Published in final edited form as:

Psychother Res. 2019 July ; 29(5): 565–580. doi:10.1080/10503307.2017.1422216.

The Multitheoretical List of Therapeutic Interventions – 30 Items (MULTI-30)

Nili Solomonov¹, Kevin S. McCarthy^{2,3}, Bernard S. Gorman¹, and Jacques P. Barber¹

¹The Derner School of Psychology, Adelphi University, Garden City, NY

²Chestnut Hill College, Philadelphia, PA

³Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

Abstract

Objective: To develop a brief version of the Multitheoretical List of Therapeutic Interventions (MULTI-60; McCarthy & Barber, 2009) in order to decrease completion time burden by approximately half, while maintaining content coverage. Study 1 aimed to select 30 items. Study 2 aimed to examine the reliability and internal consistency of the MULTI-30. Study 3 aimed to validate the MULTI-30 and ensure content coverage.

Method: In Study 1, the sample included 186 therapist and 255 patient MULTI ratings, and 164 ratings of sessions coded by trained observers. Internal consistency (Chronbach's alpha and McDonald's omega) was calculated and Confirmatory Factor Analysis was conducted. Psychotherapy experts rated content relevance. Study 2 included a sample of 644 patient and 522 therapist ratings, and 793 codings of psychotherapy sessions. In Study 3, the sample included 33 codings of sessions. A series of regression analyses was conducted to examine replication of previously published findings using the MULTI-30.

Results: The MULTI-30 was found valid, reliable and internally consistent across 2564 ratings examined across the three studies presented.

Conclusion: The MULTI-30 a brief and reliable process measure. Future studies are required for further validation.

Keywords

Integrative Treatment Models; Outcome Research; Process Research; Test development; Statistical methodology

Introduction

The 60-item Multitheoretical List of Interventions (MULTI, McCarthy & Barber, 2009) was originally developed to help clinicians, researchers, and supervisors achieve a better understanding of the therapeutic process occurring in different psychotherapies. It includes three forms: (a) Therapist – Self-report; (b) Patient Self-report; (c) Observer Rating Scale.

One of the important advantages of the MULTI is that the three forms include the same item wordings and thus allow for direct comparison across perspectives. Each item of the comprehensive MULTI is rated on a 5-point Likert scale, based on the quality and intensity of the use of interventions in a given session. These 60 interventions are clustered into eight main therapeutic orientations: psychodynamic, process-experiential, person-centered, cognitive, behavioral, dialectical-behavioral, and common factors. The latter describing interventions shared by all approaches (e.g. attentive listening, expression of warmth, empathy and support; Frank & Frank, 1991; Rosenzweig, 1936; Wampold, 2001). Previous studies have indicated that the MULTI-60 subscales are internally consistent (Cronbach's $\alpha > .70$; Shrout, 1995) and demonstrate 'good' to 'excellent' inter-rater reliability ($0.60 > \rho_i > 0.90$; Cicchetti, 1994) (see McCarthy & Barber, 2009, McCarthy, Keefe, & Barber, 2016; Solomonov, Kuprian, Zilcha-Mano, Gorman, & Barber, 2016; Solomonov, McCarthy, Keefe, Gorman, Blanchard, & Barber, 2017; Zickgraf, Chambless, McCarthy, Gallop, Sharpless, Milrod, & Barber, 2016). Since published, the psychometric properties of the MULTI-60 have been examined in multiple settings and found reliable across samples of patients, therapists, and observers (McCarthy & Barber, 2009; McCarthy et al., 2016; Solomonov et al., 2016; 2017). The use of the MULTI-60 has produced findings with important training and clinical practice implications (Boswell, Castonguay & Wasserman, 2010; Castonguay et al., 2017; McAleavey & Castonguay, 2014; McCarthy & Barber, 2009; McCarthy et al., 2016; Solomonov et al., 2017; Zickgraf et al., 2016).

Researchers have used the MULTI-60 to describe therapists' use of interventions and examine the relationships between specific interventions and in-session processes. For example, Zickgraf et al. (2016) found that therapists tended to show lower adherence to CBT manual and higher integration of other therapeutic approaches with panic disorder patients with Axis II traits and higher in-session resistance. McAleavey and Castonguay (2014) reported that sessions with high levels of exploratory (i.e., psychodynamic, process-experiential, and person-centered) techniques were less facilitative of insight in treatments for mood and anxiety disorders. Additionally, when therapists used high levels of directive (i.e., cognitive, behavioral, dialectical-behavioral) techniques without incorporating high levels of exploratory techniques, their patients reported higher alliances.

Research conducted using the MULTI-60 has also demonstrated the importance of differentiating between specific theory-driven and common factors techniques. Boswell and colleagues (2010) showed that 'common factors' techniques were used more frequently than theory-driven techniques. Additionally, patients whose therapists used more common factors techniques than average found sessions less helpful when their therapists reported using more cognitive behavioral techniques than usual. More recently, Solomonov et al. (2017) reported a bi-directional relation between use of common factors techniques and the development of the working alliance in supportive-expressive psychodynamic therapy (SET) for depression. Reportedly, high use of common factors techniques was predictive of later improvement in the alliance. Additionally, therapists tended to use more common factors techniques with patients with whom they established higher alliance early on. Another study showed that even psychotherapy experts conducting prototypical demonstrations of their therapeutic approach integrated moderate to high levels of techniques from other orientations (Solomonov et al., 2016). Taken together, these findings demonstrate that the

MULTI can be used to describe therapists' use of unique and shared techniques across treatment modalities.

Researchers have also utilized the MULTI-60 to assess the relation between use of prescribed techniques and treatment outcome. McAleavey, Castonguay, and Xiao (2014) reported that use of cognitive techniques predicted better outcome in a community sample with mixed diagnoses only when both the therapist and the supervisor's declared orientation was cognitive therapy. McCarthy and colleagues (2016) found that use of moderate levels of psychodynamic and process-experiential techniques was most predictive of subsequent symptomatic relief in SET for depression. Lastly, Castonguay et al. (2017) showed that clinicians in a naturalistic setting accurately predicted and recalled their use of techniques with specific clients. Overall, previous studies have shown that the MULTI-60 is a reliable, psychometrically sound and clinically applicable measure that can be implemented to assess therapists' use of techniques in naturalistic and experimental settings across a range of clinical populations.

The purpose of this study was to develop a short version of the MULTI-60. The need for an abbreviated version emerged from the increasing demands to decrease the time burden of routine data collection (Boswell, Kraus, Miller, & Lambert, 2015), which was found to be one of the main barriers to clinicians' participation in routine outcome assessment (Borkovec et al., 2001; Castonguay et al., 2010). The MULTI-60 is already fairly brief and completion time is up to 10 minutes. However, when used as a session-by-session repeated measure, it becomes time-consuming, especially for clinicians, who typically have only a few minutes between patients. Thus, we aimed to develop a valid and reliable short version of the MULTI, decreasing the time burden approximately by half (5 minutes completion time), while maintaining the comprehensive and clinically useful 8-subscale structure. Reducing the number of items by half (resulting in a 30-item measure) was determined to be an arbitrary but desirable outcome as it allowed us to have a point at which we considered the short version complete. We assumed it would lead to a decrease of the time burden by at least half, given the reduction in completion time per item and elimination of items with less clear wording and lower variance, and reduce by half the amount of time required for training and supervision of reliable coders. Previous studies focused on development of short forms have decreased the number of items by half (Barkham, Hardy, & Startup, 1996; Thompson, 2007) or two-thirds (e.g. Hatcher & Gillapsy, 2006; Tracey & Kokotovic, 1989) based on the assumption that this decrease would lead to a significant decrease in the time-burden.

Following previous research on item reduction strategies (Widaman, Little, Preacher, & Sawalani, 2011), we decreased the number of items based on combination of quantitative methodologies and a content-focused approach. The development and evaluation of the psychometric properties of the MULTI-30 were conducted in three steps: Study 1 was aimed to select 30 items that show the best psychometric features across perspectives and clinical populations. Study 2 focused on examining the psychometric properties of the MULTI-30 in samples collected at a single data point as well as samples with longitudinal data. Study 3 focused on validating the content coverage of the MULTI-30 by replicating previous findings reported by McCarthy and colleagues (2016; see above).

