

Endbericht

Evaluierung Einzel- versus Gruppenbehandlungen fit2work klinisch-psychologische und psychotherapeutische Behandlung im 4. Förderjahr (1.1.2017 bis 31.12.2019)

Reinhold Jagsch

Institut für Angewandte Psychologie:
Gesundheit, Entwicklung und Förderung

Fakultät für Psychologie
Universität Wien

Wien, April 2020

Inhaltsverzeichnis

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	7
2.2.2. Beck-Depressions-Inventar BDI-II (Hautzinger, Keller & Kühner, 2006)	8
2.2.3. Brief Symptom Inventory BSI-53 (Franke, 2000).....	8
2.2.4. Zufriedenheit mit Behandlung und Beziehung: Helping Alliance Questionnaire HAQ (Bassler, Potratz & Krauthauser, 1995)	10
2.3. Statistische Analyse	10
3. Beteiligung an der Evaluation	12
4. Darstellung der Ausgangssituation (Stichprobe der Ersttestung)	13
4.1. Vergleich soziodemographischer Charakteristika	13
4.2. Vergleich krankheitsspezifischer Charakteristika.....	14
5. Ergebnisse zur Veränderungsmessung sowie zur Behandlungszufriedenheit	16
5.1. Veränderungen in der Inanspruchnahme von Gesundheitsleistungen und im Gesundheitsverhalten	17
5.2. Veränderungen bezüglich der validierten klinischen Fragebögen.....	18
5.3. Bewertung der Behandlungszufriedenheit mittels HAQ.....	19
6. Dropout-Analyse	19
7. Zusammenfassung	22
Literatur	27
Anhang A. Abbildungen	28
Anhang B. Tabellen	34

Abstract

Theoretischer Hintergrund. Auf Basis der österreichischen Gesetzeslage mit einerseits Psychologengesetz und andererseits Psychotherapiegesetz sind Klinische Psychologinnen und Psychologen und Psychotherapeutinnen und Psychotherapeuten sowie Behandlerinnen und Behandler mit Doppelqualifikation berechtigt, Personen mit psychischen Störungen zu behandeln. Nachdem bereits belegt werden konnte, dass im Rahmen der Evaluation von fit2work Behandlungen durch die drei Teilgruppen hoch wirksam und gleichwertig sind, lag der Schwerpunkt der aktuellen Analyse auf dem Vergleich von Behandlungen, die im Einzel- bzw. Gruppenformat angeboten wurden. Obwohl im Gruppenformat Behandlungen für drei verschiedene Zielgruppen zur Verfügung standen („Psychosoziale und Reha-Jahres-“, „Schmerz-“ und „Soziale und Emotionale Kompetenztrainingsgruppen“), lagen lediglich für letztere ausreichend hohe Fallzahlen vor, um einen direkten Vergleich mit den Einzelbehandlungen vornehmen zu können.

Methode. Österreichweit nahmen insgesamt 2100 Personen (1740 in Einzel-, 360 in Gruppenbehandlung), deren Arbeitsplatz entweder durch gesundheitliche Probleme gefährdet war bzw. bei denen der Arbeitsplatz bereits verlorengegangen war, an dieser Phase des Projekts teil und wurden nach einer umfangreichen Anamnese einer der Behandlungsformen zugewiesen. Die Klientinnen und Klienten erhielten zu Beginn der Behandlung eine Testbatterie, die Fragebögen zu Depression (Beck-Depressions-Inventar-II, BSI-II), zur psychischen Beeinträchtigung (Brief Symptom Inventory-53, BSI-53) und einen Fragebogen zum arbeits- und gesundheitsbezogenen Status enthielt. Nach Abschluss der Behandlung wurden sie gebeten, diese Testbatterie noch einmal zu bearbeiten 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 online durchgeführt wurde und die auch eine Beurteilung mittels GAF-Skala¹ (Global Assessment of Functioning) beinhaltete.

¹ Fremdeinschätzungsskala von 1 bis 100 für Funktionsfähigkeit in Fertigkeiten des täglichen Lebens

Ergebnisse. Insgesamt absolvierten 613 Klientinnen und Klienten die Behandlung und füllten zu beiden Zeitpunkten die Fragebögen aus (497 in Einzel- und 116 in Gruppenbehandlung). Laut den Daten vor der Behandlung sind die Klientinnen und Klienten der beiden Behandlungsformen als zwei distinkte Zielgruppen der Maßnahme zu sehen, wobei beide zwar einen ähnlichen soziodemographischen und arbeitsbezogenen Hintergrund aufweisen, sich aber gesundheitsbezogen deutlich unterscheiden. Klientinnen und Klienten in Einzelbehandlung zeigten sich vor der Behandlung deutlich stärker belastet. Nach der Behandlung wiesen beide Behandlungsformen maßgebliche Verbesserungen auf: In allen klinischen Skalen des BSI-53 lagen Klientinnen und Klienten beider Behandlungsformen (ausgenommen Subskala Aggressivität/Feindseligkeit) über dem Cutoff-Wert² für Behandlungsbedürftigkeit, nach Abschluss der Behandlung in keiner der Subskalen. Neben einem hoch signifikanten Effekt der Dauer der Behandlung konnten im Rahmen der Veränderungsmessung simultan weder signifikante Gruppeneffekte noch signifikante Interaktionen gefunden werden. Das statistische Modell für das Depressionsausmaß hingegen zeigte, dass die Reduktionen im Rahmen der Einzelbehandlung bei zwar höherem Ausgangsniveau ungleich stärker waren als diejenigen der Gruppenbehandlung. Bezüglich Reintegration in den Arbeitsprozess kann vermerkt werden, dass die Quote des Wiedereinstiegs relativ zu der Quote der Klientinnen und Klienten, die in der Zeit der Behandlung ihre Arbeit verloren hatten, signifikant höher lag, und zwar für beide Behandlungsformen. Die Interventionsmaßnahmen führten zusätzlich, ohne Unterschiede zwischen den Behandlungsformen, zu einer massiven Reduktion der Inanspruchnahme von Gesundheitsleistungen. So konnten bedeutsame Reduktionen bei Arztbesuchen, bei ärztlichen Untersuchungen, bei Krankenhaus-aufenthalten und bei der Einnahme von Psychopharmaka verzeichnet werden. Auch das Gesundheitsverhalten (Sport, Bewegung, Ernährung) konnte positiv verändert werden. Beziehungs- und Erfolgsszufriedenheit scorten in beiden Behandlungsformen hoch, sowohl in der Selbsteinschätzung der Klientinnen und Klienten als auch in der Fremdeinschätzung durch die Behandlerinnen und Behandler.

² Trennwert, bei dessen Überschreitung Behandlungsbedürftigkeit angenommen wird

Schlussfolgerungen. Die Evaluation konnte belegen, dass die beiden Behandlungsformen gezielt distinkte Zielgruppen ansprechen und sich dabei als hoch wirksam erweisen. 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 beitragen. Gleichzeitig berichteten die Klientinnen und Klienten von bedeutsamen Verbesserungen im arbeits- und gesundheitsbezogenen Bereich, verbunden mit einer hohen Behandlungs- und Erfolgsszufriedenheit, sowohl auf Seiten der Klientinnen und Klienten als auch auf Seiten der Behandlerinnen und Behandler.

