Psychotherapy Research

ISSN: 1050-3307 (Print) 1468-4381 (Online) Journal homepage: www.tandfonline.com/journals/tpsr20

Assessing the psychotherapist's affective reactions toward their patient: validation of the *Clinician Affective REsponse (CARE) scales*

Alberto Stefana, Eduard Vieta, Paolo Fusar-Poli & Eric A. Youngstrom

To cite this article: Alberto Stefana, Eduard Vieta, Paolo Fusar-Poli & Eric A. Youngstrom (2026) Assessing the psychotherapist's affective reactions toward their patient: validation of the *Clinician Affective REsponse (CARE) scales*, Psychotherapy Research, 36:2, 220-236, DOI: 10.1080/10503307.2025.2465432

To link to this article: https://doi.org/10.1080/10503307.2025.2465432

Published online: 25 Feb 2025.

Psychotherapy Research, 2026
Vol. 36, No. 2, 220–236, https://doi.org/10.1080/10503307.2025.2465432

RESEARCH ARTICLE

Assessing the psychotherapist's affective reactions toward their patient: validation of the *Clinician Affective REsponse (CARE) scales*

ALBERTO STEFANA [1,2], EDUARD VIETA [3], PAOLO FUSAR-POLI [2,4,5], & ERIC A. YOUNGSTROM [6,7,8]

[1]*Department of Psychiatry and Behavioral Health, Ohio State University, Columbus, OH, USA;* [2]*Department of Brain and Behavioral Sciences, University of Pavia, Pavia, Italy;* [3]*Bipolar and Depressive Disorders Unit, Hospital Clinic, IDIBAPS, CIBERSAM, University of Barcelona, Barcelona, Catalonia, Spain;* [4]*OASIS Service, South London and Maudsley NHS Foundation Trust, London, UK;* [5]*Early Psychosis: Interventions and Clinical-detection (EPIC) Lab, Department of Psychosis Studies, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, United Kingdom;* [6]*Institute for Mental and Behavioral Health Research, Nationwide Children's Hospital, Division of Child and Family Psychiatry, The Ohio State University, Columbus, OH, USA;* [7]*Department of Psychology and Neuroscience, University of North Carolina at Chapel Hill, USA &* [8]*Helping Give Away Psychological Science, Chapel Hill, NC, USA*

(Received 26 August 2024; revised 5 February 2025; accepted 6 February 2025)

Abstract

Background: This study aims to evaluate the factor structure, reliability, and validity of the Clinician Affective REsponse (CARE) scales, a 15-item self-report measure designed for practical use in psychotherapy settings. **Methods:** Validation data were gathered from 151 mental health clinicians. These clinicians completed the CARE scales alongside measures capturing sociodemographic and professional details, patient demographics and clinical details, therapeutic intervention characteristics, therapeutic relationship elements, and session outcomes. **Results:** The CARE scales had a three-factor structure: positive engagement ($k = 5$, $\omega = .78$), enmeshed ($k = 5$, $\omega = .72$), and stuck ($k = 5$, $\omega = .71$). Confirmatory factor analysis (CFA) yielded the following fit indices for the three-factor model: $\chi^2_{(87)} = 120.41$, CFI = .94; TLI = .93, RMSEA = .05, and SRMR = .08. Multigroup CFA (which pooled two samples for a total of 607 subjects) showed that the CARE scales were invariant across remote and in-person session formats. The scales showed meaningful correlations with concurrent measures of working alliance, real relationship, countertransference, patient's experience of the therapeutic relationship, and session outcome. **Discussion:** The CARE scales are a valuable instrument in clinical, training, and research contexts, adept at capturing clinicians' session-level affective responses and perceptions of the therapeutic relationship. Quantifying these reactions facilitates statistical analysis and empirical research, while their monitoring can guide therapeutic interventions and inform clinical supervision.

