Journal of Counseling Psychology

Beyond Symptom Reduction: Development and Validation of the Complementary Measure of Psychotherapy Outcome (COMPO)

Harold Chui, Eddie S. K. Chong, Dana Atzil-Slonim, Zeynep Sahin, Nili Solomonov, Mary V. Minges, Nadia Kuprian, and Jacques P. Barber

Online First Publication, October 22, 2020. http://dx.doi.org/10.1037/cou0000536

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Counseling Psychology

ISSN: 0022-0167

http://dx.doi.org/10.1037/cou0000536

Beyond Symptom Reduction: Development and Validation of the Complementary Measure of Psychotherapy Outcome (COMPO)

Harold Chui
The Chinese University of Hong Kong

Eddie S. K. Chong
University of Maryland, College Park

Dana Atzil-Slonim
Bar-Ilan University

Zeynep Sahin
Adelphi University

Nili Solomonov
Weill Cornell Medical College

Mary V. Minges
Albert Einstein College of Medicine

Nadia Kuprian and Jacques P. Barber
Adelphi University

Most measures of psychotherapy outcome focus on symptomatic change. However, clients often report other changes through therapy, such as increased self-acceptance. This study reports on the development and validation of the Complementary Measure of Psychotherapy Outcome (COMPO) that assesses different areas of psychological functioning deemed important by clients and therapists. Items were written based on a literature review of client-reported change and feedback from experienced therapists. Exploratory factor analysis was conducted on the initial 42-item COMPO administered to 264 psychotherapy clients. Iterative item reduction resulted in the final 12-item, four-factor solution, with factors named self-acceptance, self-knowledge, relationship quality, and consideration of others. This factor structure, along with a bifactor model that contains a general factor and the four domain-specific factors, was replicated on a sample of 571 adults in the community. The 12-item COMPO exhibits convergent validity with measures of self-esteem, insight, social support, and empathy; demonstrates 2-week test–retest reliability; and predicts life satisfaction. The 12-item COMPO was further administered to 28 clients in short-term psychodynamic therapy for depression. Except for consideration of others, COMPO subscales and total scale scores improved from pre- to posttherapy. Posttherapy COMPO scores were also higher among clients who experienced clinically significant change compared to those who did not. The COMPO was negatively associated with depressive symptoms and impairments in functioning across the three samples. The brevity of the COMPO makes it a convenient tool to supplement symptom-based measures for a more comprehensive assessment of outcome in psychotherapy.

Public Significance Statement
The goal of psychotherapy does not only include the reduction of psychological symptoms, yet most outcome measures focus on symptom change. A new measure is developed to assess nonsymptomatic changes that patients and therapists care about in therapies across theoretical orientations.

Keywords: psychotherapy outcome, self-report, psychological functioning, exploratory factor analysis, confirmatory factor analysis

Supplemental materials: http://dx.doi.org/10.1037/cou0000536.supp

Harold Chui, Department of Educational Psychology, The Chinese University of Hong Kong; Eddie S. K. Chong, Department of Psychology, University of Maryland, College Park; Dana Atzil-Slonim, Department of Psychology, Bar-Ilan University; Zeynep Sahin, Derner School of Psychology, Adelphi University; Nili Solomonov, Weill Cornell Institute of Geriatric Psychiatry, Weill Cornell Medical College; Mary V. Minges, Department of Psychiatry and Behavioral Sciences, Montefiore Medical Center, Albert Einstein College of Medicine; Nadia Kuprian and Jacques P. Barber, Derner School of Psychology, Adelphi University.

Eddie S. K. Chong is now at Counseling and Psychological Services, University of Pennsylvania. Nadia Kuprian is now at an Independent practice in New York, New York.

Results from this study were presented in part at the 49th Society for Psychotherapy Research International Meeting in Amsterdam, the Netherlands, in June 2018. We thank the graduate students, early-career psychologists, and experienced psychotherapists for their comments and feedback on the initial pool of items. We also thank David Kraus and Kimberlee Trudeau at Outcome Referrals for permission to use and scoring the Treatment Outcome Package.

Correspondence concerning this article should be addressed to Harold Chui, Department of Educational Psychology, The Chinese University of Hong Kong, Room 302 Ho Tim Building, Shatin, N. T., Hong Kong. E-mail: haroldchui@cuhk.edu.hk

Research in psychotherapy outcome primarily has focused on changes in symptoms, often overlooking other changes that patients may experience (e.g., Hill, Chui, & Baumann, 2013). However, the goals of psychotherapy are often broader. For example, Strupp and Hadley (1979) noted that many patients go to therapy to find meaning in life and maximize their potential. Patients also reported numerous other changes, such as improvements in various aspects of the self (Binder, Holgersen, & Nielsen, 2010; Connolly & Strupp, 1996; Göstas, Wiberg, & Kjellin, 2012; Hasler, Moergeli, & Schnyder, 2004; Klein & Elliott, 2006; Levitt, Butler, & Hill, 2006). Thus, a narrow focus on symptomatic change in research may limit our understanding of the functions and benefits of psychotherapy. In particular, changes in positive aspects of psychological functioning, such as self-acceptance, through psychotherapy are not assessed by widely used, symptom-based outcome measures.

Previous attempts in measuring nonsymptomatic change in therapy often have come from the psychoanalytic tradition. For example, the Personality Health Index was developed from the clinician-rated Shedler-Westen Assessment Procedure (Westen & Shedler, 1999) to quantify patients' levels of personality functioning during psychoanalysis (Waldron et al., 2011). The Scales for Psychological Capacities (DeWitt, Hartley, Rosenberg, Zilberg, & Wallerstein, 1991) also were developed to assess the psychoanalyst's perspective of changes in personality structure in patients. A common thread across these measures is the reliance on clinical interviews. Although clinicians' training places them in a unique position to assess patient functioning, and their judgment may be less prone to the effects of self-deception and social desirability compared to patients' (Waldron et al., 2011), observer-rated measures are time-consuming and require extensive training. Such time and labor commitment render these measures impossible to administer in large-sample outcome studies involving repeated measurements.

Outside of the psychotherapy research literature, a number of self-report measures have been developed in the field of positive psychology to assess psychological well-being. For instance, Ryff's (1989) Scales of Psychological Well-Being assess six theory-guided areas of well-being, including self-acceptance, autonomy, positive relations with others, environmental mastery, purpose in life, and personal growth. Keyes (2005, 2007) developed the model of mental health flourishing to include Ryff's six psychological well-being dimensions, in addition to two emotional well-being dimensions (e.g., positive affect) and five social well-being dimensions (e.g., social acceptance). In addition, Keyes (2007) conceptualized complete mental health as not only the absence of mental illness but also having high levels of at least one dimension of emotional well-being and six dimensions of psychological or social well-being. Frisch (1998) integrated literatures in subjective well-being and clinical psychology to highlight patients' quality of life as a criterion of psychotherapy outcome. In particular, fulfillment in valued life domains contributes to mental health in addition to the absence of symptoms.

Although the models by Ryff, Keyes, and Frisch have strong theoretical and empirical foundations, clients' views about changes in therapy were not typically considered. Because clients are the recipients of care, their experience is a crucial measure of treatment usefulness. The APA Presidential Task Force on Evidence-Based Practice (2006) recognizes that consideration of patients' experience improves the connection between research and practice. Patients and therapists also do not always agree on what is helpful in therapy (e.g., Chui, Palma, Jackson, & Hill, 2020); a self-report measure of nonsymptomatic change constructed based on patients' experience in therapy will complement theoretically driven measures of well-being.

Patient's Perspective of Psychotherapy Outcome

We located six studies that examined psychotherapy patients' perceived changes in therapy. These studies focused on areas of change (i.e., outcome domains) as opposed to psychotherapy processes, which are much more frequently studied in qualitative research (Hill et al., 2013). Three of these studied patients who received psychotherapy from heterogeneous orientations. In a survey of psychiatric outpatients (Hasler et al., 2004), patients reported that improvements in symptoms, interpersonal domain, well-being, meaning of life, and self-concept were important outcomes. Binder et al. (2010) interviewed former therapy patients and reported four valued categories of outcomes: establishing new ways of relating to others, less symptomatic distress or changes in behavioral patterns contributing to suffering, better self-understanding and insight, and accepting and valuing oneself. In contrast to these two studies, Levitt et al. (2006) noted that few patients brought up symptom reduction as a major change. Instead, they discussed improved ability to relate with others, understand themselves and others, and feel better about themselves and others as important outcomes.

