die hier angebotenen Korrekturvorschläge (Therapie verkürzen bzw. verlängern) aus klinikinternen, organisatorischen Gründen befolgt werden konnten, ist fraglich. Möglicherweise konnte bei der knappen Einführung in die Interpretation der Rückmeldung und deren empirische Rechtfertigung die verbreitete Annahme, das früher „nonresponse“ prognostisch ungünstig für den weiteren Therapieverlauf ist, nicht nachhaltig genug geändert werden, oder die Darlegung der Einschränkungen dieser Empfehlung hat den gegenteiligen Effekt hervorgerufen. Möglicherweise hat eine wahrgenommene „Bevormundung“ durch den Computer bei einem Teil der Therapeuten psychologische Reaktanz hervorgerufen.

Dennoch konnte ein positiver Einfluss auf den Therapieerfolg beobachtet werden. Dieser beruht möglicherweise darauf, dass Therapeuten den Patienten, zu denen sie Rückmeldungen erhalten, mehr Aufmerksamkeit schenken. Möglicherweise ließe sich dieser Effekt gezielter einsetzen, wenn Rückmeldungen nur zu „problematischen“ (roter Farbcode) Patienten gegeben würden. Nicht in dieses Bild passt allerdings, dass sich keine Einflüsse auf die therapeutische Arbeitsbeziehung zeigen. Falls die Annahme fallengelassen wird, dass „Zensurierung“, also das Verlassen der Therapie ohne klinisch bedeutsame Besserung zu zeigen, unabhängig von der Rückmeldung ist, bietet sich eine alternative Interpretation an: bei gegebener Rückmeldung erkennt der Therapeut eher welche Patienten von einer Weiterbehandlung nicht mehr profitieren werden (zumindest nicht von einer Weiterbehandlung mit den Mitteln und der Länge, wie sie in dieser Klinik üblich ist). Diese Patienten würden dann früher entlassen werden. Beide der ausgeführten Interpretationen favorisieren die Rückmeldung an Therapeuten. Ein direkter Einfluss des vorausgegangenen computergestützten Testens im Vergleich zum Papier/Bleistift-Testen kann zwar nicht ausgeschlossen werden, es scheint jedoch nicht plausibel, dass das Therapieergebnis tatsächlich auf diese Weise derart zu beeinflussen ist. Übertragungseffekte der Testvorgabeformen kommen als Erklärung ebenfalls kaum in Betracht, wenn schon beim direkten Vergleich computergestützte Testvorgabe und Papier/Bleistift Testvorgabe äquivalent sind.

Interessanterweise zeigt sich in der gesamten Stichprobe eine deutliche Orientierung der Therapiedauer am Ergebnis. Dies ist besonders erstaunlich, da in den Jahren vor dieser Studie kein solcher Effekt zu erkennen war. Möglicherweise kam es zu einer Übertragung der Ergebnisorientierung auf alle Patienten. Allerdings können auch andere, organisatorische Änderungen, die in dieser Zeit aufgetreten sind, u.a. der Übergang auf feste Kostenpauschalen für einzelne diagnostisch relevante Gruppen, diesen Effekt verursacht haben.

ZUSAMMENFASSUNG UND AUSBLICK

Diese Studie trägt zur Klärung des Einflusses der Gestaltung computergestützter Testdarbietung, des Einflusses personenbezogener Unterschiede sowie der praktischen Umsetzung entscheidungsorientierter Strategien bei. Die vergleichsweise große klinische Stichprobe und das Panel-Designs mit 3 Erhebungen sowie die häufig eingesetzten, bisher jedoch im Hinblick auf computergestützte Vorgabe ungeprüften Testverfahren EB45 und GBB ergänzen die vorhandene Literatur zu Äquivalenzfragen bei Selbstbeurteilungsverfahren.

Einschränkend gilt, dass die Vorgabevariante, bei welcher keine Änderungen möglich waren, vorzeitig abgesetzt wurde. Die vorgesehene Stichprobengröße wurde daher nicht erreicht und Stichprobenvergleichbarkeit ist möglicherweise nicht gegeben. Weiter ist der Vergleich der computergestützten Vorgabe mit der Papier/Bleistift Vorgabe wegen des relativ hohen Anteils fehlender Zwischenerhebungen bei Papier/Bleistift Vorgabe nur unter der zwar plausiblen aber nicht weiter überprüfbaren Annahme gesichert, dass die Beeinträchtigung zur Zwischenerhebung nicht mit fehlenden Bögen zusammenhängt. Ebenfalls problematisch beim Vergleich der computergestützten Vorgabe mit der Papier/Bleistift-Vorgabe ist die Ungenauigkeit mit welcher bei Papier/Bleistift-Vorgabe der Zeitpunkt der Zwischenerhebung bestimmt werden konnte. Die Stichprobengröße für den Vergleich der Reliabilität und der Validität war

eventuell zu klein, da die Stichprobenplanung nicht auf diese Frage ausgerichtet war. Ferner konnte das Problem steigender α -Fehler bei multiplen Vergleichen nur durch die Suche nach Mustern und Trends berücksichtigt werden.

Trotz dieser Einschränkungen deuten die Befunde dieser Studie darauf hin, dass Selbstbeurteilungsverfahren im hohen Maße robust gegenüber der Vorgabeform sind. Im Hinblick auf die Entwicklung und den praktischen Einsatz computergestützten Testens ist der Befund weitgehender psychometrischer Vergleichbarkeit verschiedener Darbietungsformen beruhigend. Aus testökonomischen Gründen, sowie wegen der Präferenz von Seiten der Patienten und der geringeren Zahl fehlender Werte, ist computergestütztes Testen vorzuziehen.

Vor dem Hintergrund einer vermehrten Umsetzung von psychometrischen Instrumenten für die Vorgabe über das Internet ist insbesondere der Befund relevant, dass HTML-Elemente für eine psychometrisch und erfahrungsbezogen angemessene Umsetzung geeignet sind. Bei umfangreichen Testbatterien oder im Fall von häufigen Messungen haben die Befunde zu entscheidungsorientierten Testen praktische Relevanz. Entscheidungsorientiertes Testen kann sinnvoll eingesetzt werden, um bei kontrollierten Einfluss auf die Entscheidungsgüte die Erhebungsdauer deutlich zu kürzen.

Entwicklung eines computergestützten Monitoringsystems (Web-AKQUASI)

Die Studie gibt Hinweise für die Entwicklung und den Einsatz computergestützter Testvorgabe in der klinischen Routineversorgung. Am Beispiel der Entwicklungen von „Web-AKQUASI“ soll dies gezeigt werden.

Web-AKQUASI ist eine Software zur Unterstützung von ergebnisorientierter Qualitätssicherung und therapiebegleitenden Ergebnismonitoring in der Psychotherapie und psychosomatischen Medizin. Die Software ermöglicht computergestützte Testvorgabe, Verlaufsbewertungen und zeitnahe Ergebnisrückmeldung und ist auf Basis von Internettechnologie entwickelt. Die Eingabeformulare für Web-AKQUASI wurden ausgehend von den hier gemachten Erfahrungen gestaltet. HTML-Standardelemente haben sich bewährt und wurden daher beibehalten. Es ist möglich, den gleichen Fragebogen flexibel aufzurufen, z. B. seitenbasiert für das „klassische“ Internet und Item für Item bei mobilen Endgeräten mit kleinem Bildschirm. Dieses Vorgehen ist von psychometrischer Seite dadurch legitimiert, dass diese Vorgabeformen sich in der vorliegenden Studie als äquivalent gezeigt haben. Es kann festgelegt werden, welche Items beantwortet werden müssen und welche nicht. Da schon ein Hinweis auf nichtbeantwortete Fragen ausreicht, um die Anzahl fehlender Werte klein zu halten, werden die Patienten nur bei entscheidenden Fragen zur Beantwortung „gezwungen“. Aufgrund der schlechten Erfahrungen

mit der Vorgabevariante, bei der kein nachträgliches Ändern der Antworten möglich war, ist bei Web-AKQUASI die Navigation durch den Fragebogen und das nachträgliche Abändern schon gegebener Antworten grundsätzlich gegeben. Sofortige Rückmeldungen an Patienten sind technisch machbar, da diese in der vorliegenden Studie jedoch keinen Nutzen brachten, manche Patienten aber möglicherweise verunsichern, gehen Rückmeldungen im Regelfall über den Therapeuten. Viele Patienten stört es, dass bei computergestützter Fragebogenvorgabe im Vergleich zur Papier/Bleistift-Vorgabe keine Kommentare zu den Fragen angefügt werden können. Daher beinhaltet der Fragebogen zu Ende der Befragung ein ausreichend großes „Kommentarfeld“¹⁶.

Items werden unabhängig voneinander definiert und unter Zuhilfenahme entsprechender XML-Definitionen zu Fragebogen zusammengefasst und verrechnet. Durch die Trennung von Items, Vorgabeform und Verrechnung ist Web-AKQUASI für eine Weiterentwicklung in Richtung „Itembank“ und „adaptive Testvorgabe“ vorbereitet. Die Ergebnisse dieser Studie rechtfertigen entscheidungsorientierte Testvorgabe. Die erwarteten Einsparungen bei der Itemanzahl können benutzt werden, um z.B. die Erhebungsdomäne zu erweitern. So können neben den Symptomen weitere therapierrelevante Bereiche erhoben werden, z.B. persönliche und soziale Ressourcen des Patienten.

¹⁶ Die 160 Zeichen welche in der berichteten Studie angeboten waren, reichten den Patienten häufig nicht.

Ferner bietet Web-AKQUASI Möglichkeiten für Testverzweigungen. So ist z. B. das detaillierte Abfragen von Suizidrisiko bei Angabe häufiger Suizidgedanken oder die Vorgabe von diagnosespezifischen Instrumenten zu realisieren. Damit kann der häufigste Kritikpunkt der Patienten im Bezug auf die Fragebogen, dass Fragen gestellt werden, welche auf ihre momentane Situation nicht zutreffen, wenigstens teilweise umgangen werden.

Als technisches Problem haben sich in der vorliegenden Studie die recht häufig nötigen Softwareupdates, Einstellungsarbeiten und Hardwareprobleme erwiesen. Probleme mit nicht gegebener Abwärtskompatibilität der eingesetzten Software (Betriebssystem, Browser, Datenbank) lassen sich nur durch kontinuierliche Weiterentwicklung und Betreuung lösen. Diese kann über einen „remote“ Zugriff auf den Server erleichtert werden, wobei auf die Datensicherheit zu achten ist. Ausgehend von den gesetzlichen Datenschutzverordnungen (vgl. <http://www.datenschutz.de>) sowie den entsprechenden Richtlinien der psychologischen Berufsverbände und medizinischen Standards (vgl. Reng, 2002) wurde Web-AKQUASI mit einer Reihe von Sicherheitsmerkmalen ausgestattet, u.a. verschlüsselter Datenübertragung und Datenablage, passwortgeschützten abgestuften Datenzugriff, Protokollierung des Zugriffs und individuell konfigurierter Firewall. Bei der Hardware muss die Zuverlässigkeit im Vordergrund stehen, was aber nicht das Erproben vielversprechender neuer Technologien,

z.B. von portablen Touchscreenrechnern („Webpads“; vgl. Fahrenberg, Leonhart, Foerster, 2002) oder SMS-basierter Befragung (vgl. Bauer et al., 2003), ausschließen soll.

Organisatorisch hat sich das Heranführen der Patienten an den Rechner durch das Klinikpersonal bewährt und soll beibehalten werden¹⁷. Wichtig ist ferner den Patienten das ungestörte Arbeiten am Rechner zu ermöglichen und den Arbeitsplatz ergonomisch zu gestalten. Beim Routineeinsatz von Web-AKQUASI werden inzwischen an mehreren Kliniken wöchentliche Erhebungen vorgenommen. Dabei hat sich gezeigt, dass bei Beachtung dieser Hinweise im Routinebetrieb ca. 90% der geplanten Erhebungen tatsächlich durchgeführt werden – ein Wert der bei Papier/Bleistift-Vorgabe nicht erreicht wird (Gallas, Percevic & Kordy, 2004). Bei wiederholter Dateneingabe ist es in den meisten Fällen ausreichend, den Patienten nur bei der ersten Dateneingabe zu betreuen, und die Eingabe erfolgt schneller (vgl. Freudenmann & Spitzer, 2001). Im ambulanten Bereich oder für Nachuntersuchungen können Patienten mit Internetanschluss auch von zu Hause aus den Fragebogen ausfüllen. Internetbenutzer verfügen in der Regel über das notwendige Wissen, sollten aber dennoch eine kurze Einführung bei ihrem Therapeuten bekommen. Entsprechende Erfahrungen wurden inzwischen im Rahmen mehrerer Studien zur

¹⁷ Viele Patienten erwähnen in ihren Kommentaren die Einführung von Seiten des Klinikpersonals lobend. Diese hat sicher auch Einfluss auf die hohe Akzeptanz des Verfahrens und das gute Zurechtkommen der Patienten mit der Software.

Nachbetreuung gesammelt (Golkaramnay et al., 2003; Wolf, Arikan & Kordy, 2004).

Die mangelnde Umsetzung der Rückmeldungen von Seiten der Therapeuten in dieser Studie zeigt, dass die Therapeuten entsprechend geschult und motiviert werden müssen um Web-AKQUASI als effektives Werkzeug für kontinuierliche Therapieadaptation einzusetzen. Sie brauchen Richtlinien zum Einsatz der Rückmeldung und müssen zumindest in der Anfangsphase bei der Umsetzung dieser Richtlinien begleitet werden. Inzwischen liegen Beobachtungen vor, welche die Bedeutung dieser Faktoren unterstreichen. An der Sonnenbergklinik Stuttgart wurde das Befinden der Patienten nach ca. dem ersten Drittel der Behandlung an das Therapeutenteam zurückgemeldet. Im Unterschied zu der vorliegenden Studie wurden diese Rückmeldungen in Qualitätszirkeln diskutiert. Als Ergebnis sank die durchschnittliche Behandlungsdauer für Patienten welche schon zum Zeitpunkt der Rückmeldung in einem guten Zustand waren deutlich, ohne das Einbussen am Behandlungsergebnis zur Entlassung gefunden wurden (Bauer et al., 2004). Dieser Befund hebt sich besonders vor dem Hintergrund hervor, dass in einer vorherigen siebenmonatigen Phase, mit Qualitätszirkeln welche das Therapieergebnis bereits entlassener Patienten zum Inhalt hatten, keine Effekte gefunden wurden. An der Panorama Fachklinik waren die Ergebnisse der vorliegenden Studie ausführlich mit den Therapeutenteam diskutiert worden. Nach

der darauffolgenden Einführung von Web-AKQUASI mit wöchentlichen Ergebnismonitoring konnte eine Zunahme der Orientierung der Behandlungsdauer am erreichten Behandlungsergebnis, ein schnellerer Rückgang der Beschwerden der Patienten, eine tendenziell kürzere durchschnittliche Behandlungszeit sowie insgesamt höhere Raten klinisch bedeutsamer Veränderung beobachtet werden (Gallas, Percevic & Kordy, 2004). Diese vielversprechenden Hinweise auf einen Nutzen therapiebegleitender Rückmeldung für die Behandlungs- und Versorgungsplanung lassen weitere Untersuchungen als lohnend erscheinen.

Erweiterung der Fragestellung

Durch die primär angewandte Ausrichtung dieser Studie wurde der aus theoretischer Sicht wichtige Bezug, die Interaktion zwischen Testanforderungen, Darstellungsform und Personeneigenschaften, nicht systematisch untersucht. Eine entsprechende Facettierung (im Sinne von Borg, 1992) und Auswertung (siehe z.B. Cattell, 1966 und Wittmann, 1985 für Datenanalysekonzepte) bleibt ein Anliegen für künftige Studien und soll vor allem die Möglichkeiten (und Grenzen) der Generalisierung über unterschiedliche Fragebögen, Darstellungsformen und Anwendergruppen klären.