Keywords: therapeutic relationship; emotional reaction; clinician; psychotherapy process research; self-report measure; CARE scale

Clinical or methodological significance of this article: The subjective affective reactions of clinicians toward their patients during individual psychological counseling and psychotherapeutic sessions play a crucial role in influencing treatment outcomes. In this article, we validated the Clinician Affective REsponse (CARE) scales, a reliable and valid self-report measure of clinicians' session-level affective responses. Quantifying these reactions facilitates statistical analysis and empirical research, and monitoring them can guide therapeutic interventions and inform clinical supervision.

Correspondence concerning this article should be addressed to Alberto Stefana Department of Brain and Behavioral Sciences, University of Pavia, Via Forlanini 6, Pavia 27100, Italy. Email: alberto.stefana@gmail.com

Psychotherapy can evoke a broad spectrum of emotions not only in patients but also in therapists (Atzil-Slonim et al., 2018; Kimerling et al., 2000; Stefana et al., 2022). Among the various affective states experienced by a therapist during sessions, those directed towards the patient are particularly significant for two primary reasons. First, these emotions represent a complex and co-constructed phenomenon that involves both members of the therapeutic dyad (Gelso & Hayes, 2007), thus providing information on the emotional states or psychological functioning of the patients. Second, countertransference reactions—where the clinician's unresolved conflicts come into play (Gelso & Hayes, 2007)—have been associated with poorer therapy outcomes if unmanaged, while successful management correlates with better outcomes (Hayes et al., 2018).

Mental health clinicians' awareness of their own emotional states is associated with improved therapy outcomes (Cologon et al., 2017). Furthermore, clinicians' patterns of negative emotional reactions are negatively correlated with the quality of the working alliance (Tanzilli et al., 2018) and are associated with more ruptures and fewer resolutions (Tishby & Wiseman, 2022). On the contrary, positive emotional patterns are associated with a higher quality of the therapeutic alliance (Tanzilli et al., 2018) and predict successful resolution when therapists reinforce these positive patterns. However, they can predict ruptures if used to repair previous negative patterns (Tishby & Wiseman, 2022).

The importance of emotions experienced during sessions is further highlighted by recent research showing that therapists' presession emotions do not predict their empathy or session quality (Chui et al., 2022). This is crucial, as an empathetic and positive therapeutic relationship is essential not only to facilitate change but also to ensure the effective delivery of psychotherapeutic interventions (Hayes & Vinca, 2017; Peluso & Freund, 2018). Therefore, the emotional experience of a therapist in session can be a predictor of certain negative outcomes.

Given the above, it is crucial for mental health clinicians to regularly monitor their own affective reactions towards their patients using brief, validated, and reliable self-report tools. However, existing self-report measures often have significant limitations. Some lack full psychometric validation (e.g., Brody & Farber, 1996)—a critical issue that usually justifies the avoidance of such tools. Others focus solely on therapists' emotional responses without exploring underlying dynamics or a specific relational situation (e.g., Breivik et al., 2020; Holmqvist & Armelius, 1994); for example, they explore "When I talk with this patient I feel enthusiastic" instead of "How much did I feel enthusiastic about working together with this patient." Additionally, some tools are lengthy and assess reactions over a series of sessions rather than a single session (e.g., Betan & Westen, 2009). This latter point is particularly important because, in counseling psychology and psychotherapy, scale length remains a crucial concern due to the complexity of therapeutic processes, which frequently demand the collection of multiple measures (Wampold & Flückiger, 2023) or the systematic use of a select few (Barkham et al., 2023).

There is a significant need for brief, theoretically, and psychometrically sound self-report tools that assess therapists' reactions toward patients during a single session. Thus, we developed the *Clinician Affective REsponse* (CARE) *scales*.

An Overview of the CARE Scales Development

The CARE scales development and preliminary validation followed best practice recommendations (DeVellis & Thorpe, 2022; Stefana et al., 2025) and are detailed elsewhere (Stefana, Fusar-Poli, Langfus, et al., 2024).