Our use of statistical methodologies and theoretical rationale for the inclusion and exclusion of items was similar to those used in the original development of the MULTI-60 (McCarthy & Barber, 2009). We collected data from therapists, patients, and observers in a wide range of settings, including training clinics, counseling centers, outpatient clinics, and online recruitment, as well as randomized clinical trials (Barber et al., 2012; Milrod et al., 2015). Following others' recommendations (Coste, Guillemin, Pouchot, & Fermanian, 1997; Kruijen, Emons, & Sijtsma, 2013), we also collected content ratings from psychotherapy experts from the eight approaches covered in the measure in order to ensure content coverage..

Study 1: Development of the MULTI-30 – Item Reduction Process

Study 1 was aimed at selecting the best items for the MULTI-60. The goal was to decrease the number of overlapping items (belonging to more than one subscale) and develop a 30-item measure with eight subscales consisting of 3–5 items per subscale.

Method

Participants

Patients.: Self-reported MULTI ratings were collected from patients ($n=322$) at four sites: (a) a university counseling center ($n = 293$); (b) a clinic specializing in behavioral treatment for Obsessive Compulsive Disorder ([OCD]; $n = 12$); (c) two community mental health centers ($n = 17$). Only complete responses were analyzed (67 partial responses were excluded). Overall, 255 MULTI ratings were analyzed. Demographic data were not available for the university counseling center. Of clients providing data, 75% were women with a mean age of 25 years ($SD = 6.9$), 7% Asian American, 14% as African American, 68% as Caucasian American, 3% as Latino/a, 4% East Indian, and 4% 'other'; 4% did not respond.

Therapists.: A total of 184 MULTI-60 complete therapist self-reports (with no missing data) were collected in two out the four patient recruitment centers described above: (a) therapists providing supportive therapy at a counseling center ($n = 160$); (b) therapists providing behavioral therapy at OCD clinic ($n = 24$). Responses from patients and therapists were collected anonymously and thus could not be nested. Demographic data were received for 144 therapists, with 54% women and most (83%) were Caucasian American. The highest degree for 40% was a Master's in social work, psychology, or education; 8% a medical degree; and 52% a doctoral degree in psychology. The average experience practicing was 13 years ($SD = 11.4$).

Observers.: Thirty-four trained advanced undergraduate and graduate students with variable levels of training coded 164 psychotherapy sessions included in the study. Coding training included at least a semester long course in Systems in Psychotherapy, with specific focus on MULTI coding, taught by a trained and reliable MULTI coder or a systematic structured MULTI training (Solomonov & McCarthy, 2016). Training for all coders included ongoing reliability discussions among raters and feedback to improve reliability. Reliability did not change when coders with various levels of training were eliminated from analyses. (a) Twenty-four coders rated 32 expert prototypical demonstrations retrieved from APA

PsycTHERAPY online database, as well as demonstrations published by the American Psychological Association [APA] (Shostrom, 1965; 1986; APA 2013); (b) Twelve coders rated 132 videotaped psychotherapy sessions of 44 patients who received SET as part of a randomized controlled trial (RCT) for depression (Barber et al., 2012). Overall, each session was rated by 3.63 coders on average (two out of the 34 coders rated sessions in both samples). All subscales showed “acceptable” to “good” interclass correlation coefficient in a two-way mixed model ($0.64 < \rho_i < 0.77$; Cicchetti, 1994; Shrout & Fleiss, 1979), with the exception of the DBT subscale, which was slightly lower ($\rho_i=.47$) due to very low variance in use of DBT techniques. Coders’ reliable ratings were averaged and the mean rating was used in further analyses.

Psychotherapy experts. In order to examine the content validity for the selected items, experts from each of the eight therapeutic approaches were contacted. Each expert was a well-known and respected researcher-practitioner with at least 20 years of research, clinical experience, strong publication history, teaching and supervision in his/her orientation. Experts were contacted via email, provided with the relevant MULTI-60 subscale and asked to select “5 interventions that most typify X therapy.” The request to choose five items was aimed to identify selected items with the highest content validity. Responses were received from four psychodynamic experts, two process-experiential experts, one interpersonal expert (others who met the selection criteria and contacted did not respond), three person-centered experts, two common factors experts, three behavioral experts, four cognitive experts, and three dialectical-behavioral experts.

Item Reduction Procedure—We implemented an item reduction strategy based on a comprehensive literature review of the methodologies employed in studies reporting development of abbreviated versions of psychotherapy process and outcome measures. We identified the item reduction approaches most commonly used in studies describing short form development of psychotherapy measures. These included: (a) confirmatory factor analysis (CFA; Barkham, Sturdy & Stratup, 1996; Foa et al., 2002; Hatcher & Gillaspay, 2006; McCarthy, Gibbons, & Barber, 2008; Thompson, 2007; Tracey & Kokotovic, 1989) (b) reliability analyses (Barkham, et al., 1996; Thompson, 2007); (c) internal consistency analyses (e.g. Foa et al., 2002; Hatcher & Gillaspay, 2006; McCarthy, Gibbons, & Barber, 2008; Mick, Faraone, Spencer, Zhang, & Biederman, 2008; Soltz, Budman, Demby, & Merry, 1995; Thompson, 2007). In order to ensure a conservative and rigorous item selection process we employed all of the methods above in our analyses (Widaman et al., 2011), across the three samples collected (i.e. therapists, patients and observers).

Confirmatory Factor Analysis (CFA; ‘maximum likelihood’ estimator) was conducted using the *lavaan* package (Rosseel, 2012) in *R* (R Core Team, 2016). We tested original 8-factor structure, assigning one factor per MULTI subscale (see McCarthy & Barber, 2009) to ensure the selected items adequately represent the therapeutic orientations assessed in the MULTI-60. We tested the overall internal consistency of the MULTI-30 subscales and the specific contribution of each of the MULTI items to subscales’ internal consistency by calculating the overall alpha when each given item is eliminated. Cronbach’s α (Cronbach, 1951) was calculated using the *Rcmdr* package (Fox & Bouchet-Valat, 2017) and

McDonald's ω (McDonald, 1970; 1999), was calculated using the *omega* function in the *psych* package (Revelle, 2017) in *R*. While Cronbach's α is most widely used, recent work suggested that McDonald's omega (ω) is more suitable when inter-item correlations are expected (see Widaman et al., 2011), as it decreases the risk of inflation and attenuation, and is especially suitable for short-form item reduction procedures (Dunn, Bguley, & Brunsden, 2014). Omega estimates were calculated using a hierarchical factor analysis, with oblimin rotation (Revelle, 2017).

Once analyses were completed, the research team reviewed the results, compared each of the items' performance across analyses and selected the items that performed optimally across the three samples. We excluded items that showed: (a) high loadings on multiple factors (i.e., not unique); (b) low loading on the factor hypothesized (i.e., not representative of the approach) (c) small variance across samples (i.e., rarely rated by therapists, patients, and observers). If multiple items showed similar loadings on the same factor, we eliminated the items that were less reliable, internally consistent, and/or were not chosen by most content experts. Priority was given to non-overlapping items (i.e. belong to only one subscale), but these items were not eliminated unless they had weak psychometric features. This item selection process is common in development of psychometric measures and was used by others (e.g., Foa et al., 2002; Tracey & Kokotovic, 1989). For the purpose of brevity we present the psychometric characteristics of the items selected for the MULTI-30 in this manuscript.

Results

Overall, the CFA model showed good fit in the patient sample (RMSEA=.07; CI[.07,.07]; CFI=.78), acceptable fit in the therapist sample (RMSEA=.09; CI[.09,.10], CFI=.63), and lower fit in the observer sample (RMSEA=.11; CI[.11,.12]; CFI=.57) (Table 1). Due to the high number of parameters that were enforced and the relatively small samples available, the covariance matrix of latent variable was not positive definite. All chosen items showed significant factor loadings with three exceptions (see Table 1). Experts' items selections are presented in Table 1. All subscales showed acceptable to excellent internal consistency (α 's>.63; ω 's>.77; Table 2). The crosswalk list between MULTI-60 and MULTI-30 is included in Appendix A. The MULTI-30 is included in Appendix B.

