

Endbericht

—

Evaluierung der Einzelbehandlungen im Rahmen des Projekts fit2work klinisch-psychologische und psychotherapeutische Behandlung 3. Förderjahr (1.1.2016 bis 31.12.2017)

Reinhold Jagsch

Institut für Angewandte Psychologie:
Gesundheit, Entwicklung und Förderung

Fakultät für Psychologie
Universität Wien

Wien, Februar 2018

Inhaltsverzeichnis

Inhaltsverzeichnis	1
Abstract	2
1. Einleitung	4
2. Methoden	5
2.1. Zu prüfende Fragestellungen	5
2.2. Erhebungsinstrumente	6
2.2.1. Fragebogen zur Erfassung des krankheitsspezifischen Status	6
2.2.2. Beck-Depressions-Inventar BDI-II (Hautzinger, Keller & Kühner, 2006)	7
2.2.3. Brief Symptom Inventory BSI-53 (Franke, 2000)	7
2.2.4. Zufriedenheit mit Behandlung und Beziehung: Helping Alliance Questionnaire HAQ (Bassler, Potratz & Krauthauser, 1995)	9
2.3. Statistische Analyse	9
3. Darstellung der Stichprobe	11
3.1. Vergleich der soziodemographischen Charakteristika	12
3.2. Vergleich der krankheitsspezifischen Charakteristika	13
4. Ergebnisse zur Veränderungsmessung und zur Behandlungszufriedenheit	15
4.1. Veränderungen in der Nutzung des Gesundheitssystems und im Gesundheitsverhalten	15
4.2. Veränderungen in den validierten klinischen Fragebögen	18
4.3. Bewertung der Behandlungszufriedenheit mittels HAQ	19
5. Dropout-Analyse	19
6. Diskussion	21
Literatur	26

Abstract

Theoretischer Hintergrund. Auf Basis der österreichischen Gesetzeslage mit einerseits Psychologengesetz (Bundesministerium für Gesundheit, 1990b, 2013) und andererseits Psychotherapiegesetz (Bundesministerium für Gesundheit, 1990a) sind Klinische Psychologinnen und Psychologen, Psychotherapeutinnen und Psychotherapeuten sowie Behandlerinnen und Behandler mit Doppelqualifikation berechtigt, Personen mit psychischen Störungen zu behandeln, bislang liegen dazu aber keine Vergleichsstudien vor. Im Zuge der Evaluation von fit2work psychologische und psychotherapeutische Behandlung wurde erstmalig ein empirischer Vergleich der Wirksamkeit dieser drei Behandlungsformen durchgeführt.

Methode. Österreichweit nahmen insgesamt 960 Personen, deren Arbeitsplatz entweder durch gesundheitliche Probleme gefährdet war bzw. bei denen der Arbeitsplatz bereits verlorengegangen war, an diesem Projekt teil und wurden nach einer umfangreichen Anamnese einer der drei Behandlungsformen zugewiesen. Die Klientinnen und Klienten erhielten zu Beginn der Behandlung eine Testbatterie, die Fragebögen zu Depression (Beck-Depressions-Inventar-II), zur psychischen Beeinträchtigung (Brief Symptom Inventory-53) und einen Fragebogen zum arbeits- und gesundheitsbezogenen Status enthielt. Nach Abschluss der Behandlung wurden sie gebeten, diese Testbatterie noch einmal auszufüllen und die Zufriedenheit der Behandlung mittels Helping Alliance Questionnaire (HAQ) zu beurteilen. Ergänzt wurde die subjektive Einschätzung der Klientinnen und Klienten durch die Fremdeinschätzung der Behandlerinnen und Behandler, die auch eine Beurteilung mittels GAF-Skala (Global Assessment of Functioning) beinhaltete.

Ergebnisse. Insgesamt konnten 318 Klientinnen und Klienten, die sich bezüglich soziodemographischer Kennwerte nicht voneinander unterschieden, in die Vergleichsanalyse einbezogen werden, von denen 154 Personen klinisch-psychologische Behandlung (48.4%), 62 Psychotherapie (19.5%) und 102 Behandlung durch

doppelqualifizierte Behandlerinnen und Behandler (32.1%) bekamen. In den klinischen Skalen des BSI-53 lagen die Klientinnen und Klienten der drei Behandlungsformen vor dem Beginn der Intervention mehrheitlich über dem Cutoff-Wert¹ für Behandlungsbedürftigkeit, nach Abschluss der Behandlung in keiner einzigen Subskala. Neben einem hoch signifikanten Effekt der Dauer der Behandlung konnten im Rahmen der Veränderungsmessung simultan weder signifikante Gruppeneffekte noch signifikante Interaktionen gefunden werden. Auch die Modelle für die Depressionsstärke, die GAF-Skala² und die Wirkungsanalysen für die Angaben von Beeinträchtigung von Leistungsfähigkeit in Beruf und Alltag erbrachten analoge Effektkonstellationen. Der Anteil von Klientinnen und Klienten, die ohne Erwerbstätigkeit waren, der anfangs bei 65.6% lag, konnte durch die Intervention nicht bedeutsam verringert werden. Die Interventionsmaßnahme führte aber mehrheitlich, ohne Unterschiede zwischen den Behandlungsformen, zu massiven Reduktionen der Inanspruchnahme von Gesundheitsleistungen, es konnten bedeutsame Reduktionen bei Arztbesuchen, bei ärztlichen Untersuchungen, bei Krankenhausaufenthalten und bei der Einnahme von Psychopharmaka verzeichnet werden, das Gesundheitsverhalten (Sport, Bewegung, Ernährung) konnte positiv beeinflusst werden. Beziehungs- und Erfolgswzufriedenheit scorten in allen drei Behandlungsformen hoch, sowohl aus Sicht der Klientinnen und Klienten als auch aus Sicht der Behandlerinnen und Behandler.

Schlussfolgerungen. Im Bereich des Projekts fit2work psychologische und psychotherapeutische Behandlung erwiesen sich die drei Behandlungsformen als gleichwertig und hoch wirksam. Die Interventionen konnten zu hoch signifikanten Reduktionen in den klinischen Skalen und den Belastungen auf individueller Ebene führen, wie auch auf gesamtgesellschaftlicher Ebene zu einer massiven Reduktion der Inanspruchnahme von Gesundheitsleistungen. Gleichzeitig wurden bedeutsame Verbesserungen im gesundheitsbezogenen Bereich berichtet, verbunden mit einer hohen Behandlungs- und Erfolgswzufriedenheit sowohl auf Seiten der Klientinnen und Klienten als auch auf Seiten der Behandlerinnen und Behandler.

¹ Trennwert, bei dessen Überschreitung Behandlungsbedürftigkeit angenommen wird

² Fremdeinschätzungsskala von 1 bis 100 für Funktionsfähigkeit in Fertigkeiten des täglichen Lebens (Global Assessment of Functioning)

1. Einleitung

Fit2work psychologische und psychotherapeutische Behandlung ist eine Initiative des Bundesministeriums für Arbeit, Soziales, Gesundheit und Konsumentenschutz sowie der Pensionsversicherungsanstalt. Das Projekt auf Grundlage des Arbeit-und-Gesundheit-Gesetzes wird aus Mitteln der Gebarung Arbeitsmarktpolitik und der Pensionsversicherung finanziert und seit Juni 2013 flächendeckend in ganz Österreich durchgeführt, die Koordination erfolgt durch das Sozialministeriumservice. Ziel dieser Initiative ist es, Personen, deren Arbeitsplatz durch gesundheitliche (psychische und physische) Probleme gefährdet ist, zu unterstützen, zu fördern und dadurch die Arbeitsfähigkeit zu erhalten, damit ein Jobverlust verhindert werden kann. Bei Personen mit langen Krankenständen bzw. bei solchen, bei denen der Arbeitsplatz bereits verloren gegangen ist, soll die Wiedereingliederung in den Arbeitsprozess unterstützt werden. Die Teilnahme an fit2work psychologische und psychotherapeutische Behandlung ist freiwillig, das Angebot niederschwellig und individualisiert auf den Einzelnen/die Einzelne zugeschnitten.

Da bei klinisch-psychologischen und psychotherapeutischen Behandlungen eine Versorgungslücke festgestellt wurde, wurde österreichweit eine professionelle Versorgungsstruktur aufgebaut, die die Klientinnen und Klienten der geeigneten Behandlung zuführen soll. Insgesamt stehen in allen Bundesländern österreichweit 18 Steuernde Psychologinnen und Psychologen (Doppelqualifikation aus Klinischer Psychologie und Psychotherapie) zur Verfügung, die mit den Klientinnen und Klienten ein Erstgespräch bezüglich der individuellen Problemlage durchführen. Nach dem Erstgespräch wird von den Steuernden Psychologinnen und Psychologen entschieden, ob im Einzelfall klinisch-psychologische Behandlung oder Psychotherapie indiziert ist, je nach Schwere der gesundheitlichen Problemlage wird ein entsprechendes Behandlungskontingent empfohlen. Für die direkte Umsetzung der Behandlung stehen insgesamt 236 Behandlerinnen und Behandler (Klinische Psychologinnen und Psychologen, Psychotherapeutinnen und Psychotherapeuten und solche mit Doppelqualifikation) zur Verfügung. Die Gesamtzahl der Klientinnen und Klienten für die aktuelle Studie war mit 960 festgelegt, für die Aufteilung

nach Bundesland lag ein österreichweiter Quotenplan vor.

2. Methoden

2.1. Zu prüfende Fragestellungen

Inhalt der Evaluierung von fit2work psychologische und psychotherapeutische Behandlung ist es, im Rahmen der Studie die Wirksamkeit der Maßnahme im Sinne einer Ergebnisevaluation zu prüfen. Basis für diesen Vergleich bilden alle Klientinnen und Klienten, die alle vorab festgelegten Behandlungseinheiten regulär absolviert und die vorgesehenen Fragebögen zu beiden Zeitpunkten (Behandlungsbeginn, Behandlungsabschluss) ausgefüllt und an das Evaluationsteam retourniert haben (Completer-Analyse). Gleichzeitig werden auch Daten der Behandlerinnen und Behandler in die Analyse miteinbezogen, die die dafür vorgesehenen Instrumente der Fremderhebung via online-Erhebung bearbeitet haben. Teil der Evaluation ist auch eine Dropout-Analyse, um Charakteristika der Ausgangsstichprobe in Bezug auf die Repräsentativität der Gesamtstichprobe zu erheben.