Schon bei der Vorbereitung dieser Studie wurde deutlich, dass für eine theoretische Fundierung der Testgestaltung und der Auswertung von Itemresponsezeiten der (psychologische)

Prozess der Testbearbeitung zwar wichtig, die Befundlage dazu aber dürftig ist. Idealerweise würde eine „psychologische Testtheorie“ (als Gegenstück zu den meist rein formalen Testtheorien) zu wünschen sein (siehe Ansätze dazu bei Holden et al., 1992; Schwarz, 1991; Rogers, 1974). Computergestütztes Testen ermöglicht das genaue und einfache Erfassen von Responsezeiten und Reaktionssequenzen, welche sich als empirisches Material zum Erarbeiten und Testen einer solchen Theorie anbieten. In Verbindung mit internetbasierten Experimenten (siehe dazu z.B. Batinic et al., 1999, speziell den Beitrag von Reips) ergibt sich als vielversprechende Forschungsstrategie das Testen von alternativen Klassen formaler Modelle kognitiver Prozesse anhand von Responsezeiten (vgl. Townsend & Ashby, 1983). Responsezeiten und Reaktionssequenzen wurden auch in der vorliegenden Studie miterhoben, deren Auswertung erfolgte allerdings zunächst vorrangig unter testökonomischen Gesichtspunkten.

Die klassische Testtheorie ist in der Anwendung immer noch am weitesten verbreitet. Auch diese Studie fügt sich in diesen Rahmen ein. Die Grenzen davon sind u.a. beim entscheidungsorientierten Testen deutlich geworden, welches zu Lasten der Messgenauigkeit geht. Dies stärkt den Wunsch nach einem Instrument, welches in einer sinnvollen Weise adaptiv vorgegeben werden kann. Dafür ist allerdings die Itembasis zu erweitern und ein entsprechendes probabilistisches Testmodell zu entwickeln und zu überprüfen. Dafür nötige

multidimensionale, polyvalente und dynamische Ansätze sind in der Literatur ausführlich dargestellt (z. B. Segall, 1996; van der Linden & Hambleton, 1997; Meiser et al., 1998; Müller, 1999; vgl. Steyer et al., 1991 für die Entwicklung probabilistischer Modelle zur Befindlichkeitsmessung). Entsprechende Itemsammlungen liegen ebenfalls vor (siehe z.B. Goldberg, 1999 für eine breite, frei verfügbare Itemsammlung).

Die Vielzahl offener Fragen und erfolgsversprechender Ansätze zur formalen Testgestaltung (d.h. „wie soll etwas erfasst werden“) soll nicht den Anwendungskontext, d.h. die Fragen „was soll erfasst werden“ und „wie soll es verwendet werden“, vergessen lassen. Im psychotherapeutischen Umfeld sind der diagnostische Urteilsprozess, der Therapieprozess und der Therapieverlauf die entscheidenden Faktoren, welche weiter untersucht werden müssen (vgl. Percevic, Lambert & Kordy, 2004). Die Forderung nach einer Evaluation der Wirksamkeit von QS-Maßnahmen im Gesundheitswesen begründet einen weiteren Forschungsschwerpunkt (vgl. Sachverständigenrat für die konzentrierte Aktion im Gesundheitswesen, 2000/2001). Die vorliegende Studie sowie die weiteren hier angesprochenen Untersuchungen zu Effekten eines therapiebegleitenden Ergebnismonitorings (Bauer et al., 2004; Gallas, Percevic & Kordy, 2004) können als erste Schritte in Richtung einer Umsetzung dieser Forderung verstanden werden. Die Möglichkeit, dass Erfahrungstransfer bei den beobachteten Effekten eine Rolle spielt, macht es erforderlich bei künftigen

Untersuchungen zwischen einer Randomisierung auf Patientenebene und einer Randomisierung auf Therapeutenebene zu unterscheiden. Ebenfalls wichtig scheint bei künftigen Studien eine Kontrolle unspezifischer Effekte (z.B. von Aufmerksamkeitseffekten), eine Erweiterung der Ergebniskriterien auf solche die nicht unmittelbar in der Rückmeldung angesprochen werden (siehe Evans, 2004; vgl. Wittmann & Schmidt, 2002), sowie die Einbeziehung von Kosten-Nutzen Aspekten (vgl. Miller & Magruder, 1999; Wittmann, 1990). Die aktuelle Debatte um die Einführung von Ergebnismonitoring in die ambulante Psychotherapie (z.B. Fydrich et al., 2003; Tschuschke, 2003) unterstreicht die Relevanz entsprechender Studien. Web-AKQUASI mit seinen in der psychotherapeutischen Routine bewährten Möglichkeiten zur computergestützten Testvorgabe und Rückmeldung kann dabei als Forschungswerkzeug wichtige Dienste leisten.

LITERATUR

- Agency for Health Care Policy and Research (1995). *Using clinical practice guidelines to evaluate quality of care. Volume I: Issues*. Rockville: U.S. Department of Health and Human Services.
- Alexander, L. B., & Luborsky, L. (1986). The Penn Helping Alliance Scales. In L. S. Greenberg (Hrsg.), *The Psychotherapeutic Process - A research Handbook*, pp. 325-366. New York: Guilford.
- American Psychological Association (1999). *Standards for educational and psychological testing*. Washington: American Educational Research Association.
- Bader, P., Hofmann, K., & Kubinger, K. D. (1993). Zur Brauchbarkeit der Normen von Papier-Bleistift-Tests für die Computer-Vorgabe: Ein Experiment am Beispiel des Gießen-Tests. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 14, 129-135.
- Bamert, T., & Heidingsfelder, M. (2001). Designeffekte in Online-Umfragen. In A. Theobald, M. Dreyer, & Starsetzki, T. (Hrsg.). *Online-Marktforschung: Theoretische Grundlagen und praktische Erfahrung*, pp. 165-178. Wiesbaden: Gabler.
- Barak, A., & English, N. (2002). Prospects and limitations of psychological testing on the internet. *Journal of Technology in Human Services*, 19, 65-89.

- Bassili, J. N., & Scott, B. S. (1996). Response latency as a signal to question problems in survey research. *Public Opinion Quarterly*, 60, 390-399.
- Batinic, B., Werner, A., Gräf, L., & Bandilla, W. (1999). *Online Research: Methoden, Anwendungen und Ergebnisse*. Göttingen: Hogrefe.
- Bauer, S., Haug, S., Percevic, R., & Kordy, H. (2004). Kontinuierliche Ergebnisrückmeldung im Alltagsbetrieb einer Psychosomatischen Fachklinik. Vortrag auf der 55. Jahrestagung des DKPM, Geldern, März 2004. [Abstract in Psychotherapie Psychosomatik Medizinische Psychologie, 54, p. 99].
<http://www.psyres-stuttgart.de/index.php/article/articleview/104/1/7>
- Bauer, S., Percevic, R., Okon, E., Meermann, R. & Kordy, H. (2003). Use of text messaging in the aftercare of patients with bulimia nervosa. *European Eating Disorders Review*, 11, 279-290.
- Beutler, L. E. (1989). Differential treatment selection: The role of diagnosis in Psychotherapy. *Psychotherapy*, 26, 271-281.
- Bicanich, E., Slivinski, T., Hardwicke, S. B., & Kapes, J. T. (1997). Internet-based testing: A vision or reality? *Tehnology Horizons in Education Journal*.
<http://www.thejournal.com/magazine/vault/A1918.cfm>.
- Biskin, B. H., & Kolotkin, R. L. (1977). Effects of computerized administration on scores on the Minnesota Multiphasic

- Personality Inventory. *Applied Psychological Measurement*, 1, 543-549.
- Bloxom, B. (1989). Adaptive testing: A review of recent results. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 10, 1-17.
- Booth, J. F. (1990). Computerdiagnostik: Errungenschaften und Möglichkeiten der Technologie. *Diagnostica*, 36, 180-203.
- Borg, I. (1992). *Grundlagen und Ergebnisse der Facettentheorie*. Bern: Huber.
- Bortz, J., & Lienert, G. A. (2003). *Kurzgefasste Statistik für die klinische Forschung*. Berlin: Springer.
- Bowker, D., & Dillman, D. A. (2000). An Experimental Evaluation of Left and Right Oriented Screens for Web Questionnaires. Vortrag auf der 55. Annual Conference der American Association for Public Opinion Research, Portland, Mai 2000.
<http://survey.sesrc.wsu.edu.dillman/papers/AAPORpaper00.pdf>.
- Brähler, E., & Scheer, J. W. (1997). Testinformation: Der Gießener Beschwerdebogen (GEB): 2. ergänzte und revidierte Auflage. *Diagnostica*, 43, 380-382.
- Broda, M., Dahlbender, R. W., Schmidt, J., von Rad, M., & Schors, R. (1993). DKPM-Basisdokumentation. Eine einheitliche Basisdokumentation für die stationäre Psychosomatik und Psychotherapie. *Zeitschrift für Psychotherapie, Psychosomatik und Medizinische Psychologie*, 43, 214-223.

- Brown, J. S., & Burton, R. R. (1978). Diagnostic models for procedural bugs in mathematical skills. *Cognitive Science*, 2, 155-192.
- Buchanan, T. (1999). Online Personality assessment: Equivalence of traditional and WWW personality measures.
<http://citeseer.ist.psu.edu/274113.html>.
- Buchanan, T., & Smith, J. L. (1999). Using the Internet for psychological research: Personality testing on the World-Wide-Web. *British Journal of Psychology*, 90, 125-144.
- Bukasa, B., Kisser, R., & Wenninger, U. (1990). Computergestützte Leistungsdiagnostik bei verkehrspsychologischen Eignungsuntersuchungen. *Diagnostica*, 36, 148-165.
- Butcher, J. N. (1985). Introduction to the special series. *Journal of Consulting and Clinical Psychology*, Vol. 53, 746-747.
- Butcher, J. N., & Pancheri, P. (1976). *Handbook of crossnational MMPI research*. Minneapolis: University of Minnesota Press.
- Butcher, J. N., Keller, L. S., & Bacon, S. F. (1985). Current developments and future directions in computerized personality assessment. *Journal of Consulting and Clinical Psychology*, Vol. 53, 803-815.
- Butcher, J. N., Perry, J., & Hahn, J. (2004). Computers in Clinical Assessment: Historical Developments, Present Status and Future Challenges. *Journal of Clinical Psychology*, 6, 331-345.

- Byrne, B. M. (1989). *A Primer of LISREL*. New York: Springer.
- Carter, J. E., & Wilkinson, L. (1984). A latent trait analyses of the MMPI. *Multivariate Behavioral Research, 19*, 385-407.
- Cates, J. A. (1999). The art of assessment in psychology: Ethics, expertise, and validity. *Journal of Clinical Psychology, 55*, 631-641.
- Cattell, R. B. (Hrsg.) (1966). *Handbook of multivariate experimental psychology*. Chicago: Rand McNally.
- Christian, L. M. (2003). *The Influence of Visual Layout on Scalar Questions in Web Surveys*. Unpublished Master Thesis. Washington State University.
- Clavelle, P. R., & Butcher, J. N. (1977) An adaptive typological approach to psychiatric screening. *Journal of Consulting and Clinical Psychology, 45*, 851-859.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, Hillsdale: Erlbaum.
- Cohen, R. C., & Moreland, K. L. (1992). Computer-Assisted Psychological Assessment. In R. J. Cohen (Hrsg.): *Psychological testing and assessment. An introduction to test & measurement*, pp. 707-738. Mountain View: Mayfield Publishing Company.
- Cronbach, L. J., Gleser, G. C., & Rajaratnam, N. (1963). Theory of generalizability: A liberalization of reliability theory. *British Journal of Statistical Psychology, 16*, 137-163.
- Dillman, D. A., Tortora, R. D., & Bowker, D. K. (1998a). *Principles for Constructing Web surveys*. SESRC Technical

Report 98-50.

<http://survey.sesrc.wsu.edu/dillman/papaers/websurveyppr.pdf>.

Dillman, D. A., Tortora, R. D., Conradt, J., & Bowker, D. (1998b). Influence of plain vs. fancy design of response rates for web surveys. Vortrag auf dem *Joint Statistical Meeting*, Dallas, August 1998.

<http://survey.sesrc.wsu.edu/dillman/papers/asa98ppr.pdf>.

Döge, M. (1997). *Intranet: Einsatzmöglichkeiten, Planung, Fallstudien*. Köln: O'Reilly Verlag.

Donebian, A. (1986). Criteria and Standards for quality assessment and monitoring. *Quality Rev. Bull.*, 12, 99-108.

Dubois, P. (1970). *A history of psychological testing*. Boston: Allyn and Bacon.

Düker, M. S., & Schneider, J. F. (1977). Untersuchungen zum Beantwortungsprozess bei psychodiagnostischen Fragebogen. *Zeitschrift für experimentelle und angewandte Psychologie*, 24, 282-302.

Dunn, T. G., Lushene, R. E., & O'Neil Jr., H. F. (1972). Complete automation of the MMPI and a study of its response latencies. *Journal of Consulting and Clinical Psychology*, 39, 381-387.

Dyck, J. L., Gee, N. R., & Smither, J. A. (1999). The changing construct of computer anxiety for younger and older adults. *Computers in Human Behavior*, 14, 61-77.

Eggen, T. J. H. M., & Straetmans, G. J. J. M. (2000). Computerized adaptive testing for classifying examinees into

- three categories. *Educational and Psychological Measurement*, 60, 713-734.
- Elwood, D. J. (1969). Automation of psychological testing. *American Psychologist*, 24, 287-289.
- Erdman, H. P., Klein, M. H., & Greist, J. H. (1985). Direct patient computer interviewing. *Journal of Consulting and Clinical Psychology*, 53, 760-773.
- Esser, C., & Schneider, J. F. (1998). Differentielle Reaktionslatenzzeiten beim Bearbeiten von Persönlichkeitsfragebogen als möglicher Indikator für Verfälschungstendenzen. *Zeitschrift für experimentelle und angewandte Psychologie*, 19, 246-257.
- Evans, C. (2004). A cautious clinician-researcher's view of within-therapy use of outcome monitoring. Vortrag auf dem *SPR 35th Annual Meeting, Rom, Juni 2004*.
- Fahrenberg, J. (1987). Zur psychophysiologischen Methodik: Konvergenz, Fraktionierung oder Synergismen? *Diagnostica*, 33, 272-287.
- Fahrenberg, J., Leonhart, R., & Foerster, F. (2002). *Alltagsnahe Psychologie*. Bern: Huber.
- Fahrenberg, J., Myrtek, M., Wilk, D., & Kreutel, K. (1986). Multimodale Erfassung der Lebenszufriedenheit: Eine Untersuchung an Herz-Kreislauf Patienten. *Psychotherapie, Psychosomatik medizinische Psychologie*, 36, 347-354.
- Falmagne, J.-C., Koppen, M., Villano, M., Doignon, J.-P. & Johannesen, L. (1990). Introduction to knowledge spaces: How

to build, test and search them. *Psychological Review*, 97, 201-224.

Farell, A. D. (1991). Computers and behavioural assessment: Current applications, future possibilities, and obstacles to routine use. *Behavioral Assessment*, 13, 159-179.

Farin, E., & Bengel, J. (2003) Qualitätssicherung, Evaluationsforschung und Psychotherapieforschung: Abgrenzung und Zusammenwirken. In: M. Härter, W. Linster, R. D. Stieglitz (Hrsg). *Qualitätsmanagement in der Psychotherapie*, pp. 47-70. Göttingen: Hogrefe.

Fiedler, K. (1983). Diagnostische Fähigkeiten und diagnostische Erfahrung. In: Jäger, R S, Mattenklott, A, Schröder, R-D. (Hrsg.). *Diagnostische Urteilsbildung in der Psychologie. Studien zur Pädagogischen Psychologie, Band 20*, pp. 303-327. Göttingen: Hogrefe.

Finger, M. S., & Ones, D. S. (1999). Psychometric equivalence of the computer and booklet forms of the MMPI: A meta-analysis. *Psychological Assessment*, 11, 58-66.

Finn, S., & Martin, H. (1997). Therapeutic assessment with the MMPI-2 in managed care settings. In J. N. Butcher (Hrsg.), *Personality assessment in managed health care*, pp. 131-152. New York: Oxford University Press.