1. Einleitung

Der aktuelle Endbericht gibt einen Überblick über die Evaluierung von fit2work klinisch-psychologische und psychotherapeutische Behandlung im Vergleich von Einzel- und Gruppenbehandlung des 4. Förderjahres. Allgemeines Ziel der Evaluation ist es, die Effektivität der Maßnahme zu überprüfen, d.h. eine Evaluierung im Sinne einer Ergebnisevaluation (Outcome Evaluation) durchzuführen. Aufbau und Konzept, Verwendung der Erhebungsinstrumente sowie Auswertungsmethoden der Evaluation entsprechen weitgehend denen der vorangegangenen Evaluationen (Jagsch, 2015, 2018). Die Besonderheit des 4. Förderjahres liegt darin, dass erstmals Einzel- und Gruppenbehandlungen innerhalb eines gemeinsamen Rahmenprojektes angeboten und durchgeführt wurden und im Rahmen der dazugehörigen Evaluation direkt miteinander verglichen werden können. Der Umfang war mit 1440 Klientinnen und Klienten (später um weitere 300 erhöht) für die Einzelbehandlung und 360 für die Gruppenbehandlung festgelegt. Insgesamt sind von 1497 Klientinnen und Klienten Fragebögen an das Evaluationsteam zugeschickt worden (d.s. 71.3% der geplanten Gesamtstichprobe), wobei einige nicht zur Analyse herangezogen wurden: 1. Insgesamt wurden 43 Fragebögen der Ersttestung zwar ausgefüllt und abgeschickt, aber darauf folgend keine Behandlung begonnen. 2. Wie bereits bei der Pilotstudie

zur Gruppenbehandlung wurden zwar unterschiedliche Gruppensettings eingerichtet, einerseits die „psychosozialen und Reha-Jahresgruppen“, andererseits eine Gruppe mit „Gruppentraining Soziale und Emotionale Kompetenz“ sowie drittens Gruppen, bei denen die Bearbeitung von Schmerzen im Vordergrund stand („Schmerzgruppen“), diese aber in deutlich unterschiedlicher Zahl. Da die Anzahl der eingelangten Fragebögen von zwei der drei Settings deutlich niedriger als diejenige mit „Gruppentraining Soziale und Emotionale Kompetenz“ war, konnten die beiden unterbesetzten Gruppen nicht in die Analyse einbezogen werden. Wenn im Folgenden von Gruppenbehandlung gesprochen wird, ist damit „Gruppentraining Soziale und Emotionale Kompetenz“ gemeint. Wie bereits beim 3. Förderjahr wurde eine Differenzierung nach dem Status der Behandlerinnen und Behandler (1. Klinische Psychologinnen und Psychologen, 2. Psychotherapeutinnen und Psychotherapeuten, 3. Doppelqualifikation) vorgenommen. Da aber auch hier die Anzahl der drei unterschiedlichen Formen in ihrer Größe stark differierte, bei der Gruppenbehandlung Psychotherapie lediglich im einstelligen Bereich vortreten war, musste auch hier eine Einschränkung zum Tragen kommen, d.h. auf eine differenzierte Analyse verzichtet werden. Da aber bereits im Rahmen der Evaluation des 3. Förderjahres belegt werden konnte, dass die drei unterschiedlichen Behandlungsformen gleichwertig und hoch wirksam sind, d.h. gleiche Effektivität zeigten (Jagsch, 2018), ist diese Einschränkung vertretbar.

Die Darstellung der Ergebnisse erfolgt wieder, wie gewünscht, in kurzer und prägnanter Weise, die inhaltliche Darstellung punktuell kompakt im Fließtext, die ergänzenden Abbildungen und Tabellen finden sich zusammengefasst in den Anhängen.

2. Methoden

2.1. Zu prüfende Fragestellungen

Ziel der Evaluation 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 die 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.

Inhaltlich soll die Evaluation folgende Fragestellungen auf Basis der erhobenen empirischen Daten 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 Einzel- und Gruppenbehandlung
3. Prüfung von Interaktionseffekten (Wechselwirkung von Behandlungsform 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 Reduktion der Inanspruchnahme von Gesundheitsleistungen

2.2. Erhebungsinstrumente

Für die 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) jeweils in Selbstbeurteilungsform. Zusätzlich wurde nach Abschluss der Behandlung der Fragebogen Zufriedenheit mit Behandlung und Beziehung: Helping Alliance Questionnaire (HAQ) im Selbst- und Fremdbeantwortungsformat 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 (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 wird, 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 werden nach der 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 wird 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 werden 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, wird 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 Konstanzhaltung der Personen, die zu beiden Zeitpunkten den gleichen Beschäftigungsgrad aufweisen). Im Rahmen einer Dropout-Analyse werden im ersten Schritt die maßgeblichen Variablen mit entsprechenden statistischen Methoden auf signifikante Unterschiede hin geprüft, im zweiten Schritt werden die dabei auffindig gemachten Variablen einer binär-logistischen Regressionsanalyse zugeführt, um Prädiktoren für die Frage auffindig 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 können, empfiehlt es sich, zusätzlich Kennwerte der inhaltlichen Relevanz (Effektstärken) ergänzend anzugeben (Bortz & Döring, 2006). Diese haben den Vorteil, dass deren Ausmaß unabhängig von

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.

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 mittels ALM für Messwiederholung ist als Effektstärkemaß Eta^2 anzuführen (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. Beteiligung an der Evaluation

Insgesamt sind 1190 ausgefüllte Fragebogensets der Ersttestung (Beginn der Behandlung) beim Evaluatorenteam der Fakultät für Psychologie der Universität Wien eingelangt, 1022 von Einzel- und 168 von Gruppenbehandlungen. Von Seiten der Behandlerinnen und Behandler lagen insgesamt 454 Online-Fremdbeurteilungen vollständig für Behandlungsbeginn und -ende vor, das entspricht einer Quote von 38.2%. Davon beziehen sich 338 auf die Einzel- (Quote: 33.1%) und 116 (Quote: 69.0%) auf die Gruppenhandlungen. Bedingt durch die unterschiedlichen Organisationsmodi für das Ausfüllen und die Übermittlung der Fragebogenbatterien (individuelle Zusendung der ausgefüllten Fragebögen auf dem Postweg bei den Einzelbehandlungen, getrennt nach prä und post; gemeinsame Versendung aller gesammelten Fragebögen (prä und post) aller Klientinnen und Klienten je Gruppe bei den Gruppenbehandlungen) ergeben sich unterschiedlich hohe Quoten, auch für die Anzahl der Abschlusstestungen. Für die Einzelbehandlungen liegen von 497 Klientinnen und Klienten vollständige Datensätze für die beiden Erhebungen vor, das

sind 48.6% der 1022 Ersttestungen. Für die Gruppenbehandlungen hingegen liegt die entsprechende Quote ungleich höher, nämlich bei 69.0% (116 vollständige Datensätze für die beiden Erhebungen, bei 168 vorliegenden Ersttestungen). Für eine vollständige Übersicht der Gesamt- sowie der beiden Teilpopulationen siehe die Venn-Diagramme in den Abbildungen 1, 2 und 3.

Ziel des Endberichts ist es, einen Vergleich der Effektivität von Einzel- und Gruppenbehandlung durchzuführen, wobei Gruppenbehandlung, wie bereits oben ausgeführt, mit „Sozialem und Emotionalem Kompetenztraining“ gleichzusetzen ist. Als Stichtag für das Ende der Aufnahme von Fragebögen wurde als Termin der 13. März 2020 gewählt, um die benötigten Vorarbeiten (Verknüpfung der selbst- und fremdberichteten Daten, Data Cleaning, Erstellung von Summenscores, Transformation der Rohdaten des BSI) für die Datenanalyse durchführen und so die Darstellung der Ergebnisse zum Termin der Abgabe des Endberichts gewährleisten zu können.

4. Darstellung der Ausgangssituation (Stichprobe der Ersttestung)

4.1. Vergleich soziodemographischer Charakteristika

Die Gesamtstichprobe der Ersttestung lässt sich folgendermaßen charakterisieren: Der Großteil ist weiblich ($n=763$, 64.1%), das Durchschnittsalter liegt bei 44.02 Jahren ($SD=9.90$; Range 18–65). Der vorherrschende Zivilstatus ist verheiratet ($n=483$, 40.9%), gefolgt von ledig ($n=431$, 36.4%), geschieden ($n=256$, 21.6%) und verwitwet ($n=13$, 1.1%). Vom Bildungsstand her herrscht abgeschlossene Ausbildung „Berufsschule/Lehre“ vor ($n=494$, 41.7%). Unterschiede zur Gesamtfallzahl von 1190 ergeben sich durch unvollständige Beantwortung der einzelnen Variablen. Bezüglich der soziodemographischen Charakteristika unterschieden sich die zwei Behandlungsformen nicht; weder in der Geschlechterproportion, im Alter noch nach dem Grad der Ausbildung konnten statistisch signifikante Unterschiede gefunden werden ($p>.05$). Lediglich im Zivilstatus konnte ein statistisch signifikanter Unterschied festgestellt werden ($Chi^2(3)=9.659$, $p=.022$), insofern als überzufällig

mehr verheiratete Klientinnen und Klienten in Einzelbehandlung und überzufällig mehr ledige Klientinnen und Klienten in Gruppenbehandlung waren.