Ziel der Evaluation ist es, folgende Fragestellungen auf Basis der erhobenen empirischen Daten zu prüfen:

1. Veränderungen definierter Outcome-Variablen (Depression, psychische Beeinträchtigung, Funktionsfähigkeit, arbeits- und gesundheitsbezogene Variablen) im prä/post-Vergleich vor und nach Abschluss der Behandlung
2. Gleichwertigkeit der Behandlungsergebnisse zwischen Klinischen Psychologinnen und Psychologen, Psychotherapeutinnen und Psychotherapeuten und Behandlerinnen und Behandlern mit Doppelqualifikation
3. Prüfung von Interaktionseffekten (Wechselwirkung von Behandlungsart und Zeit)
4. Vergleich der Beziehungs- und Erfolgsszufriedenheit mit der Behandlung zwischen den Behandlungsformen auf Ebene der Behandelten und der Behandlerinnen und Behandler
5. Prüfung der Gesamteffekte bezüglich Inanspruchnahme von Gesundheitsleistungen

2.2. Erhebungsinstrumente

Für die aktuelle Erhebung kamen vier verschiedene Fragebögen zum Einsatz, ein speziell für diese Studie von der Academy for Value in Health GmbH erstellter Fragebogen zur Erfassung des krankheitsspezifischen Status, der in einer Selbstbeurteilungsform für die Klientinnen und Klienten und einer Fremdbeurteilungsform für die Behandlerinnen und Behandler vorliegt, und zwei validierte klinische Fragebögen zur Erhebung der Stärke der Depression (Beck-Depressions-Inventar) und der psychischen Beeinträchtigungen (Brief Symptom Inventory) in Selbstbeurteilungsform. Zusätzlich wurde nach Abschluss der Behandlung der Fragebogen Zufriedenheit mit Behandlung und Beziehung: Helping Alliance Questionnaire (HAQ) in Selbst- und Fremdbeantwortungsform den Klientinnen und Klienten und den Behandlerinnen und Behandlern vorgegeben, um die Zufriedenheit mit dem Behandlungsergebnis von beiden Seiten abzuklären.

2.2.1. Fragebogen zur Erfassung des krankheitsspezifischen Status

Dieser Fragebogen dient dazu, einerseits soziodemographische (Alter, Geschlecht, Familienstand, höchste abgeschlossene Ausbildung, aktuelles Arbeitsverhältnis etc.) und andererseits arbeits- und krankheitsspezifische Daten der Klientinnen und Klienten (Zeitpunkt der Diagnosestellung, Beurteilung der Schwere der psychischen Belastung für den Bereich Arbeit und alltägliche Tätigkeiten, Umfang der Inanspruchnahme von Gesundheitsleistungen vor Aufnahme in die aktuelle Behandlung, Einnahme von Psychopharmaka etc.) zu erheben. Die Fremdbeurteilungsform, die in Form einer online-Befragung vorgegeben wurde, dient dazu, die durch die Behandlerinnen und Behandler erhobenen Diagnosen zu erfragen, die Bereiche der Beeinträchtigungen zu beschreiben, die verschriebenen Medikamente aufzulisten sowie den Grad der Funktionseinschränkungen zu objektivieren.

Diagnosen wurden nach der aktuellen Version 10 der International Classification of Diseases (ICD) der Weltgesundheitsorganisation vergeben, die die Vergabe von zusätzlichen komorbiden Diagnosen erlaubt, d.h. den zu Diagnostizierenden können mehrere Diagnosen zugeordnet werden. Der Grad der Funktionseinschränkungen der Klientinnen und Klienten

wurde von den Behandlerinnen und Behandlern mittels des Global Assessment of Functioning (GAF-Skala) eingeschätzt. Mit Hilfe normierter Vorgaben von Beschreibungen von Einschränkungen können so Werte auf einer Skala von 1 bis 100 zur Beurteilung des aktuellen Funktionsniveaus getätigt werden, wobei 1 den höchsten Grad der Einschränkung und 100 keinerlei Einschränkung des Funktionsniveaus bedeutet.

2.2.2. Beck-Depressions-Inventar BDI-II (Hautzinger, Keller & Kühner, 2006)

Das BDI-II ist ein objektives, reliables und valides Instrument zur Erfassung der Schwere depressiver Symptomatik in Stichproben aus der Normalpopulation sowie aus klinischen Populationen. Die interne Konsistenz des Verfahrens wird für eine klinische Population mit Cronbach's Alpha .93 und für eine gesunde mit .90 angegeben. Das BDI-II als Selbstbeurteilungsverfahren beinhaltet 21 verschiedene Gruppen von Aussagen, die aus Sätzen bestehen, die typische depressive Symptome beschreiben. Die Aussagen umfassen jeweils vier (in den Inhaltsbereichen Appetit und Schlaf sieben) Sätze pro Gruppe und sind nach aufsteigender Schwere der Symptome von Aussage eins (0 = nicht vorhanden) bis Aussage vier (3=stark vorhanden) geordnet.

Die Befragten sollen in jeder der 21 Gruppen die Aussage ankreuzen, die am besten beschreibt, wie sie sich in den vergangenen zwei Wochen einschließlich des Testtages selbst gefühlt haben. Kreuzen sie bei einer Gruppe von Aussagen mehrere Aussagen an, so geht die Aussage mit der höchsten Zahl in den Summenwert ein. Der Summenwert aller Items, der zwischen 0 und 63 Punkten liegen kann, bildet den Gesamtwert der Stärke der depressiven Verstimmung. Summenwerte unter 14 Punkten werden als unauffällig interpretiert, Werte zwischen 14 und 19 weisen auf eine leichte Depression, Werte von 20 bis 28 auf eine mittelschwere Depression und Werte von 29 bis 63 Punkten auf eine schwere Depression hin.

2.2.3. Brief Symptom Inventory BSI-53 (Franke, 2000)

Das BSI ist eine der Kurzformen der Symptom-Checkliste SCL-90-R. Mit Hilfe dieses Selbstbeurteilungsverfahrens können psychische Beeinträchtigungen über die letzten sieben Tage durch die Beurteilung von insgesamt 53 Items erfasst werden. Die interne

Konsistenz der Subskalen liegt mit Cronbach's Alpha zwischen .71 bis .85 im guten bis sehr guten Bereich. Die individuelle Stärke der psychischen Beeinträchtigung wird dabei mittels fünfstufiger Likert-Skala (0=überhaupt nicht, 1=ein wenig, 2=ziemlich, 3=stark, 4=sehr stark) beurteilt.

Der Fragebogen enthält insgesamt neun Subskalen:

- 1) *Somatisierung* (7 Items): Ausmaß von psychischen Beeinträchtigungen, die durch die Wahrnehmung von körperlichen Dysfunktionen entstehen
- 2) *Zwanghaftigkeit* (6 Items): Umfasst Gedanken, Impulse und Handlungen, die von der Person als konstant vorhanden, unveränderbar sowie als ich-fremd erlebt werden
- 3) *Unsicherheit im Sozialkontakt* (4 Items): Umfasst Fragen von leichter sozialer Unsicherheit bis hin zum Gefühl völliger persönlicher Unzulänglichkeit
- 4) *Depressivität* (6 Items): Umfasst Traurigkeit bis hin zu schwerer Depressivität
- 5) *Ängstlichkeit* (6 Items): Umfasst körperlich spürbare Nervosität bis hin zu tiefer Angst
- 6) *Aggressivität/Feindseligkeit* (5 Items): Umfasst Reizbarkeit und Unausgeglichenheit bis hin zu aggressiven Durchbrüchen
- 7) *Phobische Angst* (5 Items): Umfasst ein leichtes Gefühl von Bedrohung bis hin zu massiven phobischen Ängsten
- 8) *Paranoides Denken* (5 Items): Umfasst Misstrauen und Gefühle der Minderwertigkeit bis hin zu starkem paranoiden Denken
- 9) *Psychotizismus* (5 Items): Umfasst das Gefühl von Isolation und Entfremdung bis hin zu Psychosen

Neben diesen einzelnen Skalen besteht auch die Möglichkeit, einen Globalwert (GSI) zu berechnen, der die grundsätzliche psychische Beeinträchtigung misst. Die Ausprägungen der Stärkegrade auf Einzelitemniveau werden jeweils aufaddiert und bilden so die Rohwerte für die Subskalen und den Globalwert. Anschließend können die Rohwerte der Subskalen und des Globalwertes, die unterschiedliche Itemanzahl und daher unterschiedliche Spannweiten

aufweisen, mittels Normierung in *T*-Werte umgewandelt werden. *T*-Werte sind Werte einer Normwertskala, die auf einen Mittelwert (*M*) von 50 und eine Standardabweichung (*SD*) von 10 hin normiert sind, d.h. dass alle Werte im Umfang von maximal drei Standardabweichungen rund um den Mittelwert von 50 im Gesamtbereich von 20 bis 80 abgebildet werden können. Für das BSI liegen darüber hinaus für jede Subskala und den Globalwert *T*-Wert-Normen getrennt für Männer und Frauen vor, d.h. die Unterschiedlichkeit der Ausprägungen auf geschlechtsspezifischem Niveau, die für einige psychopathologische Bereiche vorliegt, kann damit ausgeschaltet werden. Laut Autorin des Verfahrens stellt ein *T*-Wert im GSI von 63 aufwärts eine Indikation für Behandlungsbedürftigkeit dar, das selbe gilt, wenn in mindestens zwei Subskalen des BSI ein *T*-Wert von 63 erreicht oder überschritten wird.

2.2.4. Zufriedenheit mit Behandlung und Beziehung: Helping Alliance Questionnaire HAQ (Bassler, Potratz & Krauthauser, 1995)

Das HAQ umfasst elf Fragen, mit deren Hilfe die Qualität der therapeutischen Beziehung von den Klientinnen und Klienten eingeschätzt werden kann. In der deutschen Übersetzung existiert darüber hinaus eine Fremdeinschätzungsform, mit der mit entsprechend umformulierten Fragen auch die therapeutenseitige Einschätzung erhoben werden kann. Auf Basis faktorenanalytischer Untersuchungen lassen sich neben dem Gesamtwert der Zufriedenheit (Cronbach's Alpha: .91) zwei Subskalen unterscheiden: Skala 1 (6 Items, Cronbach's Alpha: .95) erhebt die Beziehungszufriedenheit, Skala 2 (5 Items) die Erfolgsmzufriedenheit (Cronbach's Alpha: .91). Jedes der elf Items kann mit Hilfe einer Likert-Skala (von 1=sehr unzutreffend bis 6=sehr zutreffend) beurteilt werden.