Föderation Deutscher Psychologinnenvereinigungen (1986). Richtlinien für den Einsatz elektronischer Datenverarbeitung in der psychologischen Diagnostik. *Psychologische Rundschau*, 37, 163-165.

- Fowler, R. D. (1985). Landmarks in computer-assisted psychological assessment. *Journal of Consulting and Clinical Psychology, 53*, 748-759.
- Franke, G. (1995). *Die Symptom Checkliste von Derogatis - Deutsche Version (SCL-90-R)*. Göttingen: Beltz Test.
- Franke, G. H. (1998). *Computerunterstützte klinisch-psychodiagnostische Selbstbeurteilungsverfahren im Äquivalenztest*. Lengerich: Pabst.
- Franke, G. H. (1999). Effekte der Computeradministration bei der Symptom-Checkliste (SCL-90-R) unter besonderer Berücksichtigung der Itemreihenfolge. *Diagnostica, 45*, 147-153.
- Freudenmann, R. W. & Spitzer, M. (2001). Computergestützte Patientenbefragungen als Grundlage für eine moderne Qualitätssicherung in der Psychiatrie. *Nervenarzt, 72*, 40-51.
- Fulgosi, A. (1988). *Faktorska Analiza*. Zagreb: Skolska Knjiga.
- Fydrich, T., Nagel, A., Lutz, W., & Richter, R. (2003). Qualitätsmonitoring in der ambulanten Psychotherapie: Modellprojekt der Techniker Krankenkasse. *Verhaltenstherapie, 13*, 291-295.
- Gallas, C., Percevic, R., Kordy, H. (2004). Putting continuous outcome monitoring into practice: Acceptance, feasibility and effects of within therapy feedback. Vortrag auf dem *SPR 35th Annual Meeting, Roma, Juni 2004*. <http://www.psyres-stuttgart.de/index.php/article/articleview/104/1/7>

- Garb, H. N. (2000). Computers will become increasingly important for psychological assessment: Not that there's anything wrong with that! *Psychological Assessment*, 12, 31-39.
- Gass, C. S., & Brown, M. C. (1992). Neuropsychological test feedback to patients with brain dysfunction. *Psychological Assessment*, 4, 272-277.
- Gerlach, F. M. (2001). *Qualitätsförderung in Praxis und Klinik: ein Chance für die Medizin*. Stuttgart: Thieme.
- Gilbert, A. R. (1986). Measuring response latency - A crucial innovation of paper and pencil testing. *Psychologia*, 29, 108-111.
- Gitzinger, I. (1990). Akzeptanz der Darbietung eines Tests auf dem Personalcomputer von stationären Patienten/-innen. *Psychother. med. Psychol.*, 40, 143-145.
- Goldberg, L. R. (1999). *A Broad-Bandwidth, Public-Domain, Personality Inventory Measuring the Lower-Level Facets of Several Five-Factor Models*.
<http://ipip.ori.org/newBroadbandText.htm>.
- Golkaramnay, V., Wangemann, T., Dogs, J., Dogs, P., & Kordy, H. (2003). Neue Brücken für Lücken in der psychotherapeutischen Versorgung durch das Internet: Hoffnungen, Herausforderungen und ein Lösungsansatz. *Psychother. med. Psychol*, 53, 399-405.
- Gräf, L., & Heidingsfelder, M. (1999). Bessere Datenqualität bei WWW-Umfragen - Erfahrungen aus einem Methodenexperiment mit dem Internet-Rogator. In B. Batinic, A. Werner, L. Gräf, & W.

- Bandilla (Hrsg.). *Online Research - Methoden, Anwendungen und Ergebnisse*, pp. 21-38. Göttingen: Hogrefe, .
- Groth-Marnat, G. (2000). Visions of clinical assessment: Then, now, and a brief history of the future. *Journal of Clinical Psychology, Vol. 56 (3)*, 349-365.
- Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment, 12*, 19-30.
- Hager, W. (2000). About some misconceptions and the discontent with statistical tests in psychology. *Methods of Psychological Research Online, 5/1*.
- Handel, R. W., Ben-Porath, Y. S., & Watt, M. (1999). Computerized adaptive assessment with the MMPI-2 in a clinical setting. *Psychological Assessment, 12*, 19-30.
- Hannöver, W. (2002). *Vorhersage von Behandlungsergebnissen stationärer Psychotherapie*. Achen: Shaker Verlag.
- Hänsgen, K. D. (1999). Computereinsatz in der Psychodiagnostik - Stand und mögliche Perspektiven. *Forschungsbericht Nr. 141. Psychologisches Institut der Universität Freiburg*.
<http://www.unifr.ch/ztd/archiv/Forschung%20Uni%20Computerdiagnostik.pdf>.
- Hänsgen, K. D., Bernasconi, M. (2000). Die Situation der Psychodiagnostik in der Schweiz. Ergebnisse einer Befragung im Auftrage der FSP. *Forschungsbericht Nr. 146. Psychologisches Institut der Universität Freiburg*.

- Harell, T. H., Honacker, L. M., Hetu, M. & Oberwager, J. (1987). Computerized versus traditional administration of the Multidimensional Aptitude Battery-Verbal Scale: An examination of reliability and validity. *Computers in Human Behaviour*, 3, 129-137.
- Härter, M., Linster, W., & Stieglitz, R. D. (2003), *Qualitätsmanagement in der Psychotherapie*. Göttingen: Hogrefe.
- Hawkins, E. J., Lambert, M. D., Vermeersch, D. A., & Slade, K. (2004). The therapeutic effects of providing client progress information to clients and therapists. *Psychotherapy Research*, 14, 308-327.
- Hawkins, E., & Lambert, M. (2002) Outcome Monitoring and Feedback: Does this Help to Optimize Inpatient Treatment? Vortrag auf dem 33th Annual Meeting der SPR, Santa Barbara, Juni 2002.
- Hedl, J. J, O'Neal, H. F., & Hansen, D. N. (1973). Affective reactions toward computer-based intelligence testing. *Journal of Consulting and Clinical Psychology*, 40, 217-222.
- Heuft, G., & Senf, W. (1998). *Praxis der Qualitätssicherung in der Psychotherapie: Das Manual zur Psy-BaDo*. Stuttgart: Georg Thieme Verlag.
- Hofer, P. B., & Green B. F. (1985). The challenge of competence and creativity in computerized psychological testing. *Journal of Consulting and Clinical Psychology*, 53, 826-838.

- Hofer, P. J. (1985). Developing standards for computerized psychological testing. *Computers in Human Behavior*, 1, 301-315.
- Holden, R. R., Kroner, D. G., Fekken, C. G., & Popham, S. M. (1992). A model of personality test item dissimulation. *Journal of Personality and Social Psychology*, 63, 272-279.
- Honaker, L. M. (1988). The equivalence of computerized and conventional MMPI administration: A critical review. *Clinical Psychology Review*, 8, 561-577.
- Hornke, L. F. (1983). Computerunterstütztes Testen - eine bewertende empirische Untersuchung. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 4, 323-344.
- Hornke, L. F. (1993). Mögliche Einspareffekte beim computergestützten Testen. *Diagnostica*, 39, 109-119.
- Hornke, L. F. (1997). Untersuchung von Itembearbeitungszeiten beim computergestützten adaptiven Testen. *Diagnostica*, 43, 27-39.
- Hornke, L. F. (2001). Benötigte Itemanzahlen beim meß- und entscheidungsorientierten adaptiven Testen. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 22, 185-193.
- Hornke, L., & Winterfeld, U. (2004). *Eignungsbeurteilung auf dem Prüfstand: DIN 33430 zur Qualitätssicherung*. Heidelberg: Spektrum Akademischer Verlag.
- Horowitz, L., Strauß, B., & Kordy, H. (2000). *Inventar zur Erfassung Interpersonaler Probleme - Deutsche Version. Manual*. Göttingen: Beltz Test.

- Houben, A. (1970). Psychodiagnostik in psychotherapeutischen und psychosomatischen Institutionen. In: Schraml, W. J. *Klinische Psychologie*, Bern: Huber.
- Jacobson, N. S., Roberts, L. J., Berns, S. B., & McGlinchey, J. B. (1999). Methods for defining and determining the clinical significance of treatment effects: Description, application, and alternatives. *Journal of Consulting and Clinical Psychology*, 67, 300-307.
- Jäger, R. S., & Krieger, W. (1994). Zukunftsperspektiven der computerunterstützten Diagnostik, dargestellt am Beispiel der treatmentorientierten Diagnostik. *Diagnostica*, 40, 217-243.
- Johnson, J. H., & Williams, T. A. (1980). Using on-line computer technology in a mental health admitting system. In J. B. Sidowski, J. H. Johnson, & T. A. Williams (Hrsg.), *Technology in mental health care delivery systems*, pp. 237-249. Norwood: Ablex.
- Joinson, A. (1999). Social desirability, anonymity, and Internet-based questionnaires. *Behavior Research Methods, Instruments & Computers*, 31, 433-438.
- Jones, M. B., Kennedy, R. S., & Bittner, A. C. (1981). A video game for performance testing. *American Journal of Psychology*, 94, 143-152.
- Jöreskog, K. G. (1978). Structural analyses of covariance and correlation matrices. *Psychometrika*, 43, 443-477.

- Kleinmuntz, B., & McLean, R. S. (1968). Computers in behavioral science: Diagnostic interviewing by digital computer. *Behavioral Science, 13*, 815-818.
- Klinck, D. (1998). Papier-Bleistift- versus computer-unterstützte Administration kognitiver Fähigkeitstests : Eine Studie zur Äquivalenzfrage. *Diagnostica, 44*, 61-70.
- Klinck, D. (2002). *Computergestützte Diagnostik*. Göttingen: Hogrefe.
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*, 254-284.
- Kolen, M. J. & Brennan, R. I. (1995). *Test equating: methods and practice*. New York: Springer.
- Kordy, H., & Hannöver, W. (1998). Beobachten, Dokumentieren, Bewerten, Steuern: Qualitätsmanagement in der stationären Psychotherapie. In: A. R. Laireiter, H. Vogel (Hrsg.), *Qualitätssicherung in der Psychotherapie und psychosozialen Versorgung*, pp. 355-399. Tübingen: DGVT.
- Kordy, H., & Hannöver, W. (2000). Die Evaluation von Psychotherapie und das Konzept der „Klinisch bedeutsamen Veränderung“. In A. R. Lairaiter (Hrsg.): *Diagnostik in der Psychotherapie*, pp. 289-294. Wien: Springer.
- Kordy, H., & Kächele, H. (1996). Ergebnisforschung in Psychotherapie und Psychosomatik. In: J. M. Adler, J. M. Herrmann, K. Köhler, O. W. Schoneke, T. von Uexhüll, W.

- Wesiach (Hrsg.), *Psychosomatische Medizin*, pp. 490-501.
München: Urban & Schwarzenberg.
- Kordy, H., & Lutz, W. (1995). Das Heidelberger Modell: Von der Qualitätskontrolle zum Qualitätsmanagement stationärer Psychotherapie. *Psychotherapie Forum*, 3, 197-206.
- Kordy, H., Hannöver, W. & Richard, M. (2001). Computer-Assisted Feedback-Driven Quality Management for Psychotherapy: The Stuttgart-Heidelberg Model. *Journal of Consulting and Clinical Psychology*, 69, 173-183.
- Krampen, G. (1993). Effekte von Bewerbungsinstruktionen und Subskalenextraktion in der Fragebogendiagnostik. *Diagnostica*, 39, 97-108.
- Krauth, J. (1995). *Testkonstruktion und Testtheorie*. Weinheim: Beltz.
- Kubinger, K. D., & Farkas, M. G. (1991). Die Brauchbarkeit der Normen von Papier-Bleistift-Tests für die Computer-Vorgabe: Ein Experiment am Beispiel der SPM von Raven als kritischer Beitrag. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 12, 257-266.
- Kuiper, N. A., MacDonald, N. R., & Derry, P. A. (1983). Parameters of a depressive self-schema. In J. Suls & A. G. Greenwald (Hrsg.), *Psychological perspectives on the self*, Vol. 2, pp. 191-217. Hillsdale: Erlbaum.
- Laireiter, A. R. (2000). *Diagnostik in der Psychotherapie*. Wien: Springer.

- Laireiter, A. R., & Vogel, H. (1998). *Qualitätssicherung in der Psychotherapie und psychosozialen Versorgung*. Tübingen: DGVT.
- Lambert, M. J., & Finch, A. E. (1999). The Outcome Questionnaire. In M. E. Maruish (Hrsg.). *The use of psychological testing for treatment planning and outcome assessment (2nd ed.)*, 831-869. Mahwah: Lawrence Erlbaum.
- Lambert, M. J., Hannöver, W., Nisslmüller, K., Richard, M., & Kordy, H. (2002a). Fragebogen zum Ergebnis von Psychotherapie: Zur Reliabilität und Validität der deutschen Übersetzung des Outcome Questionnaire 45.2 (OQ-45.2). *Zeitschrift für Klinische Psychologie und Psychotherapie*, 31, 40-47.
- Lambert, M. J., Hansen, N. B., & Finch, A. E. (2001a). Patient-focused research: Using patient outcome data to enhance treatment effects. *Journal of Consulting and Clinical Psychology*, 69, 159-172.
- Lambert, M. J., Whipple, J. L., Smart, D. W., Vermeersch, D. A., Nielsen, S. L., & Hawkins, E. J. (2001b). The effects of providing therapists with feedback on patient progress during psychotherapy: Are outcomes enhanced? *Psychotherapy Research*, 11, 49-68.
- Lambert, M. J., Whipple, J. L., Vermeersch, D. A., Smart, D. W., Hawkins, E. J., Nielsen, S. L., & Goates, M. (2002b). Providing Therapists with feedback on patient progress as a method of enhancing psychotherapy outcomes: A replication. *Clinical Psychology and Psychotherapy*, 9, 91-103.

- Lienert, G. A., Raatz, U. (1994). *Testaufbau und Testanalyse*. Weinheim: Beltz.
- Lumsden, J. A., Sampson, J. P., Reardon, R. C., Lenz, J. G. (2002). A Comparison Study of the Paper, Personal Computer (PC), and Internet Versions of Holland's Self-Directed Search. *Technical Report No. 30. The Center for the Study of Technology in Counseling and Career Development*.
- Lushene, R. E., O'Neal, O. F., & Dunn, T. (1974). Equivalent validity of a completely computerized MMPI, *Journal of Personality Assessment*, 38, 353-361.
- Maiwald, J., & Conrad, W. (1993). Entwicklung und Evaluation des MTP-C: Mannheimer Test zur Erfassung des physikalisch-technischen Problemlösens als Computerversion. *Diagnostica*, 39, 352-367.
- Marco, G. L. (1981). Equating tests in an era of test disclosure. In B.F. Green (Hrsg.), *Issues in testing: Coaching, disclosure, and ethnic bias*, pp. 105-122. San Francisco: Jossey-Bass.
- Maschke, P. (1989). Die Bearbeitungszeit von Persönlichkeitsfragebogen in der Eignungsauswahl: Ein Indikator für Verfälschung? *Zeitschrift für experimentelle und angewandte Psychologie*, 10, 121-127.
- Mattson, J. D. (1983). *Effects of item parameter error and other factors on trait estimation in latent trait based adaptive testing*. Unveröffentlichte Doktorarbeit, University of Minnesota.

- McMinn, M. R., Buchanan, T., Ellens, B. M., & Ryan, M. K. (1999). Technology, professional practice, and ethics survey findings and implications. *Professional Psychology: Research and Practice, 30*, 165-172.
- Mead, A. D., & Drasgow, F. (1993). Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. *Psychological Bulletin, 114*, 449-458.
- Meehl, P. E. (1954). *Clinical vs. statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis: University of Minnesota Press.
- Meiser, T., Stern, E. & Langeheine, R. (1998). Latent Change in Discrete Data: Unidimensional, Multidimensional, and Mixture Distribution Rasch Models for the Analysis of Repeated Observations. *Methods of Psychological Research Online, 3/2*.
- Mertens, T. (1999). Konventionelle und computer-gestützte Durchführung von Leistungstests: der Benton-Test. *Zeitschrift für Differentielle und Diagnostische Psychologie, 20*, 97-115.
- Meyer, F., & Schulte, D. (2002) Zur Validität der Beurteilung des Therapieerfolges durch Therapeuten. *Zeitschrift für Klinische Psychologie und Psychotherapie, 31*, 56-61.
- Miles, E. W., & King, W. C. (1998). Gender and administration mode effects when paper-and-pencil personality tests are computerized. *Educational & Psychological Measurement, 58*, 68-76.
- Miller, N. E., & Magruder, K. M. (1999). *Cost-Effectiveness of Psychotherapy*. New York: Oxford University Press.