The phenomenon we aim to measure is the pattern of positive and negative emotional, cognitive, and behavioral reactions experienced by a therapist toward their patient during a counseling or psychotherapy session—reactions that stem from the patient-therapist interaction. This construct includes countertransference (i.e., reactions based on the clinician's unresolved conflicts; Gelso & Hayes, 2007; Stefana, 2017), but it is much more comprehensive and implies that not all reactions are countertransferential (Gelso & Kline, 2022).

Central to the CARE scales' design were the hierarchical model of affect (Tellegen et al., 1999a, 1999b), the social and interpersonal dimension of the dominance behavioral system (Johnson et al., 2012), the tripartite model of the psychotherapy relationship (Gelso, 2014), and the contextual model (Wampold, 2015). These last two evidence-based therapeutic process models served as guideposts, ensuring the scale's applicability and relevance in real-world clinical settings.

We employed both inductive and deductive approaches to generate the initial item pool. Existing scales (Betan et al., 2005; Friedman & Gelso, 2000; Gelso et al., 2005; Horvath & Greenberg, 1989; Izard et al., 1993; Latts, 1997; Meehan et al., 2012; Najavits et al., 1995; Pallagrosi et al., 2014) were reviewed, relevant items were selected, and, when necessary, they were adapted to reflect the affective, cognitive, and behavioral responses that psychotherapists might experience during "good/not-so-difficult" or "difficult" sessions. Additionally,

the authors and two experienced psychotherapists—one with a psychological background and another with a psychiatric background—generated new items, emphasizing emotions and affect. The content validity of the initial item pool was reviewed by six experienced clinicians from different theoretical orientations (three cognitive–behavioral, two eclectic, and one psychodynamic).

The initial pool consisted of 116 items, which were administered to 556 licensed clinicians. The statistical analysis began with descriptive statistics to eliminate items that exhibited very low or no variability—specifically, items where fewer than 5% of responses fell within the "somewhat to very much" range were dropped. A parallel analysis on the remaining 38 items suggested a four-factor structure. Thus, exploratory factor analysis (EFA) was conducted by extracting four factors. Items with unique loadings < .40 and those that cross-loaded without a distinct dominating factor were eliminated. After a second round of EFA, 21 items remained, resulting in a three-dimensional model.

The graded response model within item response theory (Samejima, 2010) was applied separately for each scale. Final item selection involved choosing items across the full range of theta levels. Confirmatory factor analysis demonstrated excellent fit indices. Internal reliability was confirmed. Criterion validity analyses showed that the positively engaged scale was weakly negatively related to cluster B personality disorders and positively related to global functioning. The enmeshed and stuck scales were weakly associated with more frequent sessions and longer therapy duration. Lastly, the stuck scale was weakly linked to poorer global functioning and to cluster C personality or trauma-related disorders.

Aim of the Study

This study aimed to more fully validate the CARE scales, a self-report measure designed to assess the patterns of in-session thoughts, feelings, and behaviors of therapists toward their patients.

We analyzed the factor structure, measurement invariance, internal consistency, score precision, and the convergent and divergent criterion validity of the three scales that make up the CARE scales. The criterion validity analysis specifically examined their relationship with a series of measures assessing the therapeutic relationship and session outcomes. We anticipated moderate-to-large positive correlations between the Engaged scale and levels of working alliance (especially the bond dimension), real relationship, patient's positive affect toward the therapist, patient's levels of insight, overall session quality, and session positivity. We also expected moderate-to-large positive correlations between the Enmeshed scale and positive transference and countertransference, with very weak or no association with session outcomes. Lastly, we expected moderate-to-large negative correlations between the Stuck scale and working alliance (especially the task dimension), patients' levels of insight, and session outcomes, as well as positive associations with negative transference and the occurrence of alliance ruptures.