Was die berufsspezifischen Daten anlangt, waren insgesamt 813 Klientinnen und Klienten (68.6%) zum Zeitpunkt der Ersterhebung arbeitslos, von denen etwa zwei Drittel auf Nachfrage angaben, aufgrund von psychischen Beschwerden arbeitslos geworden zu sein. Die Durchschnittsdauer der Arbeitslosigkeit wurde mit 22.35 Monaten (SD=35.42; Median=9 Monate, Spannweite 1–233 Monate) angegeben. In Bezug auf die arbeitslosenspezifischen Daten zeigten sich keine statistisch signifikanten Unterschiede zwischen den zwei Behandlungsformen, weder in der Häufigkeit noch im Grund noch in der Dauer der Arbeitslosigkeit ($p > .05$).

4.2. Vergleich krankheitsspezifischer Charakteristika

Von insgesamt 585 Klientinnen und Klienten lagen Daten zu 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 Doppel- bzw. Mehrfachdiagnosen ($n=269$, 46.0%) lag knapp unter demjenigen mit lediglich einer einzigen Diagnose. Den Angaben zufolge können drei Hauptgruppen von Diagnosen unterschieden werden: Störungen aus der F4-Sektion (Neurotische, Belastungs- und somatoforme Störungen, $n=329$, 56.3%) stehen an erster Stelle, gefolgt von Affektiven Störungen (F3-Diagnosen, $n=325$, 55.7%) und Störungen bezüglich Z73 (Probleme bei der Lebensbewältigung, Burnout, $n=44$, 7.5%). Es konnten dabei zwei statistisch signifikante Unterschiede zwischen den zwei Behandlungsformen eruiert werden, es zeigten sich signifikant mehr Doppel- bzw. Mehrfachdiagnosen (52.7% vs. 23.7%; $Chi^2(1)=35.071$, $p < .001$) und signifikant mehr F4-Diagnosen (61.8% vs. 38.1%; $Chi^2(1)=23.613$, $p < .001$) bei Klientinnen und Klienten in Einzelbehandlung

Die Frage, wie lange die Erstdiagnose zurückliegt, wurde sehr unterschiedlich beantwortet. Die Minimaldauer lag bei einem Monat, die Maximaldauer bei 468 Monaten, der Median bei zehn Monaten. Ähnlich inhomogen war das Antwortmuster bezüglich der Latenz vom Wahrnehmen der ersten Symptome bis zur Stellung der

Erstdiagnose. Der Median lag bei 24 Monaten (Spannweite 0–600). Auch bezüglich dieser beiden Zeitmaße ließen sich keine Unterschiede zwischen den beiden Behandlungsformen nachweisen ($p>.05$).

Zu Variablen, die die gesundheitlichen 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 65.5% an, mindestens einmal in den letzten drei Monaten im Krankenstand gewesen zu sein. Insgesamt 733 aller Befragten (62.7%) antworteten, Medikamente wegen psychischer Beschwerden einzunehmen, 72.3% hatten im Zeitraum der letzten drei Monate mindestens einmal einen Arzt aufgesucht (Spannweite: 1 bis 30 Besuche), 66.7% ließen in diesem Zeitraum medizinische Untersuchungen vornehmen (Spannweite: 1 bis 25 Untersuchungen), und 10.8% hatten in den zurückliegenden drei Monaten Krankenaufenthalte (Spannweite: 1 bis 6 Aufenthalte). Die Auswirkungen der psychischen Beschwerden auf die täglichen Aktivitäten, die auf einer Skala von 0 (kein Einfluss) bis 10 (maximaler Einfluss) abgefragt wurden, zeigten bei Beginn der Behandlung einen mittleren Wert von 6.27 (SD=2.29) für die Gesamtstichprobe. Vergleicht man diese abgefragten Maße (Krankenstände, Medikamenteneinnahme, Arztbesuche, medizinische Untersuchungen, Krankenhausaufenthalte) zwischen den Behandlungsarten, so konnten keine signifikanten Unterschiede festgestellt werden ($p>.05$). Lediglich in der Selbstbewertung der psychischen Belastung zeigten Klientinnen und Klienten in Einzelbehandlung mit einem Mittelwert 6.37 (SD=2.23) einen signifikant höheren Wert als diejenigen in Gruppenbehandlung ($M=5.66$, $SD=2.54$; $t(186.451)=3.246$, $p=.001$, $d=0.28$).

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 abgefragt. Bezüglich Stärke der Depression zeigte sich für die Gesamtstichprobe ein durchschnittlicher Summenscore von 28.08 (SD=11.42, Range 0–61), der als Schwellenwert für eine schwere depressive Verstimmung interpretiert werden kann. Der Globalwert des BSI (GSI)

erbrachte mit einem mittleren *T*-Wert von 71.63 (SD=10.14, Range 26–80) für die Gesamtstichprobe zu Beginn der Behandlung einen Score, der mehr als zwei Standardabweichungen vom theoretischen Mittelwert von 50 abweicht. Mittels Online-Erhebung bewerteten die Behandlerinnen und Behandler die Funktionsfähigkeit ihrer Klientinnen und Klienten mittels GAF-Skala, die einen Mittelwert von 55.53 (SD=10.87, Range 23–91) erbrachte, ein Wert, der laut Manal „moderaten Funktionsschwierigkeiten“ entspricht. Die Gruppenvergleiche zeigten in allen drei klinischen Instrumenten statistisch signifikante Unterschiede zwischen den zwei Behandlungsformen, alle zulasten der Klientinnen und Klienten in Einzelbehandlung, sowohl in der Depressionsstärke ($t(1142)=5.143, p<.001, d=0.44$), in der psychischen Beeinträchtigung ($t(201.059)=3.541, p<.001, d=0.30$) als auch in der Einschätzung der globalen Funktionsfähigkeit durch die Behandlerinnen und Behandler ($t(592)=-3.770, p<.001, d=0.35$).

5. Ergebnisse zur Veränderungsmessung sowie zur Behandlungszufriedenheit

Die Steuernden Psychologinnen und Psychologen wiesen den Behandlerinnen und Behandlern für die Klientinnen und Klienten der Einzelbehandlung ein Kontingent von jeweils 30 Behandlungseinheiten und für diejenigen der Gruppenbehandlung ein Kontingent von jeweils 25 Behandlungseinheiten zu, das je nach Behandlungsstand flexibel angepasst werden konnte. Die Quote der Abbrüche lag insgesamt bei 16.5%, bei den Einzelbehandlungen mit 15.1% aber signifikant niedriger als bei den Gruppenbehandlungen mit 25.0% ($Chi^2(1)=10.344, p=.001$). Die insgesamt durchschnittlich verbrauchte Anzahl an Einheiten lag bei 25.49 (SD=8.12, Range 1–40), die statistische Analyse zeigte dabei einen bedeutsamen Unterschied zwischen den zwei Behandlungsformen (26.07 Einheiten (SD=8.35) für die Einzelbehandlungen, 21.49 Einheiten (SD=4.67) für die Gruppenbehandlung; $t(259.741)=9.082, p<.001, d=0.87$).

5.1. 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 Variablen (Arztbesuche, medizinische Untersuchungen, Krankenhausaufenthalte, Medikamenteneinnahme) positive Effekte zu verzeichnen waren. Tabelle 1 zeigt die Änderungen in Richtung eines niedrigeren Bedarfs an Gesundheitsleistungen für die Gesamtstichprobe kompakt zusammengefasst im Überblick. Unterschiedlich hohe Fallzahlen ergeben sich wiederum durch unvollständige Beantwortung einzelner Items zum Fragenkomplex. In keiner der vier Variablen konnten signifikante Unterschiede zwischen den Behandlungsformen gefunden werden ($p > .05$).

Die Frage nach generellen Veränderungen im Gesundheitsverhalten während der Zeit der Behandlung beantwortete der Großteil der befragten Klientinnen und Klienten (91.6%) als positiv. Tabelle 2 zeigt die Bewertungen für die einzelnen Bereiche Sport, Bewegung, Ernährung und Rauchverhalten im Überblick. Auch hier überwiegen die positiven Bewertungen in den abgefragten Teilbereichen, insbesondere bei Sport und Bewegungsverhalten. Beim Rauchverhalten zeigt die Kategorie „keine Änderung“ die größten Häufigkeiten, bedingt auch dadurch, dass ein Großteil der Klientinnen und Klienten nicht rauchte. Etwa zwei von drei aller Klientinnen und Klienten (64.7%) berichteten zudem von einer allgemeinen Verbesserung ihrer Leistungsfähigkeit durch die Behandlung. Auch in den Variablen zum Gesundheitsverhalten konnten keine signifikante Unterschiede zwischen den Behandlungsformen belegt werden ($p > .05$).