2.3. Statistische Analyse

Zur Beschreibung der Stichprobe werden deskriptivstatistisch für kontinuierliche Daten Mittelwert (*M*), Standardabweichung (*SD*) und Spannweite (Minimum; Maximum) verwendet, Häufigkeitsdaten werden mittels absoluten und relativen Häufigkeiten beschrieben. Für die Vergleiche von Gruppenmittelwerten kommen statistische Verfahren zum Einsatz, die auf Basis von Varianzzerlegungen Aussagen darüber erlauben, ob

Unterschiede zwischen verschiedenen Gruppen als bedeutsam (signifikant) beurteilt werden können. Für die Veränderungsmessung bei kontinuierlichen Daten wurde nach Abklärung der notwendigen Voraussetzungen (Normalverteilung der Teilstichproben, Homogenität der Varianzen und der Kovarianzen) das Allgemeine Lineare Modell (ALM) für Messwiederholungen angewendet. Dieses ist in der Lage, gleichzeitig den Haupteffekt Zeit (Veränderungen prä/post), den Haupteffekt Gruppe (Effekt der Behandlungsform) sowie die Wechselwirkung zwischen Zeit und Gruppe zu überprüfen. Um einer Kumulation des Alphafehlers³ entgegenzuwirken, kam bei signifikanten Ergebnissen die Alphafehler-Korrektur nach Bonferroni-Holm zum Einsatz.

Vergleiche bei Häufigkeitsdaten wurden mit dem χ^2 -Test nach Pearson oder mittels Fisher's Exact Test durchgeführt. Um Veränderungen von dichotomen Daten (z.B. beschäftigt vs. beschäftigungslos) analysieren zu können, wurde der McNemar-Test verwendet, der die Häufigkeit von Wechslern in gegenläufige Richtungen (z.B. von beschäftigungslos zu in Beschäftigung vs. von in Beschäftigung zu beschäftigungslos) in Beziehung setzen kann (bei Konstanthaltung der Personen, die zu beiden Zeitpunkten den gleichen Beschäftigungsgrad aufweisen). Im Rahmen einer Dropout-Analyse wurden im ersten Schritt die maßgeblichen Variablen mit entsprechenden statistischen Methoden auf signifikante Unterschiede hin geprüft, im zweiten Schritt wurden die dabei ausfindig gemachten Variablen einer binär-logistischen Regressionsanalyse zugeführt, um Prädiktoren für die Frage ausfindig zu machen, welche charakteristischen Variablen vorhersagen können, welche Klientinnen und Klienten trotz vorhandener Ersttestung nicht an der Abschlusstestung teilgenommen haben.

Für alle statistischen Analysen wurde ein Signifikanzniveau von $p < .05$ festgelegt. Aufgrund der Tatsache, dass Angaben im Sinne des ausschließlichen Anführens von Kennwerten zur Absicherung statistischer Signifikanz bei stark unterschiedlich großen Teilstichproben zu verzerrten Ergebnissen führen kann, empfiehlt es sich, zusätzlich Kennwerte der inhaltlichen Relevanz (Effektstärken) ergänzend anzugeben (Bortz & Döring, 2006). Diese haben den

³ Kumulation des Alphafehlers bedeutet, dass aufgrund von mehrfachen Analysen (z.B. bei mehreren Subskalen eines Testverfahrens) die Tendenz Richtung Signifikanz potentiell ansteigt, der mittels Methoden zur Alphafehler-Korrektur entgegengewirkt werden muss.

Vorteil, dass deren Ausmaß völlig unabhängig von den vorhandenen Stichprobengrößen interpretiert werden kann. Für Vergleiche von zwei Stichproben wird Cohen's d verwendet (Cohen, 1988). Für diese Art der Effektstärke gilt:

Werte bis 0.20	kein Effekt
Werte von 0.21 bis 0.50	kleiner Effekt
Werte von 0.51 bis 0.80	moderater Effekt
Werte ab 0.81	großer Effekt

Für Vergleiche von mehr als zwei Stichproben wird als Effektstärkemaß Eta^2 verwendet (Cohen, 1988). Für diese Art der Effektstärke gilt:

Werte bis 0.010	kein Effekt
Werte von 0.011 bis 0.060	kleiner Effekt
Werte von 0.061 bis 0.140	moderater Effekt
Werte ab 0.141	großer Effekt

3. Darstellung der Stichprobe

Insgesamt sind 648 ausgefüllte Fragebogensets der Ersttestung (Beginn der Behandlung) beim Evaluatorenteam der Fakultät für Psychologie der Universität Wien eingelangt. Die Quote der Beteiligung an der Evaluation der Ersttestung beträgt damit 67.5%. Die Zahl an ausgefüllten Fragebogensets der Zweittestung, also nach Ende einer abgeschlossenen Behandlung, die eingelangt sind, betrug 372 (38.8%, gemessen an der Gesamtstichprobe von 960 Klientinnen und Klienten). Die gemeinsame Schnittmenge an Fragebogensets von Erst- und Abschlusstestungen umfasste insgesamt 318 Datensätze, das entspricht etwa einem Drittel (33.1%) der vollständigen Stichprobe von 960 Personen. Von Seiten der Behandlerinnen und Behandler lagen insgesamt 504 Online-Fremdbeurteilungen vor, das entspricht einer Quote von 52.5%. Die gemeinsame Schnittmenge dieser Erhebungsarten (verwertbare Fragebögen der Klientinnen und Klienten zu Erst- und Abschlussbefragung sowie dazugehörige Fremdbeschreibung durch die Behandlerinnen und Behandler zu Beginn

der Behandlung) beträgt 181 (18.9%, gemessen an der Zahl der Gesamtstichprobe). Die einzelnen Teilmengen und deren Überschneidungen sind zusammengefasst in Abbildung 1 (Anhang A) dargestellt.

Bezüglich des differentiellen Vergleiches nach dem Status der Behandlerinnen und Behandler ist zu sagen, dass die ursprünglich vorgesehene Teilgruppe 4 (Behandlerinnen und Behandler mit Doppelqualifikation, aber erst in Ausbildung) mit $n=29$ und Teilgruppe 5 (Psychotherapeutinnen und Psychotherapeuten in Ausbildung) mit $n=12$ als zu gering besetzt waren, als dass ein Fünf-Gruppen-Vergleich sinnvoll gewesen wäre. Die beiden Teilgruppen wurden daher entsprechend der Gruppe mit Doppelqualifikation bzw. der Gruppe der Psychotherapeutinnen und Psychotherapeuten zugeordnet. Alle statistischen Analysen dieses Endberichts beziehen sich daher auf den Drei-Gruppen-Vergleich Klinische Psychologinnen und Psychologen versus Psychotherapeutinnen und Psychotherapeuten versus Behandlerinnen und Behandler mit Doppelqualifikation. Im ersten Schritt wurde ein Vergleich der Ausgangslage (Ersterhebung) mit dem umfangreicheren Datensatz durchgeführt, der mit den Analysen zur Veränderungsmessung im zweiten Schritt komplettiert wurde. Die Gruppengrößen betragen für die Ersttestung für Behandlungen durch Klinische Psychologinnen und Psychologen $n=281$, durch Psychotherapeutinnen und Psychotherapeuten $n=137$ und durch Behandlerinnen und Behandler mit Doppelqualifikation $n=203$. Die analogen Gruppengrößen für die Veränderungsmessungen betragen für Variablen, die von der gesamten Stichprobe beantwortet wurden, $n=154$, $n=63$ und $n=102$ in oben angeführter Reihenfolge.

3.1. Vergleich der soziodemographischen Charakteristika

Nach Abzug derjenigen Bögen, die entweder durch mangelnde oder unvollständige Codierung für diese Analyse ausgeschlossen werden mussten, verblieben insgesamt 590 Datensätze. Die Stichprobe der Ersttestung lässt sich folgendermaßen charakterisieren: Der Großteil war weiblich ($n=372$, 62.9%), das Durchschnittsalter lag bei 44.31 Jahren ($SD=9.77$, Range 19–67). Der vorherrschende Zivilstatus war verheiratet ($n=259$, 44.1%), gefolgt von ledig ($n=193$, 32.8%), geschieden ($n=127$, 21.6%) und verwitwet ($n=9$, 1.5%). Vom höchsten Bildungsstand herrschte „Berufsschule/Lehre“ vor ($n=236$, 40.1%). In Bezug auf drei

soziodemographische Charakteristika (Geschlecht, Alter, Ausbildung) unterschieden sich die Behandlungsformen nicht signifikant voneinander ($p > .05$). Lediglich bezüglich Familienstand ($\chi^2(6) = 16.953, p = .009$) zeigte sich, dass verheiratete Klientinnen und Klienten häufiger der klinisch-psychologischen Behandlungsform zugeordnet wurden und ledige häufiger der Psychotherapie.

Was die berufsspezifischen Daten anlangt, waren insgesamt 387 Klientinnen und Klienten (65.6%) zum Zeitpunkt der Ersterhebung arbeitslos, von denen etwa zwei Drittel angaben, aufgrund von psychischen Beschwerden arbeitslos geworden zu sein. Die Durchschnittsdauer der Arbeitslosigkeit wurde mit 21.25 Monaten ($SD = 32.92$, Median = 9 Monate, Spannweite 1–220 Monate) angegeben. Auch in Bezug auf die arbeitslosenspezifischen Daten zeigten sich keine statistisch signifikanten Unterschiede zwischen den drei Behandlungsformen ($p > .05$).

3.2. Vergleich der krankheitsspezifischen Charakteristika

Von insgesamt 372 validen Datensätzen lagen Diagnosen aus den Online-Fragebögen der Behandlerinnen und Behandler vor (Diagnosestellung laut International Classification of Diseases, ICD-10). Der Anteil der Klientinnen und Klienten mit Mehrfachdiagnosen ($n = 188$, 50.5%) lag schwach höher als derjenige mit lediglich einer einzigen Diagnose. Den Angaben zufolge konnten drei Gruppen von Primärdiagnosen unterschieden werden: Affektive Störungen (F3-Diagnosen) standen an erster Stelle ($n = 205$, 55.1%), gefolgt von Störungen aus der F4-Sektion (Neurotische, Belastungs- und somatoforme Störungen, $n = 163$, 43.8%) und Störungen bezüglich Z73 (Probleme bei der Lebensbewältigung, Burnout, $n = 44$, 11.8%). Es konnten keine statistisch signifikanten Unterschiede zwischen den drei Behandlungsformen eruiert werden ($p > .05$). Die Frage, wie lange die Erstdiagnose zurückliegt, wurde von den Klientinnen und Klienten sehr unterschiedlich beantwortet. Die Minimaldauer lag bei einem Monat, die Maximaldauer bei 399 Monaten, der Median bei neun Monaten. Ähnlich heterogen war das Antwortmuster bezüglich der Latenz vom Wahrnehmen der ersten Symptome bis zur Stellung der Erstdiagnose. Der Median lag bei 18 Monaten (Spannweite 1–420). Auch bezüglich dieser beiden Zeitmaße ließen sich keine Unterschiede zwischen den Behandlungsformen ausmachen ($p > .05$).