Method

Participants

The sample consisted of 151 mental health clinicians, predominantly women (71%, $n = 108$). The age group represented the most was 60 years and older (57%, $n = 86$), followed by those 50–59 years (23%, $n = 35$). Regarding psychotherapeutic approaches, most were psychoanalytically/ psychodynamically oriented (40%, $n = 61$), with behavior therapy accounting for one-fifth of the sample (19%, $n = 28$). Approximately half of the respondents had more than 20 years of post-license clinical experience (55%, $n = 83$) and spent more than 21 h a week practicing therapy (46%, $n = 70$). Each of the 151 clinicians reported on a single patient. Among the patients, 65% ($n = 98$) were women and 47% ($n = 71$) were between 30 and 49 years old. Most of the patients (75%, $n = 113$) had a diagnosis of psychiatric disorder and were rated as mildly (33%, $n = 50$) or moderately (25%, $n = 37$) ill on the Clinical Global Impression scale. The average Global Assessment of Functioning score was 67.4 ($SD = 12.4$). About half of the patients had been in psychotherapy for more than 2 years (48%, $n = 73$), typically attending one session per month (48%, $n = 72$). Table I presents the demographics and professional characteristics of the therapists, while Table II provides patient demographics, clinical data, and treatment characteristics.

Measures

A comprehensive range of self-report instruments gathered extensive data on the individual characteristics of the clinicians and patients, the dynamics of the therapeutic relationship, and the outcomes of the session.

Demographic and therapy domain

Clinician demographic and professional data form. Clinicians filled out a sociodemographic and professional data form, which captured the information presented in Table I.

Table I. Therapist demographic and professional characteristics ($N = 151$).

Demographic characteristics	% (*n*)
Age (years)	
30–39	5% (8)
40–49	15% (22)
50–59	23% (35)
60 or above	57% (57)
Gender	
Woman	72% (108)
Man	28% (43)
Ethnicity	
White	91% (138)
Other	9% (13)
Professional Characteristics	% (*n*)
Professional background	
Psychologist	36% (54)
Social Worker	21% (31)
Psychiatrist	13% (20)
Other	31% (46)
Psychotherapeutic approach	
Psychoanalysis or psychoanalytic therapy	27% (41)
Behavior therapy	19% (28)
Psychodynamic therapy	13% (20)
Integrative or holistic therapy	9% (13)
Cognitive therapy	7% (11)
Humanistic therapy	7% (11)
Other	18% (27)
Post-licensed experience (years)	
1–6	6% (10)
7–10	9% (14)
11–15	15% (22)
16–20	15% (22)
More than 20	55% (83)
Time spent practicing therapy (hours per week)	
1–5	7% (10)
6–10	11% (17)
11–15	13% (19)
16–20	23% (35)
21 or more	46% (70)

Table II. Patient demographic, clinical, and treatment characteristics ($N = 151$).

Demographic characteristics	% (*n*)
Age (years)	
18–22	6% (9)
23–29	17% (25)
30–39	22% (34)
40–49	24% (37)
50–59	17% (25)
60 or above	14% (21)
Gender	
Woman	64% (98)
Man	32% (48)
Non-binary / third gender	3% (5)
Ethnicity	
White	82% (123)
Asian	5% (8)
Other	13% (20)
Clinical Characteristics	% (*n*)
Presence of any psychiatric mental disorder	75% (113)
Any anxiety disorder	42% (64)
Any trauma- and stressor-related disorders	21% (31)
Any (unipolar) depressive disorder	20% (30)
Any neurodevelopmental disorder	8% (12)
Any personality disorder	6% (9)
Any eating disorder	5% (8)
Any bipolar or related disorder	5% (7)
Any other disorder	6% (9)
CGI	
Normal, not at all ill	21% (32)
Borderline mentally ill	19% (28)
Mildly ill	33% (50)
Moderately ill	25% (37)
Markedly/Severely ill	2% (4)
GAF, *M (SD)*	67.4 (12.4)
Treatment Characteristics	% (*n*)
Therapy length (months)	
0–3	13% (20)
4–6	8% (12)
7–12	13% (19)
13–24	18% (27)
24 or more	48% (73)
Session frequency	
≤1 per month	10% (15)
2–3 per month	30% (46)
1 per week	48% (72)
≥2 per week	12% (18)
Session attendance	
In person	52% (78)
Video call	46% (69)
Telephone call	2% (4)
Location	
Private practice	89% (134)
Private health institution	7% (11)
Other	4% (6)