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

Folgetestung zeigte, dass 97 vorher arbeitslose Klientinnen und Klienten zwischenzeitlich ein neues Arbeitsverhältnis aufgenommen hatten, dem standen 44 Klientinnen und Klienten gegenüber, die in der Zwischenzeit ihren Arbeitsplatz verloren hatten ($p < .001$). Dieses positive Ergebnis konnte für beide Behandlungsformen beobachtet werden, bei der Einzelbehandlung war das Verhältnis von „positiven“ und „negativen“ Wechslern 81:40 ($p < .001$) und bei der Gruppenbehandlung 16:4 ($p = .012$).

5.2. Veränderungen bezüglich der validierten klinischen Fragebögen

Im BDI-II zeigte sich für die Klientinnen und Klienten der Einzelbehandlung eine Reduktion der Depressionswerte um 11.83 Punkte ($SD = 11.32$) und für diejenigen der Gruppenbehandlung eine Reduktion um 9.17 Punkte ($SD = 9.02$). Es war ein signifikanter Gruppeneffekt ($F(1, 577) = 6.725, p = .010, \eta^2 = .012$) zu beobachten, der Behandlungseffekt ($F(1, 577) = 329.681, p < .001, \eta^2 = .364$) war hoch signifikant, und auch die Wechselwirkung aus Behandlungsdauer und Gruppe erreichte Signifikanzniveau ($F(1, 577) = 5.283, p = .022, \eta^2 = .009$). Das heißt, dass sich unter Berücksichtigung der unterschiedlich hohen Depressions-Ausgangswerte für Klientinnen und Klienten der Einzelbehandlung statistisch gesehen ungleich größere Verbesserungen durch die Behandlung zeigten (siehe Tabelle 3).

In der GAF-Skala kam es für die Einzelbehandlung zu einer Zunahme der Werte um 13.63 Punkte ($SD = 11.50$) und für die Gruppenbehandlung zu einer solchen um 12.28 Punkte ($SD = 13.20$). Der statistische Vergleich zeigte einen hoch signifikanten Effekt für die Dauer der Behandlung ($F(1, 443) = 400.100, p < .001, \eta^2 = .475$), einen ebenfalls signifikanten Effekt für die Behandlungsform ($F(1, 443) = 4.695, p = .031, \eta^2 = .010$) sowie eine nicht signifikante Wechselwirkung zwischen Zeit und Behandlungsform ($F(1, 443) = 1.085, p = .298, \eta^2 = .002$).

Im Global Severity Index des BSI-53 zeigte sich für die Einzelbehandlung eine Reduktion des T -Werts um 8.92 Punkte ($SD = 12.31$) und für die Gruppenbehandlung eine um 7.73 Punkte ($SD = 11.82$). Der statistische Vergleich erbrachte einen hoch signifikanten Effekt der Dauer der Behandlung ($F(1, 581) = 166.766, p < .001,$

$Eta^2=.223$), einen nicht signifikanten Effekt für die Behandlungsform ($F(1, 581)=3.043, p=.082, Eta^2=.005$) sowie eine nicht signifikante Wechselwirkung zwischen Zeit und Behandlungsform ($F(1, 581)=0.846, p=.358, Eta^2=.001$).

Um diese Ergebnisse für die einzelnen Subskalen des BSI-53 zu verdeutlichen, sind in Abbildung 4 alle T -Werte ($M=50, SD=10$) aller neun Subskalen und des Globalwerts für beide Behandlungsformen dargestellt. Alle gemittelten T -Werte (ausgenommen Subskala 6 (Aggressivität/Feindseligkeit) für die Gruppenbehandlung) erreichten oder überschritten vor der Behandlung den Cutoff-Wert für Behandlungsbedürftigkeit. Abbildung 5 zeigt die durchschnittlichen Verbesserungen (Differenz aus t_1 und t_2 , daher die positive Ausrichtung) in jeder der einzelnen neun Subskalen und im Gesamtwert GSI im direkten Vergleich zwischen Einzel- und Gruppenbehandlung. Abbildung 6 stellt das Niveau der zwei Behandlungsformen in den Skalen nach Abschluss der Intervention dar. Für jede einzelne Skala ist in der aktuellen Analyse zu erkennen, dass die gemittelten T -Werte nach der abgeschlossenen Behandlung unter dem Cutoff-Wert für Behandlungsbedürftigkeit liegen.

5.3. 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 Gesamtwufriedenheit zwischen Klientinnen und Klienten der zwei Behandlungsformen verglichen. Sowohl diejenigen der Einzelbehandlung als auch jene der Gruppenbehandlung zeigten dabei Ergebnisse am höheren Ende bei Beziehungs- und Erfolgswufriedenheit wie auch bei der Gesamtwufriedenheit mit der Therapie ($p>.05$). Auch die Behandlerinnen und Behandler zeigten ein gleich hohes Maß an Zufriedenheit in allen drei Skalen des Instruments und unterschieden sich nicht signifikant voneinander (siehe Tabelle 4).

6. Dropout-Analyse

Eine wichtige Frage, die sich im Anschluss an die Darstellung der Ergebnisse stellte, war 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 nach Abschluss der Behandlung nicht mehr getan? Es kann zwar nicht erwartet werden, dass die anfängliche Quote der Beteiligung an der Evaluation von 56.7% (1190 Klientinnen und Klienten) nach einer durchschnittlichen Behandlungsdauer von etwa einem halben Jahr beibehalten werden kann, eine Beteiligungsquote von 29.2% (613 Klientinnen und Klienten) für beide Testungen ist jedoch als niedrig zu bewerten. 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 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:

Form der Behandlung (Einzel- vs. Gruppenbehandlung): Completerraten für Einzelbehandlung lag bei 48.6%, bei Gruppenbehandlung hingegen bei 69.0% ($\chi^2(1)=24.081, p<.001$)

Erhaltene Behandlungseinheiten (Completer: $M=25.94, SD=7.72$, Dropouts: $M=20.21, SD=10.28$; $t(1066.684)=10.821, p<.001$, Cohen's $d=0.63$)

Aktuelles Arbeitsverhältnis (Quote der Completer: 33.9%, Quote der Dropouts: 28.7%; $\chi^2(1)=3.621, p=.057$)

Auswirkung der psychischen Beschwerden auf die täglichen Aktivitäten (Completer: $M=6.13, SD=2.32$, Dropouts: $M=6.42, SD=2.24$; $t(1151)=2.174, p=.030$, Cohen's

$d=0.13$)

Summenwert des Depressionsscores (Completer: $M=27.19$, $SD=11.45$, Dropouts: $M=29.03$, $SD=11.32$; $t(1142)=2.729$, $p=.006$, Cohen's $d=0.16$)

T-Wert GSI (Completer: $M=70.90$, $SD=10.46$, Dropouts: $M=72.42$, $SD=9.73$; $t(1142.797)=2.542$, $p=.011$, Cohen's $d=0.15$)

F3-Diagnose (Quote der Completer: 52.4%, Quote der Dropouts: 60.3%; $\chi^2(1)=3.550$, $p=.060$)

F4-Diagnose (Quote der Completer: 59.9%, Quote der Dropouts: 51.1%; $\chi^2(1)=4.562$, $p=.033$)

Von diesen acht Variablen, die von insgesamt 536 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:

Behandlungsform: $\beta=-1.175$, $\text{Wald}(1)=20.821$, $p<.001$, $\text{Exp}(\beta)=0.309$

Erhaltene Behandlungseinheiten: $\beta= 0.066$, $\text{Wald}(1)=42.822$, $p<.001$, $\text{Exp}(\beta)=1.079$

T-Wert GSI: $\beta=-0.034$, $\text{Wald}(1)= 9.750$, $p=.002$, $\text{Exp}(\beta)=0.967$

F4-Diagnose: $\beta=-0.532$, $\text{Wald}(1)= 7.380$, $p=.007$, $\text{Exp}(\beta)=0.587$

Diese Prädiktoren weisen darauf hin, dass Klientinnen und Klienten in Einzelbehandlung, mit höheren psychischen Beeinträchtigungen, die aber keine F4-Diagnose erhalten hatten, sowie solche, die signifikant weniger Behandlungseinheiten bekommen hatten, sich mit höherer Wahrscheinlichkeit nicht an der Abschlusserhebung beteiligt haben (Trefferquote: 66.8%). Die präsentierten Gesamtergebnisse der Evaluation weisen daher tendenziell einen gewissen Bias durch diese Personengruppe auf, die zwar an der Ersttestung, nicht aber an der Abschlusserhebung teilgenommen hat.