Moser, K., Selig, J. G., & Rebstock, M. (1990).

Durchführungsvarianten bei einem computerunterstützten Leistungstest. *Diagnostica*, 36, 321-328.

Müller, H. (1999). *Probabilistische Testmodelle für diskrete und kontinuierliche Ratingskalen: Einführung in die Item-Response-Theorie für abgestufte und kontinuierliche Items*. Bern: Huber.

Nelson-Grey, R. O., Herbert, J. D., Herbert, D. L., Sigmon, S. T., & Brannon, S. E. (1989) Effectiveness of matched, mismatched and package treatments of depression. *Journal of Behavior Therapy and Experimental Psychiatry*, 20, 281-294.

Olson, K. R. (2001). Computerized psychological test usage in APA-Accredited training programs. *Journal of Clinical Psychology*, 57, 727-736.

Oseas, L. (1966). Give the first natural answer. *Journal of Counseling Psychology*, 13, 454-458.

Panorama Fachklinik, (1996). *1. Jahresbericht der Forschungsstelle für Psychotherapie in Stuttgart über Qualitätssicherung durch Qualitätsmanagement*. Unveröffentlichter Bericht.

Panorama Fachklinik, (1997). *2. Jahresbericht der Forschungsstelle für Psychotherapie in Stuttgart über Qualitätssicherung durch Qualitätsmanagement*. Unveröffentlichter Bericht.

Panorama Fachklinik, (1998). *3. Jahresbericht der Forschungsstelle für Psychotherapie in Stuttgart über*

Qualitätssicherung durch Qualitätsmanagement.

Unveröffentlichter Bericht.

Panorama Fachklinik, (1999). *4. Jahresbericht der Forschungsstelle für Psychotherapie in Stuttgart über Qualitätssicherung durch Qualitätsmanagement.*

Unveröffentlichter Bericht.

Panorama Fachklinik, (2000). *5. Jahresbericht der Forschungsstelle für Psychotherapie in Stuttgart über Qualitätssicherung durch Qualitätsmanagement.*

Unveröffentlichter Bericht.

Parshall, C. G., Spray, J. A., Kalohn, T. D., & Davey, T. (2002). *Practical consideration in computer-based testing.* New York: Springer.

Percevic, R. (2003). AKQUASI 2002: Qualitätssicherung durch Ergebnismonitoring. Vortrag auf dem Qualitätsmanagement Workshop, Göppingen, Februar, 2003. <http://www.psyres-stuttgart.de/index.php/article/articleview/104/1/7>

Percevic, R., & Wesiak, G. (2001). Adaptive-Sequential Testing in Psychotherapy Outcome Monitoring. Vortrag auf dem 32nd European Mathematical Psychology Group Meeting, Lisbon, September 2001. <http://www.psyres-stuttgart.de/index.php/article/articleview/104/1/7>

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.
- Percevic, R., Herrmann, A., Arian, L., & Kordy, H. (2000). *AKQUASI 2002*. Tätigkeitsbericht der Forschungsstelle für Psychotherapie Stuttgart, pp. 13-14. <http://psyres-stuttgart.de>.
- Percevic, R., Herrmann, A., Hannöver, W., & Sigle, J. (1999). *Technische Weiterentwicklung von AKQUASI*. Tätigkeitsbericht der Forschungsstelle für Psychotherapie Stuttgart, p. 27. <http://psyres-stuttgart.de>.
- Percevic, R., Lambert, M., & Kordy, H. (Eingereicht). What is the predictive value of responses to psychotherapy for the future course? Empirical explorations and consequences for outcome monitoring. Eingereicht bei *Psychotherapy Research*.
- Percevic, R., Lambert, M., Kordy, H. (2004). Computer Supported Monitoring of Patient Treatment Response. *Journal of clinical psychology*, 60, 285-300.
- Perrez, M., & Reicherts, M. (1989). Belastungsverarbeitung: Computerunterstützte Selbstbeobachtung im Feld. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 10, 129-139.
- Piotrowski, C. (1999). Assessment practices in the era of managed care: Current status and future directions. *Journal of Clinical Psychology*, 55, 787-796.
- Pomerich, M. (2004). Developing Computerized Versions of Paper-and-Pencil Tests: Mode Effects for Passage-Based Tests. *Journal of Technology, Learning, and Assessment*, 2/6.

- Pope, K. S. (1992). Responsibilities in providing psychological test feedback to clients. *Psychological Assessment*, 4, 268-271.
- Postmann, L., & Zimmermann, C. (1945). Intensity of attitude as a determinant of decision time. *Amer. J. Psychol.*, 58, 510-518.
- Rammstedt, B., Holzinger, B., Rammsayer, T. (2004). Zur Äquivalenz der Papier-Bleisift- und einer computergestützten Version des NEO-Fünf-Faktoren-Inventars (NEO-FFI). *Diagnostica*, 50, 88-97.
- Reng, M. (2002). Datenschutzkonzept der Kompetenznetze für Medizin. Vortrag auf der 13. Jahrestagung der Deutschen Gesellschaft für klinische Informatik e.V., Regensburg, September 2002.
- Robinson, R., & West, R. (1992). A Comparison of Computer and Questionnaire Methods of History-Taking in a Genito-Urinary Clinic. *Psychology and Health*, 6, 77-84.
- Rogers, T. B. (1974). An analysis of the stages underlying the process of responding to personality items. *Acta Psychologica*, 38, 205-213.
- Rome, H. P., Swenson, W. M., Mataya, P., McCarthy, C. E., Pearson, J. S., Keating, F. R., & Hathaway, S. R. (1962). Symposium on automation techniques in personality assessment. *Proceedings of the Staff Meetings of the Mayo Clinic*, 37, 61-61.

- Roper, B. L., Ben-Porath, Y. S., & Butcher, J. N. (1991). Comparability and validity of computerized adaptive testing with the MMPI-2. *Journal of Personality Assessment*, 65, 358-371.
- Rosen, L. D., & Weil, M. M. (1996). Psychologists and technology: A look at the future. *Professional Psychology: Research and Practice*, 27, 635-638.
- Rost, D. H., & Hoberg, K. (1997). Itempositionsveränderungen in Persönlichkeitsfragebogen: Methodischer Kunstfehler oder tolerierbare Praxis? *Diagnostica*, 43, 97-112.
- Sachverständigenrat für die konzentrierte Aktion im Gesundheitswesen (2000/2001). *Gutachten 2000/2001: Bd II: Qualitätssicherung in Medizin und Pflege*. <http://www.svr-gesundheit.de/gutacht/gutalt/gutaltle.htm>.
- Sawyer, R. (1966). Measurement and Prediction, clinical and statistical. *Psychological Bulletin*, 66, 178-200.
- Schepank, H. H. (1995). *Der Beeinträchtigungsschwere Score (BSS)*. Göttingen: Belz Test.
- Schmidt, J., Lamprecht, F., & Wittmann, W. W. (1989). Zufriedenheit mit der stationären Versorgung. Entwicklung eines Fragebogens und erste Validitätsuntersuchungen. *Psychotherapie, Psychosomatik medizinische Psychologie*, 39, 248-255.
- Schmidt, J., Steffanowski, A., Nübling, R., Lichtenberg, S., & Wittmann, W. W. (2003). *Ergebnisqualität stationärer psychosomatischer Rehabilitation*. Regensburg: Roderer.

- Schneider, J. F., & Hübner, R. (1980). Einfluß von Verfälschungsinstruktionen auf die Bearbeitungszeit von Persönlichkeitsfragebogen. *Zeitschrift für experimentelle und angewandte Psychologie*, 27, 565-579.
- Schulte, D. (1991). *Therapeutische Entscheidungen*. Göttingen, Hogrefe.
- Schulte, D. (1993). "Wie soll Therapieerfolg gemessen werden?". *Zeitschrift für Klinische Psychologie*, 4, 374-393.
- Schulte, D., & Meyer, F. (2002) woran orientieren sich Therapeuten bei ihrer Sitzungsbeurteilung und Erfolgsprognose. *Zeitschrift für Klinische Psychologie und Psychotherapie*, 31, 257-265.
- Schwarz, N., Strack, F., & Hippler, H. J. (1991). Kognitionspsychologie und Umfrageforschung: Themen und Befunde eines interdisziplinären Forschungsgebietes. *Psychologische Rundschau*, 42, 175-186.
- Schwenkmezger, P., & Hank, P. (1993). Papier-Bleistift- versus computerunterstützte Darbietung von State-Trait-Fragebogen: eine Äquivalenzüberprüfung. *Diagnostica*, 39, 189-210.
- Segall, D. O. (1996). Multidimensional adaptive testing. *Psychometrika*, 61, 331-354.
- Shao, J., & Tu, D. (1996). *The Jackknife and bootstrap*. New York: Springer.
- Skinner, H. A., & Allen, B. A. (1983). Does the computer make a difference? Computerized versus face-to-face versus self-

- report assessment of alcohol, drug, and tobacco use. *Journal of Consulting and Clinical Psychology*, 51, 267-275.
- Smith, M. A., & Leight, B. (1997). Virtual Subjects: Using the Internet as an alternative source of subjects and research environment. *Behavior Research Methods, Instruments & Computers*, 29, 496-505.
- Stanton, J. M. (1998). An empirical assessment of data collection using the Internet. *Personnel Psychology*, 51, 709-725.
- Steck, P. (1997). Aus der Arbeit des Testkuratoriums. *Diagnostica*, 43, 267-284.
- Stein, S. J. (1987). Computer-assisted Diagnosis in Children's Mental Health. *Applied Psychology*, 36, 343-357.
- Steyer, R., Schwenkmezger, P., Eid, M., & Notz, P. (1991) *Befindlichkeitsmessung und Latent-State-Trait-Modelle*.
<http://www.uni-jena.de/svw/metheval/materialien/ges7/ZwischenberichtStufel.html>.
- Stieglitz, R. D. (2000). Technologisch orientierte Diagnostik als Rahmen einer orientierungsübergreifenden Diagnostik in der Psychotherapie. In A. R. Lairaiter (Hrsg.): *Diagnostik in der Psychotherapie*, pp. 289-294. Wien: Springer.
- Stieglitz, R. D. (2003). Psychodiagnostische Verfahren. In: M. Härter, W. Linster, R. D. Stieglitz (Hrsg.), *Qualitätsmanagement in der Psychotherapie*, pp. 97-118. Göttingen: Hogrefe.

- Styles, I. (1991). Clinical assessment and computerized testing. *International Journal of Man-Machine Studies*, 35, 133-150.
- Sutherland, B. V., & Spilka, B. (1964). Social Desirability, Item-Response Time and Item Significance. *Journal of Consulting Psychology*, 28, 447-451.
- Theobald, A. (2001). Sinn und Unsinn von Incentives in der Online-Marktforschung. In A. Theobald, M. Dreyer, & Starsetzki, T. (Hrsg.). *Online-Marktforschung: Theoretische Grundlagen und praktische Erfahrung*, pp. 179-190. Wiesbaden: Gabler.
- Thurner, F. (1965). Zur Entscheidungsgeschwindigkeit von Selbsturteilen. *Zeitschrift für experimentelle und angewandte Psychologie*, 12, 161-171.
- Tourangeau, R., & Smith, T. W. (1996). Asking sensitive questions: the impact of data collection mode, question format, and question context. *Public Opinion Quarterly*, 60, 275-304.
- Townsend, J. T. & Ashby, F. G. (1983). *The stochastic modelling of elementary psychological processes*. New York: Cambridge University Press.
- Troche, S., Rammstedt, B., & Rammsayer, T. (2002). Vergleich einer Papier-Bleistift- und einer computergestützten Version des Leistungsprüfsystems (LPS). *Diagnostica*, 48, 115-120.
- Tschuschke, V. (2003). *Fachliche Stellungnahme zum "Modellvorhaben der Techniker-Krankenkasse - Qualitätsmonitoring in der ambulanten Psychotherapie"*. Verband

psychologischer Psychotherapeuten. http://bdp-vpp.de/verband/lfv/nrw/31004_tschuschke.shtml

- Tucker, G. J., & Rosenberg, S. D. (1975). Computer content analysis of schizophrenic speech: A preliminary report. *American Journal of Psychiatry*, 132, 611-616
- Vale, C. D. (1981). Design and implementation of a microcomputer-based adaptive testing system. *Behavior Research Methods and Instrumentation*, 13, 399-406.
- van der Linden, W. J. & Hambleton, R. K. (1997). *Handbook of modern Item Response Theory*. New York: Springer.
- Vanger, P., Hönlinger, R., Strauß, M., v. Ballestrem, C. L., Kächele, H., & Haken, H. (1997). *Entwicklung einer teil-automatisierten Prozedur für die Kodierung mimischen Ausdrucks in der psychotherapeutischen Interaktion*. Tätigkeitsbericht der Forschungsstelle für Psychotherapie Stuttgart, pp. 19-20. <http://psyres-stuttgart.de>.
- Vansickle, T. R. & Kapes, J. T. (1993). Comparing paper-pencil and computer-based versions of the Strong Campbell Interest Inventory. *Computers in Human Behaviour*, 9, 441-449.
- Vehovar, V., Lozar, M. K., & Batagelj, Z. (2000). Design issues in web surveys. Vortrag auf der 55. Annual Conference der American Association for Public Opinion Research, Portland. <http://surveys.over.net/method/nase/aapor2000slides.ppt>.
- Watkins, C. E., Campbell, V. L., Nieberding, R., & Hallmark, R. (1995). Contemporary practice of psychological assessment by

- clinical psychologists. *Professional Psychology: Research and Practice*, 26, 54-60.
- Weiss, D. J. (1974). Strategies of adaptive ability measurement. *Research Rep. No. 74-5*, Minneapolis: University of Minnesota.
- Weiss, D. J. (1985). Adaptive testing by computer. *Journal of Consulting and Clinical Psychology*, 53, 774-789.
- Weiss, D. J., & Suhadolnik, D. (1985). Robustness of adaptive testing to multidimensionality. In D.J. Weiss (Hrsg.), *Proceedings of the 1982 Item Response Theory and Computerized Adaptive Testing Conference*, pp. 248-280. Minneapolis: University of Minnesota, Department of Psychology, Computerized Adaptive Testing Laboratory.
- Wild, B. (1989a). Neue Simulationsstudien zur Effizienz verschiedener Parameterschätz- und Itemauswahl-strategien beim "tailored-testing". In Kubinger, K. D. (Hrsg.), *Moderne Testtheorie*, pp. 163-178. Weinheim: Beltz-Verlag.
- Wild, B. (1989b). Neue Erkenntnisse zur Effizienz des "tailoerd"-adaptiven Testens. In Kubinger, K. D. (Hrsg.), *Moderne Testtheorie*, pp. 179-186. Weinheim: Beltz-Verlag.
- Wildgrube, W. (1990). Computergestützte Diagnostik in einer Großorganisation. *Diagnostica*, 36, 127-147.
- Wilkinson, L. & the Task Force on Statistical Inference (1999). Statistical methods in psychology journals. Guidelines and explanations. *American Psychologist*, 54, 594-604.
- Wittmann, W. W. (1985). *Evaluationsforschung: Aufgaben, Probleme und Anwendungen*. Berlin: Springer.