Patient demographic and clinical data form. Clinicians were asked to review their appointment records, identify the last adult patient (18 years or above) they saw for an individual session, and complete a demographic and clinical data form for that patient, which included the information reported in Table I, as well as the Clinical Global Impressions (CGI) (Busner & Targum, 2007) and Global Assessment of Functioning (GAF) (Jones et al., 1995) scales.

Therapeutic relationship domain

Clinician Affective REsponse (CARE) scales. The CARE scales (Stefana, Fusar-Poli, Langfus, et al., 2024) are a 15-item self-report instrument designed to assess therapists' emotional, cognitive, and behavioral patterns towards their patients during an individual adult psychotherapy session. Responses are recorded on a 3-point Likert scale: 0 = "Not at all," 1 = "A little," and 2 = "Somewhat-to-Very much." The CARE scales do not have a total score but consist of three scales, each containing five items: positive engagement, enmeshed, and stuck. The positive engagement scale includes items

that reflect feelings of appreciating the patient as an individual, contentment in seeing them, comfort and enthusiasm about working with them, and interest in their work. The enmeshed scale contains items that indicate a desire to give love, an urge to act for the patient, concern for their feelings, needs, and wishes, tenderness, and a tendency to protect them, which are notably higher compared to those for other patients. The stuck scale consists of items describing the therapist's difficulty in entering the patient's inner world, frustration in their efforts to help, feelings of incompetence, hopelessness, and annoyance towards the patient. In their development and preliminary validation study, the CARE scales demonstrated McDonald's ω_{total} values of .78 for the positive engagement scale, .75 for the enmeshed scale, and .77 for the stuck subscale.

Working Alliance Inventory–Short Revised (WAI-SR). The therapist version of WAI-SR (Hatcher et al., 2020) is a 12-item measure that evaluates the quality of the working alliance within a therapy session from the clinician's perspective. It includes three subscales, each with four items, that evaluate agreement on therapy tasks and goals and the affective bond between the patient and the clinician. Responses are given on a 7-point Likert scale from 0 ("Not at all") to 5 ("Completely"). The WAI-SR has shown good convergent validity (r = .71) with the Helping Alliance Questionnaire (Munder et al., 2010). In this study, the WAI-SR total scale achieved a Cronbach's α of .88, while the tasks, goals, and bond subscales showed McDonald's ω_{total} values of .84, .79, and .80, respectively.

Real Relationship Inventory–Therapist (RRI-T). The therapist form of the RRI (Gelso et al., 2005) contains 24 items that assess the strength of the real relationship, asking therapists to rate themselves, the client, and their relationship on a five-point scale from 1 ("Strongly disagree") to 5 ("Strongly agree"). It comprises two subscales of 12 items each: realism and genuineness. Realism is the degree to which one realistically perceives the other, while genuineness is the degree of authenticity demonstrated towards the other and the extent to which one is genuinely themselves. Higher scores indicate stronger real relationships. Construct validity is supported by associations with working alliance, negative transference, and session/treatment outcomes (e.g., Bhatia & Gelso, 2018; Gelso et al., 2005). In our study, the internal consistency McDonald's ω_{total} values were .83 for realism and .84 for genuineness, while Cronbach's α for the total scale was .84.

Inventory of Countertransference Behavior (ICB). The ICB was originally developed as an observer-rated measure of countertransference behavior (Friedman & Gelso, 2000) and subsequently validated as a self-report measure (Bhatia & Gelso, 2018). It consists of 21 items rated on a five-point scale from 1 ("To little or no extent") to 5 ("To a great extent"). Higher scores indicate greater countertransference behavior in the session. The ICB provides scores on positive, negative, and overall countertransference. ICB has been validated through various studies showing its reliability and theoretical relationships with countertransference management, therapist attachment, working alliance, and treatment outcomes (e.g., Bhatia & Gelso, 2018; Mohr et al., 2005). In this study, the ICB showed McDonald's ω_{total} values of .66 for positive countertransference and .68 for negative countertransference, and Cronbach's α of .81 for the total measure.