7. Zusammenfassung

Für die Evaluation von fit2work im 4. Förderjahr standen 1190 vollständige Datensets der Ersttestung (Beginn der Behandlung) zur Analyse zur Verfügung (56.7%). Nach Abschluss der Behandlung langten von insgesamt 825 Klientinnen und Klienten der Studie ausgefüllte Fragebogensets ein. Davon mussten 212 Sets ausgeschieden werden, da von diesen Klientinnen und Klienten keine Aufnahmetestung zu Beginn der Behandlung vorlag. Daher ergab sich für die Gesamtevaluation eine Anzahl von insgesamt 613 Personen (29.2%), deren Daten in die statistische Analyse einbezogen werden konnten. Durch das Hinzufügen von Fremdbeurteilung durch die Behandlerinnen und Behandler lagen unter Zugrundelegung von Beurteilungen von Behandlungsbeginn und -ende insgesamt 270 Datensätze vollständig vor (12.9%).

Im Mittelpunkt der Evaluierung stand diesmal die Frage, ob Interventionen, die in Form von Einzel- bzw. Gruppenbehandlungen im Rahmen des Projekts fit2work angeboten wurden, das Befinden der Klientinnen und Klienten insgesamt signifikant verbessern können. Aufgrund stark unterschiedlicher Fallzahlen der drei angebotenen Arten konnten lediglich die Klientinnen und Klienten des „Sozialen und Emotionalen Kompetenztrainings“ als Vertreter der Gruppenbehandlungen in die Analyse aufgenommen werden. Operationalisiert wurde „Befinden“ im Rahmen dieses Projekts durch die Testverfahren BDI-II (Depression), BSI-53 (psychische Beeinträchtigung) sowie durch die GAF-Skala (Funktionsfähigkeit), beim BSI durch einen globalen Summenwert (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 erhoben 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 Zeit (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 beiden 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 waren überzufällig oft in der Gruppenbehandlung vertreten, verheiratete hingegen in der Einzelbehandlung ($p=.022$). Bezüglich arbeitsbezogener Variablen konnten keine statistisch signifikanten Unterschiede zwischen den beiden Behandlungsformen erhoben werden, wohl aber in den gesundheitsbezogenen Variablen: Klientinnen und Klienten in Einzelbehandlung wiesen signifikant mehr Doppeldiagnosen und F4-Diagnosen auf, auch die Ergebnisse der Fragebögen zeigten, dass es sich bei den Klientinnen und Klienten in Einzelbehandlung um eine stärker belastete Population handelte. Sowohl die Depressionsscores als auch Werte für psychische Belastung und Funktionseinschränkung zeigten signifikant höhere Ausgangswerte, wie auch die Variable der subjektiven Einschätzung der durchschnittlichen täglichen Belastung (alle $ps<.001$). Damit kann festgestellt werden, dass es sich bei Klientinnen und Klienten, die im Rahmen von fit2work ein „Soziales und Emotionales Kompetenztraining“ in Form von Gruppenbehandlung bekamen, um eine grundsätzliche andere Ausgangspopulation handelte als diejenige, die eine Behandlung im Einzelsetting erhielten, nicht vom arbeitsbezogenen Hintergrund her,

sondern was deren gesundheitsbezogenen Hintergrund und deren psychosoziale Problematiken betrifft.

Die Veränderungsmessung erbrachte für die klinischen Fragebögen und den GAF-Wert eindeutige und konsistente Ergebnisse, was die Veränderungen durch die Behandlung betrifft, übergreifend waren die erzielten Verbesserungen hoch signifikant mit sehr hohen Effektstärken. Was die Gruppeneffekte betrifft, zeigten diese wie in der univariaten Auswertung auch in der Prüfung des Gesamtmodells signifikante Ergebnisse für BDI-II und GAF und zumindest tendenziell signifikante Werte für den Summenscore des BSI-53. Was aber in dieser Modellprüfung hervorsteht, ist die signifikante Wechselwirkung im BDI-II, was bedeutet, dass die Reduktionen der depressiven Symptomatik in Einzelbehandlung bei zwar höherem Ausgangswert signifikant stärker ausfallen und dass zumindest in diesem Bereich die Klientinnen und Klienten von ihrer Behandlung stärker profitieren.

Was die Subskalen der psychischen Beeinträchtigungen im BSI-53 betrifft, lagen alle Verbesserungen im Bereich von etwa 4 bis 9 *T*-Wert-Punkten, also Verbesserungen im Bereich von knapp einer halben bis fast einer ganzen Standardabweichung. Alle Mittelwerte der Subskalen, die vor der Intervention über dem Cutoff-Wert ($T=63$) für Behandlungsbedürftigkeit lagen, befanden sich nach der Abschlusserhebung unter diesem Trennwert. Zweitens zeigte sich, dass sich die beiden Behandlungsformen in keiner der Subskalen signifikant voneinander unterschieden, d.h. dass sie gleich gute Ergebnisse erzielten (Haupteffekt Gruppe), und drittens, dass auch die Wechselwirkungen zwischen Zeit und Gruppe (nach Durchführung der Alphafehler-Korrektur nach Bonferroni-Holm) keine signifikanten Ergebnisse mehr aufwiesen.

75% 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 62% kam es zu einer Reduktion der Krankenstände. Als besonderes Ergebnis zeigte sich im Rahmen dieser Evaluation, dass sich infolge der Behandlungen positive Auswirkungen auf die Beschäftigungssituation zeigten: Die Quote der Wechsler von Klientinnen und

Klienten, die vor der Behandlung arbeitslos und während der Behandlung in den Arbeitsprozess zurückgekehrt waren, war signifikant größer als diejenige, die den umgekehrten Verlauf zeigte, d.h. von in Arbeit vorher zu arbeitslos danach. Dieses Ergebnis zeigte sich nicht nur für die Gesamtgruppe, sondern auch für die beiden Behandlungsformen getrennt. Dieser Erfolg deutet einen ersten Schritt in Richtung des übergeordneten Ziels der beruflichen Reintegration durch die Teilnahme an den Angeboten von fit2work an.

Für das Gesamtkollektiv der Klientinnen und Klienten zeigten sich maßgebliche Veränderungen im Sinne einer umfassenden Reduktion der Inanspruchnahme von Gesundheitsleistungen, wobei insbesondere folgende Punkte zu nennen sind:

- Reduktion von Krankenhausaufenthalten (bei 57%), von Arztbesuchen (bei 47%), von ärztlichen Untersuchungen (bei 46%) und von Einnahme von Psychopharmaka (bei 35%),
- Positive Veränderungen im Bewegungsverhalten (bei 78%), im Sportverhalten (bei 56%) und im Ernährungsverhalten (bei 44%),
- 63% der Klientinnen und Klienten schätzten die erhaltene Behandlung 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 Erfolgswzufriedenheit 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.

Interessant erscheint beim Vergleich der Behandlungsformen, dass es bei der Gruppenbehandlung zu mehr expliziten Abbrüchen kam als bei der Einzelbehandlung, insgesamt brach dort jeder Vierte die Behandlung vorzeitig ab, bei der Einzelbehandlung war das nur jeder Siebente. Zu empfehlen wären hier verstärkte Maßnahmen zur Erhöhung der Adhärenz durch Behandlerinnen und Behandler der Gruppenbehandlung, um die Quote der Abschlüsse anzugleichen. Abschließend muss

im Sinne von Limitationen darauf hingewiesen werden, dass im Rahmen der Evaluation nicht alle Klientinnen und Klienten für die Abschlusstestung in gleicher Weise erreicht werden konnten: Die Resultate der Dropout-Analyse zeigen, dass die hier präsentierten Gesamtergebnisse der Evaluation tendenziell einen gewissen Bias durch eine umschriebenen Personengruppe aufweisen, die sich zwar an der Ersttestung, nicht aber mehr an der Abschlusserhebung beteiligt hatte. Klientinnen und Klienten mit höheren psychischen Beeinträchtigungen, die aber nicht eine F4-Diagnose aufwiesen, die signifikant weniger Behandlungseinheiten absolvierten, vor allem aber solche in Einzelbehandlung beteiligten sich an der Abschlusstestung weniger häufig. Bezüglich Ausschöpfung der Behandlungseinheiten wäre es für beide Behandlungsformen anzuraten, nicht allzu schnell den Abschluss der Behandlung anzustreben, sondern diesen langsam ausschleichend vorzunehmen. Die niedrigere Beteiligung von Klientinnen und Klienten der Einzelbehandlung hingegen kann aber auch mit einer logistischen Besonderheit zu tun haben, nämlich dass jene in Einzelbehandlung bislang ihre Fragebögen einzeln vor und nach Abschluss der Behandlung direkt verschickt haben, während bei der Gruppenbehandlung alle ausgefüllten Fragebögen, anonymisiert gesammelt in zugeklebten Kuverts, von den Behandlerinnen und Behandlern erst nach Abschluss der Behandlung gemeinsam verschickt wurden. Unter dem Aspekt der weitaus höheren Beteiligung der Gruppenbehandlung an der Evaluation wäre diese Art der Übermittlung der Fragebogen zukünftig auch für die Klientinnen und Klienten der Einzelbehandlung anzuraten.