Three other studies investigated patient-reported outcome in specific therapies. Connolly and Strupp (1996) found four types of outcomes after psychodynamically oriented therapy: improvements in symptoms, self-understanding, self-confidence, and self-definition. Klein and Elliott (2006) categorized patient-perceived changes in process-experiential therapy into changes within the self and changes in life situation. Göstas et al. (2012) found no difference in patient-reported change among patients who received cognitive–behavioral therapy and those who received psychodynamic therapy. Emotional balance, access to positive thoughts about self and others, and ability to adapt one's behavior to meet life demands were considered to be changes after both types of psychotherapy.

In sum, across the six studies reviewed, three major domains of outcome were noted in at least five studies: symptoms, interpersonal relationships, and various aspects of the self (e.g., understanding, acceptance). This categorization coincides with what Horowitz (1979) noted as problems that motivate people to seek psychotherapy: symptoms, interpersonal issues, and disturbing cognitions about the self. Because symptoms are already assessed in traditional measures, our measure will focus on changes in self and relationships.

The goal of the present study was to develop and validate a self-report measure of outcome of psychotherapy that assesses patient change beyond symptom reduction (i.e., Complementary Measure of Psychotherapy Outcome; COMPO). We strived to develop a measure that captures what patients and therapists both care about. As explained, patients are care recipients, and what they consider to be important changes in therapy should be reflected in the measure. At the same time, a measure will have clinical utility only if it contains information that therapists value. We next describe the development

of domains and items of the COMPO, followed by a series of studies that examine the measure's factor structure, validity, reliability, and sensitivity to change.

Domain Development of the COMPO

As reviewed, patients reported changes in the self and in their relationships as a result of psychotherapy. The COMPO domains are therefore designed to fall under these two broad areas. Definitions of each domain and the rationale for their inclusion are described.

Among the various changes that patients reported, changes in the interpersonal domain correlated significantly with patient satisfaction with treatment (Hasler et al., 2004). Some patients also talked about improvement in interpersonal relationships more frequently than symptom reduction at the end of psychotherapy (Connolly & Strupp, 1996). Hence, we define the first domain of interest, relationship/support, as the following: Patients feel understood and that they are supported by others outside of therapy. They perceive improvements and are more satisfied in their relationships.

The second category of change that patients report to experience relates to the self. In particular, patients considered increases in self-acceptance as an important outcome of psychotherapy across studies. For example, the most frequently endorsed cluster of change was self-confidence, which included liking of the self and reduction in self-criticism (Connolly & Strupp, 1996). Binder et al. (2010) also used the theme "accepting and valuing oneself" (p. 285) to describe increased valuing of the self. In addition, Klein and Elliott (2006) categorized positive feelings toward self, self-confidence, and allowing oneself to feel as signs of self-esteem. We chose the more inclusive term "self-acceptance" as the name of this domain and defined it as follows: Patients are able to accept different aspects of the self. They feel that they are lovable and respectable.

Another self-related change that patients report is increase in self-understanding. Although often associated with the psychodynamic construct of insight, self-understanding also may describe gains in process-experiential and cognitive–behavioral therapies, such as the awareness of one's internal experience (Klein & Elliott, 2006) and the identification of thoughts and the associated feelings (Binder et al., 2010). Hence, the third domain of interest is self-understanding, defined as follows: Patients have clarity about themselves and their situation. This may include awareness of multiple perspectives and the presence of different feelings in a given situation, as well as insight into one's feelings and actions.

Besides the three outcome domains derived from patients' perspectives of change, we drew from patients' less commonly reported outcomes and from the clinical literature on therapeutic change to conclude that three additional outcome domains should be considered. These are freedom, being true to self, and balancing different aspects of life and experience.

One of the goals of psychodynamic therapy is to bring unconscious materials to consciousness so that people are free to choose how they respond to life circumstances without repeating maladaptive ways of being (Summers & Barber, 2009). In behavioral therapy, expanding one's behavioral repertoire allows one to choose from more behavioral options in response to the same stimulus (Farmer & Chapman, 2008). Thus, patients receiving different forms of psychotherapies may experience increased freedom, defined as the following: Patients experience freedom in choosing how they live their lives. They are not trapped by past experiences. They make more conscious decisions in relationships and in life without falling into old, automatic, maladaptive patterns.

In the humanistic tradition, healthy personality development involves living authentically, where one behaves and expresses emotions in ways that are congruent with their values and beliefs (Rogers, 1961). In addition, from a psychodynamic perspective, people who have internalized harsh external standards during development may have the subjective experience of not knowing oneself or be out of touch with their true experience (Horney, 1951). Being true to self is defined as the following: Patients live closely to their true selves. They are not affected by the opinions of others or by internalized standards ("shoulds"). Patients experience the self as consistent across different situations, and they desire things that are consistent with their values. There is low discrepancy among patients' thoughts, feelings, and behaviors.

Finally, the ideal of balance and moderation came from Eastern and Western philosophies. In the Eastern theory of yin and yang, two opposing forces in nature need to be balanced to maintain physical and psychological health (Lu, 2002). Measures in the West also attempt to evaluate a person's ability to have a balanced approach in life (e.g., DeWitt et al., 1991). Interpersonally satisfying relationships are characterized by reciprocity, balancing between give and take (Buunk & Schaufeli, 1999). Hence, an indicator of psychological health is achieving balance in various life domains. Balancing different aspects of life and experience is defined as follows: Patients achieve balance in their approach toward life. This may manifest in the cognitive domain (e.g., less black and white thinking), emotional domain (e.g., affect regulation without inhibition), behavioral domain (e.g., work-life balance), and interpersonal domain (e.g., balancing the needs of self and others).

Study 1: Item Development and Review

Item Generation

Based on the six domain definitions, the first author generated the initial pool of items. The fourth to eighth authors reviewed the items and offered suggestions to add or revise items during weekly meetings. The item construction process took 3 months and resulted in 48 items. The items and domains were then sent for review, first by graduate students and early-career psychologists and then by experienced clinicians, as detailed below.

Method

Graduate students/early career psychologists. Eight (seven female; seven White, one Middle Eastern) doctoral students in counseling/clinical psychology and two (one female, one male; one Black, one White) early-career psychologists were recruited through personal contact and asked to comment on the clarity of items. They were chosen for being native English speakers and for their familiarity with psychotherapy. Revisions were made based on their feedback (no item added or deleted).

Experienced therapists. Thirty-four (18 female; 30 White, one each for Latino, Asian, biracial, and did not report) experienced therapists were asked to rate each of the 48 items in the initial pool. These therapists had diverse theoretical and education backgrounds (17 integrative/eclectic, 11 psychodynamic, five cognitive–behavioral, one other; 33 psychology, one social work;

32 doctoral level, two master's level) and had between 5 and 45 years of experience providing psychotherapy (M = 23.5 years, SD = 11.0). Therapists were asked whether each item is clear (yes/no), is relevant to psychotherapy outcome (yes/no), and belongs to one or more of the six domains. In addition, therapists were asked to provide written comments for the instrument.

Results

Based on therapists' ratings, three items were deemed unclear, two were deemed irrelevant, and five were not assigned correctly to the intended domain. These items were labeled as such when over 30% of therapists did not provide the desired response and were deleted. To create a scale of reasonable length, if a subscale still had over seven items after initial deletion, items that had between 20% and 30% of therapists wrongly assigning them to unintended domains also were deleted. Four additional items were deleted this way, resulting in a 34-item scale.

In addition to item evaluation, the six domains were examined based on experienced therapists' feedback. Four domains (relationship/support, self-understanding, self-acceptance, and freedom) were considered satisfactory by most reviewers and kept. Being true to self was thought of as vague and overlapping with other domains and was dropped. Balancing different aspects of life was thought of as too broad ("conflate work/life balance with object constancy and affect regulation") and was redefined more narrowly to only include items in the interpersonal realm: balancing the needs of self and others. Two domains, empathy (i.e., being able see things from another person's perspective) and equanimity (i.e., being nonreactive), were added as at least four expert reviewers independently suggested the inclusion of these domains. For the two new domains, items were again written by the first author and reviewed and revised by the fourth to eighth authors. The final COMPO consists of 42 items representing seven theoretical domains and was used in Study 2.