Discussion

Overall, almost all 30 items showed acceptable to excellent psychometric properties across perspectives, thus ensuring that the MULTI-30 could perform reliably across settings. Five of the eight MULTI-30 subscales are comprised of unique items that do not belong to more than one orientation: psychodynamic (items 2, 6, 12, 14, 15); common factors (items 5, 11, 16, 1); person centered (items 4, 21, 22); process-experiential (items 5, 7, 18, 23); interpersonal (items 25, 26, 27, 30). The remaining three subscales (cognitive, behavioral and DBT) include both items that are unique to those orientations and shared items, which belong to more than one orientation (i.e. subscale). The behavioral and cognitive subscales are comprised of three *unique* items each (behavioral items: 8, 9, 19; cognitive items: 13, 20, 24), as well as two *shared* CBT items (item 1 [setting an agenda]; item 10 [assigning homework]). The DBT subscale includes two *unique* items (28, 29), the behavioral items (8,

9, 19) and the *shared* CBT items (1, 10). This modular structure allows for flexibility and personalized adjustments of the MULTI-30.

Overall, the fit indices reported were not ideal but comparable to those previously found for the MULTI-60 (McCarthy & Barber, 2009). While we found that the covariance matrix was non-positive definite, we decided to present these results as exploratory in nature and used for the sole purpose of item selection (see Wothke, 1993). The relatively low fit in the therapist and observer samples was likely largely due to inter-correlations among the MULTI-60 subscales, and the large number of parameters enforced to maintain the 8-factor structure, as shown previously (McCarthy & Barber, 2009). Maintaining the original factor structure was a top priority as it ensured content coverage and clinical utility of the measure. Notably, fit could be improved substantially by including the correlated errors among items, a commonly used practice in short form development. However, we maintained a conservative approach and avoided this method, since others have found that it could result in capitalization on chance, artificially improve model fit, and produce misleading results (see Hermida, 2015 for review of this issue).

Study 2: The Psychometric Features of the MULTI-30

The aim of Study 2 was to examine the psychometric features of the MULTI-30, developed in the previous study, across a range of diverse clinical populations. We included data collected from therapists, patients, and observers at a single time point, as well as longitudinal data (multiple time points per patient) in order to ensure that the MULTI-30 is suitable for use in studies including both types of data. We predicted that the psychometric features of the MULTI-30 would be comparable to those of the MULTI-60.

Methods

Participants

Patients.: A total of 644 patient ratings were collected: (a) Seventy-seven patients (a total of 191 psychotherapy sessions) who received panic-focused psychodynamic therapy and CBT for panic disorder completed the MULTI-60 after sessions 2, 5, and 10, as part of a larger RCT (Milrod et al., 2015). Analyses were conducted on the 30 items selected. Mean patient age was 37.49 years ($SD=13.27$), with 62% females ($n=48$), 74% Caucasian ($n=57$), 17% African American ($n=13$), 3% Latino/a ($n=6$), and 1% Asian ($n=1$). (b) A diverse sample of 453 patients completed the MULTI-30 after one session. Patients were recruited through online websites, listservs, social media, and local community clinics in Israel and the US (See Appendix C for demographic data).

Therapists.: A total of 552 therapist ratings were collected from 376 therapists: (a) Thirteen therapists who provided panic-focused psychotherapy and CBT for panic disorder completed the MULTI-60 after sessions 2, 5 and 10 as part of a larger RCT ($n=188$; see Milrod et al., 2015 for details). The MULTI-30 items were analyzed. (b) A diverse sample of 363 therapists (Master's level and above with at least one year of clinical experience) from Israel and the US completed the MULTI-30 following one session with one patient. Therapists were recruited through listservs, social media, professional websites, local community

clinics and online websites. Therapists' demographics and declared therapeutic orientations are described in Appendix C.

Observers.: A total of 793 videotaped psychotherapy sessions (2 and 10) of 147 patients who participated in a RCT for panic disorder (Milrod et al., 2015) were coded by 25 trained observers ($M=2.58$ raters per session) using MULTI-60 (the MULTI-30 items were analyzed). Coders were advanced doctoral students for Clinical Psychology who received at least 20 hours of coding training, as well as ongoing supervision, which included discussions on agreements and disagreements and personalized feedback on coding. Once coders were found reliable ($ICC>.60$), they began coding sessions included in this sample. Intra-class correlation in a two-way mixed model for reliability for the average score of three raters ($ICC [2,3]$; Shrout & Fleiss, 1979) was calculated using *lme4* package (Bates, Maechler, Bolker, & Walker, 2015) in *R* (R Core Team, 2016). All MULTI-30 subscales showed 'excellent' inter-rater reliability ($\rho_i>.84$; Cicchetti's, 1994), with the exception of the DBT subscale, which was in the 'good' range ($\rho_i=.74$), and the common factors subscale which was in the 'fair' range ($\rho_i=.50$).

Statistical Analyses—A series of statistical analyses was conducted in order to assess the psychometric properties of the MULTI-30. For data collected with the MULTI-60, only complete responses were included in the analyses. For new data collected using the MULTI-30, partial responses with missing items (≤ 5) were imputed using the *MICE* package (Van Buuren & Groothuis-Oudshoorn, 2011) in *R* (R Core Team, 2016). We tested the overall internal consistency of the MULTI-30 subscales, as well as the specific contribution of each of the MULTI items to subscales' internal consistency using Cronbach's α , calculated in the *Rcmdr* package (Fox & Bouchet-Valat, 2017), and McDonald's ω (McDonald, 1970; 1999), calculated in the *psych* package (Revelle, 2017) in *R*.

Confirmatory factor analyses were conducted using two methodologies: (a) The new MULTI-30 data (one session per patient) was analyzed using the traditional CFA in *R lavaan* package (Rosseel, 2012); (b) Longitudinal MULTI-60 data (multiple sessions per patient), was analyzed using a Multilevel Confirmatory Factor Analysis (MCFA; Muthén, 1994) in *Mplus* software, version 7.4 (Muthén, & Muthén, 1998–2015). The advantage of MCFA is that it enables disentangling the between-person variation (i.e. variance due to differences between persons' overall means) from the within-person variation (i.e. persons' individual changes over time). This method enables isolating the within-patient time-lagged fluctuations from confounding external variables. Our model specification included eight latent variables (representing the MULTI subscales), with 3–5 observed variables (MULTI items) per latent variable. A multilevel model with random intercepts and random slopes was estimated (TYPE=TWOLEVEL RANDOM; CLUSTER = patient [ID]), with Maximum-likelihood estimation method. The two-level random model (including random effects) estimates (rather than just controls for) the between-level confounding variables, thus decreasing the risk of increased Type-I error rate, by explicitly modeling the nested data structure (Falkenström, Finkel, Sandell, Rubel, & Holmqvist, 2017). Unfortunately, fit indices are not yet available for this type of MCFA (Muthén, & Muthén, 2017). We report

the within-level MCFA standardized factor loadings using the *StdYX* standardization method for continuous covariates and latent variables (Muthén, & Muthén, 2007). The analyses were repeated using the standard CFA in *lavaan* package (without accounting for between-within variances) to ensure results are replicated using a more traditional and well-established method.

Results

Descriptive statistics are presented in Appendix D (Table D1). Overall, the CFA results of the new MULTI-30 data were comparable to those found with the MULTI-60 data (see Study 1), and in the original MULTI-60 study (McCarthy & Barber, 2009). The CFA model showed good fit in the patient sample (RMSEA=.08; CI[.09,.08]; CFI=.83) and in the therapist sample (RMSEA=.08; CI[.09,.07]; CFI=.81), and acceptable fit in the observer sample (RMSEA=.09; CI[.09;.10]; CFI=.82). The vast majority of the MULTI-30 items showed high CFA and MCFA factor loadings across all samples (Table 3), with only four items (16, 31, 47, and 58) showing high variance in magnitude of factor loading across samples. Results from MCFA were comparable to those found using the standard CFA, without accounting for within-between variance, with almost all items showing higher loadings in CFA models. Despite the smaller number of items, almost all MULTI-30 subscales showed acceptable to excellent internal consistency across all five samples (α 's>.61; ω 's>.60), with only two exceptions (out of 40) (see Appendix D; Table D2).

Discussion

Our analyses show that within-person fluctuations in use of techniques over time could be assessed reliably using the MULTI-30. Our results demonstrate the stability and clinical applicability of the MULTI-30, as the factor structure as well as the measure's psychometric properties were strong across samples and comparable when assessed using a traditional CFA and an innovative MCFA approach. Maintaining the 8-factor structure in the MULTI-30 enables investigation of techniques from a range of approaches using a brief and time efficient measure.

Study 3: Validation of the MULTI-30

Study 3 was aimed at investigating the validity and content coverage and clinical utility of the MULTI-30. In a previous study, McCarthy et al. (2016), found a curvilinear relationship between use of process-experiential and psychodynamic techniques (as coded using the MULTI-60) and subsequent decrease in depressive symptoms. We predicted that these results will be replicated using the MULTI-30, thus demonstrating its validity and adequate content coverage.