Danksagung. Dank gebührt der Wissenschaftlichen Mitarbeiterin Vera Noll, MSc sowie Studienassistentin Alina Dorsch, BSc BFA 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.
- 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.
- Jagsch, R. (2015). *Endbericht – Evaluierung des arbeitsbezogenen Therapieerfolgs der klinisch-psychologischen und psychotherapeutischen Behandlung im Rahmen des Projekts „fit2work“*. Unveröffentlichter Forschungsbericht.
- Jagsch, R. (2018). *Endbericht – Evaluierung der Einzelbehandlungen im Rahmen des Projekts fit2work klinisch-psychologische und psychotherapeutische Behandlung 3. Förderjahr*. Unveröffentlichter Forschungsbericht.
- Saß, H., Wittchen, H.-U., Zaudig, M. & Houben, I. (2003). *Diagnostische Kriterien des Diagnostischen und Statistischen Manuals psychischer Störungen. DSM-IV-TR*. Göttingen: Hogrefe.

Anhang A. Abbildungen

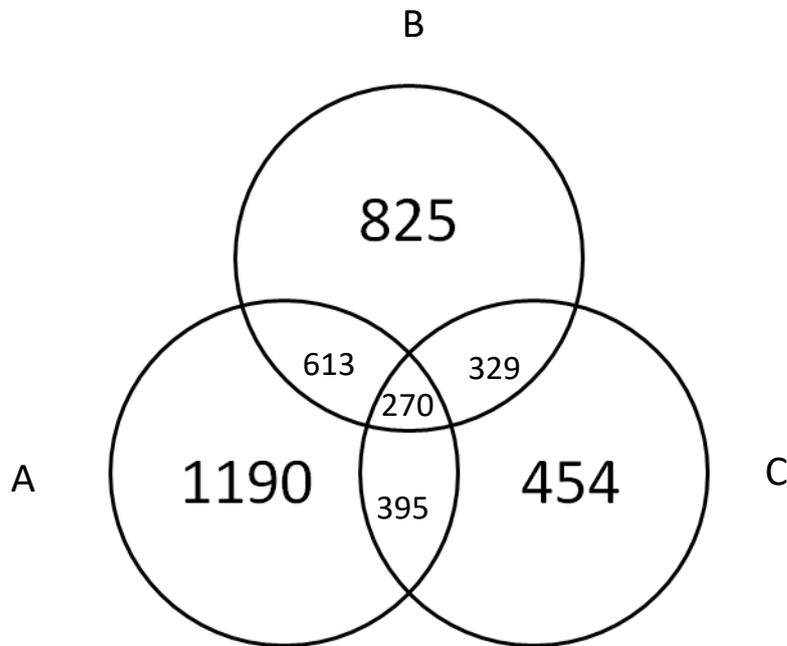


Abbildung 1. Venn-Diagramm zu den vorliegenden Datensätzen der beiden Teilstichproben (Einzel- und Gruppenbehandlungen).

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 und -ende

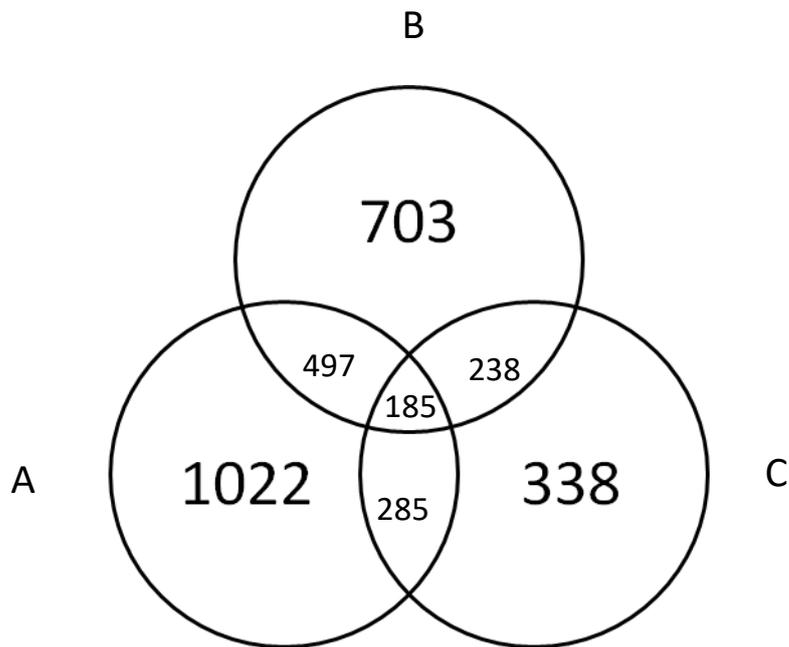


Abbildung 2. Venn-Diagramm zu den vorliegenden Datensätzen der Einzelbehandlungen.
Anmerkung. A Fragebogensets der Klientinnen und Klienten zu Behandlungsbeginn, B Fragebogensets der Klientinnen und Klienten zu Behandlungsende, C Online-Fragebogensets der Behandlerinnen und Behandler zu Behandlungsbeginn und -ende

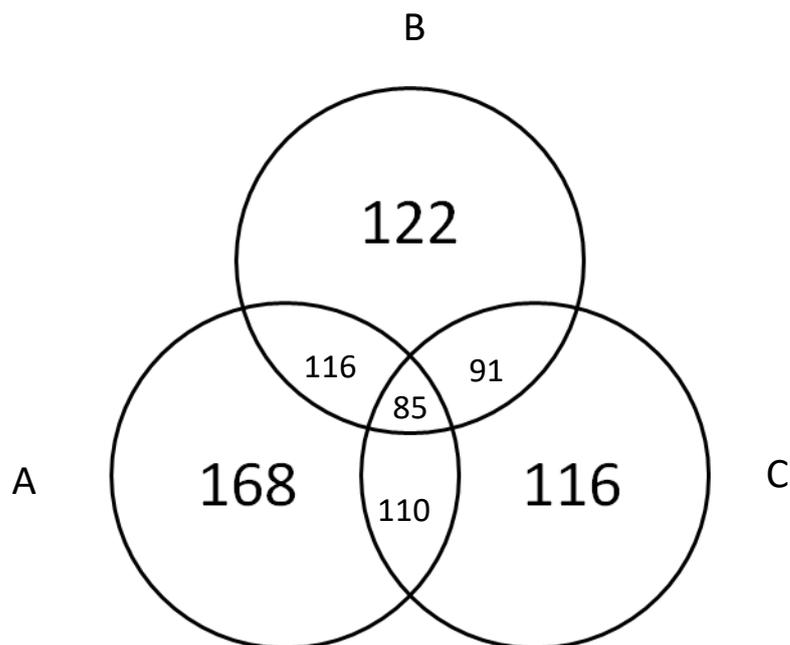


Abbildung 3. Venn-Diagramm zu den vorliegenden Datensätzen der Gruppenbehandlungen.
Anmerkung. A Fragebogensets der Klientinnen und Klienten zu Behandlungsbeginn, B Fragebogensets der Klientinnen und Klienten zu Behandlungsende, C Online-Fragebogensets der Behandlerinnen und Behandler zu Behandlungsbeginn und -ende