Zu Variablen, die die Belastungen der Behandlungspopulation vor Beginn der Behandlung charakterisieren, lassen sich folgende Befunde nennen: Von den Klientinnen und Klienten, die zur Zeit des Beginns der Behandlung in einem aufrechten Arbeitsverhältnis waren, gaben 67.5% an, mindestens einmal in den letzten drei Monaten im Krankenstand gewesen zu sein. Insgesamt 367 aller befragten Klientinnen und Klienten (63.7%) antworteten, Medikamente wegen psychischer Beschwerden einzunehmen, 74.7%, hatten im Zeitraum der letzten drei Monate mindestens einmal einen Arzt aufgesucht (Spannweite: 1 bis 30 Besuche), 68.1% ließen in diesem Zeitraum medizinische Untersuchungen vornehmen (Spannweite: 1 bis 30 Untersuchungen), und 10.9% hatten in den zurückliegenden drei Monaten Krankenaufenthalte (Spannweite: 1 bis 9 Aufenthalte). Die Auswirkungen der psychischen Beschwerden auf die täglichen Aktivitäten, die auf einer Skala von 0 (kein Einfluss) bis 10 (maximaler Einfluss) abgefragt wurden, erbrachte für die Erhebung zu Beginn der Behandlung einen mittleren Wert von 6.28 (SD=2.33) für die Gesamtstichprobe. Vergleicht man diese abgefragten Maße zwischen den Behandlungsformen, so konnten auch hier keine signifikanten Unterschiede festgestellt werden (Krankenstände: $Chi^2(2)=5.566$, $p=.062$; Medikamenteneinnahme: $Chi^2(2)=1.308$, $p=.520$; Arztbesuche: $F(2)=0.360$, $p=.698$; medizinische Untersuchungen: $F(2)=0.284$, $p=.753$; Krankenhausaufenthalte: $F(2)=1.983$, $p=.140$; Selbstbewertung der psychischen Belastung: $F(2)=0.628$, $p=.534$).

Durch Vorgabe der Fragebogenverfahren Beck-Depressions-Inventar (BDI-II) sowie des Brief Symptom Inventory (BSI-53) wurde die Stärke der psychischen Beeinträchtigung vor Beginn der Erhebung erhoben. Bezüglich Stärke der Depression zeigte sich für die Gesamtstichprobe ein Summenscore von 27.83 (SD=11.07, Range 0–57), der als Indikator für eine mittelschwere Form einer depressiven Verstimmung bewertet werden kann. Der Globalwert des BSI (GSI) erbrachte mit einem mittleren T -Wert von 71.60 (SD=10.07, Range 31–80) für die Gesamtstichprobe zu Beginn der Behandlung einen Score, der mehr als zwei Standardabweichungen vom theoretischen Mittelwert von 50 abweicht (Scores über $T=63$ gelten als behandlungsbedürftig). Mittels Online-Erhebung bewerteten die Behandlerinnen und Behandler die Funktionsfähigkeit ihrer Klientinnen und Klienten mittels Global Assessment of Functioning (GAF-Skala), die einen Mittelwert von 53.69 (SD=11.04, Range 18–89) erbrachte, ein Wert, der laut Manual „moderaten Funktionsschwierigkeiten“

entspricht. Auch hier zeigte der Gruppenvergleich keine statistisch signifikanten Unterschiede zwischen den drei Behandlungsformen, weder im BDI-II ($F(2)=0.011, p=.989$) noch im GSI ($F(2)=0.083, p=.920$) noch in der GAF-Skala ($F(2)=1.864, p=.157$).

Zusammenfassend kann zur Charakterisierung der Ausgangsstichprobe gesagt werden, dass es sich um eine stark belastete und behandlungsbedürftige Population handelt. Für die differentielle Überprüfung der Effektivität der Behandlung ist die Ausgangslage insofern positiv zu bewerten, als die Zuordnung der Klientinnen und Klienten zu den drei zum Vergleich herangezogenen Behandlungsformen in den überprüften soziodemographischen (ausgenommen Familienstand) und krankheitsspezifischen Variablen insignifikante Ergebnisse erbracht hat, also von als gleich zu bewertenden Ausgangsniveaus gesprochen werden kann.

4. Ergebnisse zur Veränderungsmessung und zur Behandlungszufriedenheit

Die Steuernden Psychologinnen und Psychologen wiesen den Behandlerinnen und Behandlern ein Kontingent von maximal jeweils 30 Behandlungseinheiten zu, das von diesen je nach Behandlungsstand flexibel angepasst werden konnte. Die durchschnittlich verbrauchte Anzahl an Einheiten lag bei 24.23 (SD=7.81, Range 1–30), die statistische Analyse zeigte diesbezüglich keine bedeutsamen Unterschiede zwischen den Behandlungsformen ($F(2)=1.854, p=.158$).

4.1. Veränderungen in der Inanspruchnahme von Gesundheitsleistungen und im Gesundheitsverhalten

Bezüglich der verschiedenen Formen der Inanspruchnahme von Gesundheitsleistungen wurden im Rahmen der Abschlusserhebung massive Änderungen berichtet, wobei in allen vier untersuchten Variablen Arztbesuche, medizinische Untersuchungen, Krankenhausaufenthalte und Medikamenteneinnahme positive Ergebnisse zugunsten einer Reduktion zu verzeichnen waren. Tabelle 1 (Anhang B) zeigt die Änderungen für die Gesamtstichprobe kompakt zusammengefasst im Überblick. Unterschiedlich hohe

Fallzahlen ergeben sich durch unvollständige Beantwortung einzelner Items zum Fragenkomplex. Zwischen den drei Behandlungsformen konnten keine signifikanten Unterschiede erhoben werden, weder für die Veränderungen der Variable Arztbesuche ($Chi^2(4)=4.301$, $p=.367$) noch bezüglich der Veränderungen der medizinischen Untersuchungen ($Chi^2(4)=5.074$, $p=.280$), der Krankenhausaufenthalte ($Chi^2(4)=3.326$, $p=.505$) oder der Medikamenteneinnahme ($Chi^2(4)=2.474$, $p=.649$).

Die Frage nach generellen Veränderungen im Gesundheitsverhalten während der Zeit der Behandlung beantworteten 361 Personen (76.5%) als positiv. Tabelle 2 (Anhang B) zeigt die Bewertungen für die einzelnen Bereiche Sport, Bewegung, Ernährung und Rauchverhalten im Überblick. Auch hier überwiegen die positiven Bewertungen in allen abgefragten Teilbereichen, einzig in der Frage Rauchverhalten zeigt die Kategorie „keine Änderung“ die höchsten Häufigkeiten, bedingt auch dadurch, dass der Großteil Nichtraucher war. 120 der arbeitstätigen Klientinnen und Klienten (79.5%) berichteten zudem von einer allgemeinen Verbesserung ihrer Leistungsfähigkeit durch die Behandlung, 90 (59.2%) von einer Reduktion der Krankenstände. In den abgefragten Variablen konnten keine Unterschiede zwischen den Behandlungsformen festgestellt werden (Sport: $Chi^2(4)=4.545$, $p=.337$; Bewegung: $Chi^2(4)=4.198$, $p=.380$; Ernährung: $Chi^2(4)=2.735$, $p=.603$; Rauchverhalten: $Chi^2(4)=2.286$, $p=.683$; Leistungsfähigkeit: $Chi^2(4)=2.443$, $p=.655$).

Betrachtet man den Bereich Arbeit und Wiedereingliederung in den Arbeitsprozess, so zeigte sich zu Beginn der Behandlung ein ungleiches Verhältnis von in Arbeit stehenden und arbeitslosen Klientinnen und Klienten im Verhältnis von etwa 2:1. Mittels McNemar-Test wurden die „Wechsler“ von ohne Arbeit vor der Behandlung in Richtung Arbeit nach der Behandlung denjenigen gegenübergestellt, deren Wechsel sich in die gegenläufige Richtung vollzog. Der Vergleich von Ersttestung und Zweittestung zeigte, dass 40 vorher arbeitslose Klientinnen und Klienten zwischenzeitlich ein neues Arbeitsverhältnis aufgenommen hatten, dem standen aber 32 gegenüber, die in der Zwischenzeit ihren Arbeitsplatz verloren hatten ($p=.420$). Getrennt nach Behandlungsformen war für die klinisch-psychologische Behandlung das Verhältnis von „positiven“ und „negativen“ Wechslern 17:14 ($p=.720$), für die Psychotherapie 7:6 ($p=1.000$) und für die Behandlerinnen und Behandler mit

Doppelqualifikation 16:12 ($p=.572$).

Bei der Analyse der Frage, inwieweit sich die psychischen Beschwerden auf die Leistungsfähigkeit im Beruf auswirken, zeigte sich für die klinisch-psychologische Behandlung (auf der Skala von 0 bis 10) eine Verbesserung der Scores um 2.79 Punkte ($SD=3.37$), im Bereich der Psychotherapie um 2.04 Punkte ($SD=3.67$) und im Bereich der Doppelqualifikation um 1.74 Punkte ($SD=2.67$). Der statistische Vergleich zeigte einen hoch signifikanten Effekt für die Dauer der Behandlung ($F(1, 92)=37.526, p<.001, \eta^2=.290$), einen nicht signifikanten Effekt für die Behandlungsform ($F(2, 92)=1.652, p=.197, \eta^2=.035$) sowie eine nicht signifikante Wechselwirkung zwischen Behandlungsdauer und Behandlungsform ($F(2, 92)=0.931, p=.398, \eta^2=.020$). Analog zur Bewertung des Einflusses der Beschwerden auf die Leistungsfähigkeit wurden die Klientinnen und Klienten auch gebeten, den Einfluss der Beschwerden auf die normalen täglichen Aktivitäten mittels dieser elfstufigen Skala (0 = überhaupt kein Einfluss, 10 = halten mich vollkommen von meinen Aktivitäten ab) einzuschätzen. Bei der Analyse dieser Frage zeigte sich für die klinisch-psychologische Behandlung eine Verbesserung der Scores um 1.95 Punkte ($SD=2.98$), im Bereich der Psychotherapie um 1.92 Punkte ($SD=2.78$) und im Bereich der Doppelqualifikation um 1.75 Punkte ($SD=2.74$). Der statistische Vergleich zeigte einen hoch signifikanten Effekt für die Dauer der Behandlung ($F(1, 309)=117.290, p<.001, \eta^2=.275$), einen nicht signifikanten Effekt für die Behandlungsform ($F(2, 309)=2.102, p=.124, \eta^2=.013$) sowie eine nicht signifikante Wechselwirkung zwischen Behandlungsdauer und Behandlungsform ($F(2, 309)=0.165, p=.848, \eta^2=.001$).