In-Session Patient Affective Reactions Questionnaire—Clinician form (SPARQ-C). The SPARQ was originally developed as a self-report measure for the evaluation of the patient's perceptions in session and affective reactions toward their therapist (Stefana et al., 2023; Stefana, Fusar-Poli, Vieta, et al., 2024). Subsequently, it has been validated as a measure rated by clinicians (Stefana et al., unpublished work). It comprises two distinct 4-item scales: positive and negative affect. The positive affect scale reflects a secure and comfortable therapeutic relationship, while the negative affect scale includes feelings of shame, shyness, fear of speaking openly, worry about inadequate help, and a sense of failure. The items are rated on a 5-point Likert scale from 0 ("Not at all true") to 4 ("Very true"). The positive affect scale demonstrated good convergent validity with the goal subscale (r = .51), task subscale (r = .57), and bond subscale (r = .68) of the WAI-SR therapist form, and with the RRI-T (r = .64). The negative affect scale showed good convergent validity with the task subscale of the WAI-SR therapist form (r = −.38), and with negative transference (r = .41). In this study, SPARQ-C showed McDonald's ω_{total} coefficients of .83 for positive affect and .78 for negative affect.

Therapy Session Checklist–Transference Items (TSC-TI). Clinicians rated the patient's amount of transference using three single items from the Therapist Session Checklist (Graff & Luborsky, 1977). Transference was defined as the displacement of material from an early significant relationship onto the therapist. Each of the items

assesses the amount, respectively, of total transference, positive transference, and negative transference using a five-point scale from 1 ("None or slight") to 5 ("Very much"). The TSC-TI ratings are associated with various phenomena such as successful analysis, counselor intentions, and a multi-item transference measure, supporting their construct validity (Gelso et al., 1991; Graff & Luborsky, 1977; Multon et al., 1996). No internal consistency is reported for single-item scales.

Session outcome domain

Session Evaluation Scale—short form (SES-3). The SES-3 (Stefana & Hill, under review) evaluates the perception of the quality of a specific therapy session, which is a key component of session outcome. It consists of three items rated on a five-point scale from 1 ("Strongly disagree") to 5 ("Strongly agree"). Validity was indicated by correlations with intermediate session outcome measures (Stefana & Hill, under review). This study used a five-item version recommended by Lent et al. (2006), including an item on overall session effectiveness. The internal consistency McDonald's ω_{total} for the SES in this study was .74.

Session Quality (SQ). Clinicians assessed the overall quality of sessions using a single item from Gelso et al. (1991), rated on a five-point scale ranging from 1 ("Very poor") to 5 ("Very good"). This single-item measure has been used effectively to assess the interaction of transference and insight in predicting session quality outcomes. No internal consistency is reported for single-item scales.

Session Progress Scale (SPS). The SPS (Kolden, 1991) is part of the broader Therapy Session Report (Orlinsky & Howard, 1966). It is a four-item measure of session impact. Items are rated on a Likert scale from 1 to 6, where lower scores indicate greater session progress. The specific words corresponding to each number vary depending on the question; for example, 1 may mean "Completely helpful" and 6 "Not at all helpful." The SPS has demonstrated a test-retest reliability of .75. Its validity has been consistently supported, with the scales showing predictive validity for treatment duration and termination outcomes (Kolden & Howard, 1992). McDonald's ω_{total} for the SPS in this study was .85.