- Wittmann, W. W. (1987). Grundlagen erfolgreicher Forschung in der Psychologie: Multimodale Diagnostik, Multiplismus, multivariate Reliabilitäts- und Validitätstheorie, *Diagnostika*, 33, 209-226.
- Wittmann, W. W. (1990). Bewertung einer psychosomatischen Fachklinik auf der Basis sozial- und wirtschaftswissenschaftlicher Kriterien. In: U. Koch, W. W. Wittmann (Hrsg.). *Evaluationsforschung*, pp. 135-140. Berlin: Springer.
- Wittmann, W. W., & Schmidt, J. (2002). A Multi-Aspect Approach to the Evaluation of Outcome Quality in Psychosomatic Treatment. Vortrag auf dem *Workshop Quality Management and Outcome Monitoring*, Stuttgart, März 2003.
http://www.psychologie.uni-mannheim.de/psycho2/publi/papers/wittmann_Schmidt.ppt.
- Wolf, M., Arikan, M., & Kordy, H. (2004). Process-outcome in computer mediated psychotherapy: What works in a psychotherapeutic maintenance aftercare via email? Vortrag auf dem 34th Annual Meeting of the SPR, Rom, Juni 2004.
- Yates, J. F. (1990). *Judgment and decision making*. New Jersey: Prentice-Hall.
- Zerr, K. (2001). Online-Marktforschung - Erscheinungsformen und Nutzenpotenziale. In A. Theobald, M. Dreyer, & T. Starsetzki (Hrsg.). *Online-Marktforschung: Theoretische Grundlagen und praktische Erfahrung*, pp. 7-26. Wiesbaden: Gabler.

ANHANG

KORRELATIONSTABELLEN

			Zwischenerhebung							
			PAE	PTBEZ	GBB	ZUF8	EB45-SYB	EB45-BEZ	EB45-SOZ	EB45-TOT
Itemweise Vorgabe	Aufn.	GBB	0,10	-0,12	0,87	-0,07	0,62	0,22	0,33	0,54
		EB45-SYB	0,25	-0,31	0,55	0,00	0,80	0,55	0,63	0,81
		EB45-BEZ	0,13	-0,34	0,06	-0,02	0,31	0,76	0,25	0,47
		EB45-SOZ	0,12	-0,17	0,18	-0,10	0,40	0,41	0,67	0,52
		EB45-TOT	0,23	-0,33	0,41	-0,03	0,69	0,66	0,63	0,77
	Entl.	GBB	-0,24	0,55	0,09	0,32	-0,07	-0,32	-0,19	-0,18
		GBB	0,05	0,18	0,76	0,06	0,51	0,09	0,20	0,41
		EB45-SYB	0,42	-0,09	0,52	0,03	0,76	0,31	0,48	0,69
		EB45-BEZ	0,42	-0,15	0,25	-0,08	0,45	0,61	0,33	0,54
		EB45-SOZ	0,31	-0,08	0,22	-0,07	0,52	0,20	0,72	0,55
	Zwerh.	EB45-TOT	0,45	-0,12	0,46	-0,02	0,73	0,41	0,55	0,71
		PAE	0,45	-0,20	-0,03	-0,12	0,28	0,23	0,25	0,30
		PTBEZ	-0,18	0,28	0,06	0,27	-0,06	0,13	-0,05	-0,01
		ZUF8	-0,15	0,20	0,19	0,57	0,03	-0,01	-0,05	0,01
		GBB	0,20	-0,16	1,00	-0,11	0,67	0,25	0,28	0,58
		EB45-SYB	0,44	-0,33	0,67	-0,02	1,00	0,58	0,64	0,96
		EB45-BEZ	0,21	-0,29	0,25	-0,05	0,58	1,00	0,41	0,75
		EB45-SOZ	0,39	-0,21	0,28	-0,04	0,64	0,41	1,00	0,75
		EB45-TOT	0,43	-0,35	0,58	-0,04	0,96	0,75	0,75	1,00
		PAE	1,00	-0,49	0,20	-0,24	0,44	0,21	0,39	0,43
PTBEZ	-0,49	1,00	-0,16	0,49	-0,33	-0,29	-0,21	-0,35		
ZUF	-0,24	0,49	-0,11	1,00	-0,02	-0,05	-0,04	-0,04		
Seitenweise Vorgabe	Aufn.	GBB	0,17	-0,18	0,75	-0,15	0,48	0,06	0,41	0,44
		EB45-SYB	0,33	-0,25	0,49	0,11	0,64	0,58	0,48	0,70
		EB45-BEZ	0,30	-0,24	0,14	0,10	0,33	0,76	0,31	0,51
		EB45-SOZ	-0,08	0,14	0,00	0,35	0,16	0,23	0,57	0,30
		EB45-TOT	0,30	-0,21	0,37	0,18	0,57	0,68	0,55	0,70
	Entl.	PTBEZ	0,02	0,21	0,05	0,24	-0,05	-0,02	-0,09	-0,06
		GBB	0,14	-0,18	0,67	-0,22	0,38	0,01	0,27	0,32
		EB45-SYB	0,33	-0,49	0,54	-0,12	0,65	0,34	0,45	0,64
		EB45-BEZ	0,25	-0,32	0,10	0,04	0,25	0,61	0,32	0,42
		EB45-SOZ	0,09	-0,05	0,27	0,04	0,41	0,13	0,62	0,45
	Zwerh.	EB45-TOT	0,31	-0,43	0,45	-0,05	0,59	0,43	0,53	0,64
		PAE	0,28	-0,27	0,23	-0,40	0,21	0,03	-0,13	0,12
		PTBEZ	-0,06	0,34	-0,16	0,49	-0,12	0,04	0,11	-0,04
		ZUF8	-0,05	0,36	-0,18	0,75	-0,02	0,11	0,23	0,07
		GBB	0,41	-0,48	1,00	-0,22	0,70	0,24	0,39	0,63
		EB45-SYB	0,64	-0,52	0,70	0,02	1,00	0,49	0,55	0,94
		EB45-BEZ	0,48	-0,28	0,24	0,14	0,49	1,00	0,43	0,72
		EB45-SOZ	0,09	-0,15	0,39	0,21	0,55	0,43	1,00	0,72
		EB45-TOT	0,59	-0,47	0,63	0,10	0,94	0,72	0,72	1,00
		PAE	1,00	-0,55	0,41	-0,13	0,64	0,48	0,09	0,59
PTBEZ	-0,55	1,00	-0,48	0,48	-0,52	-0,28	-0,15	-0,47		
ZUF	-0,13	0,48	-0,22	1,00	0,02	0,14	0,21	0,10		

			Zwischenerhebung							
			PAE	PTBEZ	GBB	ZUF8	EB45-SYB	EB45-BEZ	EB45-SOZ	EB45-TOT
Fehlende Werte nicht Erlaubt	Aufn.	GBB	-0,07	-0,05	0,80	-0,02	0,58	0,16	0,33	0,49
		EB45-SYB	0,25	-0,35	0,50	0,04	0,74	0,58	0,62	0,78
		EB45-BEZ	0,33	-0,43	0,10	-0,13	0,38	0,77	0,42	0,56
		EB45-SOZ	0,19	-0,03	0,05	0,21	0,39	0,36	0,65	0,49
		EB45-TOT	0,31	-0,36	0,36	0,03	0,68	0,69	0,68	0,78
	Entl.	PTBEZ	-0,25	0,42	0,04	0,27	-0,11	-0,18	-0,03	-0,13
		GBB	-0,07	0,09	0,78	0,02	0,50	0,12	0,24	0,41
		EB45-SYB	0,27	-0,29	0,63	-0,02	0,78	0,50	0,54	0,76
		EB45-BEZ	0,17	-0,24	0,13	-0,14	0,35	0,70	0,41	0,51
		EB45-SOZ	0,02	0,06	0,26	0,06	0,48	0,32	0,63	0,53
	Zwerh.	EB45-TOT	0,23	-0,25	0,50	-0,04	0,71	0,61	0,60	0,76
		PAE	0,36	-0,32	0,04	-0,20	0,29	0,24	0,11	0,28
		PTBEZ	-0,22	0,40	-0,07	0,44	-0,08	-0,15	0,12	-0,07
		ZUF8	-0,17	0,32	-0,04	0,75	0,05	-0,16	0,15	0,02
		GBB	-0,01	-0,15	1,00	-0,02	0,65	0,26	0,34	0,57
		EB45-SYB	0,38	-0,36	0,65	0,07	1,00	0,55	0,70	0,95
		EB45-BEZ	0,28	-0,40	0,26	-0,13	0,55	1,00	0,52	0,74
		EB45-SOZ	0,25	-0,24	0,34	0,20	0,70	0,52	1,00	0,81
		EB45-TOT	0,38	-0,39	0,57	0,05	0,95	0,74	0,81	1,00
		ZUF	-0,19	0,43	-0,02	1,00	0,07	-0,13	0,20	0,05
Fehlende Werte Erlaubt	Aufn.	GBB	0,31	-0,19	0,82	-0,16	0,54	0,19	0,41	0,51
		EB45-SYB	0,34	-0,18	0,51	0,12	0,72	0,61	0,51	0,77
		EB45-BEZ	0,09	-0,12	0,10	0,23	0,26	0,75	0,15	0,43
		EB45-SOZ	-0,11	0,06	0,10	0,02	0,21	0,36	0,60	0,38
		EB45-TOT	0,22	-0,14	0,37	0,15	0,59	0,71	0,51	0,71
	Entl.	PTBEZ	0,13	0,22	0,04	0,32	-0,02	-0,10	-0,29	-0,11
		GBB	0,20	-0,03	0,65	-0,10	0,40	0,03	0,21	0,33
		EB45-SYB	0,44	-0,27	0,47	-0,05	0,66	0,17	0,40	0,58
		EB45-BEZ	0,51	-0,28	0,29	0,07	0,41	0,50	0,25	0,48
		EB45-SOZ	0,34	-0,15	0,22	-0,08	0,46	0,05	0,71	0,48
	Zwerh.	EB45-TOT	0,50	-0,28	0,44	-0,03	0,64	0,25	0,48	0,61
		PAE	0,36	-0,18	0,16	-0,36	0,21	0,04	0,02	0,16
		PTBEZ	-0,05	0,29	-0,04	0,37	-0,10	0,28	-0,02	0,01
		ZUF8	-0,07	0,37	0,01	0,56	-0,05	0,33	0,06	0,07
		GBB	0,56	-0,44	1,00	-0,28	0,72	0,28	0,34	0,64
		EB45-SYB	0,67	-0,47	0,72	-0,07	1,00	0,56	0,49	0,95
		EB45-BEZ	0,38	-0,15	0,28	0,25	0,56	1,00	0,34	0,74
		EB45-SOZ	0,21	-0,09	0,34	-0,03	0,49	0,34	1,00	0,65
		EB45-TOT	0,61	-0,38	0,64	0,01	0,95	0,74	0,65	1,00
		ZUF	-0,21	0,55	-0,28	1,00	-0,07	0,25	-0,03	0,01

			Zwischenerhebung							
			PAE	PTBEZ	GBB	ZUF8	EB45-SYB	EB45-BEZ	EB45-SOZ	EB45-TOT
Änderungen nicht möglich	Aufn.	GBB	0,13	-0,04	0,86	-0,25	0,48	0,23	0,44	0,45
		EB45-SYB	0,34	-0,08	0,52	0,15	0,84	0,83	0,51	0,87
		EB45-BEZ	0,38	-0,09	-0,04	0,36	0,40	0,73	-0,05	0,46
		EB45-SOZ	0,22	-0,04	0,37	0,24	0,70	0,55	0,73	0,75
		EB45-TOT	0,38	-0,08	0,38	0,26	0,79	0,85	0,46	0,84
	Entl.	PTBEZ	0,05	0,26	0,05	0,33	0,12	0,12	0,07	0,12
		GBB	0,16	0,09	0,70	0,05	0,70	0,54	0,50	0,69
		EB45-SYB	0,32	0,00	0,38	0,20	0,80	0,60	0,62	0,80
		EB45-BEZ	0,39	-0,03	0,14	0,28	0,62	0,79	0,24	0,67
		EB45-SOZ	0,09	0,15	0,33	0,04	0,43	0,00	0,78	0,42
	Zwerh.	EB45-TOT	0,34	0,03	0,37	0,22	0,80	0,62	0,66	0,82
		PAE	0,40	-0,14	0,04	0,00	0,37	0,33	0,08	0,34
		PTBEZ	0,03	0,00	0,25	0,20	0,15	0,09	0,05	0,13
		ZUF8	0,19	-0,13	-0,02	0,48	0,01	0,11	0,06	0,05
		GBB	0,23	-0,16	1,00	-0,26	0,61	0,34	0,53	0,59
		EB45-SYB	0,47	-0,03	0,61	0,17	1,00	0,74	0,70	0,98
		EB45-BEZ	0,43	-0,02	0,34	0,27	0,74	1,00	0,32	0,82
		EB45-SOZ	0,30	0,05	0,53	0,10	0,70	0,32	1,00	0,73
		EB45-TOT	0,48	-0,02	0,59	0,21	0,98	0,82	0,73	1,00
		PAE	1,00	-0,40	0,23	-0,11	0,47	0,43	0,30	0,48
PTBEZ	-0,40	1,00	-0,16	0,45	-0,03	-0,02	0,05	-0,02		
ZUF	-0,11	0,45	-0,26	1,00	0,17	0,27	0,10	0,21		
Änderungen möglich	Aufn.	GBB	0,12	-0,15	0,81	-0,07	0,57	0,12	0,35	0,50
		EB45-SYB	0,26	-0,28	0,53	0,05	0,69	0,45	0,59	0,73
		EB45-BEZ	0,14	-0,29	0,13	-0,03	0,28	0,77	0,41	0,48
		EB45-SOZ	-0,03	0,02	0,03	0,05	0,19	0,25	0,58	0,32
		EB45-TOT	0,21	-0,27	0,40	0,04	0,59	0,59	0,65	0,70
	Entl.	PTBEZ	-0,10	0,31	0,07	0,25	-0,07	-0,20	-0,18	-0,14
		GBB	0,05	0,01	0,72	-0,07	0,42	-0,06	0,17	0,31
		EB45-SYB	0,37	-0,32	0,56	-0,07	0,69	0,22	0,42	0,63
		EB45-BEZ	0,34	-0,29	0,19	-0,09	0,30	0,55	0,35	0,43
		EB45-SOZ	0,22	-0,11	0,22	-0,03	0,49	0,24	0,64	0,54
	Zwerh.	EB45-TOT	0,38	-0,32	0,47	-0,08	0,63	0,35	0,50	0,64
		PAE	0,35	-0,27	0,11	-0,34	0,22	0,06	0,05	0,18
		PTBEZ	-0,15	0,38	-0,11	0,43	-0,12	0,11	0,05	-0,05
		ZUF8	-0,17	0,39	0,00	0,70	0,02	0,07	0,13	0,06
		GBB	0,30	-0,34	1,00	-0,13	0,70	0,21	0,28	0,60
		EB45-SYB	0,54	-0,48	0,70	-0,03	1,00	0,48	0,58	0,95
		EB45-BEZ	0,30	-0,33	0,21	-0,01	0,48	1,00	0,47	0,70
		EB45-SOZ	0,21	-0,21	0,28	0,09	0,58	0,47	1,00	0,74
		EB45-TOT	0,50	-0,47	0,60	0,00	0,95	0,70	0,74	1,00
		PAE	1,00	-0,56	0,30	-0,20	0,54	0,30	0,21	0,50
PTBEZ	-0,56	1,00	-0,34	0,48	-0,48	-0,33	-0,21	-0,47		
ZUF	-0,20	0,48	-0,13	1,00	-0,03	-0,01	0,09	0,00		