Abschließend wurden die Klientinnen und Klienten nach einer Gesamtbeurteilung der Behandlung befragt. Der Großteil ($n=350, 95.1\%$) bewertete die Behandlung als sehr bzw. eher hilfreich, lediglich 18 Klientinnen und Klienten (4.9%) betrachteten die Behandlung als nicht hilfreich. Zwischen den Behandlungsformen konnten diesbezüglich keine Unterschiede erhoben werden ($\chi^2(4)=3.044, p=.550$).

4.2. Veränderungen in den validierten klinischen Fragebögen

Tabelle 3 (Anhang B) zeigt die Ergebnisse der zwei Fragebögen im Überblick. Im BDI-II zeigte sich für die Klientinnen und Klienten der Klinischen Psychologinnen und Psychologen eine Reduktion der Depressionswerte um 12.92 Punkte (SD=11.50), für diejenigen der Psychotherapeutinnen und Psychotherapeuten um 11.87 Punkte (SD=9.53) und für solche von Behandlerinnen und Behandler mit Doppelqualifikation um 10.50 Punkte (SD=11.80). Der statistische Vergleich erbrachte einen hoch signifikanten Effekt der Dauer der Behandlung ($F(1, 302)=294.307, p<.001, \eta^2=.494$), einen nicht signifikanten Effekt für die Behandlungsform ($F(2, 302)=1.046, p=.353, \eta^2=.007$) sowie eine nicht signifikante Wechselwirkung zwischen Dauer der Behandlung und Behandlungsform ($F(2, 302)=1.363, p=.257, \eta^2=.009$).

Im Globalwert des BSI-53 zeigte sich für Klientinnen und Klienten der Klinischen Psychologinnen und Psychologen eine Reduktion des *T*-Werts um 9.58 Punkte (SD=12.85), für diejenigen der Psychotherapeutinnen und Psychotherapeuten um 8.55 Punkte (SD=9.77) und für solche von Behandlerinnen und Behandler mit Doppelqualifikation um 9.12 Punkte (SD=12.20). Der statistische Vergleich erbrachte auch hier einen hoch signifikanten Effekt der Dauer der Behandlung ($F(1, 314)=157.873, p<.001, \eta^2=.335$), einen nicht signifikanten Effekt für die Behandlungsform ($F(2, 314)=1.326, p=.267, \eta^2=.008$) sowie eine nicht signifikante Wechselwirkung zwischen Behandlungsdauer und Behandlungsform ($F(2, 314)=0.166, p=.847, \eta^2=.001$). Damit weisen die Hauptindikatoren der beiden klinischen Fragebögen ein übereinstimmendes Ergebnismuster in dem Sinne auf, dass die Klientinnen und Klienten nach Behandlungsende hoch signifikante Verbesserungen aufzeigten; welcher Form die jeweiligen Behandlerinnen und Behandler angehörten, hatte hingegen auf das Behandlungsergebnis keinen nachweisbaren Einfluss.

Um diese Ergebnisse für die einzelnen Subskalen des BSI-53 zu verdeutlichen, sind in Abbildung 2 (Anhang A) alle *T*-Werte aller neun Subskalen und des Globalwerts für alle drei Behandlungsformen zu Behandlungsbeginn dargestellt. Alle gemittelten *T*-Werte (ausgenommen Subskala 6 Aggressivität/Feindseligkeit) erreichten oder überschritten vor

der Behandlung den Cutoff-Wert für Behandlungsbedürftigkeit. Abbildung 3 (Anhang A) zeigt die durchschnittlichen Verbesserungen (Differenz aus t1 und t2, daher die positive Ausrichtung) in jeder der einzelnen neun Subskalen und im Gesamtwert GSI im direkten Vergleich zwischen den drei Behandlungsformen, Abbildung 4 (Anhang A) das Niveau der drei Behandlungsformen in den Skalen nach Abschluss der Intervention. Für jede einzelne Skala (ausgenommen für eine Gruppe im Globalwert GSI) ist in der aktuellen Analyse zu erkennen, dass die gemittelten *T*-Werte nach der Behandlung unter dem Cutoff-Wert für Behandlungsbedürftigkeit liegen.

4.3. Bewertung der Behandlungszufriedenheit mittels HAQ

Mit der Selbstbeurteilungsform des HAQ wurde die Zufriedenheit mit der Behandlung in den zwei Subskalen Beziehungs- und Erfolgswufriedenheit sowie der Gesamtzufriedenheit zwischen Klientinnen und Klienten der drei Behandlungsformen verglichen. In keinem der drei Zufriedenheitsindikatoren konnte ein signifikantes Ergebnis gefunden werden, d.h. Klientinnen und Klienten aller drei Behandlungsformen zeigten ein ähnliches Niveau an Zufriedenheit. Da für dieses Instrument keine Normen vorliegen, wurde für jede einzelne Skala der Wertebereich auf 1 bis 6 standardisiert. Aus Tabelle 4 (Anhang B) lässt sich ersehen, dass jeder einzelne angeführte Mittelwert im oberen Bereich der Skala (Minimum der angeführten Mittelwerte: 4.73; Maximum der angeführten Mittelwerte: 5.48) gelagert ist, d.h. dass die Klientinnen und Klienten aller drei Behandlungsformen im Durchschnitt mit dem Behandlungsprozess zufrieden waren. Analog dazu erbrachte auch die Analyse der Behandlerinnen und Behandler selbst keine signifikanten Unterschiede zwischen den drei Behandlungsformen.

5. Dropout-Analyse

Eine wichtige Frage, die sich im Anschluss an die Darstellung der Ergebnisse stellt, ist die Frage nach der Repräsentativität der Ergebnisse. Warum haben Klientinnen und Klienten, die an der Ersttestung zu Beginn der Erhebung teilgenommen und die Fragen der

Testbatterie beantwortet haben, dies bei Abschluss der Behandlung nicht mehr getan? Es kann zwar nicht erwartet werden, dass die anfängliche Quote der Beteiligung an der Evaluation von 67.5% (648 Klientinnen und Klienten) nach einer durchschnittlichen Behandlungsdauer von etwa einem halben Jahr beibehalten werden kann, eine Beteiligungsquote von 33.1% (318 Klientinnen und Klienten) zum Zeitpunkt des Abschlusses ist jedoch niedrig. Da im Rahmen dieses Projektes keine Nachbefragung möglich war, lässt sich die Frage nach der Motivation, an der Evaluation des Projektes mitzuarbeiten, nicht direkt beantworten. Es ist jedoch möglich, auf Basis der Daten aus der Erstbefragung Variablen zu extrahieren, die es ermöglichen, einen Überblick über spezifische Charakteristika der Klientinnen und Klienten zu bekommen. So können Prädiktoren ermittelt werden, die es zulassen vorherzusagen, wer sich an der Abschlusserhebung nicht mehr beteiligt hat, und zwar mittels binär-logistischer Regressionsanalyse (Backhaus, Erichson, Plinke & Weiber, 2000).

An soziodemographischen und krankheitsspezifischen Variablen, in denen sich Completer (Klientinnen und Klienten, die sich sowohl an der Erst- wie auch der Abschlusserhebung der Evaluation beteiligt haben) von Dropouts (Klientinnen und Klienten, die nur an der Ersterhebung der Evaluation teilgenommen haben) unterschieden haben, konnten folgende gefunden werden:

Ausbildung: Completerrate variierte zwischen 36.4% für Klientinnen und Klienten ohne Ausbildung und 68.3% für solche mit Abschluss einer berufsbildenden mittleren Schule ($Chi^2(6)=13.250, p=.039$)

Einnahme von Medikamenten (Quote der Completer: 60.0%, Quote der Dropouts: 68.4%; $Chi^2(1)=4.299, p=.038$)

Aktuelles Arbeitsverhältnis (Quote der Completer: 39.5%, Quote der Dropouts: 28.0%; $Chi^2(1)=8.594, p=.003$)

Erhaltene Behandlungseinheiten (Completer: $M=25.71, SD=6.46$, Dropouts: $M=21.94, SD=9.09$; $t(429.213)=5.577, p<.001, Cohen's d=0.47$)

Auswirkung der psychischen Beschwerden auf die täglichen Aktivitäten (Completer: $M=6.09,$

SD=2.35, Dropouts: $M=6.52$, $SD=2.28$; $t(567)=2.219$, $p=.027$, Cohen's $d=0.19$)

Summenwert des Depressionsscores (Completer: $M=26.87$, $SD=10.794$ Dropouts: $M=29.00$, $SD=11.12$; $t(573)=2.298$, $p=.022$, Cohen's $d=0.19$)

T-Wert GSI (Completer: $M=70.58$, $SD=10.44$, Dropouts: $M=72.86$, $SD=9.47$; $t(575)=2.716$, $p=.007$, Cohen's $d=0.23$)

Von diesen sieben Variablen, die von insgesamt 529 Klientinnen und Klienten vollständig vorlagen und die in eine binär-logistische Regression (Methode: Schrittweise Rückwärts (Wald)) eingeschlossen wurden, erwiesen sich vier als signifikante Prädiktoren für die Nicht-Teilnahme an der Abschlusserhebung:

Ausbildung:	Wald(6)=13.922, $p=.031$
Erhaltene Behandlungseinheiten:	$\beta=.066$, Wald(1)=28.426, $p<.001$, $Exp(\beta)=1.068$
Arbeitsverhältnis:	$\beta=-.415$, Wald(1)= 4.220, $p=.040$, $Exp(\beta)=0.660$
T-Wert GSI:	$\beta=-.020$, Wald(1)= 4.183, $p=.041$, $Exp(\beta)=0.981$

Diese Prädiktoren weisen darauf hin, dass Klientinnen und Klienten mit geringer Bildung, ohne aufrechtes Arbeitsverhältnis, mit höheren psychischen Beeinträchtigungen, die signifikant weniger Behandlungseinheiten bekommen hatten, mit höherer Wahrscheinlichkeit keine Daten für die Abschlusstestung geliefert haben (Trefferquote: 66.2%). Die präsentierten Gesamtergebnisse der Evaluation weisen daher tendenziell einen gewissen Bias in Richtung der Personengruppe auf, die sich zwar an der Ersttestung, nicht aber an der Abschlusserhebung beteiligt hat.