Participants

Patients.—This sample included 33 patients who received 16 weeks of SET, as part of a larger RCT for depression (See Barber et al. [2012]). Sixty-one percent of patients were female ($n = 20$). The mean age was 35.5 years ($SD = 12$), with 49% African-American ($n = 16$), 42% identified as Caucasian ($n = 14$), 6% as Asian ($n = 2$), and 3% as Latino/a (see McCarthy et al. [2016] for details on this sample).

Therapists.—Four therapists (3 females) with PhD in Clinical Psychology and over 15 years of experience conducting psychodynamic therapy provided SET. Therapists received ongoing supervision and adherence was established (see Barber et al., 2012)

Observers.—Five advanced doctoral students for Clinical Psychology received at least 20 hours of MULTI coding training, as well as ongoing supervision and periodic group discussions focused on identifying agreements and disagreements in coding in order to maintain high interrater reliability. ICC [2,3] (Shrout & Fleiss, 1979) was calculated and all subscales tested (process-experiential, psychodynamic, person-centred, and common factors) showed “good” to “excellent” inter-rater reliability ($0.66 > \rho_i > 0.90$; Cicchetti, 1994).

Measures

Multitheoretical list of Therapeutic Interventions (MULTI; McCarthy & Barber, 2009)—See Study 1 for details.

Hamilton Rating Scale for Depression (HRSD; Hamilton, 1960)—This diagnostician-based rating scale was used to assess patients’ depressive symptoms at intake, Week 4 and termination. Diagnosticians demonstrated excellent interrater reliability (ICC [2, 1], $\rho I = .92$).

Replication Procedure

Following the analyses performed by McCarthy and colleagues (2016), we conducted a series of regression analyses. First, we assessed the *linear* relation, testing whether use of techniques at Week 4 as measured by the MULTI-60 and the MULTI-30 predicted depressive symptoms (as measured by the Hamilton Rating Scale for Depression (HRSD) at termination. In order to account for the symptom change occurring prior to Week 4, we covaried prior symptom change (residuals of intake HRSD regressed on Week 4 HRSD) in the regression models. Second, we assessed the curvilinear relation between use of MULTI-60 and MULTI-30 subscales and subsequent symptom change, by repeating the regression models while including the raw scores as well as the *quadratic* terms for each of the MULTI subscales. We predicted that results using the MULTI-60 and MULTI-30 would be comparable.

Results

The results reported by McCarthy and colleagues (2016) were fully replicated (see Appendix D; Table D3). Linear relations between use of techniques (as measured by the MULTI subscales) and outcome were not significant. However, significant curvilinear relations were found between the psychodynamic and process-experiential subscales and outcome, indicating that moderate use of these techniques predicted subsequent symptomatic relief. Additionally, high significant ($p < .01$) correlations were found between the MULTI-60 and MULTI-30 subscales ($r = .80$ for all subscales, with the exception of DBT ($r = .66$).

Discussion

The replication of the previously reported MULTI-60 findings shows that the MULTI-30 demonstrates adequate content coverage, despite the decrease in the number of items. We employed the statistical method used in the original study (McCarthy et al., 2016), in order to conduct a direct replication of the previously reported findings. Future studies could expand on the replication, by assessing use of techniques across sessions (i.e. multiple time points), while accounting for within- and between- patient variability or examining curvilinear techniques-outcomes relationships in other clinical populations.

General Discussion—The current report focused on developing a short version MULTI, while maintaining maximum content coverage and preserving the reliability and internal consistency of the original measure. Indeed, we demonstrated that the psychometric properties of the MULTI-30 are and the MULTI-60 (McCarthy & Barber, 2009) are comparable, and replicated previous findings (McCarthy et al., 2016) to ensure content coverage. By reducing most of the overlapping items, we reduced the inter-item and inter-subscale correlations and increased subscales' independence, while maintaining the clinical utility of the 8-subscale structure. Thus, the MULTI-30 is likely to provide researchers and clinicians with similar depth and breadth of information, while decreasing the time and resources required in administration and analysis. Its modular structure allows flexibility and adjustments based on researchers and clinicians' specific needs. Additionally, previous studies suggested that the 8-factor structure could be reduced to a 3–5 factor structure (e.g. directive vs. exploratory interventions) and provide a higher-level analysis of use of interventions (see McAleavy & Castonguay, 2014; Castonguay et al., 2017). Future studies will determine whether previously found higher-level factors could be replicated using the MULTI-30.

Psychometric experts assert that the process of developing a short version for an existing measure is highly complex and requires extensive resources (Coste et al., 1997; Kruey et al., 2013; Smith et al., 2000; Widaman et al., 2011). In order to ensure that the items selected provide maximum content coverage with strong psychometric features we employed a relatively conservative approach by combining statistical and content-focused approaches. Specifically, previous short-form development studies in psychotherapy research based their item selection process on Cronbach's α value and inter-item correlations (Mick et al., 2008), and factor analytic approaches (e.g., IIP-32; Barkham, Sturdy & Stratup, 1996; OCI-R; Foa et al., 2002; PANAS; Thompson, 2007; WAI-R; Tracey & Kokotovic, 1989; Hatcher & Gillasp, 2006; Q-LES-QSF; Mick et al., 2008; IIP-62; Soltz et al., 1995). In this study, we incorporated these approaches along with a content-focused approach in a range of clinical samples and across different perspectives (i.e., therapists, patients, observers), thus, providing strong rationale for our item selections and validating our measure across a range of heterogeneous samples.

Another important strength of our study is the inclusion of longitudinal data (multiple sessions per patient) collected in various naturalistic and experimental settings. Analyses using multilevel modeling provided us with the opportunity to investigate the reliability of the measure over time, which is more representative of psychotherapy processes in the real

world (e.g., Denny et al., 2012). Our use of longitudinal data collected from therapists, patients and observers in different settings is rather innovative, as the field of longitudinal psychometrics is relatively new. Finally, we also replicated existing MULTI-outcome findings, using our short version, thus ensuring that the measure we developed is representative of the original form.

Following Tracey and Kokotovic (1989), we conducted separate analyses for the patient and therapist data. Additionally, similar to Soltz, Budman, Demby, & Merry (1995), we included patient populations from diverse community settings. We also included analyses of data collected from independent trained coders, which is rarely included in short-form development, likely because the coding process is highly labor-intensive and time consuming. To the best of our knowledge, other short-form development studies in psychotherapy research did not include such data, with the exception of the WAI-observer scale. However, in contrast to the MULTI-30, the WAI-observer has a completely different psychometric structure than the patient and therapist versions (Darchuk, Wang, Weibel, Fende, Anderson, & Horvath, 2000). This demonstrates an important feature of the MULTI – a process measure that captures multiple perspectives.

In terms of limitations, in some cases, we examined the psychometric properties of the MULTI-30, using longitudinal data collected using the MULTI-60. While this is not ideal, we included those samples in order to ensure that data collected in naturalistic settings was comparable to data collected in a highly controlled experimental environment. Indeed, the psychometric properties of the MULTI-30 were comparable to those of the MULTI-60 suggesting generalizability of our results and decrease in the likelihood that our findings are sample-specific. Another important limitation is that in some cases we could not account for nesting in our data due to confidentiality agreements. Nevertheless, the large majority of our samples were independent, which is a strength of our study. Moreover, while our samples were larger or similar to those of other studies developing short versions of psychotherapy measures, they could be considered small given the high number of parameters enforced in an 8-factor confirmatory factor analysis (CFA), especially in multilevel CFA (MCFA). Given that fit indices are not yet available for MCFA, we were unable to assess the fit of our models in these analyses. With regard to our CFA analyses, in Study 1, when a large number of parameters were enforced in order to test the MULTI-60 factor structure, our covariance matrix was non positive-definite and findings should be considered exploratory. It is highly likely that the goodness of fit and model convergence could be significantly improved in larger samples and future studies will be needed to support this assumption. Specifically, future studies could focused on assessing the within- and between- variability in ratings of patients, therapists, and observers, and their effect on psychotherapy outcomes using the MULTI-30. Additionally, given that our a-prior goal was to conduct a systematic item-reduction procedure through elimination of several items per subscale, we decided to conduct a confirmatory, rather than exploratory factor analysis. Future studies could include further explorations of the MULTI-30 factor structure. Finally, given that the MULTI is a relatively new measure, developed less than a decade ago, future studies would be needed to assess the measure's psychometric properties and clinical applications.