			Zwischenerhebung							
			PAE	PTBEZ	GBB	ZUF8	EB45-SYB	EB45-BEZ	EB45-SOZ	EB45-TOT
Keine Rückmeldung an Pat.	Aufn.	GBB	0,24	-0,32	0,84	-0,19	0,59	0,29	0,31	0,55
		EB45-SYB	0,28	-0,34	0,48	-0,08	0,64	0,50	0,63	0,72
		EB45-BEZ	0,02	-0,23	0,19	0,04	0,23	0,81	0,36	0,45
		EB45-SOZ	0,00	-0,09	0,04	0,09	0,25	0,33	0,67	0,40
		EB45-TOT	0,19	-0,30	0,37	-0,01	0,54	0,63	0,67	0,68
	Entl.	PTBEZ	-0,26	0,49	-0,04	0,32	-0,25	-0,26	-0,27	-0,30
		GBB	0,25	-0,10	0,75	-0,13	0,46	0,02	0,16	0,36
		EB45-SYB	0,51	-0,27	0,51	-0,05	0,76	0,18	0,39	0,67
		EB45-BEZ	0,27	-0,09	0,24	0,08	0,35	0,54	0,26	0,44
		EB45-SOZ	0,24	-0,04	0,10	0,10	0,48	0,15	0,64	0,51
	Zwerh.	EB45-TOT	0,46	-0,21	0,43	0,01	0,71	0,30	0,47	0,67
		PAE	0,62	-0,29	0,15	-0,25	0,33	0,04	0,16	0,27
		PTBEZ	-0,19	0,38	-0,12	0,51	-0,20	0,16	0,05	-0,09
		ZUF8	-0,14	0,34	-0,05	0,68	-0,06	0,18	0,12	0,03
		GBB	0,40	-0,42	1,00	-0,27	0,67	0,28	0,25	0,60
		EB45-SYB	0,61	-0,58	0,67	-0,12	1,00	0,43	0,61	0,95
		EB45-BEZ	0,23	-0,30	0,28	0,02	0,43	1,00	0,48	0,66
		EB45-SOZ	0,29	-0,30	0,25	0,09	0,61	0,48	1,00	0,77
		EB45-TOT	0,55	-0,55	0,60	-0,06	0,95	0,66	0,77	1,00
		PAE	1,00	-0,55	0,40	-0,18	0,61	0,23	0,29	0,55
Rückmeldung an Patienten	Aufn.	PTBEZ	-0,55	1,00	-0,42	0,56	-0,58	-0,30	-0,30	-0,55
		ZUF	-0,18	0,56	-0,27	1,00	-0,12	0,02	0,09	-0,06
		GBB	-0,03	0,06	0,78	0,00	0,51	0,03	0,42	0,42
		EB45-SYB	0,29	-0,20	0,60	0,19	0,83	0,62	0,49	0,82
		EB45-BEZ	0,43	-0,33	0,04	0,05	0,46	0,75	0,25	0,58
	Entl.	EB45-SOZ	0,07	0,08	0,24	0,09	0,41	0,36	0,60	0,50
		EB45-TOT	0,34	-0,22	0,46	0,16	0,78	0,72	0,53	0,83
		PTBEZ	0,06	0,20	0,15	0,23	0,11	-0,08	-0,03	0,04
		GBB	-0,23	0,18	0,68	0,08	0,46	0,11	0,35	0,40
		EB45-SYB	0,17	-0,27	0,54	-0,02	0,64	0,46	0,53	0,66
	Zwerh.	EB45-BEZ	0,41	-0,37	0,09	-0,14	0,36	0,65	0,36	0,50
		EB45-SOZ	0,13	-0,07	0,42	-0,15	0,45	0,18	0,70	0,49
		EB45-TOT	0,26	-0,30	0,46	-0,09	0,61	0,53	0,59	0,68
		PAE	0,01	-0,21	0,04	-0,30	0,15	0,25	-0,06	0,16
		PTBEZ	-0,01	0,29	0,08	0,23	0,10	0,00	0,06	0,08
	ZUF8	-0,10	0,36	0,06	0,66	0,10	-0,08	0,11	0,06	
	GBB	0,13	-0,18	1,00	-0,01	0,69	0,21	0,43	0,59	
	EB45-SYB	0,40	-0,24	0,69	0,15	1,00	0,65	0,57	0,96	
	EB45-BEZ	0,42	-0,24	0,21	0,07	0,65	1,00	0,38	0,80	
	EB45-SOZ	0,13	-0,01	0,43	0,09	0,57	0,38	1,00	0,69	
	EB45-TOT	0,41	-0,23	0,59	0,14	0,96	0,80	0,69	1,00	
	PAE	1,00	-0,51	0,13	-0,21	0,40	0,42	0,13	0,41	
	PTBEZ	-0,51	1,00	-0,18	0,42	-0,24	-0,24	-0,01	-0,23	
	ZUF	-0,21	0,42	-0,01	1,00	0,15	0,07	0,09	0,14	

			Zwischenerhebung							
			PAE	PTBEZ	GBB	ZUF8	EB45-SYB	EB45-BEZ	EB45-SOZ	EB45-TOT
Papier/Bleistift-Vorgabe	Aufn.	GBB	0,24	-0,24	0,75	-0,17	0,55	0,24	0,40	0,51
		EB45-SYB	0,27	-0,26	0,65	-0,18	0,75	0,55	0,56	0,75
		EB45-BEZ	0,17	-0,08	0,22	-0,02	0,36	0,69	0,30	0,49
		EB45-SOZ	0,13	-0,26	0,44	-0,15	0,52	0,47	0,65	0,61
		EB45-TOT	0,25	-0,25	0,58	-0,15	0,70	0,67	0,60	0,77
	Entl.	PTBEZ	-0,23	0,47	-0,11	0,31	-0,25	-0,14	-0,19	-0,24
		GBB	0,22	-0,40	0,83	-0,33	0,63	0,31	0,40	0,58
		EB45-SYB	0,32	-0,50	0,67	-0,45	0,76	0,51	0,55	0,75
		EB45-BEZ	0,21	-0,29	0,30	-0,29	0,42	0,73	0,31	0,54
		EB45-SOZ	0,19	-0,33	0,55	-0,34	0,62	0,41	0,75	0,68
	Zwerh.	EB45-TOT	0,30	-0,46	0,62	-0,44	0,72	0,61	0,59	0,76
		PAE	0,44	-0,40	0,43	-0,36	0,36	0,15	0,26	0,33
		PTBEZ	-0,31	0,53	-0,40	0,45	-0,40	-0,10	-0,22	-0,34
		ZUF8	-0,24	0,35	-0,31	0,61	-0,27	-0,17	-0,24	-0,27
		GBB	0,42	-0,45	1,00	-0,26	0,80	0,41	0,52	0,74
		EB45-SYB	0,53	-0,56	0,80	-0,30	1,00	0,61	0,68	0,96
		EB45-BEZ	0,30	-0,21	0,41	-0,14	0,61	1,00	0,44	0,77
		EB45-SOZ	0,27	-0,33	0,52	-0,19	0,68	0,44	1,00	0,78
		EB45-TOT	0,48	-0,49	0,74	-0,28	0,96	0,77	0,78	1,00
		PAE	1,00	-0,53	0,42	-0,25	0,53	0,30	0,27	0,48
Computergestützte Vorgabe	Aufn.	PTBEZ	-0,53	1,00	-0,45	0,46	-0,56	-0,21	-0,33	-0,49
		ZUF	-0,25	0,46	-0,26	1,00	-0,30	-0,14	-0,19	-0,28
		GBB	0,12	-0,14	0,81	-0,11	0,56	0,15	0,37	0,49
		EB45-SYB	0,28	-0,26	0,52	0,05	0,73	0,56	0,56	0,76
		EB45-BEZ	0,20	-0,26	0,10	0,04	0,32	0,76	0,28	0,49
	Entl.	EB45-SOZ	0,02	0,01	0,11	0,09	0,31	0,34	0,62	0,43
		EB45-TOT	0,25	-0,24	0,39	0,06	0,64	0,67	0,59	0,74
		PTBEZ	-0,09	0,31	0,06	0,27	-0,06	-0,15	-0,13	-0,11
		GBB	0,07	0,02	0,72	-0,05	0,46	0,06	0,23	0,37
		EB45-SYB	0,36	-0,28	0,53	-0,03	0,71	0,32	0,46	0,67
	Zwerh.	EB45-BEZ	0,35	-0,25	0,18	-0,03	0,36	0,60	0,32	0,48
		EB45-SOZ	0,20	-0,06	0,24	-0,01	0,47	0,17	0,67	0,50
		EB45-TOT	0,38	-0,27	0,45	-0,03	0,67	0,42	0,53	0,68
		PAE	0,36	-0,24	0,10	-0,27	0,25	0,13	0,06	0,21
		PTBEZ	-0,12	0,34	-0,05	0,40	-0,09	0,08	0,05	-0,03
	ZUF8	-0,12	0,34	-0,01	0,66	0,00	0,06	0,11	0,04	
	GBB	0,29	-0,32	1,00	-0,16	0,69	0,25	0,34	0,60	
	EB45-SYB	0,52	-0,42	0,69	0,00	1,00	0,54	0,60	0,95	
	EB45-BEZ	0,33	-0,27	0,25	0,05	0,54	1,00	0,42	0,74	
	EB45-SOZ	0,23	-0,16	0,34	0,09	0,60	0,42	1,00	0,73	
	EB45-TOT	0,49	-0,39	0,60	0,03	0,95	0,74	0,73	1,00	
	PAE	1,00	-0,54	0,29	-0,19	0,52	0,33	0,23	0,49	
	PTBEZ	-0,54	1,00	-0,32	0,48	-0,42	-0,27	-0,16	-0,39	
	ZUF	-0,19	0,48	-0,16	1,00	0,00	0,05	0,09	0,03	

KOMPONENTENLADUNGEN

	1	2
PAE	0,50	-0,55
PTBEZ	-0,33	0,82
GBB	0,71	-0,06
ZUF8	-0,12	0,81
EB45-SYB	0,95	0,09
EB45-BEZ	0,72	0,40
EB45-SOZ	0,79	0,15
EB45-TOT	0,97	0,20

Itemweise Vorgabe

	1	2
PAE	0,56	-0,46
PTBEZ	-0,46	0,74
GBB	0,65	0,01
ZUF8	-0,17	0,82
EB45-SYB	0,92	0,14
EB45-BEZ	0,74	0,13
EB45-SOZ	0,65	0,44
EB45-TOT	0,96	0,22

Seitenweise Vorgabe

	1	2
PAE	0,46	-0,52
PTBEZ	-0,45	0,76
GBB	0,65	0,18
ZUF8	-0,29	0,75
EB45-SYB	0,92	0,20
EB45-BEZ	0,74	0,08
EB45-SOZ	0,74	0,30
EB45-TOT	0,97	0,22

Fehlende Werte nicht Erlaubt

	1	2
PAE	0,58	-0,51
PTBEZ	-0,34	0,81
GBB	0,70	-0,12
ZUF8	0,01	0,87
EB45-SYB	0,95	0,04
EB45-BEZ	0,74	0,34
EB45-SOZ	0,72	0,25
EB45-TOT	0,97	0,18

Fehlende Werte Erlaubt

	1	2
PAE	0,39	-0,46
PTBEZ	0,20	0,89
GBB	0,71	-0,15
ZUF8	0,13	0,79
EB45-SYB	0,96	-0,03
EB45-BEZ	0,72	0,00
EB45-SOZ	0,73	0,06
EB45-TOT	0,98	-0,01

Änderungen nicht möglich

	1	2
PAE	0,56	-0,51
PTBEZ	-0,51	0,71
GBB	0,66	-0,01
ZUF8	-0,20	0,82
EB45-SYB	0,93	0,12
EB45-BEZ	0,72	0,31
EB45-SOZ	0,73	0,35
EB45-TOT	0,96	0,24

Änderungen möglich

	1	2
PAE	0,58	-0,42
PTBEZ	-0,58	0,71
GBB	0,67	0,06
ZUF8	-0,30	0,81
EB45-SYB	0,92	0,18
EB45-BEZ	0,68	0,23
EB45-SOZ	0,70	0,41
EB45-TOT	0,95	0,27

Keine Rückmeldung an Pat.

	1	2
PAE	0,44	-0,60
PTBEZ	-0,18	0,84
GBB	0,67	-0,06
ZUF8	0,05	0,80
EB45-SYB	0,94	0,04
EB45-BEZ	0,79	0,18
EB45-SOZ	0,75	0,17
EB45-TOT	0,98	0,11

Rückmeldung an Patienten

	1	2
PAE	0,60	-0,32
PTBEZ	-0,56	0,64
GBB	0,77	0,07
ZUF8	-0,41	0,72
EB45-SYB	0,93	0,16
EB45-BEZ	0,65	0,41
EB45-SOZ	0,72	0,17
EB45-TOT	0,95	0,26

Papier/Bleistift-Vorgabe

	1	2
PAE	0,53	-0,51
PTBEZ	-0,39	0,79
GBB	0,68	-0,01
ZUF8	-0,14	0,82
EB45-SYB	0,94	0,11
EB45-BEZ	0,73	0,25
EB45-SOZ	0,73	0,29
EB45-TOT	0,97	0,21

Computergestützte Vorgabe

ITEMKENNWERTE

Item	computergestützte Testvorgabe			Papier/Bleistift-Testvorgabe		
	M	SD	r _{is}	M	SD	r _{is}
THBEZ1	1,99	0,98	0,62	2,17	0,89	0,63
THBEZ2	1,68	1,01	0,57	1,83	0,94	0,64
THBEZ3	1,99	0,95	0,39	2,16	0,84	0,48
THBEZ4	0,90	1,54	0,54	1,19	1,47	0,52
THBEZ5	1,18	1,29	0,56	1,15	1,42	0,58
THBEZ6	2,27	0,94	0,51	2,49	0,80	0,52
THBEZ7	2,30	0,98	0,57	2,39	0,89	0,58
THBEZ8	2,49	0,74	0,52	2,59	0,82	0,28
THBEZ9	2,23	0,95	0,56	2,22	0,93	0,63
THBEZ10	1,99	0,93	0,61	1,87	1,19	0,50
THBEZ11	0,40	1,76	0,31	0,30	1,77	0,37
GBB1	1,25	1,10	0,50	1,21	1,13	0,65
GBB2	0,99	1,04	0,48	0,81	1,04	0,57
GBB3	1,16	1,25	0,53	0,97	1,15	0,48
GBB4	1,37	1,22	0,47	1,12	1,23	0,38
GBB5	1,41	1,33	0,50	1,10	1,25	0,46
GBB6	0,92	0,96	0,56	1,00	1,05	0,62
GBB7	1,60	1,24	0,41	1,46	1,22	0,45
GBB8	1,88	1,19	0,46	1,66	1,25	0,47
GBB9	0,14	0,45	0,39	0,11	0,57	0,27
GBB10	0,46	0,88	0,47	0,47	0,85	0,38
GBB11	0,66	0,98	0,44	0,48	0,94	0,59
GBB12	0,47	0,84	0,43	0,41	0,81	0,31
GBB13	0,46	0,92	0,42	0,36	0,79	0,43
GBB14	1,09	1,07	0,48	1,23	1,15	0,56
GBB15	1,66	1,19	0,66	1,59	1,21	0,70
GBB16	1,77	1,16	0,50	1,65	1,10	0,64
GBB17	0,98	1,04	0,63	0,86	0,99	0,56
GBB18	1,15	1,21	0,55	0,95	1,10	0,62
GBB19	1,47	1,12	0,68	1,26	1,09	0,71
GBB20	0,60	0,92	0,54	0,61	0,94	0,65
GBB21	0,65	0,98	0,48	0,71	0,97	0,51
GBB22	0,42	0,78	0,43	0,39	0,77	0,56
GBB23	1,03	1,13	0,61	1,09	1,22	0,64
GBB24	0,47	0,80	0,50	0,41	0,81	0,56
EB1	0,54	0,70	0,38	0,68	0,72	0,29
EB2	2,05	1,00	0,44	2,05	1,04	0,42
EB3	1,54	1,06	0,56	1,44	0,98	0,65
EB4	1,97	1,14	0,43	2,09	1,23	0,40
EB5	2,40	0,93	0,51	2,40	0,96	0,60
EB6	1,76	0,87	0,52	1,70	0,90	0,53
EB7	1,72	1,36	0,38	1,69	1,42	0,40
EB8	0,51	0,86	0,48	0,35	0,80	0,56
EB9	1,86	0,97	0,64	1,80	1,01	0,73
EB10	2,11	1,08	0,69	1,97	1,03	0,71
EB11	0,03	0,26	0,15	0,03	0,14	-0,03
EB12	1,82	1,36	0,24	1,74	1,41	0,46
EB13	1,92	0,92	0,67	1,78	0,91	0,66
EB14	2,14	1,15	0,10	2,21	1,18	0,01

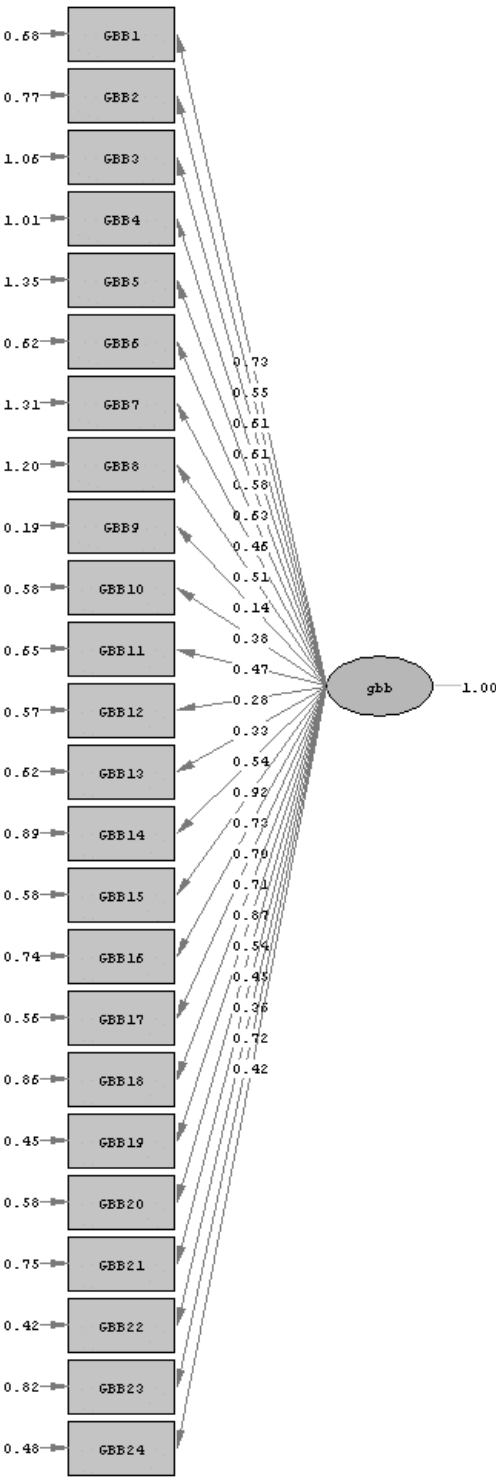
...

