

# Die Erfassung allgemeiner Wirkfaktoren in der Psychotherapie

## Eine Validierung des Mainzer Stundenbeurteilungsbogens (MSB)

Anne-Kathrin Bräscher und Michael Witthöft

Abteilung für Klinische Psychologie, Psychotherapie und experimentelle Psychopathologie, Johannes Gutenberg-Universität Mainz

**Zusammenfassung: Theoretischer Hintergrund:** Stundenbögen dienen einer standardisierten Prozessdiagnostik und Verlaufsdokumentation. Bislang vorliegende Instrumente werden in der Praxis häufig nicht eingesetzt, haben einen eingeschränkten Geltungsraum oder weisen psychometrische Schwächen auf. Der Mainzer Stundenbeurteilungsbogen (MSB) versucht allgemeine Wirkfaktoren ökonomisch zu erfassen. Ziel der Studie ist die Untersuchung des MSB in der Patienten- und Therapeutenversion hinsichtlich Faktorenstruktur, Itemstatistiken und Gütekriterien. **Methoden:** Der MSB wurde zu zwei Zeitpunkten (5. und 20. therapeutische Sitzung) in einer Stichprobe ambulanter Psychotherapiepatienten ( $N = 4.608$ ,  $N = 3.909$ ) und deren Therapeuten ( $N = 4.650$  und  $N = 3.980$  Bögen) untersucht. Die Faktorenstruktur wurde mittels explorativer und konfirmatorischer Faktorenanalysen untersucht und es wurden Reliabilitäts- und Validitätsanalysen vorgenommen. **Ergebnisse:** Ein Bifaktorenmodell mit drei Faktoren spiegelte weitgehend die theoretisch postulierte Struktur mit den Subskalen *Therapiebeziehung*, *Problemaktualisierung* und *Problembewältigung* wider, bei moderatem Modellfit. Maße der diskriminanten, konvergenten und prädiktiven Validität fielen zufriedenstellend aus. **Diskussion:** Der MSB ist ein reliables, valides und ökonomisches Instrument zu Erfassung der allgemeinen Wirkfaktoren *Therapiebeziehung*, *Problembewältigung* und *Problemaktualisierung* aus Patienten- und Therapeutensicht.

**Schlüsselwörter:** Prozessdiagnostik, Fragebogenvalidierung, therapeutische Wirkfaktoren, therapeutische Beziehung, Stundenbeurteilungsbogen

### Assessment of Common Impact Factors in Psychotherapy – Validation of the Mainz Hourly Assessment Form (MSB)

**Abstract: Introduction:** Hourly assessment forms enable standardized formative evaluation and documentation during psychotherapy. Currently available assessment instruments, however, are rarely used in practice and have a limited scope or psychometric shortcomings. The Mainz Hourly Assessment Form (MSB) strives to economically assess common factors of psychotherapy. This study explores the factorial structure, item statistics, and psychometric properties of the Mainz Hourly Assessment Form in both the patient and therapist versions. **Methods:** We applied the Mainz Hourly Assessment Form at two points in time (5th and 20th therapeutic session) in a sample of psychotherapy outpatients ( $N = 4,608$ ,  $N = 3,909$  questionnaires) and their therapists ( $N = 4,650$  and  $N = 3,980$  questionnaires). The factorial structure was explored via exploratory and confirmatory factor analyses, and reliability and validity were assessed. **Results:** A bifactor model with three specific factors largely mirrored the theoretically postulated structure with the subscales *therapy alliance*, *problem activation*, and *mastery*, with a moderate model fit. The measures of discriminant, convergent, and predictive validity were satisfactory. **Discussion:** The Mainz Hourly Assessment Form is a reliable, valid, and economic instrument for the assessment of the common factors *therapy alliance*, *problem activation*, and *mastery* from both the patient and therapist perspective.

**Keywords:** formative evaluation, test construction, common factors, therapeutic alliance, hourly assessment form

Die Therapieprozessforschung belegt konsistent den Zusammenhang von Therapieprozessmerkmalen (z. B. therapeutische Beziehung) und dem Erfolg einer Psychotherapie im Sinne von Symptomreduktion (Fluckiger et al., 2018; Horvath & Symonds, 1991; Martin et al., 2000) und Therapieabbruch (Sharf et al., 2010). Die kontinuierliche Erhebung und Rückmeldung von Prozessmerkmalen im Verlauf der Therapie ermöglicht in diesem Sinne eine adaptive Indikation (Krampen and Wald, 2001) und ständige An-

passung des therapeutischen Vorgehens. Fragebögen dienen hierbei als ökonomische Instrumente zur standardisierten Prozessdiagnostik und Verlaufsdokumentation. Zur Sicherung von Prozess- und Ergebnisqualität wird nicht erst seit der Qualitätsmanagement-Richtlinienänderung des gemeinsamen Bundesausschusses in 2015 der Einsatz psychometrischer Instrumente zu Beginn, im Verlauf und am Ende einer Psychotherapie empfohlen (Gemeinsamer Bundesausschuss, 2015). Während jedoch bereits die Er-

hebung von symptomorientierten standardisierten Fragebögen (z. B. BSI, Franke, 2000) in der ambulanten Versorgung selten zu sein scheint (Heider & Zaby, 2012) ist von einer noch selteneren Erfassung von Therapieprozessmerkmalen (z. B. mittels Stundenbögen) auszugehen.

Verfügbare Instrumente unterscheiden sich neben ihrer Länge hinsichtlich verschiedener Aspekte; so liegen für einige Instrumente nicht nur Fragebogenversionen für Patienten, sondern auch für Therapeuten vor. Hierbei können sich die Therapeutenversionen in Bezug auf die eingenommene Perspektive unterscheiden, zum Beispiel Selbsteinschätzung des Therapeuten beim Berner Therapeutenbogen 2000 (Flückiger et al., 2010) oder auf den Patienten bezogene Fremdeinschätzung beim Helping Alliance Questionnaire (HAQ; Bassler et al., 1995). Es zeigt sich, dass die Patienteneinschätzung meist stärker mit dem Therapieerfolg zusammenhängt als die Einschätzung der Therapeuten (Horvath & Symonds, 1991). Während ältere Fragebögen häufig in einer schulenspezifischen Tradition entwickelt wurden (z. B. psychodynamische Konzeption des HAQ, Bassler et al., 1995; gesprächspsychotherapeutischer Hintergrund des BIKEP, Höger & Eckert, 1997), beziehen sich neuere Instrumente auf schulübergreifende, eklektische oder integrative Konzepte (z. B. WAI-SR, Wilmers et al., 2008; STEP, Krampen & Wald, 2001; für einen Überblick vgl. Hank & Krampen, 2008). Entsprechend existieren wesentliche Unterschiede zwischen den Instrumenten im Hinblick auf die erfassten Merkmale und Konzepte (z. B. Definition und Bestandteile der therapeutischen Beziehung).