Conclusion

The MULTI-30 is a brief measure for use of therapeutic interventions in psychotherapy sessions. It is valid and reliable and could be used to assess patients, therapists, and observers' perceptions of use of interventions from eight major therapeutic approaches. The MULTI-30 is not only likely to be a useful research tool, but it is also highly clinically relevant and could be used for training and supervision purposes. Specifically, supervisors could discuss the efficacy of use of techniques with their supervisees using MULTI coding, and professors could demonstrate typical/ideal examples of use of techniques, as coded using the MULTI, in prototypical demonstrations or sessions conducted by psychotherapy experts.

Acknowledgments:

We would like to thank Andrew McAleavey whose thoughtful comments helped us significantly improve this manuscript.

APPENDIX A: Crosswalk List - MULTI-60 to MULTI-30

MULTI subscale	Item # in MULTI-60	Item # in MULTI-30
CBT*	1	1
PD	2	2
CF	7	3
PC	10	4
PE	11	5
PD	12	6
PE	13	7
BT	15	8
BT	16	9
CBT*	17	10
CF	18	11
PD	20	12
CT	21	13
PD	22	14
PD	24	15
CF	28	16
CF	31	17
PE	34	18
BT	35	19
CT	37	20
PC	40	21
PC	46	22
PE	47	23
CT	49	24
IPT	50	25

MULTI subscale	Item # in MULTI-60	Item # in MULTI-30
IPT	51	26
IPT	54	27
DBT **	56	28
DBT **	58	29
IPT	60	30

Note. MULTI = Multitheoretical list of therapeutic interventions.

* item should be added to the CT, BT and DBT subscales;

** A full DBT subscale should consist these items in addition to the CBT and BT items; CBT = Cognitive-Behavioral Therapy; CT = Cognitive; BT = Behavioral; PC = Person Centered; PD= Psychodynamic; PE = Process-experiential; IPT= Interpersonal; DBT = Dialectical Behavioral; CF = Common Factors. MULTI subscale = indicates the subscale each item belong to. The DBT subscale is created by combining the CBT, DBT and BT items. The cognitive and behavioral subscales are created by combining the CT or BT items (respectively), with the two CBT items.

APPENDIX B: MULTITHEORETICAL LIST OF THERAPEUTIC INTERVENTIONS (MULTI) – 30 ITEMS

<i>Instructions: The following items represent actions that may or may not have occurred in the session which you just observed. Please rate each item using the scale provided. There are no right or wrong answers.</i>						
1	2	3	4	5		
Not at All Typical of the Session	Slightly Typical of the Session	Somewhat Typical of the Session	Typical of the Session	Very Typical of the Session		
1. The therapist set an agenda or established specific goals for the therapy session.	T:	1	2	3	4	5
2. The therapist made connections between the client's current situation and his/her past.	T:	1	2	3	4	5
3. The therapist worked to give the client hope or encouragement.	T:	1	2	3	4	5
4. The therapist repeated back to the client (paraphrased) the meaning of what the client was saying.	T:	1	2	3	4	5
5. The therapist encouraged the client to identify or label feelings that he/she had in or outside of the session.	T:	1	2	3	4	5
6. The therapist encouraged the client to talk about feelings he/she had previously avoided or never expressed.	T:	1	2	3	4	5
7. The therapist pointed out times when the client's behavior seemed inconsistent with what the client was saying, like when he/she: <ul style="list-style-type: none"> • suddenly shifted his/her moods or topics; • was silent a long time; • laughed, smiled, looked away, or was uncomfortable; • avoided talking about specific topics or people. 	T:	1	2	3	4	5
8. The therapist taught the client specific new skills or behaviors, like how to: <ul style="list-style-type: none"> • relax his/her muscles; • how to control his/her emotions; • how to be assertive with others; • how to act in social situations. 	T:	1	2	3	4	5
9. The therapist encouraged the client to think about, view, or touch things that the client is afraid of.	T:	1	2	3	4	5
10. The therapist reviewed or assigned homework exercises, like:	T:	1	2	3	4	5

Instructions: The following items represent actions that may or may not have occurred in the session which you just observed. Please rate each item using the scale provided. There are no right or wrong answers.									
1	2	3	4	5					
Not at All Typical of the Session	Slightly Typical of the Session	Somewhat Typical of the Session	Typical of the Session	Very Typical of the Session					
<ul style="list-style-type: none">• writing down certain thoughts or feelings outside the session;• practicing certain behaviors.									
11. The therapist was warm, sympathetic, and accepting.	T:	1	2	3	4	5			
12. The therapist talked about the function or purpose that the client's problem might have, like how it: <ul style="list-style-type: none">• lets him/her avoid responsibility;• keeps others away from him/her.	T:	1	2	3	4	5			
13. The therapist encouraged the client to explore explanations for events or behaviors other than those that first came to the client's mind.	T:	1	2	3	4	5			
14. The therapist made connections between the way the client acts or feels towards the therapist and the way that the client acts or feels in his/her other relationships.	T:	1	2	3	4	5			
15. The therapist and the client discussed the client's dreams, fantasies, or wishes.	T:	1	2	3	4	5			
16. The therapist and the client worked together as a team.	T:	1	2	3	4	5			
17. The therapist listened carefully to what the client was saying.	T:	1	2	3	4	5			
18. The therapist focused on how disagreements between certain parts of the client's personality have caused the client's problems.	T:	1	2	3	4	5			
19. The therapist encouraged the client to change specific behaviors.	T:	1	2	3	4	5			
20. The therapist encouraged the client to look for evidence in support of or against one of the client's beliefs or assumptions.	T:	1	2	3	4	5			
21. The therapist encouraged the client to explore the personal meaning of an event or a feeling.	T:	1	2	3	4	5			
22. The therapist seemed interested in trying to understand what the client was experiencing.	T:	1	2	3	4	5			
23. The therapist encouraged the client to focus on his/her moment-to-moment experience.	T:	1	2	3	4	5			
24. The therapist encouraged the client to question his/her beliefs or to discover flaws in his/her reasoning.	T:	1	2	3	4	5			
25. The therapist focused on a specific concern in the client's relationships, like: <ul style="list-style-type: none">• disagreements or conflicts;• major changes;• loss of a loved one;• loneliness.	T:	1	2	3	4	5			
26. The therapist encouraged the client to explore ways in which the client could make changes in his/her relationships, like ways to: <ul style="list-style-type: none">• resolve a conflict in a relationship;• fulfill a need;• establish new relationships or to contact old friends;• ways to avoid problems the client had experienced in previous relationships.	T:	1	2	3	4	5			
27. The therapist encouraged the client to examine his/her relationships with others, like: <ul style="list-style-type: none">• positive and negative aspects of his/her relationships;• what the client wants and others want from him/her;• the way the client acts in relationships.	T:	1	2	3	4	5			

Instructions: The following items represent actions that may or may not have occurred in the session which you just observed. Please rate each item using the scale provided. There are no right or wrong answers.

1	2	3	4	5				
Not at All Typical of the Session	Slightly Typical of the Session	Somewhat Typical of the Session	Typical of the Session	Very Typical of the Session				
28. The therapist both accepted the client for who he/she is and encouraged him/her to change.	T:	1	2	3	4	5		
29. The therapist encouraged the client to think about or be aware of things in his/her life without judging them.	T:	1	2	3	4	5		
30. The therapist tried to help the client better understand how the client's problems were due to difficulties in his/her social relationships.	T:	1	2	3	4	5		

APPENDIX C

Study 2

Participants.

Patients.—(b) A sample of 285 patients filled out the optional English version demographic survey. The mean age was 33 years ($SD = 12.67$), 81% females ($n = 228$), 71% Caucasian ($n = 201$), 12% African American ($n = 33$), 7% Latino/a ($n = 21$), 5% Asian ($n = 14$), 1% Indian/East Indian ($n = 4$), and 4% ‘Other’ ($n = 11$). A sample of 106 patients completed the optional Hebrew version demographic survey. The mean age was 32.88 years ($SD=9.32$), 90% females ($n = 95$) and 10% ($n = 11$) males participated, with 45% of respondents from European-American descent ($n=47$), 10% from Asian-African descent ($n=10$) and 31% “Other” ($n=33$). These ethnic categories are commonly used in demographic surveys in Israel.