Item	computergestützte Testvorgabe			Papier/Bleistift-Testvorgabe		
	M	SD	r _{is}	M	SD	r _{is}
EB15	1,63	1,17	0,73	1,67	1,16	0,71
EB16	1,86	1,19	0,38	1,63	1,10	0,39
EB17	1,99	1,34	0,40	1,90	1,46	0,31
EB18	2,06	1,08	0,62	1,85	1,14	0,63
EB19	1,19	0,76	0,23	1,12	0,77	0,07
EB20	1,70	1,04	0,58	1,62	1,10	0,62
EB21	1,66	1,13	0,56	1,61	1,05	0,51
EB22	2,25	0,93	0,52	2,13	0,97	0,61
EB23	1,82	1,15	0,67	1,72	1,16	0,76
EB24	1,61	1,02	0,67	1,53	1,10	0,59
EB25	2,11	1,06	0,52	1,85	1,06	0,47
EB26	0,10	0,43	0,02	0,03	0,31	-0,12
EB27	1,50	1,16	0,22	1,54	1,23	0,36
EB28	2,22	1,03	0,43	2,16	1,06	0,44
EB29	1,39	1,06	0,36	1,13	1,05	0,50
EB30	0,98	0,92	0,37	0,95	0,79	0,32
EB31	1,78	1,05	0,65	1,66	1,02	0,70
EB32	0,02	0,21	-0,05	0,02	0,15	-0,13
EB33	1,32	1,12	0,62	1,35	1,05	0,58
EB34	1,60	1,26	0,21	1,30	1,23	0,31
EB35	0,69	1,01	0,34	0,63	1,00	0,46
EB36	2,14	0,98	0,52	2,08	0,97	0,65
EB37	2,18	1,38	0,35	1,93	1,42	0,31
EB38	1,54	1,09	0,49	1,59	1,13	0,56
EB39	1,03	0,89	0,23	1,00	0,92	0,42
EB40	0,87	1,04	0,52	0,84	1,06	0,52
EB41	2,15	1,25	0,33	2,16	1,27	0,37
EB42	2,11	0,92	0,70	2,06	0,93	0,75
EB43	1,35	0,97	0,55	1,29	0,95	0,60
EB44	0,63	0,80	0,35	0,75	0,80	0,22
EB45	1,50	1,10	0,27	1,53	1,10	0,40
ZUF1	3,32	0,52	0,63	3,45	0,51	0,54
ZUF2	3,29	0,55	0,55	3,45	0,52	0,66
ZUF3	3,76	0,45	0,64	3,84	0,42	0,49
ZUF4	3,35	0,66	0,59	3,43	0,64	0,67
ZUF5	3,48	0,58	0,52	3,53	0,65	0,46
ZUF6	3,61	0,61	0,69	3,73	0,51	0,49
ZUF7	3,48	0,54	0,69	3,57	0,53	0,70
ZUF8	3,27	0,63	0,65	3,43	0,56	0,74
PAE1	2,21	1,00	0,76	2,24	0,92	0,64
PAE2	2,32	1,06	0,67	2,10	0,86	0,60
PAE3	2,18	0,97	0,83	2,17	0,86	0,79
PAE4	2,31	0,90	0,72	2,33	0,82	0,71

MESSMODELLE

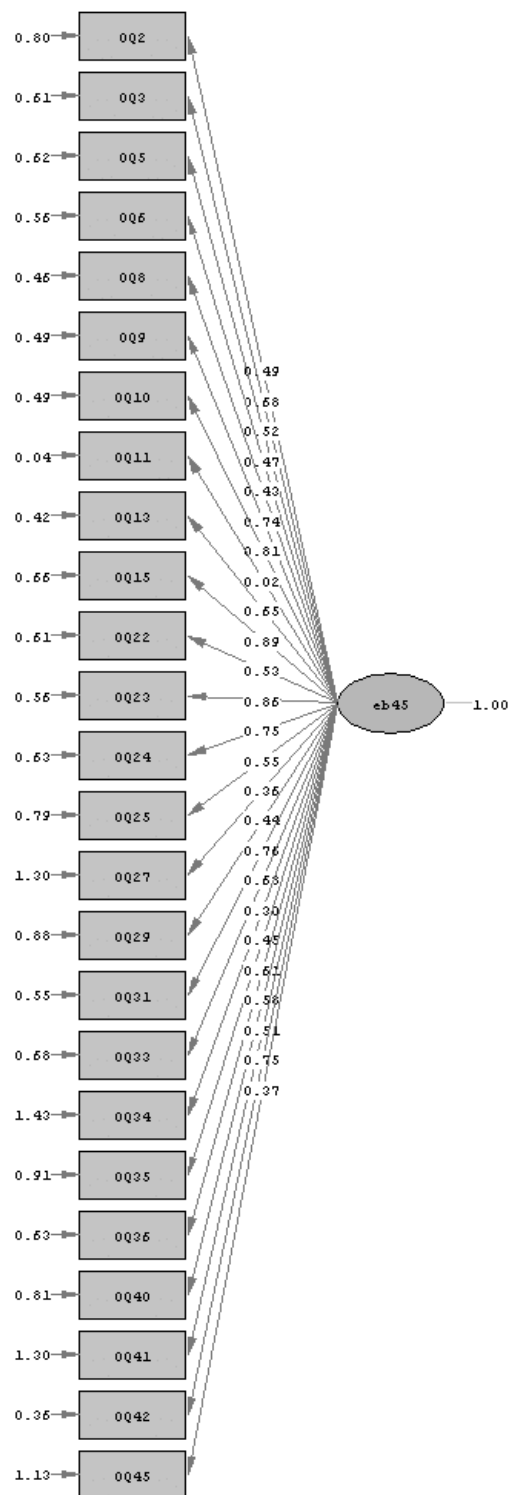
(Parameterschätzung über Gesamtstichprobe)

GBB



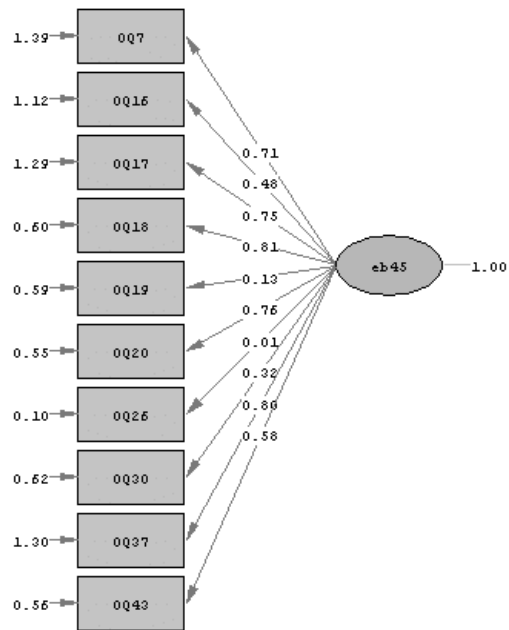
Chi-Square=2815.94, df=552, P-value=0.00000, RMSEA=0.160

EB45-SYB



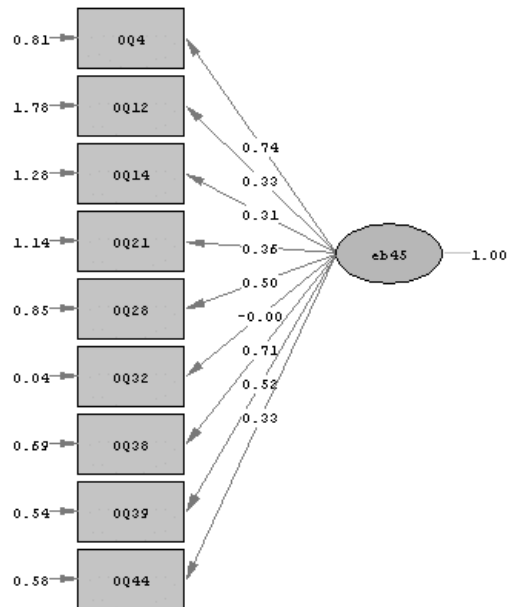
Chi-Square=2239.25, df=600, P-value=0.00000, RMSEA=0.112

EB45-BEZ



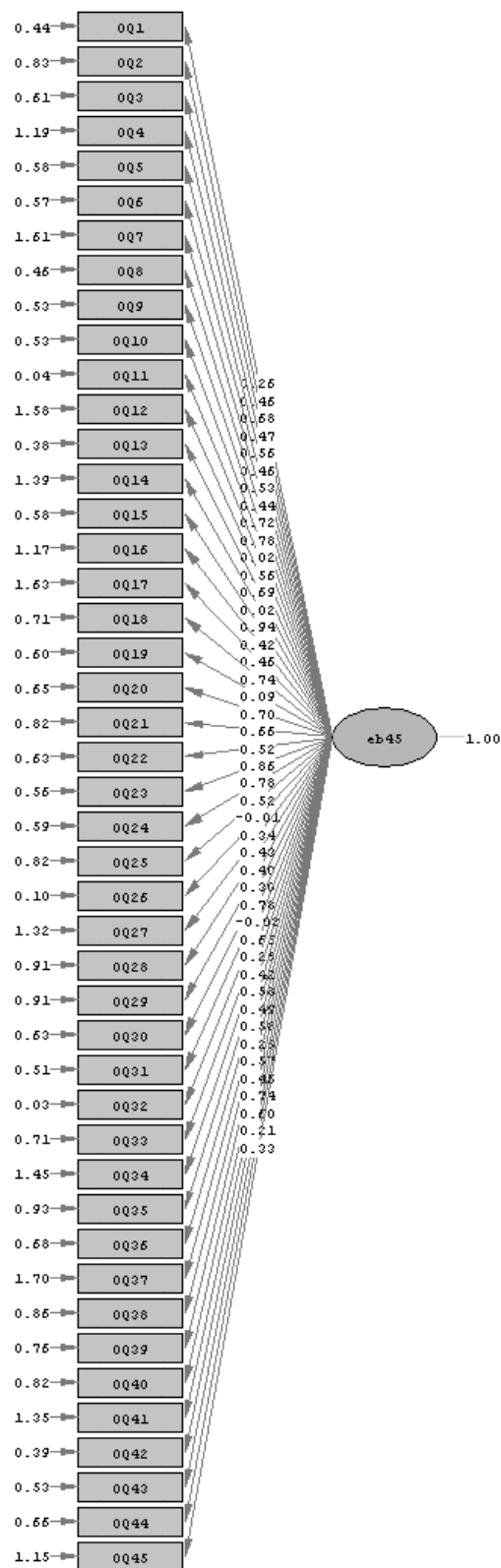
Chi-Square=499.24, df=90, P-value=0.00000, RMSEA=0.144

EB45-SOZ



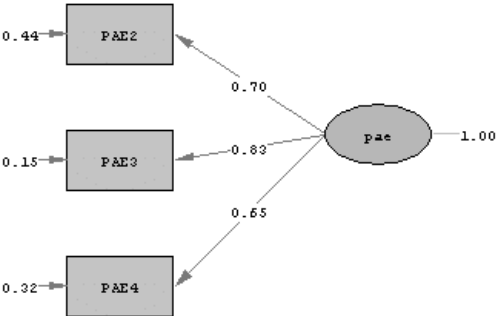
Chi-Square=205.54, df=72, P-value=0.00000, RMSEA=0.108

EB45-TOT



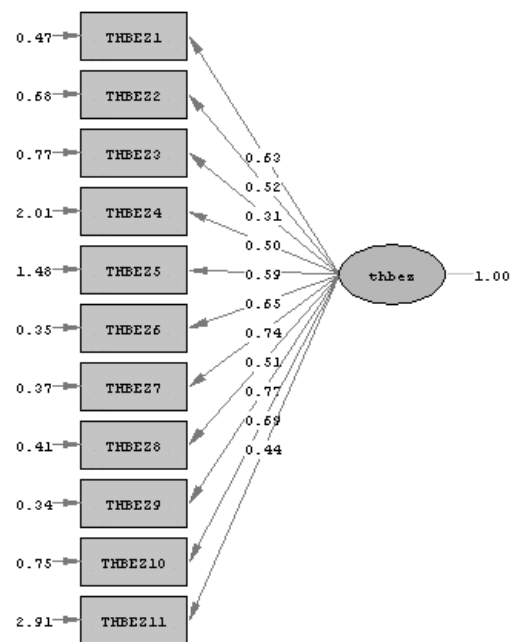
Chi-Square=6855.02, df=1980, P-value=0.00000, RMSEA=0.106

PAE



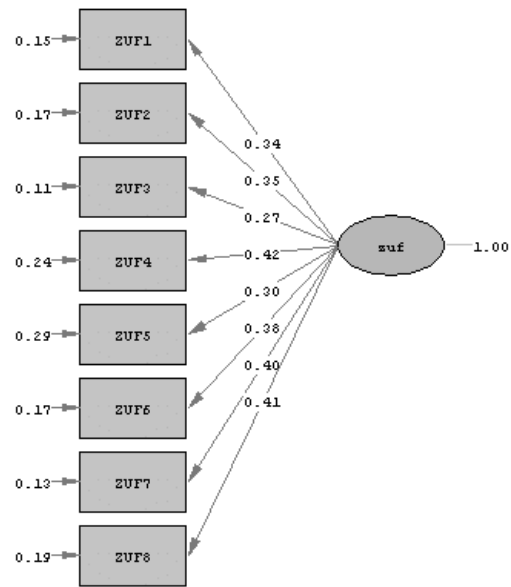
Chi-Square=11.44, df=6, P-value=0.07568, RMSEA=0.064

THBEZ



Chi-Square=803.12, df=110, P-value=0.00000, RMSEA=0.169

ZUF8



Chi-Square=159.70, df=56, P-value=0.00000, RMSEA=0.092

KURZANLEITUNG COMPUTERUNTERSTÜTZTE ZWISCHENERHEBUNG

STARTEN:

- Icon "ZW" Doppelklicken

PASSWORT:

- *****

EINGABE:

- 6-stelligen Patientencode eingeben, OK-Button anklicken oder Return Taste drücken.
- Falls dieser Code schon existiert, wird das angezeigt. Wenn dieser Patient schon Daten eingegeben hat und die Eingabe nur unterbrochen hat, weiter mit "eine alte Eingabe fortsetzen", sonst mit "einen neuen Code eingeben" abbrechen und Code ändern.
- Instruktionen mit Patienten durchgehen. Den Patienten alle Funktionen ausprobieren lassen. Verhalten bei Klicken der rechten Maustaste erklären ("dann einfach mit der linken Taste weitermachen wie bisher"). Auf Einschränkungen (keine fehlenden Werte erlaubt, kein Ändern möglich) aufmerksam machen (bei Fragen warum oder warum andere Patienten das konnten: "das wird so vom System vorgegeben, wir haben darauf keinen Einfluß"). Wenn Patient die computerunterstützte Vorgabe ablehnt oder damit überhaupt nicht klarkommt: Papierfragebogen anbieten, mit OK weiter und dann mit ALT-F4 abbrechen.
- Patienten alleine Fragebogen ausfüllen lassen, aber für Rückfragen erreichbar sein. Wenn Patient eine Frage nicht beantworten will oder kann und sie nicht offenlassen darf: "geben Sie die am besten passende Antwort. Falls die Frage nicht auf Sie zutrifft, Antworten sie mit NICHT oder NEIN". Wenn Patient etwas ändern will und das nicht geht: "Wir können das leider nicht ändern, zum Glück beziehen sich mehrere Fragen auf diesen Punkt so das eine falsche Antwort kein so hohes Gewicht hat". Wenn Patient abbrechen will: versuchen zu motivieren, wenn es nicht klappt Papierfragebogen anbieten.
- Wenn Patient fertig ist, mit ALT-F4 beenden.
- Fragen, Anmerkungen und außergewöhnliche Geschehnisse (wie Abbruch, Ablehnung...) notieren.