Der Erfassung der Wirkfaktoren nach Grawe widmen sich bislang nur wenige Fragebögen: Beim WIFA-k (von Consbruch et al., 2013) etwa handelt es sich um ein Instrument zur Fremdeinschätzung der vier Wirkfaktoren (*Problembewältigung*, *motivationale Klärung*, *Ressourcenaktivierung* und *Problemaktualisierung*, Grawe et al., 1994) und der *therapeutischen Beziehung*. Der Berner Patienten- und Therapeutenstundenbogen 2000 (Flückiger et al., 2010) misst unter anderem mit Hilfe von 22 (Patientenversion) bzw. 27 (Therapeutenversion) Items die vier postulierten allgemeinen Wirkfaktoren, hat dabei jedoch das Ziel vielseitige Aspekte des Therapieprozesses zu erfassen, weshalb die Maßgaben der klassischen Fragebogenkonstruktion zugunsten inhaltlich-konzeptioneller Aspekte nicht beachtet wurden. Anders der STEP (Krampen & Wald, 2001), der inhaltlich überlappende Items formuliert und die Subskalen *motivationale Klärung und Entwicklung*, *Hilfe zur Problembewältigung* sowie *therapeutische Beziehung und Unterstützung* umfasst. Die ursprünglich intendierte Unterscheidung von *Ressourcenaktivierung* und *Problemaktualisierung* gelang hier nicht, zudem wurden zwei ursprünglich vorgesehene Items wegen zu geringer Kommunalitäten ausgeschlossen. Daher wurden die

jeweils verbliebenen Items zur ad-hoc gebildeten Subskala *therapeutische Beziehung und Unterstützung* zusammengefasst. Unter inhaltlichen Gesichtspunkten scheint es fraglich, ob die drei verbliebenen Items die therapeutische Beziehung in geeigneter Weise repräsentieren, zudem erscheinen manche Items aufgrund von Relativierungen in der Formulierung in Verbindung mit der verwendeten Antwortskala mehrdeutig (Lienert & Raatz, 1998).

Mit dem Ziel im Rahmen der Routineevaluation einer universitären Hochschulambulanz allgemeine Wirkfaktoren von Psychotherapie wiederholt im Therapieverlauf zu erfassen, wurde auf der Basis vorhandener Instrumente ein Stundenbogen in einer Patienten- und einer Therapeutenversion entwickelt, der auf die Bedürfnisse der Routineversorgung angepasst ist und versucht Schwächen vorhandener Instrumente zu verbessern. Ziel dieser Studie ist es, auf einer breiten Datenbasis die Faktorenstruktur des Mainzer Stundenbeurteilungsbogens (MSB) sowie die Itemstatistiken und Gütekriterien zu zwei Messzeitpunkten während einer ambulanten kognitiv-verhaltenstherapeutischen Psychotherapie zu überprüfen.

## Methoden

### Stichprobenbeschreibung

Die vorliegenden Analysen beziehen sich auf eine Patientenstichprobe (s. Alter, Geschlecht, Familienstand, Bildung und Diagnosestatus in Tab. S1 im elektronischen Supplement 1), die an einer universitären Hochschulambulanz zwischen den Jahren 2004 und 2019 mit kognitiv-behavioraler Psychotherapie behandelt wurde.

Die Stichprobengrößen umfassen  $N = 4.608$  Patienten zum ersten Messzeitpunkt und  $N = 3.909$  Patienten zum zweiten Messzeitpunkt. Parallel dazu wurden von Therapeuten zum ersten Messzeitpunkt  $N = 4.650$  Stundenbeurteilungsbögen und zum zweiten Messzeitpunkt  $N = 3.980$  Stundenbeurteilungsbögen ausgefüllt.

Bei den Therapeuten handelte es sich um verhaltenstherapeutische Ausbildungskandidaten (87%) und approbierte Psychotherapeuten (12%) im Alter von 23 bis 53 Jahren ( $M = 31$ ,  $SD = 4.8$  Jahre, 77% weiblich).

### Verwendete Instrumente

Mainzer Stundenbeurteilungsbogen (MSB). Der MSB liegt in einer Therapeuten- (MSB-T) und einer Patientenversion (MSB-P, vgl. Fragebogen im elektronischen Supplement 2) vor und besteht aus jeweils 15 Items. Drei Subskalen messen die Wirkfaktoren *therapeutische Beziehung*, *Problem-*

*aktualisierung* und *Problembewältigung*. Bislang vorliegende Stundenbögen sehen eine Beurteilung nach jeder einzelnen Sitzung vor. Aufgrund des hohen Aufwands, die eine Erhebung nach jeder Therapiesitzung mit sich bringt, bezieht sich die Instruktion des MSB auf die letzten 5 Sitzungen, so dass lediglich alle fünf Sitzungen eine Erhebung stattfinden muss. Die Instruktion der Patientenversion lautet: „Bitte lesen Sie die folgenden Aussagen genau und beurteilen Sie jeweils, in welchem Maße Sie die letzten 5 Sitzungen in der angegebenen Art erlebt haben. Markieren Sie jeweils die Antwortkategorie, die Ihrem Empfinden am besten entspricht. Bedenken Sie, dass nur eine realistische Einschätzung für Ihre\_n Therapeuten\_in hilfreich ist.“ Bei der Instruktion der Therapeutenversion wurde auf den letzten Satz verzichtet.