Therapists.—(b) A total of 175 therapists reported their therapeutic orientation. The sample included 41% CBT therapists ($n = 85$); 32% psychodynamic therapists ($n=57$); 8% person-centered therapists ($n = 15$); 6% interpersonal therapists ($n = 10$); and 4% clinicians practicing acceptance-based therapy ($n = 8$). A sample of 165 therapists completed the demographic survey in English. The mean age was 43.47 years ($SD = 14.92$), 65% females ($n = 111$) and 35% males ($n = 61$), with 86% Caucasian ($n = 149$), 5% Latino/a ($n = 8$), 4% African-American ($n = 7$), 2% Asian ($n = 4$), and 3% ‘Other’ ($n = 5$). Sixty-six therapists completed the optional demographic survey in Hebrew. The mean age was 39.21 years ($SD = 10.13$), 89% ($n = 59$) females and 11% ($n = 7$) males, with 82% ($n = 55$) therapists from European-American descent, 13% ($n = 8$) from Asian-African descent, and 4.5% ($n = 3$) “Other.”

APPENDIX D

Table D1

Study 2: Means and SD's for Patients, Therapists, and Observers on the MULTI-30 subscales

	Patients N=644	Therapists N=522	Observers N=280
	<u>M (SD)</u>	<u>M (SD)</u>	<u>M (SD)</u>
Psychodynamic	3.25 (0.99)	3.12 (1.06)	1.79 (0.70)
Process-experiential	3.19 (1.00)	3.36 (0.94)	1.95 (0.67)
Interpersonal	3.09 (1.22)	3.09 (1.24)	1.75 (0.83)
Person Centered	3.85 (0.87)	3.96 (0.80)	3.00 (0.74)
Common Factors	4.02 (0.79)	4.43 (0.55)	4.03 (0.48)
Behavioral	2.98 (1.08)	3.18 (1.16)	2.54 (0.96)
Cognitive	3.07 (0.94)	3.17 (1.03)	2.70 (1.01)
Dialectical-Behavioral	3.16 (0.93)	3.39 (0.93)	2.40 (0.76)

Note. MULTI = Multitheoretical List of Therapeutic Interventions; Subscale scores are on 5-point Likert scale: 1 = Not at all typical to the session; 5 = Very typical to the session.

Table D2

Study 2: Internal Consistency for MULTI-30 subscales

Subscale	Estimate	Patients		Therapists		Observers
		<u>Single^a N=315</u>	<u>Repeated^b N=191</u>	<u>Single^a N=363</u>	<u>Repeated^b N=188</u>	<u>Repeated^b N=280</u>
Psychodynamic	α	.74	.81	.72	.82	.87
	ω	.78	.78	.82	.75	.79
Process-Experiential	α	.71	.69	.69	.70	.67
	ω	.78	.71	.75	.85	.78
Interpersonal	α	.83	.89	.78	.88	.80
	ω	.88	.88	.82	.91	.86
Person-Centered	α	.59	.60	.41	.68	.75
	ω	.75	.70	.60	.80	.84
Common Factors	α	.87	.71	.66	.61	.72
	ω	.89	.79	.72	.75	.84
Behavioral	α	.72	.68	.79	.70	.69
	ω	.83	.79	.88	.95	.80
Cognitive	α	.69	.74	.64	.82	.79
	ω	.83	.86	.81	.36	.87

Note.

^a single session per patient;

^b multiple sessions per patient.;

Table D3

Study 3: Replication of McCarthy, Keefe, & Barber, 2016 –Effect Sizes (r) for Linear and Curvilinear Relationships of the MULTI-30 and MULTI-60 Subscales to Subsequent Outcome

	MULTI-60			MULTI-30			r(MULTI-60 – MULTI-30)
		Curvilinear ^b			Curvilinear ^b		
MULTI Subscale	Linear ^a	Linear	Quadratic	Linear ^a	Linear	Quadratic	
Psychodynamic	.01	-.43 [*]	.44 [*]	.03	-.43	.44 [*]	.94 ^{**}
Process-experiential	-.03	-.56 [*]	.57 ^{**}	-.09	-.38	.37 [*]	.91 ^{**}
Interpersonal	-.05	-.12	.12	-.05	-.09	.08	.98 ^{**}
Person Centered	.10	-.18	.19	.09	.04	-.04	.92 ^{**}
Common Factors	.26	.02	.01	.25	.05	-.03	.95 ^{**}
Behavioral	-.12	.17	-.16	-.02	-.02	.02	.88 ^{**}
Cognitive	-.12	.05	-.04	-.04	.06	-.07	.80 ^{**}
Dialectical-Behavioral	.32	.14	-.11	.08	-.01	.02	.66 ^{**}

HRSD = Hamilton Rating Scale for Depression. Table includes effect sizes (r).

^a Semi-partial correlations (df = 30) of termination HRSD scores and MULTI subscale scores, controlling for prior HRSD change.

^b Semi-partial correlations (df = 29) of termination HRSD scores and either linear or quadratic MULTI subscale scores, controlling for prior HRSD change and the linear or quadratic term for that MULTI subscale score.

* p < .05;

** p < .01

References

- American Psychological Association. Brown Y, & Tullos JM Producers, Tullos JM Director. (2013). Three approaches to psychotherapy with a male client: The next generation [DVD]. Available from <http://www.apa.org/pubs/videos/4310890.aspx>
- American Psychological Association. Brown Y, & Tullos JM Producers, Tullos JM Director. (2013). Three approaches to psychotherapy with a female client: The next generation [DVD]. Available from <http://www.apa.org/pubs/videos/4310889.aspx>
- Barber JP, Barrett MS, Gallop R, Rynn MA, & Rickels K (2012). Short-term dynamic psychotherapy versus pharmacotherapy for major depressive disorder: a randomized, placebo-controlled trial. *Journal of Clinical Psychiatry*, 73, 66–73. doi: 10.4088/JCP.11m06831. [PubMed: 22152401]
- Barkham M, Hardy GE, & Startup M (1996). The IIP-32: A short version of the Inventory of Interpersonal Problems. *British Journal of Clinical Psychology*, 35, 21–35. doi: 10.1111/j.2044-8260.1996.tb01159.x [PubMed: 8673033]
- Bates D, Maechler M, Bolker B, & Walker S (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67, 1–48. doi:10.18637/jss.v067.i01
- Borkovec TD, Echemendia RJ, Ragusea SA, & Ruiz M (2001). The Pennsylvania Practice Research Network and future possibilities for clinically meaningful and scientifically rigorous psychotherapy effectiveness research. *Clinical Psychology: Science and Practice*, 8, 155–167. doi: 10.1093/clipsy/8.2.155
- Boswell JF, Castonguay LG, & Wasserman RH (2010). Effects of psychotherapy training and intervention use on session outcome. *Journal of Consulting and Clinical Psychology*, 78, 717–723. doi: 10.1037/a0020088 [PubMed: 20873906]