Study 2: Measurement Structure, Replication, Validity, and Reliability

In Study 2, our overarching goal was to examine the psychometric properties of the COMPO. We conducted a series of factor analyses to determine and validate a factor solution that fit our data best. Specifically, we conducted exploratory factor analysis (EFA) on a development sample and then tried to replicate the factor solution on a validation sample using confirmatory factor analysis (CFA). In this process, given our goal to develop a scale in which a general score can be used to evaluate treatment outcomes, we also evaluated the utility and replicability of a bifactor structure, where items were uniquely related to specific factor domains as well as to a general factor (Reise, 2012). To examine replicability, the two samples were recruited close in time and were administered the same initial COMPO items in the same order so that the context of responding was consistent. In terms of recruitment, we recruited clients who were currently in psychotherapy for the development sample and adults in the community for the validation sample. We broadened the recruitment of the validation sample to include anyone who had experience with psychotherapy and who may be potential psychotherapy clients in the future so as to examine the broader relevance of the measure. Finally, we examined the evidence for convergent validity, criterion validity, and test–retest reliability of the COMPO.

Method

Participants.

Development sample. Four hundred twenty-five potential participants visited the website link to the study and consented to participate in the study online. The link was posted on Craigslist webpages in the United States, and the advertisement invited adults currently in outpatient psychotherapy to participate. Those who completed the questionnaires could win one of five US$30 Amazon gift cards. Of those who consented to the study, 63 (14.8%) were excluded because they responded "no" or did not respond to the question on whether they were currently in psychotherapy. Of the remaining 362 participants who were in psychotherapy, 264 (72.9%; 207 female, 52 male, five other) completed the COMPO and answered a validity check question (e.g., "Please answer '2' for this item") correctly and were included in the analysis. Their mean age was 36.1 years (SD = 12.7). In terms of race/ethnicity, 211 were White, 20 Latino/a, 17 Asian/Pacific Islander, six Black, four Native American, and one East Indian; five reported to be "other." In terms of experience in therapy, 229 (86.7%) reported that they had therapy before, whereas 35 (13.3%) noted that this was their first time in therapy. With respect to duration of therapy received, 98 (37.1%) reported to have been in therapy for over 6 years, 38 (14.4%) between 4 and 6 years, 41 (15.5%) between 2 and 4 years, 28 (10.6%) between 1 and 2 years, 23 (8.7%) under 1 year, and 36 (13.6%) did not report this information.

Validation sample. Eight hundred ten potential participants visited the website link to the study and consented to participate. They were presented with the 42-item COMPO and other measures used in the validation study. Of the potential participants, 212 (26.2%) did not respond to any item of the 42-item COMPO and were excluded. Nine (1.1%) had eight or more missing items (i.e., over 15% missing) and also were excluded. In addition, 18 (2.2%) completed the COMPO and validity measures in under 5 min and also were excluded. The remaining 571 participants (442 female, 122 male, six other, one did not report gender; 359 White, 74 Latino/a, 72 Black, 36 Asian/Pacific Islander, six Native American, three East Indian, three Middle Eastern/Arab, 16 other races/ethnicity, two did not report race/ethnicity; age M = 32.9 years, SD = 13.3) were included in the sample. In terms of their experience with psychotherapy, 225 (39.4%) reported that they had ever received psychotherapy, of which 183 (81.3%) had at least 6 months of therapy. This sample therefore represents adults in the community who are potential recipients of psychotherapy. Participants who completed the questionnaire could enter an e-mail to win one of five US$30 Amazon gift cards. This study was approved by the university institutional review board.

Measures.

Complementary Measure of Psychotherapy Outcome (COMPO). The initial COMPO has 42 items and is used to assess participants' psychological functioning. Participants report on a Likert-type scale, with response options of 1 = *strongly disagree*, 2 = *disagree*, 3 = *slightly disagree*, 4 = *neither agree nor disagree*, 5 = *slightly agree*, 6 = *agree*, and 7 = *strongly agree*, their level of agreement with each item. No time frame is

specified, so it is assumed that judgments are made based on the present moment. The COMPO has seven theoretical domains, and a sample item for each domain is as follows: relationship/support (seven items; e.g., "Other people care about me"), self-understanding (seven items; e.g., "I know why I keep finding myself in similar situations"), self-acceptance (eight items; e.g., "I accept myself for who I am"), freedom (six items; e.g., "I feel free to be myself"), balancing the needs of self and others (four items; e.g., "I am good at balancing my needs and the needs of others"), empathy (five items; e.g., "People tell me that I am understanding"), and equanimity (five items; e.g., "I can stand back and think about what's happening before reacting").

Self-esteem. The Rosenberg Self-Esteem Scale (RSE; Rosenberg, 1965) is a widely used 10-item self-report measure of self-esteem. Items are rated from 0 = *strongly disagree* to 3 = *strongly agree*. A higher score indicates a higher level of self-esteem. The RSE has been shown to be positively associated with other measures of self-esteem and is considered to be a test of convergent validity for the COMPO self-acceptance subscale. The Cronbach's alphas were .91 for both development and validation samples.

Empathy. The Interpersonal Reactivity Inventory (IRI; Davis, 1980) is a 28-item self-report measure of empathy. Items are rated on a Likert-type scale from 1 = *does not describe me well* to 5 = *describes me very well*. The seven-item perspective taking subscale (IRI-PT) was used in this study because it assesses a person's sensitivity toward others (Davis, 1980), which is deemed closest to the COMPO empathy subscale. A higher score indicates better ability to consider others' perspective. The Cronbach's alphas were .82 and .79 for the development and validation samples, respectively.

Social support. The Multidimensional Scale of Perceived Social Support (MSPSS; Zimet, Dahlem, Zimet, & Farley, 1988) is a 12-item measure of social support. The total scale represents the overall sense of social support from significant others, family, and friends. Items are rated from 1 = *very strongly disagree* to 7 = *very strongly agree*, and a higher score indicates greater level of perceived social support. Cronbach's alphas were .92 for both development and validation samples. Positive association with the MSPSS total scale is thought to provide evidence of convergent validity for the COMPO relationship/support subscale.

Insight. The Self-Reflection and Insight Scale (SRIS; Grant, Franklin, & Langford, 2002) is a self-report measure of "self-reflection, the inspection and evaluation of one's thoughts, feelings and behavior, and insight, the clarity of understanding of one's thoughts, feelings and behavior" (p. 821). The eight-item insight subscale (SRIS-IN) was used. Items are rated on a Likert-type scale from 1 = *strongly disagree* to 6 = *strongly agree*, and a higher score indicates a higher level of insight. The SRIS-IN subscale has been shown to be positively associated with cognitive flexibility and self-control, both of which arise from awareness of one's cognition, emotion, and action (Grant et al., 2002). Positive association with the SRIS-IN subscale is thought to provide evidence of convergent validity for the COMPO self-understanding subscale. The Cronbach's alphas were .88 and .86 for the development and validation samples, respectively.

Life satisfaction. The Satisfaction With Life Scale (SWLS; Diener, Emmons, Larsen, & Griffin, 1985) is a five-item measure of global life satisfaction. Items are rated on a Likert-type scale from 1 = *strongly disagree* to 7 = *strongly agree*. A higher total score indicates greater life satisfaction. This scale has been shown to be positively associated with many measures of subjective well-being and positive affect (Diener et al., 1985). Cronbach's alphas were .87 and .91 for the development and validation samples, respectively. Positive association between the COMPO total scale and the SWLS was thought to provide evidence of criterion validity for the COMPO.

Symptoms. The Treatment Outcome Package (TOP; Kraus, Seligman, & Jordan, 2005) is a 58-item measure of symptoms of psychopathology and life functioning. Items are rated on a 6-point frequency scale from 1 = *all* to 6 = *none*. The TOP has 12 subscales focusing on different symptoms and functional impairments, such as depression and substance abuse. Scoring of the TOP is based on weightings from CFA studies, and subscale scores are obtained from *Z* transformation based on population means and standard deviations (Kraus et al., 2005). A higher score indicates higher problem severity. In this study, we focused on depressive symptoms as measured by the Depression (DEPRS) subscale. The Cronbach's alpha for the DEPRS subscale was .93 in their study (Kraus et al., 2005). Negative association between COMPO total score and the TOP-DEPRS was thought to provide evidence of criterion validity for the COMPO.