Die verwendete Antwortskala ist 7-stufig und drückt das jeweilige Ausmaß des Zutreffens der Aussagen (*überhaupt nicht* bis *sehr stark*) aus. Die fünf Items zur Messung der *therapeutischen Beziehung* wurden in Anlehnung an Konzepte nach Rogers (1957) konstruiert und sollen Empathie, Akzeptanz und Wertschätzung messen. Die Itemgenerierung der Skalen *Problemaktualisierung* und *Problembewältigung* erfolgte basierend auf Grawes Beschreibung der Wirkfaktoren (Grawe et al., 1994). Zwei der fünf Items zur Erfassung der Skala *Problembewältigung* wurden modifiziert aus dem STEP (Krampen & Wald, 2001) übernommen (Items 01 und 08). Die fünf Items der Skala *Problemaktualisierung* ähneln ebenfalls teilweise Items des STEP (Items 12 und 15) und erfassen das emotionale und kognitive Problemlernen.

Working Alliance Inventory – revidierte Kurzversion (WAI-SR). Das Working Alliance Inventory (WAI-SR) ist ein Selbstbeurteilungsinstrument, das die aktuelle therapeutische Arbeitsbeziehung beschreibt (Hatcher & Gillaspay, 2006; deutschsprachige Version von Wilmers et al., 2008). Das WAI-SR orientiert sich hierbei an Bordins therapieübergreifender Konzeption der therapeutischen Allianz (Bordin, 1979). Das Instrument liegt in einer Patienten- (WAI-SR-P) und einer Therapeutenversion (WAI-SR-T) vor und besteht jeweils aus 12 Items, die den Skalen *Bindung* (emotionale Bindung; die gemeinsame Arbeit basiert auf gegenseitigem Vertrauen und gegenseitiger Verpflichtung), *Aufgaben* (Abstimmung angewandter Interventionen oder Techniken und Vereinbarung der Rahmenbedingungen und Regeln der Therapie) und *Ziele* (Verständigung und Übereinstimmung bezüglich anzustrebende Therapieziele) zugeordnet werden können. Die Items werden auf einer 5-stufigen Likert-Skala bewertet (1, *selten* bis 5, *immer*). Die Reliabilität ist gut (Patientenversion:  $\alpha = .81-.91$ ; Therapeutenversion:  $\alpha = .82-.92$ ) und die konvergente Validität zeigte sich durch hohe Korrelationen mit dem HAQ (Luborsky, 1976; deutschsprachige Version von Bassler et al., 1995).

Brief Symptom Inventory. Beim Brief Symptom Inventory (BSI; Derogatis & Spencer, 1993; deutsche Version von Franke, 2000) handelt es sich um ein Selbstbeurteilungsinstrument mit 53 Items, das die subjektiv empfundene Beeinträchtigung durch körperliche und psychische Symptome erfasst. Die Items werden auf einer 5-stufigen Likert-Skala eingeschätzt (0, *überhaupt nicht*, bis 4, *sehr stark*) und lassen sich 9 Subskalen zuordnen. In der vorliegenden Arbeit wird insbesondere auf den Global Severity Index (GSI) Bezug genommen, der die grundsätzliche psychische Belastung erfasst, indem der Summenwert aller Items durch die Anzahl beantworteter Items dividiert wird. Die interne Konsistenz des GSI ist mit  $\alpha = .92-.96$  als exzellent einzustufen.

## Design und Datenerhebung

Die vorgestellten Analysen basieren auf Daten der Stundenbeurteilungsbögen von der 5. und der 20. Sitzung einer ambulanten Psychotherapie. Die Daten wurden im Rahmen der Routineevaluation zur Qualitätssicherung erhoben. Eine Erhebung des MSB fand ab der Probatorik alle 5 Sitzungen wiederholt statt. Die Messzeitpunkte zum Zwecke der Validierung wurden ausgewählt, indem der frühestmögliche Zeitpunkt während der Therapie (5. Sitzung) ausgewählt wurde sowie ein möglichst später Zeitpunkt, zu dem noch möglichst viele Daten vorhanden waren (Ende der Kurzzeittherapie bzw. fortgeschrittener Zeitpunkt einer Langzeittherapie, 20. Sitzung). Die Datenerhebung fand in der Regel elektronisch über Tablet-Computer statt und die Ergebnisse wurden den Therapeuten in elektronischer Form rückgemeldet. Der BSI wurde im Rahmen der Routineevaluation noch vor dem Erstgespräch, zur ersten probatorischen Sitzung und je nach Therapiedauer zur 10., 20., 40., 55., 75., 90., 95. und 115. therapeutischen Sitzung erhoben. Der Baseline-Wert des GSI ergab sich entsprechend aus dem ersten vorliegenden Wert (vor dem Erstgespräch bzw., falls dieser Wert fehlte, erste probatorische Sitzung), der Post-Wert des GSI ergab sich aus der letzten vorhandenen Messung (last-observation-carried-forward mit intention-to-treat-Analyse). Der WAI-SR wurde zum Zweck der Validierung des Stundenbogens für eine kurze Zeitspanne jeweils gleichzeitig mit dem MSB erhoben.

Demographische Angaben der Patienten wurden vor dem Erstgespräch mit Hilfe eines Anamnesebogens erfasst. Fünf Sitzungen vor Therapieende wurden die Patienten im Rahmen der Abschlussdiagnostik hinsichtlich ihrer Zufriedenheit mit der Behandlung befragt (6-stufige Skala von *völlig zufrieden* bis *völlig unzufrieden*). Zudem sollten sie einschätzen, wie sich die Therapie auf ihr spezifisches Problem ausgewirkt hat, weswegen sie

**Tabelle 1.** Übersicht über die Faktorladungen der Items für die verschiedenen explorativen Faktorenanalysen im Bifaktormodell (vgl. Tab. S1 – S4). Alle Items luden zudem jeweils auf dem g-Faktor