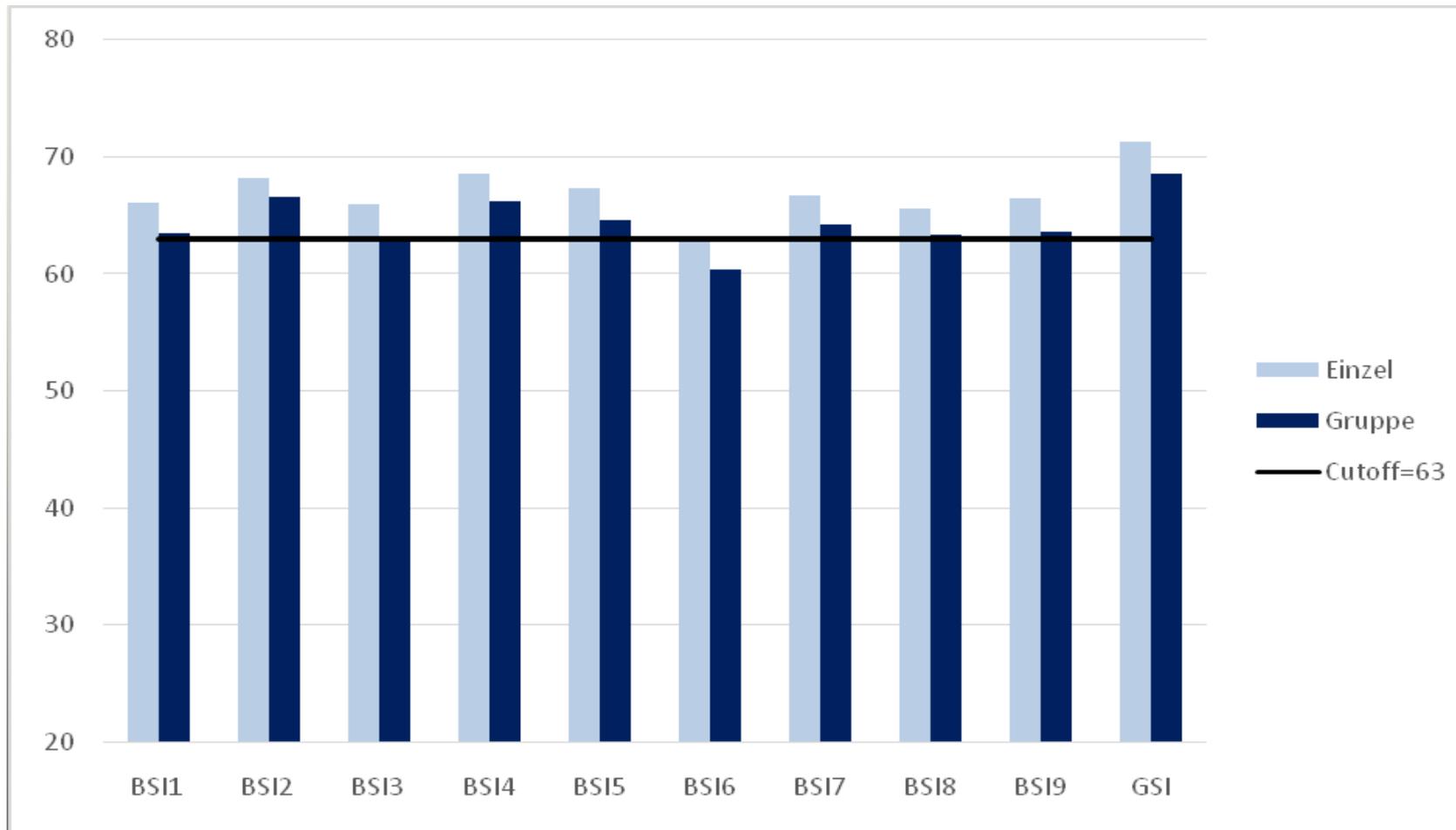


Abbildung 4. Vergleich der zwei Behandlungsformen im Brief Symptom Inventory (BSI) zu Behandlungsbeginn. Entsprechend den Angaben des BSI-Manuals weist ein T-Wert ab 63 (waagrechte Linie) auf Behandlungsbedürftigkeit hin.

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)

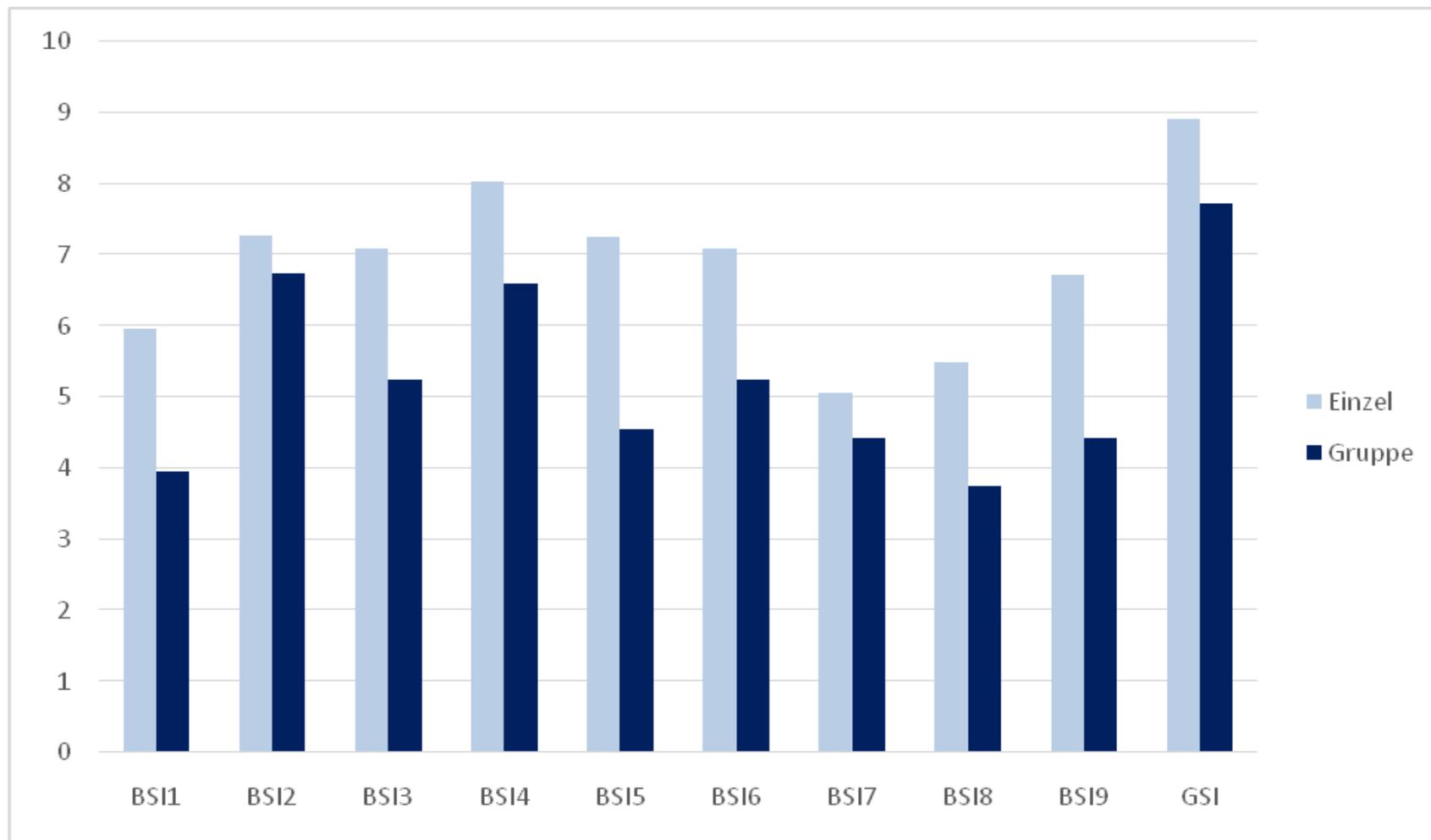


Abbildung 5. Vergleich der Verbesserungen (Differenzen in T-Werten) im Brief Symptom Inventory (BSI) zwischen den zwei Behandlungsformen. Anmerkung. BSI1=Somatisierung, BSI2=Zwanghaftigkeit, BSI3=Unsicherheit im Sozialkontakt, BSI4=Depressivität, BSI5=Ängstlichkeit, BSI6=Aggressivität/Feindseligkeit, BSI7=Phobische Angst, BSI8=Paranoides Denken, BSI9=Psychotizismus, GSI=Globalwert (Global Severity Index)