Results

Step 1: Determining measurement structure. An initial EFA was conducted on the 42 items using the development sample ($N = 264$). We first examined the item-level descriptive statistics (see online Supplemental Table S1) as well as the interitem correlation matrix using principal axis factoring extraction method in SPSS 23.0 to determine its appropriateness for factor analysis. Results indicated that the matrix was adequate (Kaiser-Meyer-Olkin index = .91; Bartlett's test of sphericity, $p < .001$). We then conducted the rest of the EFA in Mplus 8. Unless otherwise specified, analyses from this point onward were conducted in Mplus 8 using robust maximum likelihood, and missing data were estimated by full information maximum likelihood. We conducted parallel analysis, where 1,000 random permutations of the original data set were generated, to help determine the optimal number of factors to be extracted (O'Connor, 2000). Results indicated extraction of five factors, which accounted for a total variance of 51.6% (see online Supplemental Table S2 for details of the parallel analysis). Since we have theorized a seven-domain structure for COMPO, we extracted the five-, six-, and seven-factor solutions in EFA. We examined the factor loadings of the respective solutions after applying a geomin rotation, a type of oblique rotation that allows factors to be correlated with one another. The goal of this step was to evaluate the interpretability of the factor structure and to choose between the three solutions. Upon examination, we selected the five-factor solution because it was most interpretable and consistent with some of our expected theoretical domains (see factors loadings in online Supplemental Table S3).

In addition, one of the extracted factors, consistently across the three solutions, contains all 18 negatively worded items that encompass the seven theoretical domains (factor loadings ranging .34–.76, .37–.79, and .39–.79 for the five-, six-, and seven-factor solutions, respectively). We suspected that these items loaded together because of their negative valence rather than due to some latent theoretical construct (i.e., a method factor). Although we

intended to include these negatively worded items to minimize acquiescence bias, recent studies suggested the ineffectiveness of such a strategy (e.g., Zhang, Noor, & Savalei, 2016). Thus, we reran the EFA with the negatively worded items removed and extracted a four-factor solution.

To achieve a simple factor structure, an iterative EFA procedure was conducted using the following retention criteria: factor loading >.40 and low cross loading (difference in loadings between two factors >.15). The iterative process of item reduction resulted in 13 items that loaded onto four factors. One of the items, "I am free to choose the role I play in relationship," did not fit conceptually under any of the four factors and was eliminated. A final EFA conducted on 12 items resulted in a four-factor solution, which accounted for a total variance of 68.4%, with the factors labeled self-knowledge, relationship quality, self-acceptance, and consideration of others. Consideration of others was related to self-knowledge, relationship quality, and self-acceptance as latent factors (βs = .39, .62, and .49, ps < .001). Self-knowledge was related to relationship quality and self-acceptance (βs = .42 and .33, ps < .01). Relationship quality was related to self-acceptance (β = .68, p < .001). The factor loadings appear in Table 1.

Step 2: Cross-validation.

CFA. To cross-validate the four-factor model of the COMPO, we conducted a CFA on a separate validation sample using the same four-factor configuration in Mplus 8. In addition, to develop a scale where a single scale score of COMPO can be used to assess therapy outcomes, we tested how well a bifactor model fit the data and compared its fit indices with those of the four-factor model. We configured our bifactor model in such a way that a general factor accounted for variance in all COMPO items and that the four domain-specific COMPO factors accounted for variance in respective subsets of COMPO items (Reise, 2012). We evaluated model fit using the following fit indices and criteria: comparative fix index (CFI; > .95 for good fit; .92 to .94 for adequate fit), standardized root mean square residual (SRMR; < .08 for acceptable fit), and root mean square error of approximation (RMSEA; < .06 for acceptable fit; Hu & Bentler, 1999). We also used these fit indices to compare the four-factor and the bifactor models. In addition, we compared the two models based on Satorra-Bentler scaled chi-square difference test (S-B chi-square test), Bayesian information criterion (BIC), and Akaike information criterion (AIC). A significant result of the S-B chi-square test and smaller BIC and AIC values suggest better model fit (Satorra & Bentler, 2010).

Based on data from the validation sample, the S-B chi-square tests indicated that the bifactor model had a better fit to the data than the four-factor model, S-B $\chi^2(5) = 24.4$, $p < .001$. The bifactor model also yielded smaller BIC and AIC values than the four-factor model, which indicated a better model fit and was thus retained. The factor loadings of the bifactor model are shown in Figure 1. Of the 12 items, 10 items loaded onto the general factor above .30. Concurrently, three out of four items loaded onto the relationship quality factor, and the rest of the items loaded onto their respective factors above .20. These findings supported our assumption that the general and domain-specific factors were conceptually meaningful (Reise, 2012).

Measurement invariance. We conducted a multigroup CFA to test for measurement invariance of the bifactor model in fitting the data between the development and validation samples. The goal was to understand to what extent the measurement structure of COMPO was equivalent across the two samples. We evaluated measurement invariance by comparing a series of models with increasing constraints of equivalence across samples, starting from a baseline configural model (no constraints), followed by a metric model (factor loadings constrained) and finally a scalar model (factor loadings and item intercepts constrained). We used the same fit indices of CFA to determine model fit. Further, we determined invariance by assessing a lack of significant changes in the fit indices as model parameters were increasingly constrained: change in CFI (ΔCFI) less than .01, change in RMSEA (ΔRMSEA) less than .015, and change in SRMR (ΔSRMR) less than .03 (Chen, 2007). Since chi-square tests are sensitive to sample size and are known to be less useful for testing measurement invariance as sample size increases, we did not refer to results of S-B chi-square tests to determine invariance (Cheung & Rensvold, 2002).

The fit indices for the invariance tests can be found in online Supplemental Table S4. The configural model of the bifactor structure had an acceptable fit across samples, suggesting that the basic configuration of the measurement structure was comparable across samples. We then compared the configural model to the metric model. The metric model had an adequate fit to the data,

Table 1
Factor Loadings of the 12-Item COMPO From Exploratory Factor Analysis Using the Development Sample (Study 2)

Item	1	2	3	4
38. I know why I keep finding myself in similar situations.	**.93**	.00	−.02	−.01
10. I know why I act the way I do in different relationships.	**.46**	.00	.12	.12
27. I feel supported in life outside of therapy.	.08	**.79**	−.07	−.04
36. Other people care about me.	−.01	**.65**	.06	.19
32. I have meaningful relationships.	−.10	**.66**	.01	.19
37. In my relationships, I find a balance between giving and taking.	.20	**.51**	.09	.00
13. I am a lovable person.	.01	−.02	**.74**	.31
1. I am a person of worth.	−.06	.22	**.65**	−.01
30. I respect myself.	.08	.24	**.59**	−.03
14. People tell me that I am understanding.	−.01	.01	.12	**.72**
5. I try to consider how others feel about my criticisms before I voice them.	−.00	−.01	−.03	**.62**
23. If my loved ones do something I cannot tolerate, I try to understand where they are coming from.	.14	.14	−.11	**.48**

Note. COMPO = Complementary Measure of Psychotherapy Outcome. 1 = self-knowledge; 2 = relationship quality; 3 = self-acceptance; 4 = consideration of others. Coefficients in bold denote items assigned to the corresponding factor.