Item <sup>1</sup>	Skala	MSB-T05	MSB-T20	MSB-P05	MSB-P20
1. Ich habe meinem/r Patienten/in praktische Hinweise zur Bewältigung ihrer/seiner Schwierigkeiten gegeben.	PB	PB	PB	PB	PB
2. Ich sehe meine/n Patienten/in als kompetente/n Partner/in.	TB	PA	PA	TB	TB
3. Mein/e Patient/in war gefühlsmäßig beteiligt.	PA	g	g	TB	TB
4. Ich konnte meinem/r Patienten/in vermitteln, dass ich sie/ihn so akzeptiere, wie sie/er ist.	TB	TB	TB	TB	TB
5. Mein/e Patient/in hat neue Verhaltensweisen ausprobiert.	PB	PB	PB	PB	PB
6. Ich habe meinem/r Patienten/in Hilfe bei der Bewältigung ihrer/seiner Schwierigkeiten gegeben.	PB	PB	PB	PB	PB
7. Ich konnte meinem/r Patienten/in vermitteln, dass ich daran interessiert bin, wie es ihr/ihm geht.	TB	TB	TB	TB	TB
8. Mein/e Patient/in hat neue Verhaltensmöglichkeiten kennen gelernt.	PB	PB	PB	PB	PB
9. Die Therapie hilft meinem/r Patienten/in, ihre/seine Stärken und Fähigkeiten zu sehen.	PA	g	PA	g	g
10. Mein/e Patient/in fühlt sich von mir angenommen.	TB	TB	TB	TB	TB
11. Mein/e Patient/in sieht ihre/seine Schwierigkeiten aus einer neuen Perspektive.	PA	g	PA	PA	PA
12. Meinem/r Patienten/in gelingt es ihr/sein Problem differenziert zu sehen.	PA	PA	PA	PA	PA
13. Mein/e Patient/in fühlt sich von mir verstanden.	TB	TB	TB	TB	TB
14. Mein/e Patient/in hat Dinge verändert.	PB	PB	PA	PA	PA
15. Mein/e Patient/in hat Einsicht in ihre/seine Person und in ihre/seine Schwierigkeiten.	PA	PA	PA	PA	PA

Anmerkungen: <sup>1</sup>Formulierungen aus der Therapeutenversion, Patientenversion s. elektronisches Supplement 2; PB, Skala *Problembewältigung*; TB, Skala *Therapiebeziehung*; PA, Skala *Problemaktualisierung*; g, g-Faktor.

zur Therapie kamen (1, *deutlich verbessert* bis 5, *deutlich verschlechtert*). Schließlich machten die Therapeuten nach Therapieende eine Angabe bezüglich der Art des Therapieendes (qualitätsrelevanter Abbruch, nicht qualitätsrelevanter Abbruch oder reguläres Therapieende; Hiller et al., 2009).

## Statistische Analysen

Deskriptive Statistiken dienten der Stichprobenbeschreibung und Beschreibung der Item- und Skalenstatistiken.

Mittels Bartlett-Test und Kaiser-Meyer-Olkin (KMO)-Test wurde überprüft, ob sich die Daten für eine Faktorenanalyse eignen. Die Anzahl der zu extrahierenden Faktoren wurde per Parallelanalyse, Wayne Velicer's Minimum Average Partial (MAP)-Kriterium und visueller Inspektion der Scree-Plots bestimmt. Zur Überprüfung der dimensional Struktur wurden explorative Hauptachsenanalysen mit Oblimin-Rotation, hierarchische Modelle sowie Bifaktormodelle berechnet. Für die hierarchischen und Bifaktormodelle wurde jeweils ein Globalfaktor (g-Faktor) und drei spezifische Faktoren zugrunde gelegt. Aufgrund der überlegenen Passung wurde das Bifaktorenmodell beibehalten und aus Gründen der Übersichtlichkeit als einziges Modell im Folgenden berichtet. Der Modellfit wurde durch konfirmatorische Faktoren-

analysen mit WLSMV-Schätzung ermittelt. Hierfür wurden  $\chi^2$ -Wert, der comparative fit index (CFI), der Tucker-Lewis Index (TLI), die Approximationsdiskrepanzwurzel (root mean square error of approximation, RMSEA) sowie die standardisierte Residualdiskrepanzwurzel (standardized root mean square residual, SRMR) berechnet. Hierbei sollte  $\chi^2/df$  Werte zwischen 2 und 3 annehmen. Für RMSEA bedeuten Werte  $\leq .05$  einen guten Fit, Werte zwischen .08 und .10 einen mittleren Fit und Werte  $> .10$  einen schlechten Fit. SRMR  $< .05$  zeigt einen guten Fit an, Werte  $< .10$  sind akzeptabel. Für CFI und TLI gelten Werte zwischen  $> .95$  und 1.0 als gut und Werte zwischen .90 und .95 als akzeptabel.

Pearson-Korrelationen wurden zur Abschätzung der Retest-Reliabilität und der diskriminanten und konvergenten Validität durchgeführt. Zur Überprüfung der prädiktiven Validität wurden Spearman-Korrelationen mit der subjektiv wahrgenommenen Symptomverbesserung aus Patientensicht und der Zufriedenheit mit der Therapie, eine Partialkorrelation mit dem GSI-Post-Wert (unter Berücksichtigung des GSI Baseline-Wertes) sowie das Richtungsmaß Eta zur Feststellung des Zusammenhangs mit qualitätsrelevantem Therapieabbruch berechnet.

Die Berechnungen wurden mit SPSS 23 und dem Softwarepaket „lavaan“ in R 3.5.0 (<http://www.r-project.org/>) durchgeführt.

**Tabelle 2.** Globale Fit-Indizes der konfirmatorischen Faktorenanalysen

	$\chi^2$ (df)	$p$	$\chi^2/df$	CFI	TLI	RMSEA	SRMR
MSB-T05	4205.50 (90)	< .01	46.73	.944	.935	.100	.120
MSB-T20	2556.63 (90)	< .01	28.41	.956	.949	.083	.121
MSB-P05	4686.22 (90)	< .01	52.07	.929	.917	.105	.123
MSB-P20	2256.60 (90)	< .01	25.07	.963	.957	.079	.108

Anmerkungen: T05, 5. Sitzung Therapeutenversion; T20, 20. Sitzung Therapeutenversion; P05, 5. Sitzung Patientenversion; P20, 20. Sitzung Patientenversion; df, Freiheitsgrade; CFI, comparative fit-Index; TLI, Tucker Lewis-Index; RMSEA, root mean square error of approximation; SRMR, standardized root mean square residual.

## Ergebnisse

### Explorative Faktorenanalysen

Sowohl der Bartlett-Test (alle:  $\chi^2(153) \leq 45348.34$ ,  $p < .001$ ) als auch das KMO-Kriterium (alle:  $KMO \geq .93$ ) weisen darauf hin, dass sich die jeweils untersuchten Variablen für eine Faktoranalyse eignen.