6. Diskussion

Nach abgeschlossener Aufnahme von Klientinnen und Klienten im Rahmen der Evaluationsstudie von fit2work standen 648 vollständige Datensets der Ersttestung (Beginn der Behandlung) zur Analyse zur Verfügung (67.5% von insgesamt 960 Klientinnen und Klienten). Nach Abschluss der Behandlung langten von insgesamt 372 Klientinnen und

Klienten ausgefüllte Fragebogensets ein. Davon mussten 54 ausgeschlossen werden, da von diesen keine Aufnahmetestung zu Beginn der Behandlung vorlag. Daher ergab sich für die Gesamtevaluation eine Anzahl von insgesamt 318 Personen (33.1% der Ausgangsstichprobe), deren Daten in die statistische Analyse einbezogen werden konnten. Für das Einbeziehen der Fremdbeurteilung durch die Behandlerinnen und Behandler lagen unter Zugrundelegung von Beurteilungen von Behandlungsbeginn insgesamt 181 Datensätze vor (18.9% der Ausgangsstichprobe).

Im Mittelpunkt der Evaluierung stand die Frage, ob Interventionen, die durch Klinische Psychologinnen und Psychologen, Psychotherapeutinnen und Psychotherapeuten oder Behandlerinnen und Behandler mit Doppelqualifikation im Rahmen des Projekts fit2work angeboten wurden, das Befinden der Klientinnen und Klienten insgesamt signifikant verbessern können. Operationalisiert wurde „Befinden“ im Rahmen dieses Projekts durch die Testverfahren BDI-II (Depression), BSI (psychische Beeinträchtigung) sowie durch die GAF-Skala (Global Assessment of Functioning), beim BSI durch einen globalen Summenwert (Global Severity Index, GSI) bzw. durch die Ergebnisse der insgesamt neun Subskalen des Tests. Zusätzlich wurden als Outcome-Variablen die Items des Fragebogens zur Erfassung des krankheitsspezifischen Status miteinbezogen, die Fragen zur Arbeitsfähigkeit und zum Gesundheitszustand und deren Veränderungen zwischen den beiden Erhebungszeitpunkten miterhoben haben. Im Rahmen der Abschlusstestung wurde auch nach der Zufriedenheit mit dem Interventionsprozess gefragt, befragt wurden sowohl die Klientinnen und Klienten als auch die Behandlerinnen und Behandler.

Die Signifikanzprüfung zwischen den Behandlungsformen musste bei den vorliegenden Fragebogendaten durch den direkten Vergleich der Testdaten zu den beiden Zeitpunkten erfolgen, da für die Evaluierung keine unbehandelte Kontrollgruppe zur Verfügung stand. Die klinischen validierten Fragebogenverfahren weisen aber Werte für die Einstufung nach „klinisch relevanter Ausprägung“ auf (entweder in Form von Cutoff- oder von *T*-Werten), sodass das Evaluationsergebnis im zeitlichen Vergleich zusätzlich auch auf Basis des Kriteriums Behandlungsbedürftigkeit dargestellt werden kann. Mit Hilfe des Allgemeinen Linearen Modells (ALM) für Messwiederholung war es möglich, gleichzeitig die Haupteffekte Behandlungsdauer (Vergleich zwischen Erst- und Abschlusstestung) und Gruppe (Vergleich

der Behandlungsformen) bei Betrachtung etwaiger Wechselwirkungen zwischen diesen beiden Haupteffekten zu prüfen.

Bei der Zuordnung der Klientinnen und Klienten zu den drei Behandlungsformen konnten keine signifikanten Unterschiede bezüglich der soziodemographischen Variablen Geschlecht, Alter oder Ausbildungsniveau gefunden werden, lediglich die Variable Familienstand wies ein signifikantes Ergebnis auf. Ledige Klientinnen und Klienten wurden überzufällig oft der psychotherapeutischen Intervention zugewiesen, verheiratete hingegen der klinisch-psychologischen Behandlung. Bezüglich arbeits- und gesundheitsbezogener Variablen konnten keine statistischen Unterschiede erhoben werden. Keine Unterschiede ergaben sich auch für die Bereiche Diagnosengruppen und Anzahl der komorbiden Störungen, auch in den Fragebögen konnten keine Unterschiede in der Ausgangstestung festgestellt werden. Damit kann festgehalten werden, dass die Zuordnung der Klientinnen und Klienten durch die Steuernden Psychologinnen und Psychologen zu den Behandlungsformen weder auf Basis einer störungsspezifischen Indikation noch auf Basis der unterschiedlichen Stärke der Ausprägung der Beschwerden getroffen wurde. Auch das Ausmaß der von den Klientinnen und Klienten absolvierten Interventionseinheiten war für alle drei Behandlungsformen gleich.

Die Veränderungsmessung erbrachte für die klinischen Fragebögen und den GAF-Wert eindeutige und konsistente Ergebnisse: Mehrheitlich konnten signifikante Verbesserungen durch die Behandlung mit hohen Effektstärken festgestellt werden, Gruppeneffekte sowie Interaktionen hingegen waren nicht signifikant. Interpretiert werden kann dieses Ergebnis dahingehend, dass erstens in allen Fragebogen und deren Subskalen enorme Verbesserungen durch die Behandlung (in jedem Fall $p < .001$, teilweise extrem hohe Effektstärken von $\eta^2 > .250$) zu verzeichnen waren. Bei den psychischen Beeinträchtigungen (BSI) lagen alle Verbesserungen im Bereich von etwa 6 bis 9 T -Wert-Punkten, also Verbesserungen im Bereich von gut einer halben bis fast einer ganzen Standardabweichung. Alle Mittelwerte der Subskalen, die vor der Intervention über dem Cutoff-Wert ($T=63$) für Behandlungsbedürftigkeit lagen, fanden sich bei der Abschluss-erhebung unter diesem. Zweitens, dass sich die drei Behandlungsformen in keinem

Fragebogen (bzw. seinen Subskalen) signifikant voneinander unterschieden, d.h. dass alle drei Behandlungsformen gleich gute Ergebnisse erbrachten (Haupteffekt Gruppe), und drittens, dass keine Wechselwirkungen zwischen Behandlungsdauer und Behandlungsform zustande kamen.

80% der Klientinnen und Klienten, deren Arbeitsplatz in Gefahr war, die den Arbeitsplatz aber halten konnten, beschrieben ihre berufliche Leistungsfähigkeit nach der Behandlung als verbessert, bei etwa 70% kam es zu einer Reduktion der Krankenstände. Obwohl sich im klinischen Bereich bzw. im Bereich der psychischen Beeinträchtigungen große Veränderungen über die Zeit darstellen ließen, war das unmittelbar nach Abschluss der Behandlung in Bezug auf die Frage nach der Beschäftigungssituation nicht so. Obwohl der Saldo der Wechsler ein positiver war (40 Klientinnen und Klienten, die zu Beginn arbeitslos waren, hatten nach Abschluss der Behandlung wieder einen Job gefunden, während 32 in der Zwischenzeit arbeitslos geworden waren), zeigte sich beim Großteil keine Änderung im Status Beschäftigung, d.h. dass auch nach Abschluss der Behandlung weiterhin rund zwei Drittel ohne Arbeit waren. Das übergeordnete Ziel einer beruflichen Reintegration scheint unmittelbar nach der Intervention nicht direkt erreichbar zu sein, immerhin werden aber die Grundlagen dafür geschaffen. Für das Gesamtkollektiv der Klientinnen und Klienten von fit2work zeigten sich große Veränderungen im Sinne einer umfassenden Reduktion der Inanspruchnahme von Gesundheitsleistungen, wobei insbesondere folgende Punkte zu nennen sind:

- Reduktion von Krankenhausaufenthalten (bei 62%), von Arztbesuchen (bei 52%), von ärztlichen Untersuchungen (bei 50%) und von Einnahme von Psychopharmaka (bei 34%),
- Positive Veränderungen im Bewegungsverhalten (bei 80%), im Sportverhalten (bei 62%) und im Ernährungsverhalten (bei 52%)
- 64% schätzten die klinisch-psychologische Behandlung oder Psychotherapie als „sehr hilfreich“ ein, 31% als „eher hilfreich“.

Zusätzlich zu diesen Beurteilungen zeigten die Ergebnisse des Fragebogens Helping Alliance Questionnaire (HAQ), dass sowohl die Beziehungs- als auch die Erfolgsszufriedenheit nach

Beendigung der Intervention hoch ausgeprägt waren, und zwar auf beiden Seiten. Sowohl die Klientinnen und Klienten, aber auch die Behandlerinnen und Behandler waren mit dem Behandlungsprozess hoch zufrieden.

Danksagung. Dank gebührt den Wissenschaftlichen Mitarbeiterinnen Sabine Kögler-Ramazan und Sophie Haller sowie Studienassistent Andreas Goreis für die Koordination der einlangenden Fragebögen, die Auswertung der Fragebögen, die Dateneingabe und die Layoutierung des Textes sowie Dr. Dorothea König für die Programmierung der Online-Erhebung.

Literatur

- Backhaus, K., Erichson, B., Plinke, W. & Weiber, R. (2000). *Multivariate Analysemethoden – eine anwendungsorientierte Einführung* (9., überarbeitete und erweiterte Auflage). Berlin: Springer.
- Bassler, M., Potratz, B. & Krauthauser, H. (1995). Der "Helping Alliance Questionnaire" (HAQ) von Luborsky. *Psychotherapeut*, 40, 23–32.
- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human und Sozialwissenschaftler* (4. Auflage). Berlin: Springer.
- Bundesministerium für Gesundheit (1990a). *Bundesgesetz vom 7. Juni 1990 über die Ausübung der Psychotherapie*. Zugriff am 25.6.2015. Verfügbar unter <http://www.psychotherapie.at/psychotherapeutinnen/rechtsinformationen/psychotherapiegesetz/psychotherapiegesetz-1-2>
- Bundesministerium für Gesundheit (1990b). *Bundesgesetz vom 7. Juni 1990 über die Führung der Berufsbezeichnung "Psychologe" oder „Psychologin" und über die Ausübung des psychologischen Berufes im Bereich des Gesundheitswesens (Psychologengesetz)*. Zugriff am 25.6.2015. Verfügbar unter https://www.ris.bka.gv.at/Dokumente/BgblPdf/1990_361_0/1990_361_0.pdf
- Bundesministerium für Gesundheit (2013). *Bundesgesetz, über die Führung der Bezeichnung „Psychologin“ oder „Psychologe“ und über die Ausübung der Gesundheitspsychologie und der Klinischen Psychologie. Psychologengesetz 2013*. Zugriff am 25.6.2015. Verfügbar unter http://www.ris.bka.gv.at/Dokumente/BgblAuth/BGBLA_2013_I_182/BGBLA_2013_I_182.html
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale: Erlbaum.
- Franke, G. H. (2000). *BSI: Brief Symptom Inventory von L.R. Derogatis (Kurzform der SCL-90-R) – deutsche Version: Manual*. Göttingen: Beltz Test.
- Hautzinger, M., Keller, F. & Kühner, C. (2006). *BDI-II; Beck Depressions-Inventar Revision*. Frankfurt am Main: Harcourt Test Services.