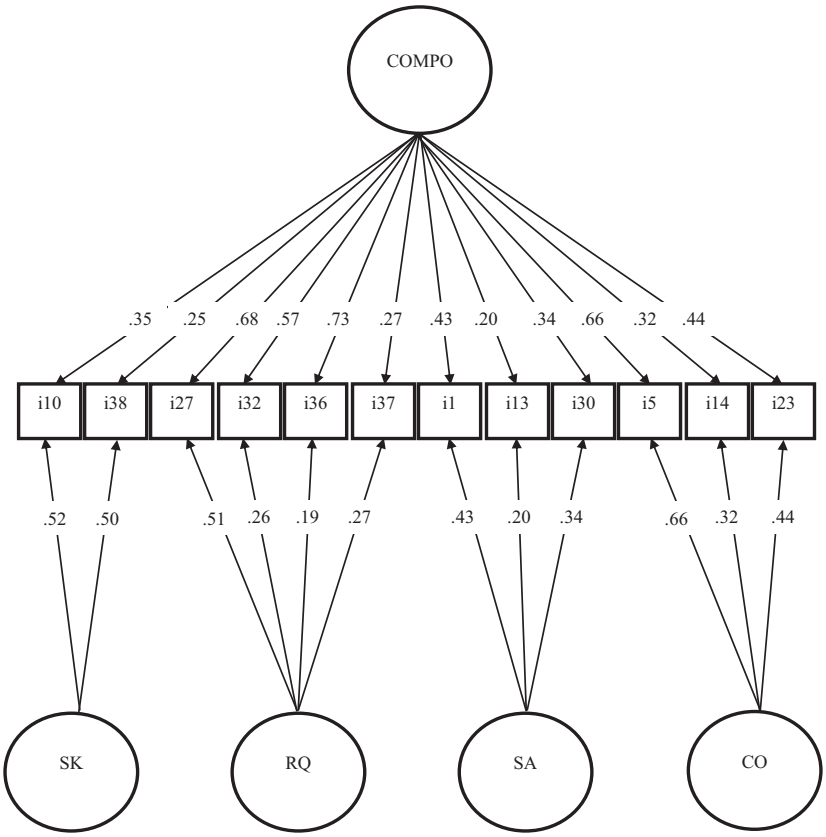

Figure 1. Confirmatory factor analysis of a 12-item bifactor model of Complementary Measure of Psychotherapy Outcome (COMPO) using the validation sample (Study 2). Complementary Measure of Psychotherapy Outcome (COMPO) = general factor; SK = self-knowledge; RQ = relationship quality; SA = self-acceptance; CO = consideration of others.

and we did not observe any significant decrement in fit (ΔCFI = .003, ΔRMSEA = .004, ΔSRMR = .013), suggesting that the factor loadings were comparable across samples. Next, we compared the metric model to the scalar model. The scalar model had an acceptable fit to the data, but the changes in CFI indicated significant decrement in fit (ΔCFI = .016, ΔRMSEA = .004, ΔSRMR = .015), suggesting that the intercepts of certain items may not be invariant across samples. We examined modification indices and found that the constrained intercepts of items 14 and 30 across the two samples contributed to the decrement in fit. We modified the scalar model such that the intercepts of items 14 and 30 were not constrained to be equivalent across the two samples and compared that against the metric model. The modified scalar model had an acceptable fit to the data, and the changes in fit indices indicated no significant decrement in fit (ΔCFI = .008, ΔRMSEA = .001, ΔSRMR = .013). In summary, our results indicated measurement invariance across the two samples, except for the intercepts of items 14 and 30.

Step 3: Validity and test-retest reliability. Finally, we examined the convergent validity, criterion validity, and test–retest reliability of the 12-item COMPO. In terms of convergent validity, we hypothesized that self-acceptance, consideration of others, relationship quality, and self-knowledge would be positively associated with measures of self-esteem, empathy, perceived social support, and insight, respectively. Given the similarity of construct used to correlate with each subscale, the effect sizes are expected to be large ($r \geq .5$; Abma, Rovers, & van der Wees, 2016). In terms of criterion validity, we hypothesized that the COMPO-12 total scale would be positively associated with life satisfaction and negatively associated with depressive symptoms, with medium to large effect sizes ($r \geq .3$). Finally, we investigated the 2-week test–retest reliability of the COMPO-12 and expected a large correlation ($r \geq .5$) between measurements given that the scores are not expected to change over a short period of time without intervention.

Table 2 shows the means, standard deviations, and Cronbach's alphas for the COMPO subscales and total scale for the development and validation samples. Internal consistency for the total scale was good ($\alpha = .84$). Cronbach's alphas for the subscales were less satisfactory ($.51 < \alpha < .81$), but understandably so given that each subscale had only between two and four items (Saucier & Goldberg, 2002).

Convergent validity. In an expected fashion, self-acceptance correlated with RSE, consideration of others with IRI-PT, and relationship quality with MSPSS, all with large effect sizes ($r \geq .5$; see online Supplemental Table S5). Self-knowledge correlated moderately (.33 and .37) with SRIS-IN. These results show that three

Table 2
Means, Standard Deviations, and Internal Consistency for COMPO-12 Subscales and Total Scale in Three Samples

Subscales and total scale	Number of items	Study 2: Development sample (n = 264)			Study 2: Validation sample (n = 571)			Study 3: Clients at pretherapy (n = 28)		
		M	*SD*	α	*M*	*SD*	α	*M*	*SD*	α
1. Self-acceptance	3	4.94	1.38	.81	5.38	1.28	.79	5.12	1.19	.85
2. Consideration of others	3	5.40	1.14	.66	5.28	1.10	.63	5.32	1.19	.77
3. Relationship quality	4	4.67	1.32	.80	4.94	1.27	.78	5.13	1.37	.87
4. Self-knowledge	2	4.63	1.39	.63	4.73	1.28	.51	4.36	1.39	.61
5. Total COMPO-12	12	4.95	0.99	.84	5.18	0.96	.84	5.05	0.98	.87

Note. COMPO = Complementary Measure of Psychotherapy Outcome.

COMPO subscales have convergent validity with existing measures of similar constructs, with less robust evidence for self-knowledge.

Criterion validity. Across the two samples, the COMPO-12 total score correlated positively with the SWLS with large effect sizes and negatively with the TOP-DEPRS with medium effect sizes (see online Supplemental Table S5). This suggests that people who score highly on the COMPO experience higher life satisfaction and fewer depressive symptoms.

Test-retest reliability. To assess test–retest reliability, participants from the development and validation samples were invited via e-mail to complete the COMPO a second time 2 weeks after initial completion. At Time 2, 125 out of 264 participants (47.3%) in the development sample and 242 out of 571 participants (42.4%) in the validation sample completed the COMPO. Table 3 shows the means, standard deviations, and correlations between Time 1 and Time 2 for the COMPO subscales and total scale. All correlations ($.49 < r < .83$) were significant at $p < .001$, suggesting good short-term test–retest reliability of the COMPO subscales and total scale.

Study 3: Sensitivity to Change in Psychotherapy

In Study 3, a sample of psychotherapy clients was recruited to examine the psychometric properties of the COMPO-12 over the course of psychotherapy. In particular, we had the following research questions: (a) How do COMPO-12 scores change from pre- to posttherapy? (b) Can scores on the COMPO-12 distinguish clients who achieved clinically significant change compared to those who did not change, as defined by established criterion for change? (c) How do COMPO-12 scores correlate with symptoms? (d) How do COMPO-12 scores correlate with psychotherapy process variables, such as working alliance?

Method

Participants.

Clients. Thirty-one (20 female; age $M = 34.7$, $SD = 9.13$) clients who sought psychotherapy for depression at a psychology department clinic were included. Inclusion criteria were (a) a primary diagnosis of major depressive disorder as indicated by the Mini-International Neuropsychiatric Interview Version 5.0 (Sheehan et al., 1998), (b) a score of 14 or more on the 17-item clinician-administered semistructured interview version of the Hamilton Rating Scale for Depression (Hamilton, 1960), and (c) age 18–67 years. Exclusion criteria were active suicidality, substance abuse or dependence, current or past bipolar disorder, presence of psychotic features, past severe head injury, pending legal proceedings, and current pregnancy or a medical condition warranting hormonal treatment. Of the 31 clients, one had a medication change during therapy, and two did not complete the pretherapy COMPO; they were excluded from the analysis.

Therapists. Nine (four female) therapists provided psychotherapy to the 28 clients in the included sample. Each therapist saw between one and six clients ($M = 3.00$, $SD = 1.76$). The therapists were master's or doctoral-level trainees with 2–6 years of clinical experience. Clients were assigned to therapists based on availability and therapist caseload.