Die Parallelanalyse empfahl für MSB-T zur 5. Sitzung die Extraktion von 4 Faktoren und zur 20. Sitzung die Extraktion von 5 Faktoren, für MSB-P jeweils 3 Faktoren. Die visuelle Inspektion der Scree-Plots und Velicer MAP legten jeweils die Extraktion von 3 Faktoren nahe. Es wurden daher jeweils 3 Faktoren extrahiert; dies entsprach auch der theoretisch postulierten Fragebogenstruktur.

Die gefundenen Faktoren entsprechen weitgehend den theoretisch postulierten Subskalen *Therapiebeziehung*, *Problembewältigung* und *Problemaktualisierung*. (Tab. S2–S5 im elektronischen Supplement 2). Für die meisten Items zeigte sich eine klare Struktur, d.h. sie luden auf einen Gruppenfaktor deutlich und auf die anderen deutlich geringer (Tab. 1; Abb. S1 A–D im elektronischen Supplement 3). Während für Item 02 in der Therapeutenversion zu beiden Messzeitpunkten die Hauptladung auf der Subskala *Problemaktualisierung* lag, lud das Item in der Patientenversion eindeutig auf *Therapiebeziehung*. Theoretisch wurde dieses Item der Subskala *Therapiebeziehung* zugeordnet. Item 03 (theoretische Zuordnung zur Skala *Problemaktualisierung*) lud in den Faktorenanalysen der Patientenversion eindeutig auf *Therapiebeziehung*, während für die Therapeutenversion keine Hauptladung  $> .3$  auf einem der spezifischen Faktoren identifiziert werden konnte, das Item jedoch auf dem g-Faktor lud. Für Item 09 ergab sich nur für die Therapeutenversion zur 20. Sitzung die theoretisch postulierte Zuordnung zur Subskala *Problemaktualisierung*, während das Item ansonsten nur auf dem g-Faktor lud. Schließlich ergab sich für Item 14 eine Hauptladung auf der theoretisch postulierten Subskala *Problembewältigung* nur in der Therapeutenversion zur 5. Sitzung; ansonsten lud dieses Item auf *Problemaktualisierung*. Alle Items laden deutlich auf den g-Faktor.

### Konfirmatorische Faktorenanalysen

Mittels konfirmatorischer Faktorenanalysen wurde jeweils das theoretisch postulierte 3-Faktorenmodell getestet (Tabelle 2). Der  $\chi^2$ -Test auf Modellfit lehnt alle Modelle ab, jedoch kann man aufgrund der großen Stichproben annehmen, dass dieser Test übermäßig sensitiv ist. Während CFI und TLI für einen akzeptablen bis guten Modellfit sprechen, erreicht der RMSEA nur für die Patientenversion zur 20. Sitzung einen akzeptablen Fit. Auch der SRMR spricht bei allen getesteten Modellen für eine unzureichende Modellpassung ( $> .10$ ). Insgesamt weisen jeweils die späteren Messzeitpunkte einen tendenziell besseren Fit auf.

### Itemanalysen und Reliabilitäten

Die Itemschwierigkeiten (Tab. S2–S5 im elektronischen Supplement 2), lagen im mittleren bis hohen Bereich (MSB-T05:  $p_i = .47$  bis  $p_i = .78$ ; MSB-T20:  $p_i = .61$  bis  $p_i = .79$ ; MSB-P05:  $p_i = .64$  bis  $p_i = .88$ ; MSB-P20:  $p_i = .64$  bis  $p_i = .88$ ), d.h. die Items differenzieren gut im Bereich höherer Ausprägungen. Die Items wiesen eine ausreichende bis hohe Trennschärfe bezüglich der Subskalen auf (MSB-T05:  $r_{it} \geq .40$ ; MSB-T20:  $r_{it} \geq .48$ ; MSB-P05:  $r_{it} \geq .35$ ; MSB-P20:  $r_{it} \geq .46$ ). Die interne Konsistenz der Gesamtskalen ist exzellent (jeweils 15 Items; MSB-T05:  $\alpha = .92$ , MSB-T20:  $\alpha = .93$ , MSB-P05:  $\alpha = .91$ , MSB-P20:  $\alpha = .94$ ); die der Subskalen im guten bis exzellenten Bereich (Tab. S6 im elektronischen Supplement 2). Die Profilreliabilität (Lienert, 1969) ist für die Patientenversionen als befriedigend einzuschätzen (MSB-P05:  $prof r_{it} = .69$ , MSB-P20:  $prof r_{it} = .62$ ), für die Therapeutenversionen als etwas schlechter (beide:  $prof r_{it} = .55$ ). Die Retest-Reliabilität zwischen beiden Erhebungen fiel erwartungsgemäß niedrig aus, wobei die höchste Konstanz jeweils in der Subskala *Therapiebeziehung* erreicht wurde (Therapeutenversion:  $r = .61$ ,  $p < .001$ ,  $N = 3.786$ ; Patientenversion:  $r = .62$ ,  $p < .001$ ,  $N = 3.733$ ) und die niedrigste in der Subskala *Problembewältigung* (Therapeutenversion:

$r = .45$ ,  $p < .001$ ,  $N = 3.786$ ; Patientenversion:  $r = .48$ ,  $p < .001$ ,  $N = 3.731$ ; *Problemaktualisierung*: Therapeutenversion:  $r = .56$ ,  $p < .001$ ,  $N = 3.786$ ; Patientenversion:  $r = .56$ ,  $p < .001$ ,  $N = 3.732$ ).

## Validität

Die Interkorrelationen zwischen Therapeuten- und Patientenversionen des MSB zeigten jeweils für dieselben Subskalen die höchsten Werte, was für die strukturelle Validität der Stundenbögen spricht (Tab. S6 im elektronischen Supplement 2).