Anhang A. Abbildungen

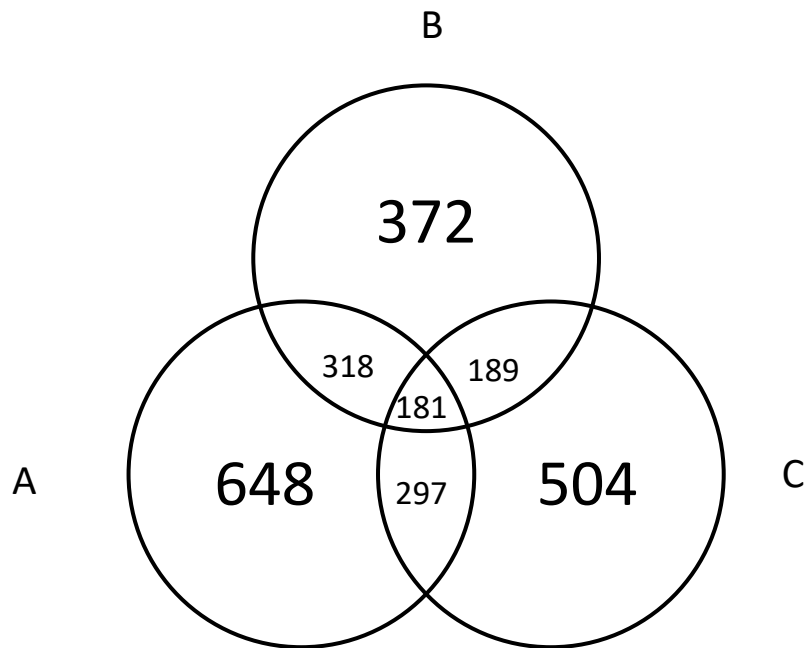


Abbildung 1. Venn-Diagramm für die Anzahl der vorliegenden Datensätze.

Anmerkung. A Fragebogensets der Klientinnen und Klienten zu Behandlungsbeginn, B Fragebogensets der Klientinnen und Klienten zu Behandlungsende, C Online-Fragebogensets der Behandlerinnen und Behandler zu Behandlungsbeginn

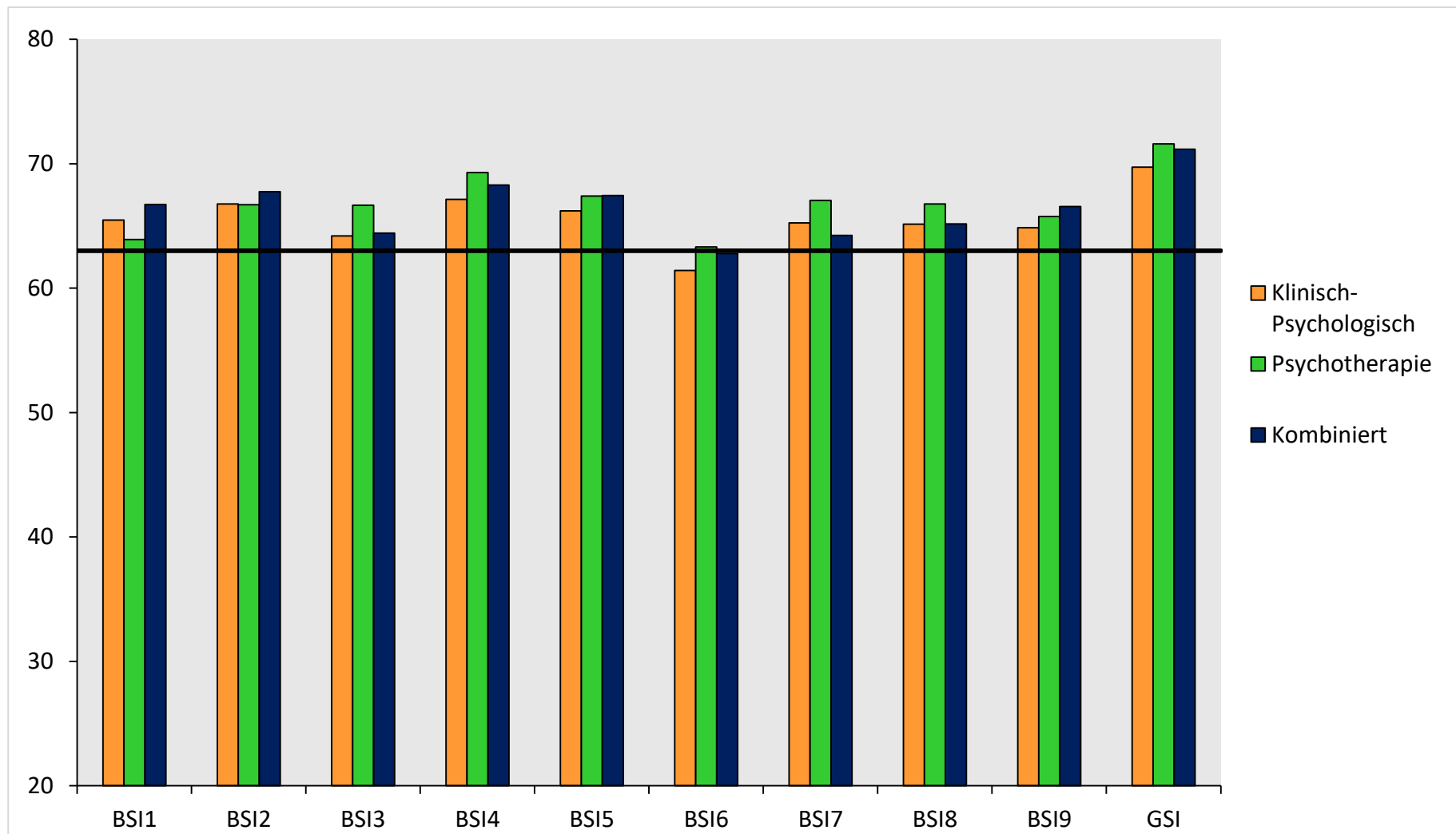


Abbildung 2. Vergleich der drei Behandlungsformen im Brief Symptom Inventory (BSI) zu Behandlungsbeginn. Entsprechend den Angaben des BSI-Manuals weist ein T-Wert ab 63 (waagrechte Linie) auf Behandlungsbedürftigkeit hin.

Anmerkung. BSI1=Somatisierung, BSI2=Zwanghaftigkeit, BSI3=Unsicherheit im Sozialkontakt, BSI4=Depressivität, BSI5=Ängstlichkeit, BSI6=Aggressivität/Feindseligkeit, BSI7=Phobische Angst, BSI8=Paranoides Denken, BSI9=Psychotizismus, GSI=Globalwert (Global Severity Index)

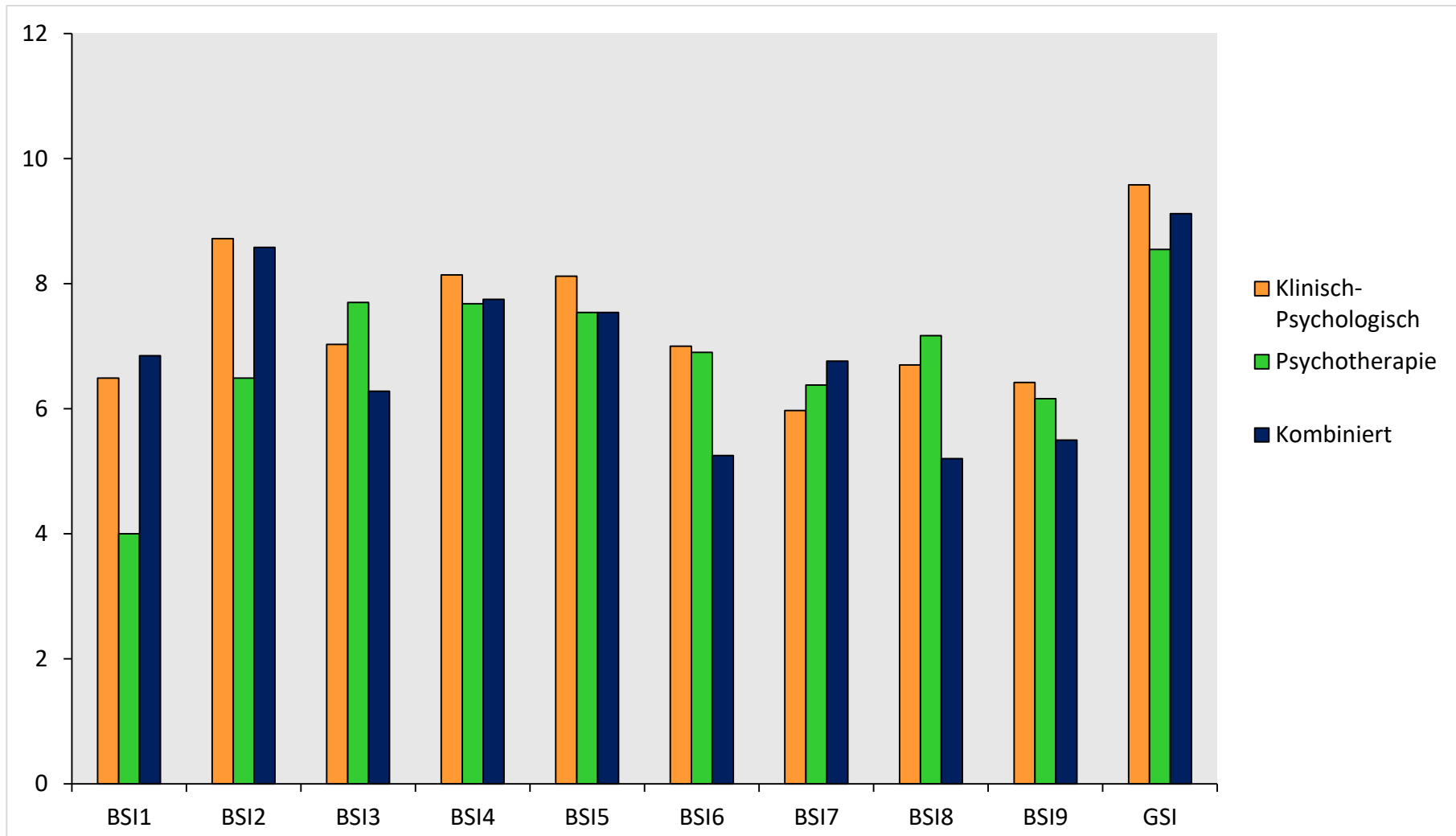


Abbildung 3. Vergleich der Verbesserungen (Differenzen in T-Werten) im Brief Symptom Inventory (BSI) zwischen den drei Behandlungsformen. Anmerkung. BSI1=Somatisierung, BSI2=Zwanghaftigkeit, BSI3=Unsicherheit im Sozialkontakt, BSI4=Depressivität, BSI5=Ängstlichkeit, BSI6=Aggressivität/Feindseligkeit, BSI7=Phobische Angst, BSI8=Paranoides Denken, BSI9=Psychotizismus, GSI=Globalwert (Global Severity Index)