Table 3
Means, Standard Deviations, and Bivariate Correlations of COMPO Scores Between Two Time Points Across Three Samples

Subscales and total scale	Development sample (n = 125)					Validation sample (n = 242)					Clients in psychotherapy (n = 28)					
	Time 1		2 weeks later			Time 1		2 weeks later			Pretherapy		Posttherapy			
	M	*SD*	*M*	*SD*	*r*	*M*	*SD*	*M*	*SD*	*r*	*M*	*SD*	*M*	*SD*	*r*	*t*
Self-acceptance	4.82	1.42	4.78	1.47	.74	5.42	1.30	5.39	1.31	.83	5.12	1.19	5.56	1.08	.57	2.22*
Consideration of others	5.35	1.10	5.31	1.04	.61	5.37	1.09	5.28	1.05	.63	5.32	1.19	5.33	1.25	.63	0.06
Relationship quality	4.77	1.38	4.69	1.47	.80	5.09	1.24	5.05	1.19	.80	5.13	1.37	5.56	1.06	.63	2.10*
Self-knowledge	4.69	1.47	4.74	1.27	.49	4.76	1.24	4.77	1.19	.49	4.36	1.39	5.04	1.15	.51	2.81**
Total COMPO-12	4.94	1.02	4.88	1.02	.83	5.26	0.93	5.14	.94	.79	5.05	0.98	5.42	0.92	.61	2.33*

Note. COMPO = Complementary Measure of Psychotherapy Outcome. t = paired t test between pre- and posttherapy.
* $p < .05$. ** $p < .01$.

Therapy. Clients received supportive-expressive therapy, a 16-session short-term psychodynamic psychotherapy (Luborsky, 1984). It includes supportive elements, such as enhancing the therapeutic alliance and coping aspects, and expressive elements designed to work on the client's repetitive maladaptive relationship patterns (i.e., core conflictual relationship theme). All therapy sessions were videotaped as part of supervision. Each therapist received an hour of individual supervision weekly from supervisors who were experts in psychodynamic psychotherapy. Supervision focused on the review of videotaped case material and appropriate use of supportive-expressive interventions. This study was approved by the university institutional review board.

Measures and procedure.

Complementary Measure of Psychotherapy Outcome (COMPO). The newly developed 12-item version of COMPO was administered as part of the evaluation of its psychometric properties. The means, standard deviations, and internal consistency (Cronbach's alphas) for the COMPO-12 subscales and total scale at pretherapy are shown on Table 2.

Depression. The Beck Depression Inventory-II (BDI-II; Beck, Steer, & Brown, 1996) is a 21-item self-report measure of depression symptoms. Participants rate between 0 and 3 on statements evaluating their level of depression. The BDI-II was administered before and after therapy, and the total score (range = 0–63) was used, with a higher score indicating more severe depression. Cronbach's alpha was .79 at pretherapy in this sample.

Psychological functioning. The Outcome Questionnaire 45.2 (OQ-45.2; Lambert et al., 2004) is a 45-item self-report measure of psychological functioning, including interpersonal relationships, symptom distress, and social role. Participants rate between 0 = *never* and 4 = *almost always* on items depicting level of functioning across these domains. The OQ-45.2 was administered before and after therapy, and the total score (range = 0–180) was used, with a higher score indicating poorer functioning. Cronbach's alpha was .92 at pretherapy in this sample.

Working alliance. The Working Alliance Inventory—Short Revised (WAI-SR; Hatcher & Gillaspy, 2006) is a 12-item measure. Participants rate from 1 = *seldom* to 5 = *always* on items assessing therapist-client agreement on tasks and goals and their emotional bond. The client and therapist versions were administered after every session of therapy. The Cronbach's alphas were .83 and .95 for the client and therapist versions, respectively, at the first administration.

Results

Table 3 shows the mean and standard deviation COMPO-12 scores at pre- and posttherapy. Paired-sample *t* tests indicate that scores on the COMPO subscales and total scale improved over therapy, with the exception of the consideration of others subscale. The COMPO-12 is therefore sensitive to change in psychotherapy on most tested domains.

To investigate whether the COMPO-12 is able to distinguish clients who had different therapy outcomes, we examined clients' pre and posttherapy COMPO scores in terms of clinically significant change according to the OQ-45.2 (Lambert et al., 2004). Table 4 shows the COMPO-12 pre- and posttherapy scores for clients who recovered (i.e., decrease of 14 or more points on the OQ and a total score < 64; $n = 11$) or improved (i.e., decrease of 14 or more points on the OQ but a total score ≥ 64; $n = 5$) and those who did not change from therapy (i.e., total OQ score did not change more than 14 points from pre- to posttherapy; $n = 12$). As evident from Table 4, the improvement in COMPO-12 subscale and total scale scores in the full client sample was driven by clients who recovered or improved according to the OQ criterion. Those who did not change on the OQ did not change on the COMPO. Moreover, posttherapy COMPO-12 total score was significantly higher for those who recovered or improved than those who did not change, $t(26) = 3.21$, $p = .003$, $d = 1.20$. Outcome groups also differed significantly on self-acceptance, $t(26) = 3.16$, $p = .004$, $d = 1.18$, consideration of others, $t(26) = 2.44$, $p = .022$, $d = 0.93$, and self-knowledge, $t(26) = 2.51$, $p = .019$, $d = 0.92$, and marginally on relationship quality, $t(26) = 1.99$, $p = .057$, $d = 0.75$, at posttherapy. The effect sizes show that the between-groups differences are moderate to large across all COMPO subscales and the total scale.

Correlation with symptom and functioning measures. The COMPO-12 total scale correlated with the BDI-II (pretherapy: $r = -.44$, $p = .02$; posttherapy: $r = -.47$, $p = .01$) and OQ-45.2 total scale (pretherapy: $r = -.52$, $p = .005$; posttherapy: $r = -.50$, $p = .006$). In terms of COMPO-12 subscales, self-acceptance correlated with BDI-II (pretherapy: $r = -.56$, $p = .002$; posttherapy: $r = -.58$, $p = .001$), and self-acceptance (pretherapy: $r = -.61$, $p = .001$; posttherapy: $r = -.66$, $p < .001$) and relationship quality (pretherapy: $r = -.49$, $p = .008$; posttherapy: $r = -.41$, $p = .03$) correlated with the OQ-45.2 total scale.

Table 4
Pre- and Posttherapy COMPO-12 Subscale and Total Scale Scores by Psychotherapy Outcome Groups According to OQ-45.2

	Recovered/improved ($n = 16$)					No change ($n = 12$)				
	Pretherapy		Posttherapy			Pretherapy		Posttherapy		
Subscales and total scale	*M*	*SD*	*M*	*SD*	*t*	*M*	*SD*	*M*	*SD*	*t*
Self-acceptance	5.40	1.19	6.04	0.83	3.73**	4.75	1.13	4.92	1.06	0.42
Consideration of others	5.85	0.87	5.79	1.13	−0.25	4.61	1.23	4.72	1.18	0.34
Relationship quality	5.03	1.24	5.89	0.93	3.66**	5.27	1.56	5.13	1.11	−0.50
Self-knowledge	4.59	1.47	5.47	0.81	2.78*	4.04	1.27	4.46	1.32	1.10
Total COMPO-12	5.26	0.94	5.83	0.67	3.55**	4.77	1.00	4.86	0.93	0.31

Note. COMPO = Complementary Measure of Psychotherapy Outcome; OQ-45.2 = Outcome Questionnaire 45.2. *t* = paired *t* test between pre- and posttherapy.
* $p < .05$. ** $p < .01$.

Correlation with working alliance. We first calculated the mean client- and therapist-rated WAI-SR scores across sessions for each client. Mean working alliance scores were not significantly associated with COMPO-12 at posttherapy (client-rated WAI-SR: $r = .22$, $p = .25$; therapist-rated WAI-SR: $r = .08$, $p = .67$). As a point of reference, mean working alliance also was not significantly correlated with posttherapy BDI-II (client-rated WAI-SR: $r = -.22$, $p = .26$; therapist-rated WAI-SR: $r = .03$, $p = .86$) or OQ-45.2 (client-rated WAI-SR: $r = -.21$, $p = .27$; therapist-rated WAI-SR: $r = -.01$, $p = .96$). Nevertheless, the comparable effect sizes (i.e., $r = .2$) indicate a modest relationship between client-rated working alliance and the three outcome variables, and the lack of statistical significance may be attributed to low power and small sample size.