Zur Überprüfung der konvergenten Validität wurden Korrelationen der Subskalen des MSB mit den Subskalen des WAI-SR berechnet (Tab. S7 im elektronischen Supplement 2). Es zeigten sich mittlere bis starke Zusammenhänge, wobei insbesondere die Subskala *Problembewältigung* mit den Subskalen *Aufgaben* und *Ziele* des WAI-SR und weniger stark mit der Subskala *Bindung* zusammenzuhängen scheint. *Problemaktualisierung* zeigte zu beiden Zeitpunkten die höchste Korrelation mit der Subskala *Aufgaben* der WAI-SR. *Therapiebeziehung* korrelierte zur 5. Sitzung am stärksten mit der Subskala *Bindung* des WAI-SR, während es zur 20. Sitzung zu hohen Korrelationen mit allen drei Subskalen des WAI-SR kommt. Die Therapeuten- und Patientenversionen zeigten ähnliche Muster. Die Ergebnisse liefern bestätigende Hinweise zur konvergenten Validität des MSB.

Um die diskriminante Validität zu beurteilen wurden Korrelationen mit dem Baseline-Wert des global severity index (GSI) des Brief Symptom Inventory (BSI) berechnet (Tab. S8 im elektronischen Supplement 2). Es ergaben sich Korrelationen zwischen  $-.01$  bis  $-.05$  mit den jeweiligen Subskalen der Patienten- und Therapeutenversionen. Aufgrund der großen Stichprobengröße wurden diese Korrelationen teilweise statistisch signifikant, die absolute Höhe der Korrelation spricht jedoch für die diskriminante Validität des MSB.

Vorhersage des Therapieerfolgs. Um die prädiktive Validität zu bestimmen wurden Partialkorrelationen mit dem Post-Wert des GSI (kontrolliert für den Baseline-Wert) und Korrelationen mit dem Patientenurteil, ob die Therapie eine Symptomreduktion herbeigeführt hat und mit der Zufriedenheit zu Therapieende berechnet (Tab. S8 im elektronischen Supplement 2). Zudem wurde der Zusammenhang der Subskalen mit qualitätsrelevantem Therapieabbruch überprüft. Die Korrelationen mit dem Postwert des GSI (unter Berücksichtigung des Baseline-Wertes) lagen zwischen  $-.07$  und  $-.27$  (alle signifikant), was für kleine bis mittlere Effekte spricht. Vermutlich aufgrund der größeren Nähe zum Therapieende liegen die Korrelationen der Subskalen mit dem Postwert des GSI

zur 20. Sitzung höher als die zur 5. Sitzung. Insbesondere die Subskala *Problemaktualisierung* scheint für den Therapieerfolg wichtig zu sein, da sich hier meist die höchsten Korrelationen zum GSI-Post-Wert zeigen. Das Patientenurteil, ob die Therapie zu einer Symptomreduktion geführt hat, korreliert signifikant mit den Subskalen der Patientenversion, wobei die Subskalen *Problemaktualisierung* und *Problembewältigung* höhere Korrelationen zeigen, im Vergleich zur Subskala *Therapiebeziehung*. Auch hier sind die Korrelationen zum zweiten Messzeitpunkt größer als die zum ersten Messzeitpunkt. Die Zufriedenheitseinschätzung der Patienten korreliert mit allen Subskalen in etwa gleich stark, auch hier sind höhere Korrelationen zum späteren Zeitpunkt zu beobachten. Schließlich spricht das Richtungsmaß *Eta* für den mittelgradig bis stark ausgeprägten Zusammenhang zwischen den Subskalen des MSB und dem Vorkommen qualitätsrelevanten Abbruchs. Hier werden ebenfalls die Zusammenhänge zum späteren Messzeitpunkt eher stärker.

## Diskussion

Die vorliegende Validierungsstudie untersucht ein neues Instrument zur Prozessdiagnostik längsschnittlich im naturalistischen ambulanten Psychotherapie-Setting anhand einer großen und diagnostisch heterogenen Stichprobe. Der Mainzer Stundenbeurteilungsbogen in der Patienten- und Therapeutenversion hat sich in der verhaltenstherapeutischen ambulanten Routineversorgung als gut einsetzbares Instrument bewährt und zeigt insgesamt gute bis ausreichende Kennwerte.

Die Ergebnisse der explorativen Faktorenanalysen sprechen für die theoretisch postulierte dreifaktorielle Struktur mit den Gruppenfaktoren *Therapiebeziehung*, *Problemaktualisierung* und *Problembewältigung* sowie einem g-Faktor, auf dem alle Items laden. Der g-Faktor kann als Ausmaß der Zufriedenheit des Patienten mit der Therapie (in der Therapeutenversion aus Therapeutesicht) interpretiert werden. Auf Ebene der Einzelitems scheint die Zuordnung von Items 02 und 03 je nach Kontext (Patient oder Therapeut) zu variieren. Item 02 (Therapeutenversion: „Ich sehe meine\_n Patienten\_in als kompetente\_n Partner\_in“, Patientenversion: „Mein\_e Therapeut\_in sieht mich als kompetente\_n Partner\_in.“) ist theoretisch der Subskala *Therapiebeziehung* zugeordnet. Während diese Zuordnung in den Faktorenanalysen der Patientenversion bestätigt wird, kommt es in der Therapeutenversion zu einer Zuordnung zur Subskala *Problemaktualisierung*. Möglicherweise führt die Betonung des Kompetenzerlebens des Items bei den Therapeuten zu einer abweichenden Zuordnung, da die restlichen Items