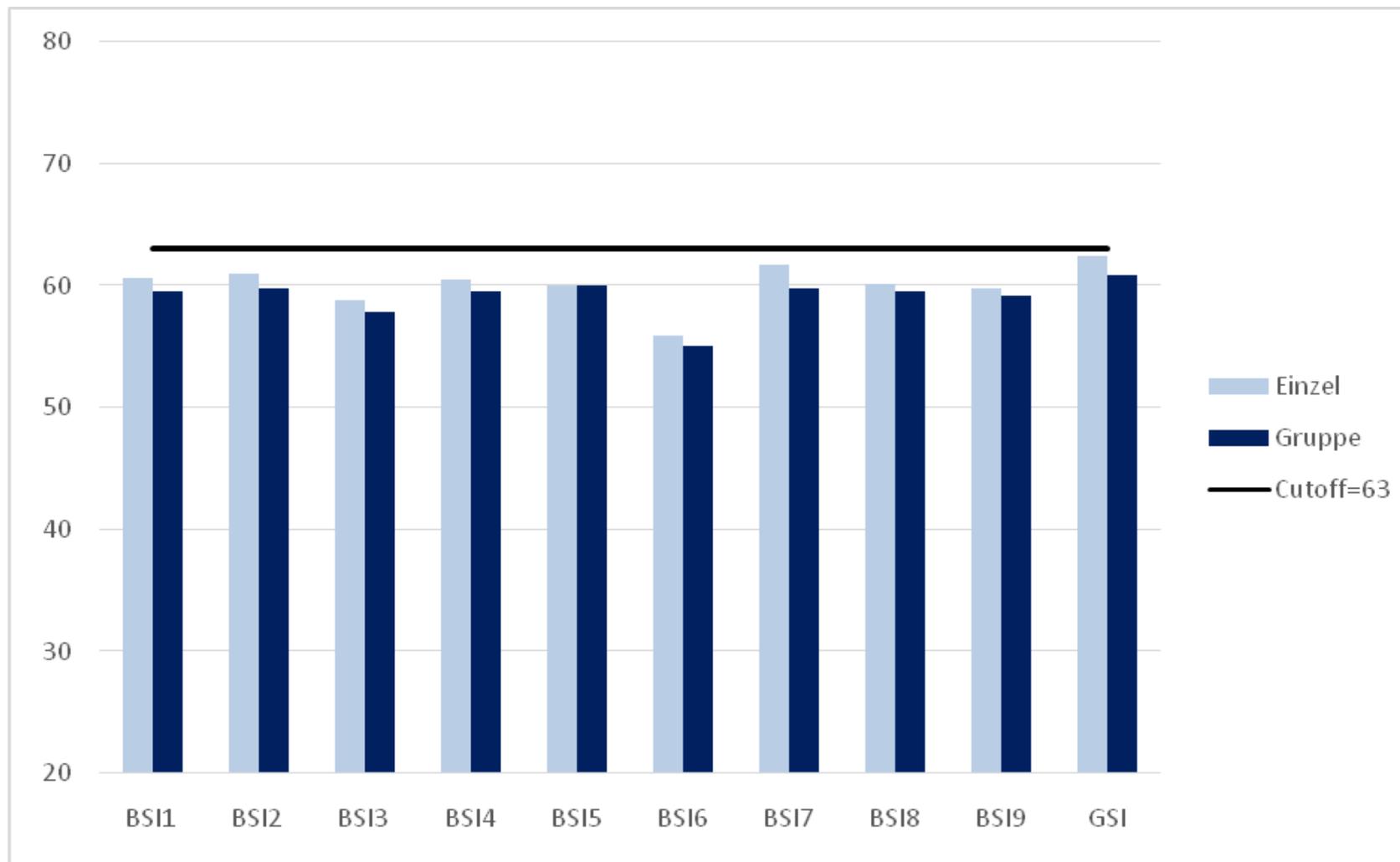


Abbildung 6. Vergleich der zwei Behandlungsformen im Brief Symptom Inventory (BSI) zu Behandlungsende. Entsprechend den Angaben des BSI-Manuals weist ein T-Wert ab 63 (waagrechte Linie) auf Behandlungsbedürftigkeit hin.

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	326 (45.9%)	335 (47.4%)	330 (57.3%)	232 (34.5%)
keine	320 (45.0%)	298 (42.1%)	220 (38.2%)	343 (51.1%)
Zunahme	65 (9.1%)	74 (10.5%)	26 (4.5%)	97 (14.4%)

Tabelle 2. *Absolute und relative Veränderungen im Bereich des Gesundheitsverhaltens*

Art der Änderung n (%)	Sportverhalten	Bewegungsverhalten	Ernährungsverhalten	Rauchverhalten
positiv	321 (55.8%)	450 (77.6%)	249 (43.9%)	87 (19.5%)
keine	226 (39.2%)	110 (19.0%)	291 (51.3%)	324 (72.4%)
negativ	29 (5.0%)	20 (3.4%)	27 (4.8%)	36 (8.1%)

Tabelle 3: Vergleich der Evaluationsergebnisse zwischen Einzel- und Gruppenbehandlung vor der und nach Abschluss der Behandlung

	Einzelbehandlung		Gruppenbehandlung		p_z	p_B	p_{z*B}
	t1 M (SD)	t2 M (SD)	t1 M (SD)	t2 M (SD)			
BDI-II	27.97 (11.47)	16.14 (12.01)	23.83 (10.63)	14.65 (11.09)	<.001	.010	.022
GAF	55.14 (10.26)	68.77 (14.00)	58.34 (11.80)	70.62 (12.94)	<.001	.031	.298
BSI (T-Werte)							
Somatisierung	66.60 (10.85)	60.64 (12.60)	63.48 (13.18)	59.53 (12.27)	<.001	.061	.083
Zwanghaftigkeit	68.23 (9.81)	60.96 (12.58)	66.58 (10.03)	59.83 (13.09)	<.001	.188	.654
Unsicherheit im Sozialkontakt	65.93 (11.59)	58.85 (13.00)	63.04 (11.28)	57.80 (12.26)	<.001	.088	.131
Depressivität	68.53 (9.92)	60.49 (12.45)	66.17 (10.41)	59.57 (12.61)	<.001	.115	.206
Ängstlichkeit	67.32 (11.92)	60.07 (13.77)	64.61 (11.71)	60.06 (13.48)	<.001	.259	.032
Aggressivität/Feindseligkeit	63.06 (11.84)	55.96 (12.44)	60.36 (10.92)	55.11 (12.63)	<.001	.111	.141
Phobische Angst	66.73 (13.05)	61.67 (13.66)	64.21 (13.58)	59.77 (13.51)	<.001	.084	.598
Paranoides Denken	65.63 (11.03)	60.13 (11.53)	63.33 (11.30)	59.58 (12.49)	<.001	.181	.102
Psychotizismus	66.46 (11.72)	59.74 (12.51)	63.56 (11.75)	59.14 (12.60)	<.001	.125	.048
Global Severity Index	71.31 (10.21)	62.39 (15.03)	68.60 (11.48)	60.87 (15.44)	<.001	.082	.358

Anmerkung. p_z Zeit, p_B Behandlungsform, p_{z*B} Interaktion Zeit *Behandlungsform

Tabelle 4. Vergleich der Ergebnisse im Helping Alliance Questionnaire (HAQ) zwischen den zwei Behandlungsformen (Einzel- versus Gruppenbehandlung) aus Sicht der Klientinnen und Klienten sowie der Behandlerinnen und Behandler

	Einzelbehandlung M (SD)	Gruppenbehandlung M (SD)	t	p
HAQ				
Klientenseitig				
Beziehungszufriedenheit	5.41 (0.69)	5.46 (0.75)	-0.759	.448
Erfolgswufriedenheit	4.76 (0.96)	4.93 (0.82)	-1.844	.066
Gesamtwufriedenheit	5.11 (0.74)	5.22 (0.68)	-1.493	.136
Behandlerseitig				
Beziehungszufriedenheit	5.33 (0.64)	5.37 (0.67)	-0.710	.478
Erfolgswufriedenheit	4.87 (0.92)	4.90 (0.93)	-0.341	.733
Gesamtwufriedenheit	5.12 (0.73)	5.16 (0.75)	-0.540	.589
(Wertebereich jeweils von 1=sehr unzufrieden bis 6=sehr zufriede)				