General Discussion

In this study, we report on the development and validation of the COMPO, which attempts to assess different aspects of psychological functioning beyond symptoms. Seven theoretical domains and 42 items were initially created from a review of the literature of what patients gain from psychotherapy and from experienced clinicians' feedback about relevant outcomes of psychotherapy. Although the seven theoretical domains could not be found, EFA and item reduction resulted in a 12-item measure with a four-factor solution: self-knowledge, self-acceptance, consideration of others, and relationship quality. A bifactor model with a general factor and four domain-specific factors, found to be superior to an oblique four-factor solution, was replicated between the development and validation samples. The factor subscales also had convergent validity with measures assessing similar constructs. In addition, the 12-item COMPO exhibits short-term test–retest reliability, and it is sensitive to change from pre- to posttherapy. Posttherapy COMPO scores also distinguished clients who had clinically significant change and those who did not.

The inability to find seven factors from the original COMPO measure is perhaps not surprising. In particular, some of the theoretical domains of psychological functioning are overlapping. For example, people who are nonreactive (i.e., scoring high on the subscale of equanimity) are likely able to do so because they can consider the perspectives of others (empathy) and are sufficiently secure about themselves (self-acceptance). On the other hand, people who have difficulty in managing reciprocity in relationships (balancing the needs of self and others) likely have poor relationship quality with others (relationship/support). Hence, items that belong to more than one domain would load onto multiple factors.

The present factor analyses of the COMPO led to a cross-validated bifactor solution with a general factor and four-specific domain factors. Interestingly, three of the retained specific factors (i.e., self-acceptance, self-knowledge, and relationship quality) were areas that clients noted to experience change through therapy (e.g., Binder et al., 2010), whereas the dropped factors (equanimity, freedom, balancing the needs of self and others) were domains added by clinicians because of their deemed relevance for outcome. This finding illustrates the importance of prioritizing the client's perspective when constructing a client self-report measure.

The good model fit found in the bifactor model suggests that clinicians and researchers can use the COMPO total score as a global measure, as well as the COMPO subscale scores to understand different aspects of clients. For instance, two of the subscales, self-acceptance and self-knowledge, pertain to areas of a client's self, whereas the other two subscales, relationship quality and consideration of others, pertain to a client's interpersonal relationship. The assessment of self and interpersonal relationship complements the assessment of symptoms in psychotherapy because disturbances in the self, interpersonal relationships, and psychological functioning are all reasons why people seek psychotherapy (Horowitz, 1979) and what patients report to improve with therapy (e.g., Connolly & Strupp, 1996).

The four factors of COMPO bear a striking resemblance of Criterion A of the Alternative Model of Personality Disorder (AMPD) in Section III of the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 2013). In particular, Criterion A refers to impairments in the sense of self and interpersonal relatedness. Although not perfect in correspondence, specific areas of concern under Criterion A match the COMPO factors well. For instance, with respect to the self, Criterion A looks at identity and self-direction, which map onto self-knowledge and self-acceptance, respectively. With respect to interpersonal relatedness, Criterion A looks at empathy and intimacy, which map onto consideration of others and relationship quality, respectively. Given the recent interest in moving from a categorical to a dimensional classification of personality disorders and the active research surrounding the AMPD (e.g., Widiger et al., 2019), the COMPO may be a useful assessment tool for this emerging framework.

Although the COMPO displays configural and metric invariance across the development and validation samples, the intercepts of two items were not equivalent between them. Caution is needed when researchers are trying to compare the mean scores of COMPO across populations. When designing this study, we wanted to replicate the factor structure obtained from a psychotherapy client sample onto a broader, "potential client" sample. As such, the validation sample consisted of adults in the community with and without experience in psychotherapy. Perhaps the variability in psychotherapy experience contributed to the lack of scalar invariance across all items. Future studies may examine measure invariance in another client sample to confirm the replicability of the COMPO. Nevertheless, because the main purpose of the COMPO is to track changes within a person during therapy, this study has gathered important evidence for test–retest reliability and sensitivity to change and provided initial support for such a purpose.

In Study 3, the size of association between working alliance and the COMPO was similar to how working alliance related to other outcome measures (i.e., OQ and BDI). The modest association between working alliance and posttherapy COMPO therefore replicates the alliance-outcome relationship consistently found in the psychotherapy research literature (Flückiger, Del Re, Wampold, & Horvath, 2018). In addition, the association between COMPO and depressive symptoms is approximately $-.5$ across the three study samples. This magnitude of association is comparable to the association between mental health and mental illness ($-.53$; Keyes, 2007). Taken together, while the COMPO may function similarly to existing outcome measures in relation to psychotherapy process (i.e., working alliance), it assesses nonsymptomatic changes in

psychological functioning and complements symptom-based outcome measures.

In this article, we developed and validated a brief, self-report measure of psychological functioning intended to complement current psychotherapy outcome measures that focus on symptomatic change. Strengths in the development of this measure include consideration of what patients report to change in psychotherapy, what therapists thought to be relevant and important areas of change, and the inclusion of therapist perspectives from different theoretical orientations. In addition, we used two large, independent samples to conduct factor analyses and examined the new measure's psychometric properties using well-established measures. We also administered the new measure to a psychotherapy client sample at pre- and posttherapy to investigate the measure's sensitivity to change and its ability to distinguish patients who experienced clinically significant change from those who did not. Nevertheless, a number of limitations need to be acknowledged.

First, the factor structure of COMPO needs to be tested in other settings and cultural contexts. In particular, the doctoral students and therapists who reviewed the items were rather homogenous in race, and the measure's cross-cultural relevance needs further investigation. Next, there are psychometric concerns about the two-item self-knowledge subscale, namely its less satisfactory reliability and factor underidentification when modeling item-to-factor measurement structure with CFA. However, given the subscale's promise of clinical utility (i.e., scores improved over therapy, and posttherapy scores differed between clients who had clinical improvement vs. those who did not), future studies should consider adding an item to this subscale to enhance its psychometric properties.

The COMPO could be administered over the course of therapy, instead of just pre- and posttherapy, to examine how clients change in the assessed domains over time. In particular, to further understand how clients change in symptoms relative to different aspects of psychological functioning, the COMPO could be administered simultaneously with traditional symptom-based measures over time. The different rates of change will inform our understanding of the sequence and mechanisms of change. For instance, if symptom change occurs before changes in the self and in relationships, intervention efforts should target relieving symptoms early on in treatment. On the other hand, if changes in the self and/or relationship occur first, clinicians should focus on a more holistic understanding of clients and address these problem areas first. Of course, with a large enough sample, we may identify client characteristics that are associated with different patterns of change. These characteristics may then be used to tailor more individualized treatment based on the expected patterns of change.

References

Abma, I. L., Rovers, M., & van der Wees, P. J. (2016). Appraising convergent validity of patient-reported outcome measures in systematic reviews: Constructing hypotheses and interpreting outcomes. *BMC Research Notes, 9,* 226. http://dx.doi.org/10.1186/s13104-016-2034-2

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

APA Presidential Task Force on Evidence-Based Practice. (2006). Evidence-based practice in psychology. *American Psychologist, 61,* 271–285. http://dx.doi.org/10.1037/0003-066X.61.4.271

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the Beck Depression Inventory-II*. New York, NY: Psychological Corporation.

Binder, P. E., Holgersen, H., & Nielsen, G. H. (2010). What is a "good outcome" in psychotherapy? A qualitative exploration of former patients' point of view. *Psychotherapy Research, 20,* 285–294. http://dx.doi.org/10.1080/10503300903376338

Buunk, B. P., & Schaufeli, W. B. (1999). Reciprocity in interpersonal relationships: An evolutionary perspective on its importance for health and well-being. *European Review of Social Psychology, 10,* 259–291. http://dx.doi.org/10.1080/14792779943000080

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14,* 464–504. http://dx.doi.org/10.1080/10705510701301834

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9,* 233–255. http://dx.doi.org/10.1207/S15328007SEM0902_5

Chui, H., Palma, B., Jackson, J. L., & Hill, C. E. (2020). Therapist-client agreement on helpful and wished-for experiences in psychotherapy: Associations with outcome. *Journal of Counseling Psychology, 67,* 349–360. http://dx.doi.org/10.1037/cou0000393

Connolly, M. B., & Strupp, H. H. (1996). Cluster analysis of patient reported psychotherapy outcomes. *Psychotherapy Research, 6,* 30–42. http://dx.doi.org/10.1080/10503309612331331558

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *Catalog of Selected Documents in Psychology, 10,* 85.