der Subskala *Therapiebeziehung* eher auf interpersonelle und empathische Komponenten abzielen, das Kompetenzerleben jedoch die Fähigkeit des Patienten, sich konstruktiv mit seinen Problemen auseinanderzusetzen meint. Aus Patientensicht scheint der im Item angesprochene partnerschaftliche Aspekt zu überwiegen. Item 03 (Therapeutenversion: „Mein\_e Patient\_in war gefühlsmäßig beteiligt.“, Patientenversion: „Ich war gefühlsmäßig beteiligt.“) fällt durch die durchweg niedrigste Kommunalität aller Items auf (zwischen .23 und .34). Während das Item theoretisch der Subskala *Problemaktualisierung* zugeordnet wird, lädt es in der Patientenversion jeweils auf dem latenten Faktor *Therapiebeziehung* und in der Therapeutenversion ausschließlich auf dem *g*-Faktor. Möglicherweise interpretieren Patienten eine gefühlsmäßige Beteiligung ihrerseits also als Zeichen einer vertrauensvollen therapeutischen Beziehung, während Therapeuten diese keiner der drei Subskalen zuordnen. Item 09 (Therapeutenversion: „Die Therapie hilft meinem\_r Patienten\_in, ihre/seine Stärken und Fähigkeiten zu sehen.“, Patientenversion: „Die Therapie hilft mir, meine Stärken und Fähigkeiten zu sehen.“) in wird lediglich in der Faktorenanalyse der Therapeutenversion zum 2. Messzeitpunkt der theoretisch postulierten Subskala *Problemaktualisierung* zugeordnet und lädt ansonsten nur auf dem *g*-Faktor. Möglicherweise wird hiermit eher ein Aspekt von Ressourcenaktivierung abgebildet. Schließlich fällt auf, dass Item 14 (Therapeutenversion: „Mein\_e Patient\_in hat Dinge verändert.“, Patientenversion: „Ich habe Dinge verändert.“) lediglich in der Therapeutenversion zum 1. Messzeitpunkt wie theoretisch postuliert auf *Problembewältigung* lädt, während es in den restlichen Analysen auf *Problemaktualisierung* lädt. Dies kommt möglicherweise durch die hohe Korrelation der beiden latenten Faktoren zustande.

Die Ergebnisse der konfirmatorischen Faktorenanalysen sind heterogen. Während die absoluten Maße des Modellfits (*RMSEA*, *SRMR*) eher für einen ungenügenden Modellfit sprechen, fällt der Vergleich zum Nullmodell mithilfe der relativen Maße der Modellanpassung (*CFI*, *TLI*) zufriedenstellend aus. Insgesamt verbessert sich der Modellfit jeweils zum späteren im Vergleich zum früheren Messzeitpunkt. Möglicherweise ließe sich der Modellfit weiter verbessern, wenn die Zuordnung einzelner Items gemäß den Ergebnissen der explorativen Faktorenanalysen modifiziert würde.

Die Reliabilitäten der Subskalen sind als exzellent zu bewerten. Hinsichtlich der Validität bestätigen mittlere bis hohe Korrelationen mit den Subskalen des WAI-SR die Konvergenz beider Instrumente. Erwartungsgemäß korreliert die Subskala *Bindung* des WAI-SR jeweils am höchsten mit der Subskala *Therapiebeziehung* des MSB. Die Subskala *Problemaktualisierung* des MSB korreliert jeweils am höchsten mit der Subskala *Aufgaben* des WAI-

SR. Die diskriminante Validität des MSB spiegelt sich in geringen Korrelationen mit der psychischen Belastung (*BSI-GSI*) der Patienten zu Therapiebeginn wider. Schließlich ergeben sich Hinweise auf die prädiktive Validität des MSB durch kleine bis mittlere Korrelationen mit der psychischen Belastung (*GSI-BSI*) zu Therapieende, der Zufriedenheit der Patienten mit der Behandlung und der Einschätzung inwiefern die Behandlung als hilfreich empfunden wurde. Die Zusammenhänge mit qualitätsrelevantem Therapieabbruch mit den Subskalen des MSB sind als mittelgradig (*Eta* zwischen 0.04 und 0.16) bzw. groß (*Eta* > .16) einzuschätzen. Die Zusammenhänge zum späteren Messzeitpunkt sind erwartungsgemäß höher im Vergleich zum ersten Messzeitpunkt. Diese Ergebnisse sprechen für eine gute prädiktive Validität des MSB.

Inhaltlich lehnt sich der MSB teilweise an den STEP (Krampen & Wald, 2001) an, der ebenfalls die Allgemeinen Wirkfaktoren nach Grawe erfasst. Beim vorliegenden MSB wurde jedoch versucht, sprachliche Ungenauigkeiten des STEP zu verbessern, indem die Items in Kombination mit der verwendeten Antwortskala eindeutig formuliert wurden (z. B. Verzicht auf Relativierungen in der Itemformulierung). Zudem wurde die Subskala *Therapiebeziehung* (im STEP: *therapeutische Beziehung und Unterstützung*) von vornherein als solche operationalisiert, während im STEP-Restitems aus den ursprünglich vorgesehenen Subskalen *Ressourcenaktivierung* und *Problemaktualisierung* zusammengefasst wurden. Die Items der Subskala *Therapiebeziehung* im MSB wurden konstruiert, um Empathie, Akzeptanz und Wertschätzung messen (Rogers, 1957). Entsprechend gibt es in den Stundenbögen eine eigene Subskala *Problemaktualisierung*, jedoch keine wie im STEP enthaltene Subskala zur *motivationalen Klärung*.

Es ergeben sich einige Limitationen aus der vorliegenden Untersuchung. Es handelt sich um eine Inanspruchnahme-Stichprobe aus der Routine-Versorgung, die insbesondere für andere Kontexte, z. B. stationäre Versorgung nicht repräsentativ ist. Problematisch, wenn auch nicht ungewöhnlich, erscheinen die hohen Zustimmungswerte auf der Subskala *Therapiebeziehung*, die zu Deckeneffekten führen können und möglicherweise durch soziale Erwünschtheit zustande kommen, obwohl insbesondere in der Patientenversion die Instruktion auf die Wichtigkeit einer realistischen Einschätzung hinweist. Ungünstig sind zudem die nicht immer konsistenten Tempi in der Itemformulierung. Weitere Studien sollten die Modellanpassung mit adaptierter Itemzuordnung untersuchen, da vor allem die Subskala *Problemaktualisierung* nicht ganz stabil zu sein scheint. Weiterhin wäre künftig eine Ergänzung um die fehlenden Graweschen Wirkfaktoren *Ressourcenaktivierung* und *motivationale Klärung* wünschenswert.

## Elektronische Supplemente

Die elektronischen Supplemente sind mit der Online-Version dieses Artikels verfügbar unter <https://doi.org/10.1026/1616-3443/a000583>

**ESM 1.** Abbildungslegende und Tabellen S1–S8 (Anhang.docx). Die Textdatei enthält die Abbildungsbeschreibung für Abbildung 1 des elektronischen Supplements sowie Tabellen, zur Stichprobenbeschreibung (S1), faktoren- und itemanalytische Befunde (S2–S5), interne Konsistenzen der Fragebögen und Subskalen-Interkorrelationen (S6), Werte zur konvergenten Validität (S7) sowie Werte zur diskriminanten und prädiktiven Validität (S8).