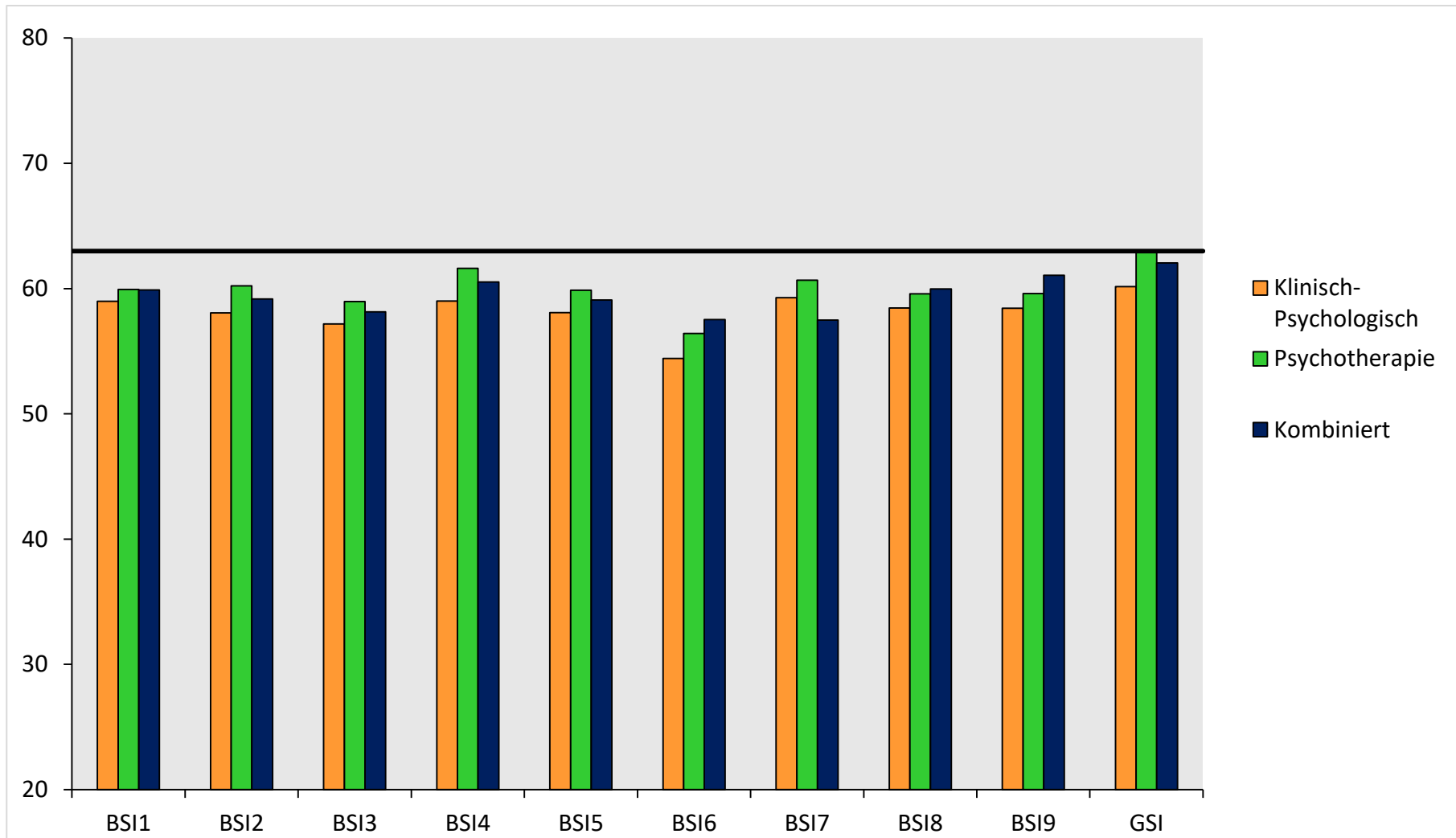


Abbildung 4. Vergleich der drei Behandlungsformen im Brief Symptom Inventory (BSI) zu Behandlungsende. Entsprechend den Angaben des BSI-Manuals weist ein T-Wert ab 63 (waagrechte Linie) auf Behandlungsbedürftigkeit hin.

Anmerkung. BSI1=Somatisierung, BSI2=Zwanghaftigkeit, BSI3=Unsicherheit im Sozialkontakt, BSI4=Depressivität, BSI5=Ängstlichkeit, BSI6=Aggressivität/Feindseligkeit, BSI7=Phobische Angst, BSI8=Paranoides Denken, BSI9=Psychotizismus, GSI=Globalwert (Global Severity Index)

Anhang B. Tabellen

Tabelle 1. Absolute und relative Veränderungen in Variablen der Inanspruchnahme von Gesundheitsleistungen

Art der Änderung n (%)	Ärztliche Untersuchungen	Arztbesuche wegen psychischer Probleme	Krankenhausaufenthalte	Medikamenteneinnahme
Abnahme	161 (50.2%)	164 (52.2%)	156 (61.5%)	105 (34.1%)
keine	134 (41.7%)	118 (37.6%)	89 (35.0%)	157 (51.0%)
Zunahme	26 (8.1%)	32 (10.2%)	9 (3.5%)	46 (14.9%)

Tabelle 2. Absolute und relative Veränderungen im Bereich des Gesundheitsverhaltens

Art der Änderung n (%)	Sportverhalten	Bewegungsverhalten	Ernährungsverhalten	Rauchverhalten
positiv	146 (62.4%)	192 (79.7%)	118 (52.0%)	44 (22.1%)
keine	71 (30.3%)	35 (14.5%)	89 (39.2%)	131 (65.8%)
negativ	17 (7.3%)	14 (5.8%)	20 (8.8%)	24 (12.1%)

Tabelle 3. Vergleich der Evaluationsergebnisse zwischen den drei Behandlungsformen zu Beginn (t1) und nach Abschluss der Behandlung (t2) in den klinischen Fragebögen BDI-II und BSI-53

	Behandlungsform1		Behandlungsform2		Behandlungsform3		p_z	p_B	p_{z*B}
	t1 M (SD)	t2 M (SD)	t1 M (SD)	t2 M (SD)	t1 M (SD)	t2 M (SD)			
BDI-II	26.62 (12.05)	13.70 (12.22)	27.64 (8.99)	15.77 (9.99)	27.14 (10.59)	16.64 (12.70)	<.001	.353	.257
BSI (T-Werte)									
Somatisierung	65.48 (10.94)	58.99 (12.24)	63.92 (11.71)	59.92 (12.54)	66.73 (10.71)	59.88 (12.47)	<.001	.645	.189
Zwanghaftigkeit	66.77 (10.75)	58.05 (13.16)	66.71 (10.65)	60.22 (11.23)	67.75 (9.48)	59.17 (12.47)	<.001	.656	.364
Unsicherheit im Sozialkontakt	64.20 (12.23)	57.17 (12.14)	66.67 (10.30)	58.97 (10.92)	64.43 (12.94)	58.15 (12.48)	<.001	.429	.725
Depressivität	67.14 (10.26)	59.00 (12.03)	69.30 (8.00)	61.62 (10.84)	68.28 (9.80)	60.53 (12.40)	<.001	.212	.938
Ängstlichkeit	66.21 (11.39)	58.09 (12.08)	67.41 (10.98)	59.87 (11.47)	67.44 (10.33)	59.09 (12.81)	<.001	.524	.906
Aggressivität/Feind- seligkeit	61.43 (11.13)	54.43 (11.12)	63.32 (10.89)	56.42 (11.52)	62.78 (12.03)	57.53 (12.77)	<.001	.171	.463
Phobische Angst	65.24 (11.86)	59.27 (13.55)	67.05 (11.09)	60.67 (11.59)	64.24 (13.00)	57.48 (13.35)	<.001	.230	.879
Paranoides Denken	65.14 (10.73)	58.44 (11.89)	66.76 (8.53)	59.59 (11.74)	65.17 (11.79)	59.97 (11.75)	<.001	.628	.355
Psychotizismus	64.85 (11.56)	58.43 (12.23)	65.76 (11.29)	59.60 (11.30)	66.57 (11.94)	61.07 (12.30)	<.001	.265	.815
Global Severity Index	69.73 (10.83)	60.15 (14.78)	71.60 (9.10)	63.05 (12.19)	71.17 (10.58)	62.05 (14.83)	<.001	.267	.847

Anmerkungen. Behandlungsform1 = Klinische Psychologinnen und Psychologen, Behandlungsform2 = Psychotherapeutinnen und Psychotherapeuten, Behandlungsform3 = Behandlerinnen und Behandler mit Doppelqualifikation
 p_z Zeit, p_B Behandlungsform, p_{z*B} Interaktion Zeit*Behandlungsform

Tabelle 4. Vergleich der Ergebnisse im Helping Alliance Questionnaire (HAQ) zwischen den drei Behandlungsformen in Selbst- (Klientinnen und Klienten) und Fremdbeurteilung (Behandlerinnen und Behandler)

	Selbstbeurteilung			p_s	Fremdbeurteilung			p_f
	Behandlung1 M (SD)	Behandlung2 M (SD)	Behandlung3 M (SD)		Behandlung1 M (SD)	Behandlung2 M (SD)	Behandlung3 M (SD)	
HAQ								
Beziehungszufriedenheit	5.48 (0.69)	5.41 (0.60)	5.39 (0.81)	.548	5.37 (0.57)	5.52 (0.52)	5.42 (0.57)	.473
Erfolgswzufriedenheit	4.88 (0.89)	4.73 (0.79)	4.77 (0.96)	.354	5.03 (0.81)	5.39 (0.58)	5.01 (0.81)	.075
Gesamtwzufriedenheit	5.21 (0.72)	5.09 (0.62)	5.10 (0.83)	.381	5.21 (0.65)	5.46 (0.52)	5.23 (0.65)	.188
(Wertebereich jeweils von 1=sehr unzufrieden bis 6=sehr zufrieden)								

Anmerkungen. Behandlung1 = klinisch-psychologische Behandlung, Behandlung2 = Psychotherapie, Behandlung3 = Behandlerinnen und Behandler mit Doppelqualifikation; p_s Selbstbeurteilung, p_f Fremdbeurteilung