- Boswell JF, Kraus DR, Miller SD, & Lambert MJ (2015). Implementing routine outcome monitoring in clinical practice: Benefits, challenges, and solutions. *Psychotherapy Research*, 25, 6–19. doi: 10.1080/10503307.2013.817696 [PubMed: 23885809]
- Castonguay LG, Janis RA, Youn SJ, Xiao H, McAleavey AA, Boswell JF, ... & Hemmelstein NA (2017). Clinicians' prediction and recall of therapeutic interventions: practice research network study. *Counselling Psychology Quarterly*, 1–15. doi: 10.1080/09515070.2017.1334628
- Castonguay LG, Nelson DL, Boutsellis MA, Chiswick NR, Damer DD, Hemmelstein NA, ... & Spayd C (2010). Psychotherapists, researchers, or both? A qualitative analysis of psychotherapists' experiences in a practice research network. *Psychotherapy: Theory, Research, Practice, Training*, 47, 345–354. doi: 10.1037/a0021165
- Castonguay LG, Youn SJ, Xiao H, Muran JC, & Barber JP (2015). Building clinicians-researchers partnerships: Lessons from diverse natural settings and practice-oriented initiatives. *Psychotherapy Research*, 25, 166–184. doi: 10.1080/10503307.2014.973923 [PubMed: 25437382]
- Cicchetti DV (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6, 284–290. doi: 10.1037/1040-3590.6.4.284
- Cronbach LJ (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297–334. doi:10.1007/BF02310555
- Coste J, Guillemin F, Pouchot J, & Fermanian J (1997). Methodological approaches to shortening composite measurement scales. *Journal of Clinical Epidemiology*, 50, 247–252. doi: 10.1016/S0895-4356(96)00363-0 [PubMed: 9120523]
- Darchuk A, Wang V, Weibel D, Fende J, Anderson T, & Horvath AO (2000). Manual for the Working Alliance Inventory—Observer Form, 4th revision. Unpublished report, Department of Psychology, Ohio University.
- Dennhag I, Gibbons MBC, Barber JP, Gallop R, & Crits-Christoph P (2012). How many treatment sessions and patients are needed to create a stable score of adherence and competence in the treatment of cocaine dependence? *Psychotherapy Research: Journal of the Society for Psychotherapy Research*, 22, 475–488. doi:10.1080/10503307.2012.674790 [PubMed: 22449079]
- Dunn TJ, Baguley T & Brunsden V (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, 105, 399–412. doi: 10.1111/bjop.12046 [PubMed: 24844115]
- Falkenström F, Finkel S, Sandell R, Rubel JA, & Holmqvist R (2017). Dynamic models of individual change in psychotherapy process research. *Journal of Consulting and Clinical Psychology*, 85, 537–549. doi:10.1037/ccp0000203 [PubMed: 28394170]
- Foa EB, Huppert JD, Leiberg S, Langner R, Kichic R, Hajcak G, & Salkovskis PM (2002). The Obsessive-Compulsive Inventory: development and validation of a short version. *Psychological Assessment*, 14, 485–496. doi: 10.1037/1040-3590.14.4.485 [PubMed: 12501574]
- Fox J, and Bouchet-Valat M (2017). Rcmdr: R Commander. R package version 2.3–2. Retrieved from <https://cran.r-project.org/web/packages/Rcmdr/Rcmdr.pdf>
- Frank JD, & Frank JB (1991). *Persuasion and healing: A comparative study of psychotherapy* (3rd ed.). Baltimore, MD: Johns Hopkins University Press.
- Hatcher RL, & Gillaspie JA (2006). Development and validation of a revised short version of the Working Alliance Inventory. *Psychotherapy Research*, 16, 12–25. doi: 10.1080/10503300500352500
- Hermida R (2015). The problem of allowing correlated errors in structural equation modeling: Concerns and considerations. *Computational Methods in Social Sciences*, 3, 5–17.
- Kruyen PM, Emons WH, & Sijtsma K (2013). On the shortcomings of shortened tests: A literature review. *International Journal of Testing*, 13, 223–248. doi: 10.1080/15305058.2012.703734
- McAleavey AA, & Castonguay LG (2014). Insight as a common and specific impact of psychotherapy: Therapist-reported exploratory, directive, and common factor interventions. *Psychotherapy*, 51, 283–294. doi: 10.1037/a0032410 [PubMed: 24059736]
- McAleavey AA, Castonguay LG, & Xiao H (2014). Therapist orientation, supervisor match, and therapeutic interventions: Implications for session quality in a psychotherapy training PRN. *Counselling and Psychotherapy Research*, 14, 192–200. doi: 10.1080/14733145.2014.929418

- McCarthy KS, & Barber JP (2009). The multitheoretical list of therapeutic interventions (MULTI): Initial report. *Psychotherapy Research*, 19, 96–113. doi: 10.1080/10503300802524343 [PubMed: 19065285]
- McCarthy KS, Gibbons MBC, & Barber JP (2008). The relation of rigidity across relationships with symptoms and functioning: An investigation with the revised Central Relationship Questionnaire. *Journal of Counseling Psychology*, 55, 346–358. doi: 10.1037/a0012578 [PubMed: 19578479]
- McCarthy KS, Keefe JR, & Barber JP (2016). Goldilocks on the couch: Moderate levels of psychodynamic and process-experiential technique predict outcome in psychodynamic therapy. *Psychotherapy Research*, 26, 307–317. doi: 10.1080/10503307.2014.973921 [PubMed: 25364984]
- McDonald RP (1970). The theoretical foundations of principal factor analysis, canonical factor analysis, and α factor analysis. *British Journal of Mathematical and Statistical Psychology*, 23, 1–21. doi: 10.1111/j.2044-8317.1970.tb00432.x
- McDonald RP (1999). *Test theory: A unified approach*. Mahwah, NJ: Erlbaum.
- Mick E, Faraone SV, Spencer T, Zhang HF, & Biederman J (2008). Assessing the validity of the quality of life enjoyment and satisfaction questionnaire—short form in adults with ADHD. *Journal of attention disorders*, 11, 504–509. doi: [PubMed: 17934183] 10.1177/1087054707308468
- Milrod B, Chambless DL, Gallop R, Busch FN, Schwalberg M, McCarthy KS, ... & Barber JP (2015). Psychotherapies for panic disorder: a tale of two sites. *The Journal of Clinical Psychiatry*, 77, 927–935. doi: 10.4088/JCP.14m09507
- Muthén BO (1994). Multilevel covariance structure analysis. *Sociological Methods & Research*, 22, 376–398.
- Muthén LK., & Muthén BO (2007). *Mplus User's Guide* (6th ed.). Los Angeles, CA: Muthén & Muthén.
- Muthén LK and Muthén BO (1998–2015). *Mplus User's Guide* (7th ed.). Los Angeles, CA: Muthén & Muthén
- Muthén LK and Muthén BO (2017). Re: Multilevel Factor Analysis [Online discussion group]. Retrieved from <http://www.statmodel.com/discussion/messages/12/101.html?1491002754>
- Soldz S, Budman S, Demby A and Merry J 1995 A short form of the Inventory of Interpersonal Problems Circumplex scales. *Assessment*, 2, 53–63. doi: 10.1177/1073191195002001006
- Shostrom EL Director/Producer. (1965). *Three approaches to psychotherapy I*. [Film]. Orange, CA: Psychological Films.
- Shostrom EL, & Shostrom SK Producer, Ratner H Director. (1986). *Three approaches to psychotherapy III*. [Film]. Orange, CA: Psychological Films.
- Shrout PE (1995). Reliability In Tsuang MT, Tohen M & Zahner GEP (Eds), *Textbook in psychiatric epidemiology* (pp. 213–227). New York: Wiley-Liss.
- Shrout PE, & Fleiss JL (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, 86, 420–428. doi: 10.1037/0033-2909.86.2.420 [PubMed: 18839484]
- Smith GT, McCarthy DM, & Anderson KG (2000). On the sins of short-form development. *Psychological Assessment*, 12, 102–111. doi: 10.1037/1040-3590.16.3.340 [PubMed: 10752369]
- Solomonov N, Kuprian N, Zilcha-Mano S, Gorman BS, & Barber JP (2016). What do psychotherapy experts actually do in their sessions? An analysis of psychotherapy integration in prototypical demonstrations. *Journal of psychotherapy Integration*, 26, 202–216. doi:10.1037/int0000021
- Solomonov N, McCarthy KS (6, 2016). Introduction to the Multitheoretical List of Intervention. Workshop given at the International Meeting of the Society for Psychotherapy Research, Israel.
- Solomonov N, McCarthy SK, Keefe RJ, Gorman SB, Blanchard M, & Barber JP (2017). Fluctuations in Alliance and Use of Techniques over Time: A Bi-directional Relation between Use of 'Common Factors' Techniques and the Development of the Working Alliance. *Journal of Clinical Psychology & Psychotherapy*. 1–10. doi: 10.1002/cpp.2143
- Tracey TJ, & Kokotovic AM (1989). Factor structure of the working alliance inventory. *Psychological Assessment: A journal of consulting and clinical psychology*, 1, 207–210. doi: 10.1037/1040-3590.1.3.207
- Thompson ER (2007). Development and validation of an internationally reliable short-form of the positive and negative affect schedule (PANAS). *Journal of Cross-cultural Psychology*, 38, 227–242. doi: 10.1177/0022022106297301