**ESM 2.** Fragebogen Patientenversion (MSB-P.docx). Die Textdatei enthält die Patientenversion des Mainzer Stundenbeurteilungsbogen (MSB-P).

**ESM 3.** Abbildung (Supp.Abb. 1.tif). Die Bilddatei enthält Abbildungen der vier Bifaktorenmodelle.

## Literatur

- Bassler, M., Potratz, B. & Krauthauser, H. (1995). Der „Helping Alliance Questionnaire“ (HAQ) von Luborsky. Möglichkeiten zur Evaluation des therapeutischen Prozesses von stationärer Psychotherapie. *Psychotherapeut*, 40 (1), 23–32.
- Bordin, E. S. (1979). The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy: Theory, Research & Practice*, 16, 252–260.
- Consruch, K. von, Flückiger, C., Stangier, U., Beutel, M. E., Herpertz, S., Hoyer, J. et al. (2013). WIFA-k: Ein neues Messinstrument zur zeitökonomischen Erfassung allgemeiner Wirkfaktoren nach Grawe. *Psychotherapie, Psychosomatik, Medizinische Psychologie*, 63 (7), 286–289.
- Derogatis, L. R. & Spencer, P. M. (1993). *Brief symptom inventory: BSI* (Vol. 18). Upper Saddle River, NJ: Pearson.
- Flückiger, C., Del Re, A. C., Wampold, B. E. & Horvath, A. O. (2018). The alliance in adult psychotherapy: A meta-analytic synthesis. *Psychotherapy (Chic)*, 55, 316–340. <https://doi.org/10.1037/pst0000172>
- Flückiger, C., Regli, D., Zwahlen, D., Hostettler, S. & Caspar, F. (2010). Der Berner Patienten- und Therapeutenstundenbogen 2000. *Zeitschrift für Klinische Psychologie und Psychotherapie*, 39, 71–79.
- Franke, G. (2000). *BSI: Brief Symptom Inventory von L. R. Derogatis (Kurzform der SCL-90-R) – deutsche Version*. Göttingen: Beltz Test GmbH.
- Gemeinsamer Bundesausschuss. (2015). *Qualitätsmanagement-Richtlinie/QM-RL*. Verfügbar unter: <https://www.g-ba.de/richtlinien/87/>
- Grawe, K., Donati, R. & Bernauer, F. (1994). *Psychotherapie im Wandel: Von der Konfession zur Profession*. Göttingen: Hogrefe.
- Hank, P. & Krampen, G. (2008). *Diagnostik der therapeutischen Beziehung* Tübingen: Dgvt-Verlag.
- Hatcher, R. L. & Gillaspay, J. A. (2006). Development and validation of a revised short version of the Working Alliance Inventory. *Psychotherapy Research*, 16 (1), 12–25.
- Heider, J. & Zaby, A. (2012). Psychotherapieevaluation: Empfehlungen für niedergelassene Psychotherapeutinnen und -therapeuten. *PID-Psychotherapie im Dialog*, 13 (1), 60–64.
- Hiller, W., Bleichhardt, G. & Schindler, A. (2009). Evaluation von Psychotherapien aus der Perspektive von Qualitätssicherung und Qualitätsmanagement. *Zeitschrift für Psychiatrie, Psychologie und Psychotherapie*, 57, 7–22.
- Höger, D. & Eckert, J. (1997). Der Bielefelder Klienten-Erfahrungsbogen (BIKEB). Ein Verfahren zur Erfassung von Aspekten des „Post-Session Outcome“ bei Psychotherapien. *Zeitschrift für klinische Psychologie*, 26 (2), 129–137.
- Horvath, A. O. & Symonds, B. D. (1991). Relation Between Working Alliance and Outcome in Psychotherapy: A Meta-Analysis. *Journal of Counseling Psychology*, 38 (2), 139–149.
- Krampen, G. & Wald, B. (2001). Kurzinstrumente für die Prozess-evaluation und adaptive Indikation in der Allgemeinen und Differenziellen Psychotherapie und Beratung. *Diagnostica*, 47 (1), 43–50.
- Lienert, G. A. & Raatz, U. (1998). *Testaufbau und Testanalyse*. Weinheim: Psychologie Verlags Union.
- Luborsky, L. (1976). *Helping alliances in psychotherapy*. New York: Brunner/Mazel.
- Martin, D. J., Garske, J. P. & Davis, M. K. (2000). Relation of the therapeutic alliance with outcome and other variables: a meta-analytic review. *J Consult Clin Psychol*, 68, 438–450.
- Rogers, C. R. (1957). The necessary and sufficient conditions of therapeutic personality change. *Journal of Consulting Psychology*, 21 (2), 95–103.
- Sharf, J., Primavera, L. H. & Diener, M. J. (2010). Dropout and therapeutic alliance: a meta-analysis of adult individual psychotherapy. *Psychotherapy (Chic)*, 47, 637–645. <https://doi.org/10.1037/a0021175>
- Wilmers, F., Munder, T., Leonhart, R., Herzog, T., Plassmann, R., Barth, J. & Linster, H. W. (2008). Die deutschsprachige Version des Working Alliance Inventory–short revised (WAI-SR) – Ein schulübergreifendes, ökonomisches und empirisch validiertes Instrument zur Erfassung der therapeutischen Allianz. *Klinische Diagnostik und Evaluation*, 1, 343–358.

### Historie

Onlineveröffentlichung: 07.05.2021

### Danksagung

Wir danken Herrn Prof. Hiller als ehemaligem Ambulanzleiter und Gaby Bleichhardt als ehemaliger Evaluationsleitung, unter deren Anleitung der MSB entwickelt wurde. Weiterhin danken wir Herrn PD Dr. Lars Hölzel für hilfreiche Informationen was die Historie des Instruments angeht.

### Förderung

Open Access-Veröffentlichung ermöglicht durch Johannes Gutenberg-Universität Mainz.

### Dr. Anne-Kathrin Bräscher

Johannes Gutenberg-Universität Mainz  
Abteilung für Klinische Psychologie, Psychotherapie und experimentelle Psychopathologie  
Wallstr. 3  
55122 Mainz  
[abraesch@uni-mainz.de](mailto:abraesch@uni-mainz.de)