DeWitt, K. N., Hartley, D. E., Rosenberg, S. E., Zilberg, N. J., & Wallerstein, R. S. (1991). Scales of psychological capacities: Development of an assessment approach. *Psychoanalysis & Contemporary Thought, 14,* 343–361.

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction With Life Scale. *Journal of Personality Assessment, 49,* 71–75. http://dx.doi.org/10.1207/s15327752jpa4901_13

Farmer, R. F., & Chapman, A. L. (2008). *Behavioral interventions in cognitive behavior therapy: Practical guidance for putting theory into action*. Washington, DC: American Psychological Association. http://dx.doi.org/10.1037/11664-000

Flückiger, C., Del Re, A. C., Wampold, B. E., & Horvath, A. O. (2018). The alliance in adult psychotherapy: A meta-analytic synthesis. *Psychotherapy, 55,* 316–340. http://dx.doi.org/10.1037/pst0000172

Frisch, M. B. (1998). Quality of life therapy and assessment in health care. *Clinical Psychology: Science and Practice, 5,* 19–40. http://dx.doi.org/10.1111/j.1468-2850.1998.tb00132.x

Göstas, M. W., Wiberg, B., & Kjellin, L. (2012). Increased participation in the life context: A qualitative study of clients' experiences of problems and changes after psychotherapy. *European Journal of Psychotherapy and Counselling, 14,* 365–380. http://dx.doi.org/10.1080/13642537.2012.734498

Grant, A. M., Franklin, J., & Langford, P. (2002). The Self-Reflection and Insight Scale: A new measure of private self-consciousness. *Social Behavior and Personality, 30,* 821–835. http://dx.doi.org/10.2224/sbp.2002.30.8.821

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery & Psychiatry, 23,* 56–62. http://dx.doi.org/10.1136/jnnp.23.1.56

Hasler, G., Moergeli, H., & Schnyder, U. (2004). Outcome of psychiatric treatment: What is relevant for our patients? *Comprehensive Psychiatry, 45,* 199–205. http://dx.doi.org/10.1016/j.comppsych.2004.02.001

Hatcher, R., & Gillaspy, J. (2006). Development and validation of a revised short version of the Working Alliance Inventory. *Psychotherapy Research, 16,* 12–25. http://dx.doi.org/10.1080/10503300500352500

Hill, C. E., Chui, H., & Baumann, E. (2013). Revisiting and reenvisioning the outcome problem in psychotherapy: An argument to include individualized and qualitative measurement. *Psychotherapy, 50,* 68–76. http://dx.doi.org/10.1037/a0030571

Horney, K. (1951). *Neurosis and human growth*. London, England: Routledge.

Horowitz, L. M. (1979). On the cognitive structure of interpersonal problems treated in psychotherapy. *Journal of Consulting and Clinical Psychology, 47,* 5–15. http://dx.doi.org/10.1037/0022-006X.47.1.5

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1–55. http://dx.doi.org/10.1080/10705519909540118

Keyes, C. L. M. (2005). Mental illness and/or mental health? Investigating axioms of the complete state model of health. *Journal of Consulting and Clinical Psychology, 73,* 539–548. http://dx.doi.org/10.1037/0022-006X.73.3.539

Keyes, C. L. M. (2007). Promoting and protecting mental health as flourishing: A complementary strategy for improving national mental health. *American Psychologist, 62,* 95–108. http://dx.doi.org/10.1037/0003-066X.62.2.95

Klein, M. J., & Elliott, R. (2006). Client accounts of personal change in process-experiential psychotherapy: A methodologically pluralistic approach. *Psychotherapy Research, 16,* 91–105. http://dx.doi.org/10.1080/10503300500090993

Kraus, D. R., Seligman, D. A., & Jordan, J. R. (2005). Validation of a behavioral health treatment outcome and assessment tool designed for naturalistic settings: The Treatment Outcome Package. *Journal of Clinical Psychology, 61,* 285–314. http://dx.doi.org/10.1002/jclp.20084

Lambert, M. J., Morton, J. J., Hatfield, D., Harmon, C., Hamilton, S., Reid, R. C., . . . Burlingame, G. M. (2004). *Administration and scoring manual for the Outcome Questionnaire (OQ45.2)*. Orem, UT: American Professional Credentialing Service.

Levitt, H., Butler, M., & Hill, T. (2006). What clients find helpful in psychotherapy: Developing principles for facilitating moment-to-moment change. *Journal of Counseling Psychology, 53,* 314–324. http://dx.doi.org/10.1037/0022-0167.53.3.314

Lu, L. (2002). A preliminary study on the concept of health among the Chinese. *Counselling Psychology Quarterly, 15,* 179–189. http://dx.doi.org/10.1080/09515070110104051

Luborsky, L. (1984). *Principles of psychoanalytic psychotherapy: A manual for supportive-expressive treatment*. New York, NY: Basic Books.

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments & Computers, 32,* 396–402. http://dx.doi.org/10.3758/BF03200807

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research, 47,* 667–696. http://dx.doi.org/10.1080/00273171.2012.715555

Rogers, C. R. (1961). *On becoming a person: A therapist's view of psychotherapy*. London, England: Constable.

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press. http://dx.doi.org/10.1515/9781400876136

Ryff, C. D. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of Personality and Social Psychology, 57,* 1069–1081. http://dx.doi.org/10.1037/0022-3514.57.6.1069

Satorra, A., & Bentler, P. M. (2010). Ensuring positiveness of the scaled difference chi-square test statistic. *Psychometrika, 75,* 243–248. http://dx.doi.org/10.1007/s11336-009-9135-y

Saucier, G., & Goldberg, L. R. (2002). Assessing the Big Five: Applications of ten psychometric criteria to the development of marker scales. In B. De Raad & M. Perugini (Eds.), *Big Five assessment* (pp. 29–58). Boston, MA: Hogrefe and Huber.

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., . . . Dunbar, G. C. (1998). The Mini-International Neuropsychiatric Interview (M. I. N. I.): The development and validation of a structured diagnostic psychiatric interview for *DSM–IV* and ICD-10. *The Journal of Clinical Psychiatry, 59*(Suppl. 20), 22–33.

Strupp, H. H., & Hadley, S. W. (1979). Specific vs nonspecific factors in psychotherapy. A controlled study of outcome. *Archives of General Psychiatry, 36,* 1125–1136. http://dx.doi.org/10.1001/archpsyc.1979.01780100095009

Summers, R. F., & Barber, J. P. (2009). *Psychodynamic therapy: A guide to evidence-based practice*. New York, NY: Guilford Press.

Waldron, S., Moscovitz, S., Lundin, J., Helm, F. L., Jemerin, J., & Gorman, B. (2011). Evaluating the outcomes of psychotherapies: The Personality Health Index. *Psychoanalytic Psychology, 28,* 363–388. http://dx.doi.org/10.1037/a0024559

Westen, D., & Shedler, J. (1999). Revising and assessing Axis II, Part I: Developing a clinically and empirically valid assessment method. *The American Journal of Psychiatry, 156,* 258–272.

Widiger, T. A., Bach, B., Chmielewski, M., Clark, L. A., DeYoung, C., Hopwood, C. J., . . . Thomas, K. M. (2019). Criterion A of the AMPD in HiTOP. *Journal of Personality Assessment, 101,* 345–355. http://dx.doi.org/10.1080/00223891.2018.1465431

Zhang, X., Noor, R., & Savalei, V. (2016). Examining the effect of reverse worded items on the factor structure of the need for cognition scale. *PLoS ONE, 11*(6), e0157795. http://dx.doi.org/10.1371/journal.pone.0157795

Zimet, G. D., Dahlem, N. W., Zimet, S. G., & Farley, G. K. (1988). The Multidimensional Scale of Perceived Social Support. *Journal of Personality Assessment, 52,* 30–41. http://dx.doi.org/10.1207/s15327752jpa5201_2

Received October 25, 2019
Revision received August 10, 2020
Accepted August 22, 2020 ■