EXPORT:

- Damit werden die Daten auf Diskette geschrieben.
- Bitte immer bei der Zusendung der Papierfragebögen an die Forschungsstelle auch den EXPORT durchführen und Diskette mitschicken. Dazu leere Diskette benutzen. Alle Papierfragebögen zuschicken, auch von Patienten welche nicht an der QS teilnehmen (auf einen gesonderten Stapel).

ANSICHT:

- Patienten sind nach Code und nach Behandlung/Entlassen sortiert, zuerst kommen Patienten die noch im Behandlung sind, dann die Entlassenen (mit Zusatz "Entlassen")
- ROT bedeutet das diese Patienten keine Verbesserung oder eine Verschlechterung während der Therapie und/oder häufige Suizidgedanken angeben.
- GRÜN bedeutet das diese Patienten in keinem der erfaßten Bereiche zu den 30% der Bevölkerung mit den höchsten Symptomwerten gehören, eine Verbesserung durch die Therapie wahrnehmen und keine häufigen Suizidgedanken haben.
- SCHWARZ bedeutet das weder die Bedingungen für ROT noch für GRÜN vorliegen, d.h. diesen Patienten geht es besser, sie haben keine häufigen Suizidgedanken, aber gehören in einen oder mehreren Bereichen zu den 30% der Bevölkerung mit den höchsten Symptomwerten.

- Der Zusatz "k.A.m" bedeutet das dieser Patient keine Angaben zur Veränderung während der Therapie und/oder zu Suizidgedanken macht, so das keine Farbcodierung möglich ist.

AUSWERTUNG:

- Anzeige von GBB (somatische Beschwerden), OQ1-OQ4 (Symptombelastung, zwischenmenschliche Beziehungen, soziale Integration, allgemeine psychische Beeinträchtigung), direkter Veränderungseinschätzung, therapeutischer Beziehung, Patientenzufriedenheit und häufigen Suizidgedanken.
- Wert: Skalenwert oder "k.a." für fehlende Angaben.
- Perzentil: Anteil der Bevölkerung mit niedrigeren Werten in 10%-Schritten.
- GRÜN: Normbereich, 70% der Bevölkerung befindet sich in diesem Bereich.
- ROT: außerhalb dieser Norm.
- HELLGRAU: 70% aller Patienten befinden sich in diesen Bereich.

FORMULAR:

- Anzeige der Patienteneingaben.

ENTLASSUNG:

- Patient wird als entlassen Markiert.

ENDE:

- Programm bitte immer über diesen Menüpunkt verlassen.

DRUCKEN:

- Bildschirminhalte können Sie durch anklicken der rechten Maustaste und Auswahl von "Drucken" in dem darauffolgenden Menü ausdrucken.

INFO:

- Robert Percevic,
E-Mail: percevic@psyres-stuttgart.de
Telefon: 0711 6781 406
Fax: 0711 6876 902
in Notfällen: *****
- vom 10.10.2000 bis 10.11.2000: Anke Herrmann,
E-Mail: herrmann@psyres-stuttgart.de
Telefon: 0711 6781 409
Fax: 0711 6876 902

INTERPRETATION DER ZWISCHENERHEBUNG

Wie auch Diagnoseverfahren in anderen Bereichen der Medizin liefert AKQUASI wichtige Informationen über den Zustand des Patienten, welche Ihren klinischen Eindruck ergänzen und bestätigen können. Das System kann Sie in Ihrer Einschätzung zum Therapiebedarf, zum Ansprechen auf die Therapie, zur therapeutischen Beziehung und zur Patientenzufriedenheit unterstützen. Dieses Manual soll Ihnen dabei helfen, die Auswertungen, welche Sie bekommen, richtig einzuschätzen.

Die Patienten werden routinemäßig etwa zwei Wochen nach Therapiebeginn befragt. Die Auswertungen beziehen sich also auf diesen Therapieabschnitt. Alle Antworten des Patienten, wie sie auch auf dem Papierformularen zu sehen sind, können Sie im System unter „FORMULAR“ abrufen.

Erster Überblick über den Zustand des Patienten

Einen ersten Überblick gibt die Farbkodierung des Patienten. Auf Schwarzweissdrucken ist Grün Hellgrau, Rot ist Dunkelgrau. Falls ein Patient häufige Suizidgedanken oder keine Verbesserung bzw. eine Verschlechterung seines Befindens im Laufe der Therapie angibt, wird er **ROT** markiert. Häufige Suizidgedanken führen zwar selten zu Suizidversuchen, sollten aber dennoch ernstgenommen und angesprochen werden. Patienten, welche während der ersten Behandlungsphase nicht auf die Therapie ansprechen, haben ein etwas höheres Risiko, dass die Gesamtbehandlung „auffällig“ endet. Sie haben aber sehr gute Chancen von weiterer Therapie zu profitieren. Möglicherweise würde diesen Patienten eine längere Behandlung helfen, ihren anfänglichen „nonresponse“ wettzumachen. Ein Blick auf die therapeutische Beziehung kann hilfreich sein: falls deren Einschätzung in den dunkelgrauen Bereich fällt, könnte die Beziehung gestört sein. Wie Sie am besten auf häufige Suizidgedanken oder auf „nonresponse“ reagieren, hängt allerdings stark vom Einzelfall ab.

Ein Patient, welcher keine häufigen Suizidgedanken und eine Verbesserung seines Befindens angibt, sowie in keinen der erfassten psychotherapeutisch relevanten Bereiche hohe Krankheitswerte zeigt, wird **GRÜN** markiert. Bei diesen Patienten kann, unter Berücksichtigung der Stabilität seines guten Zustandes, eventueller besonderer Indikationen, welche das System nicht erfasst, und eventueller Verfälschungen, z.B. durch sozial erwünschte Antworten, eine Beendigung der Therapie in Erwägung gezogen werden.

Alle übrigen Patienten sind in schwarzer Farbe dargestellt. Diese Patienten geben also an, dass es ihnen besser geht, sie haben keine Suizidgedanken, sind aber nicht auf allen Skalen im funktionalen Bereich. Bei diesen Patienten werden keine Änderungen der Behandlung nahegelegt.

Sollte der Patient keine Angaben zu Suizidgedanken oder zur Ergebniseinschätzung gemacht haben, und kann deshalb keinen Farbcode erhalten, wird er mit dem Zusatz „k.A.m.“ (keine Auswertung möglich) angezeigt.

Weiterführende Informationen über die Patienten

Die Auswertung zeigt im ersten Teil die Ergebnisse an den Symptomskalen:

GBB: der Giesener Beschwerdebogen ist ein Verfahren zur Erfassung des körperlichen Beschwerdebildes von psychoneurotischen und psychosomatischen Patienten. Er dient dazu, zwischen organmedizinisch begründbarer, objektivierbarer

Symptomatik und subjektiven Beschwerden zu unterscheiden, da Diskrepanzen zwischen diesen Bereichen wichtige diagnostische Hinweise geben können.

OQ4: die deutsche Übersetzung des Outcome Questionnaire (Ergebnisbogen, EB45) ist ein speziell für die therapiebegleitende Veränderungsmessung konzipiertes Instrument, das als Entscheidungshilfe zur adaptiven Behandlungszuweisung herangezogen werden kann. Er erfasst drei wesentliche Aspekte psychotherapeutischer Veränderung:

1. **OQ1** – Symptombelastung – erfasst Angst- und Depressions- sowie Abhängigkeitserkrankungen mit engem Bezug zu DSM-III-R Kriterien.
2. **OQ2** – Interpersonelle Beziehungen – erfasst den Bereich interpersoneller Probleme, wie sie häufig in psychotherapeutischen Behandlungen geschildert werden.
3. **OQ3** – Soziale Integration – erfasst Unzufriedenheit, Belastung und Unzulänglichkeit mit den sozialen Rollen im Arbeits- und Familienleben sowie der Freizeit.

Für jede dieser Skalen wird der Percentilwert in 10%-Schritten angezeigt. Dieser informiert darüber, wo sich ein Patient im Vergleich zu der Bevölkerung befindet. Liegt der Percentilwert z.B. bei 6, so klagen 40% der Bevölkerung über mehr Beschwerden in diesem Bereich, und 60% geben weniger Beschwerden an. In welchem Percentil sich der Patient befindet, wird durch einen schwarzen Punkt markiert. Die Symptomskalen sind in einen grünen und einen roten Bereich unterteilt. Der grüne Bereich umfasst die „gesünderen“ 70% der Bevölkerung, der rote Bereich die 30% der Bevölkerung mit mehr Beschwerden. In der Spalte „Wert“ wird zusätzlich zu den Percentilen der Skalenrohwert angegeben. Dieser ist mit Hilfe einer Normtabelle interpretierbar. Falls der Patient nicht genügend Fragen beantwortet hat, damit der Skalenwert bestimmt werden kann, wird als Wert „k.a.“ angezeigt.

Unter den Symptomskalen werden Indikatoren für den Therapeutischen Prozess angezeigt:

Patienteneinschätzung: die allgemeine Ergebniseinschätzung aus Patientensicht ist ein Summenindex, welcher die vom Patienten eingeschätzte Zustandsveränderung im Vergleich zum Behandlungsbeginn in seiner seelischen Verfassung, seinem Allgemeinbefinden und seiner Leistungsfähigkeit abbildet.

Therapeutische Beziehung: als Maß für den psychotherapeutischen Prozess wird die Therapeutische Arbeitsbeziehung aus Sicht des Patienten erhoben.

Patientenzufriedenheit: erfragt wird die vom Patienten erlebte Behandlungsqualität.

Im Unterschied zu den Symptomskalen werden diese Skalen nicht als Percentile dargestellt. Die Skalierung ist gleichmäßig vom kleinstmöglichen bis zum größtmöglichen Wert. Die allgemeine Ergebniseinschätzung ist in einen grünen und einen roten Bereich unterteilt. Im grünen Bereich liegen Patienten, welche eher eine Verbesserung ihres Zustandes im Laufe der Therapie angeben, im roten Bereich solche die eine Verschlechterung oder keine Verbesserung angeben. Die therapeutische Arbeitsbeziehung und die Behandlungsqualität sind in einen hellgrauen und einen dunkelgrauen Bereich unterteilt. Im hellgrauen Bereich geben 70% aller Patienten Ihre Antworten ab, der dunkelgraue Bereich markiert die 30% der negativsten Antworten.

Falls der Patient auf die Frage „Ich habe Gedanke, mir das Leben zu nehmen“ mit „häufig“ oder „fast immer“ geantwortet hat, wird am Ende der Auswertung „SUIZIDGEDANKEN!“ angegeben.

Neben diesem Teil der Rückmeldung, welchen nur Sie erhalten, bekommt ein Teil der Patienten ebenfalls Rückmeldungen. Dabei handelt es sich um eine Verbalisierung der Skalenwerte. Diese Rückmeldung ist nicht ausgelegt, um einen therapeutischen Effekt zu erzielen, sie soll lediglich den Patienten für die aufgewendete Mühe belohnen und ihre

Motivation erhöhen, an der Entlassungsbefragung teilzunehmen. Dennoch kann es vorkommen, und durchaus stimulierend auf den therapeutischen Prozess wirken, dass Patienten diese Rückmeldung mit Ihnen besprechen. Damit Sie sich darauf vorbereiten können, bekommen Sie die Patientenrückmeldung ebenfalls ausgedruckt. Sie bekommen diese Rückmeldung für alle Patienten, beachten Sie bitte, dass nur Patienten mit einem schwarzen Häkchen hinter „Patientenauswertung“ auch eine erhalten haben.

TEXTBAUSTEINE FÜR DIE PATIENTENRÜCKMELDUNG

Hier erhalten Sie die Ergebnisse aus der Auswertung Ihres Fragebogens. Diese Auswertung gibt Ihnen nur grobe Hinweise, besprechen Sie deshalb Unklarheiten mit Ihrem Therapeuten.

Eins vorweg: Sie haben sehr gute Chancen, vom weiteren Verlauf der Behandlung zu profitieren

- Ø Ihre Gedanken, sich das Leben zu nehmen, sollten Sie unbedingt mit Ihrem Therapeuten besprechen.
- Ø Ihre Belastung durch Symptome konnte nicht ermittelt werden. Sie haben zu viele Fragen nicht beantwortet.
- Ø Die Auswertung zeigt, dass Sie sich stark durch Symptome belastet fühlen.
- Ø Die Auswertung zeigt, dass Sie sich nicht außergewöhnlich stark durch Symptome belastet fühlen.
- Ø Ihre psychischen Beschwerden konnten nicht ermittelt werden. Sie haben zu viele Fragen nicht beantwortet.
- Ø Sie geben viele psychische Beschwerden an.
- Ø Sie geben nicht außergewöhnlich viele psychische Beschwerden an.
- Ø Probleme bei zwischenmenschlichen Beziehungen konnten nicht ermittelt werden. Sie haben zu viele Fragen nicht beantwortet.
- Ø Zwischenmenschliche Beziehungen sind ein Problempunkt.
- Ø Zwischenmenschliche Beziehungen sind kein Problempunkt.
- Ø Probleme in Ihrem sozialen und beruflichen Umfeld konnten nicht ermittelt werden. Sie haben zu viele Fragen nicht beantwortet.
- Ø Sie geben an, im sozialen und beruflichen Umfeld Probleme zu haben.
- Ø Im sozialen und beruflichen Umfeld haben Sie keine außergewöhnlichen Probleme.
- Ø Es konnte nicht ermittelt werden wie sehr Sie unter körperlichen Beschwerden leiden. Sie haben zu viele Fragen nicht beantwortet.
- Ø Sie leiden unter vielen körperlichen Beschwerden.
- Ø Sie leiden nicht unter außergewöhnlich vielen körperlichen Beschwerden.
- Ø Veränderungen im Vergleich zum Behandlungsbeginn konnten nicht ermittelt werden. Sie haben zu viele Fragen nicht beantwortet.
- Ø Im Vergleich zum Behandlungsbeginn fühlen Sie sich nicht besser.
- Ø Im Vergleich zum Behandlungsbeginn fühlen Sie sich besser.