- R Core Team. (2016). R: A Language and Environment for Statistical Computing (Version 3.3.1). Vienna, Austria: R Foundation for Statistical Computing Retrieved from <https://www.R-project.org/>
- Revelle W (2017) psych: Procedures for Personality and Psychological Research. Evanston, IL: Northwestern University Retrieved from <https://CRAN.R-project.org/package=psych>
- Rosenzweig S (1936). Some implicit common factors in diverse methods of psychotherapy. American Journal of Orthopsychiatry, 6, 412–415. doi:10.1111/j.1939-0025.1936.tb05248.x
- Rosseel Y, (2012). lavaan: An R Package for Structural Equation Modeling. Journal of Statistical Software, 48, 1–36. Retrieved from <http://www.jstatsoft.org/v48/i02/>.
- Van Buuren S & Groothuis-Oudshoorn K (2011). Mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software, 45, 1–67. Retrieved from <http://www.jstatsoft.org/v45/i03/>.
- Wampold BE (2001). The great psychotherapy debate: Models, methods and findings. Mahwah, NJ: Lawrence Erlbaum. doi: 10.1002/pits.10115
- Widaman KF, Little TD, Preacher KJ, & Sawalani GM (2011). On creating and using short forms of scales in secondary research In Trzesniewski KH, Donnellan MB, & Lucas RE (Eds.), Secondary data analysis: An introduction for psychologists (pp. 39–62). Washington, DC: American Psychological Association.
- Wothke W (1993). Nonpositive definite matrices in structural modeling In Bollen KA & Long JS (Eds.), Testing structural equation models (pp. 256–293). Newbury Park, CA: Sage
- Zickgraf HF, Chambless DL, McCarthy KS, Gallop R, Sharpless BA, Milrod BL, & Barber JP (2016). Interpersonal Factors Are Associated with Lower Therapist Adherence in Cognitive-Behavioural Therapy for Panic Disorder. Clinical Psychology & Psychotherapy, 23, 272–284. doi: 10.1002/cpp.1955 [PubMed: 25882924]

Clinical and Methodological Significance of this Article

The MULTI-30, developed and validated in this study, is a valid, reliable, and cost-effective brief measure which could be used to assess patients, therapists, and observers' perceptions of use of interventions from eight major therapeutic approaches. The MULTI-30 could be used to examine the role of use of specific interventions on process and outcome of different treatment modalities. It could also be used as a clinical tool in teaching, training, and supervision.

Table 1

Study 1: Confirmatory Factor Analysis for 30 selected items from MULTI-60

MULTI Subscale	Selected Items for MULTI-30	Factor Loading		Ratings	
		Patients N=255	Therapists N=184	Observers N=164	Experts
Psychodynamic	2. Making connections between past and present	.66	.61	.63	2
	12. Exploring avoided emotions	.67	.59	.65	4
	20. Identifying the function of symptoms	.77	.62	.34	4
	22. Making transference interpretations	.49	.73	.18 ^a	4
Process-experiential	24. Exploring dreams, wishes, and fantasies	.54	.61	.52	4
	11. Identifying and labeling emotions	.69	.55	.34	1
	13. Identifying defenses	.53	.67	.36	2
	34. Identifying conflict splits and consequences	.65	.57	.94	0
Interpersonal	47. Focusing on moment-to-moment experience	.41	.23	.30	2
	50. Focusing on a specific relationship problem	.71	.75	.80	1
	51. Encourage change in relationships	.76	.78	.58	1
	54. Exploring specific interpersonal behaviors	.85	.83	.86	0
Person-Centered	60. Making connections between relationship problems and symptoms	.62	.68	.87	1
	10. Paraphrasing	.48	.48	.40	1
	40. Exploring personal meaning	.64	.65	.58	3
	46. Demonstrating interest in patient's experience	.67	.63	.83	3
Common Factors	7. Providing hope and encouragement	.78	.66	.62	2
	18. Being warm, sympathetic and accepting	.73	.33	.74	2
	28. Team work and collaboration	.85	.54	.83	2
	31. Listening carefully	.69	.11 ^b	.73	1
Behavioral	15. Teaching new skills and behaviors	.66	.63	.78	3
	16. Exposure to thing the patient is afraid of	.57	.55	.45	2
Cognitive	35. Encouraging behavioral change	.71	.82	.85	0
	21. Exploring alternative explanations to behaviors	.70	.34	.57	3
	37. Evidence search	.68	.82	.86	4
	49. Challenging irrational thoughts	.67	.78	.92	3

MULTI Subscale	Selected Items for MULTI-30	Factor Loading		Ratings	
		Patients N=255	Therapists N=184	Observers N=164	Experts
Cognitive-Behavioral	1. Setting an agenda	.66	.73	.63	2
	17. Assigning/reviewing homework	.71	.84	.61	3
Dialectical-Behavioral	56. Accepting and encouraging change	.71	.54	.58	3
	58. Encouraging mindfulness	.72	.43	.73	2

Note. MULTI = Multitheoretical List of Therapeutic Interventions; Selected Item=Item descriptions are brief topical summaries; N= number of responses collected;
J CBT subscale included 4 MULTI-6-items (1, 17, 33, 39). All factor loadings are significant ($p<.01$), unless marked:
a $p=.02$
b non-significant. Experts = the number of experts who rated each item as “most representative/typical of the therapeutic approach” are indicated.

Table 2

Study 1: Means, SDs and Internal Consistency Estimates for Patients, Therapists, and Observers on the 30 items Selected from the MULTI-60 subscales

	Patients N=255			Therapists N=184			Observers N=164		
	<u>M (SD)</u>	<u>α</u>	<u>ω</u>	<u>M (SD)</u>	<u>α</u>	<u>ω</u>	<u>M (SD)</u>	<u>α</u>	<u>ω</u>
Psychodynamic	2.97 (0.86)	.87	.89	2.84 (0.93)	.93	.94	2.62 (0.57)	.72	.86
Process-experiential	2.89 (0.85)	.83	.87	2.94 (0.74)	.84	.90	2.57 (0.54)	.72	.86
Interpersonal	3.12 (0.96)	.83	.89	2.84 (0.93)	.83	.90	2.55 (0.64)	.83	.91
Person Centered	3.42 (0.89)	.76	.84	3.50 (0.83)	.78	.86	3.31 (0.57)	.63	.77
Common Factors	3.98 (0.83)	.89	.92	4.02 (0.63)	.78	.86	3.92 (0.49)	.87	.93
Behavioral	3.06 (0.81)	.89	.91	2.99 (0.86)	.88	.92	2.10 (0.52)	.87	.93
Cognitive	3.20 (0.83)	.91	.93	3.17 (0.84)	.90	.93	2.34 (0.57)	.90	.93
Dialectical-Behavioral	3.29 (0.91)	.85	.86	3.06 (0.98)	.76	.86	2.17 (0.47)	.71	.84

Note. N= number of responses collected; MULTI = Multitheoretical List of Therapeutic Interventions; Subscale scores are on 5-point Likert scale: 1 = Not at all typical to the session; 5 = Very typical to the session; α = Cronbach's alpha; ω = McDonald's omega

Table 3

Study 2 – Confirmatory Factor analysis for MULTI-30 Subscales – Patient, Therapist and Observer samples

Subscale	Item	Patients		Therapists		Observers
		CFA ^a N= 268	MCFA ^b N=191	CFA ^a N=321	MCFA ^b N=188	MCFA ^b N=279
Psychodynamic	2	.60	.56	.58	.55	.60
	12	.66	.79	.52	.80	.79
	20	.72	.72	.44	.51	.45
	22	.50	.61	.81	.67	.47
	24	.61	.47	.68	.61	.43
Process-experiential	11	.65	.50	.51	.54	.71
	13	.58	.69	.63	.62	.43
	34	.62	.66	.65	.75	.54
	47	.67	.27	.53	.36	.10 ^a
Interpersonal	50	.74	.56	.59	.71	.79
	51	.85	.79	.71	.60	.38
	54	.85	.72	.73	.79	.75
	60	.64	.47	.68	.77	.54
Person-Centered	10	.35	.42	.30	.47	.52
	40	.64	.67	.58	.81	.75
	46	.77	.42	.54	.42	.51
Common Factors	7	.73	.77	.58	.67	.64
	18	.81	.74	.62	.55	.88
	28	.78	.86	.62	.59	.58
	31	.84	.68	.42	.03 ^a	.54
Behavioral	15	.76	.67	.88	.76	.61
	16	.65	.48	.55	.32	.50
	35	.67	.69	.84	.80	.89
Cognitive	21	.74	.58	.22	.34	.39
	37	.61	.71	.90	.91	.91
	49	.60	.73	.83	.93	.87
Dialectical-Behavioral	56	.68	.61	.67	.47	.81
	58	.79	.66	.39	.27 ^a	.44
Cognitive-Behavioral	1	.66	.68	.75	.79	.72
	17	.72	.82	.96	.86	.98

Note.

^a single session per patient, CFA=Confirmatory Factor Analysis CFA;^b multiple sessions per patient.; MCFA =Multilevel Confirmatory Factor Analysis (MCFA), within-patient standardized factor loadings are presented and significant unless marked: ^a non-significant.