Post-Session Questionnaire (PSQ), Section B. The PSQ (Samstag et al., 1998) is a four-item measure that evaluates alliance ruptures and their resolution during a specific therapy session. The first item checks for the occurrence of any conflict, misunderstanding, disagreement, or tension between the patient and the clinician during a session. If the response is affirmative, the next three items assess the highest degree of tension experienced from 1 ("Low") to 5 ("High"), the extent to which the problem was addressed by the session's end from 1 ("Not at all") to 5 ("Very much"), and the degree to which the problem was resolved from 1 ("Not at all") to 5 ("Very much"). Lower intensity and higher resolution of ruptures have been associated with better ratings of the working alliance and session quality (Muran et al., 2009). No internal consistency is reported for single-item scales.

Session Evaluation Questionnaire (SEQ). The SEQ (Stiles, 1980, 2002; Stiles et al., 1994) measures the impact of counseling or psychotherapy sessions. It consists of 21 bipolar adjective scales, presented in a 7-point semantic differential format. Respondents are instructed to select a number from 1 to 7 that best represents their feelings about the session or their post-session mood, with higher scores indicating a greater degree of the measured construct. The SEQ assesses four primary dimensions: depth, smoothness, positivity, and arousal. These dimensions are categorized into session evaluation factors (depth and smoothness) and post-session mood variables (positivity and arousal). Depth reflects the clinician's perception of the value and impact of the session, while smoothness refers to the perceived ease and comfort of the session. The post-session mood variables measure the clinician's emotional state, with positivity indicating levels of happiness, satisfaction, and confidence, and arousal measuring the level of calmness or excitement. In this study, the internal consistency for the four subscales were McDonald's $\omega_{total} = .83$ for depth, $\omega = .87$ for smoothness, $\omega = .84$ for positivity, and $\omega = .65$ for arousal.

Insight. Clinicians rated the level of insight of patients using single items developed by Gelso et al. (1991). Insight was defined as the patient's accurate understanding of the therapy material, including their relationship, functioning outside of therapy, and personal dynamics. Intellectual insight reflects an understanding of cause–effect relationships, whereas integrative insight connects affect and intellect. Clinicians rated integrative, intellectual, and overall insight on a five-point scale from ("None or slight") to 5 ("Very much"). Although single-item measures can be problematic, their use is justified by their practicality and previous validation in studies (Bhatia & Gelso, 2018; Fuertes

et al., 2013; Markin et al., 2013). No internal consistency is reported for single-item scales.

Procedures

The Institutional Review Board at the University of North Carolina at Chapel Hill (UNC-CH) reviewed and exempted all procedures and materials for this study (IRB number 23-2656). Mental health clinicians were recruited between February and May 2024 through email invitations sent to members of various professional associations and registries. Eligible participants were licensed psychologists, psychotherapists, psychoanalysts, and counselors, with the exclusion criterion being the absence of at least one adult patient (aged 18 years or older) currently under their psychological care. Participants were asked to refer to their appointment records and select the most recent adult patient they had seen for an individual session. The survey, which focused on this specific patient and the therapeutic relationship, was administered online through Qualtrics and took an average of approximately 17 min to complete. Each clinician was allowed to participate only once. The development and validation of the CARE scales are part of a larger research project funded by the European Commission, details of which are available on the Open Science Framework website (OSF.io) at https://osf.io/amzqk.

Sample Size

The sample size requirements in confirmatory factor analysis (CFA) are influenced by various factors such as the number of items, the strength of the relationships between items and factors, model complexity, reliability, and estimation methods (Kline & Little, 2023; Streiner et al., 2015). In conducting a CFA for a scale validation study involving a scale with moderate complexity and reasonable indicators per factor (like in the case of the CARE scale), a sample size of 150 is sufficient to achieve reliable and valid parameter estimates, ensuring robust model fit (Brown, 2015; Kline & Little, 2023; Kyriazos, 2018). When assessing the correlations between the scale and other measures for different types of validity, similar considerations apply. A sample size of 100–200 is generally recommended to ensure sufficient statistical power to detect meaningful correlations and establish the validity of the scale (Hair et al., 2023; Tabachnick et al., 2019). Taking into account the above, a sample size of 150 can be deemed adequate for the present study and to overcome the problems derived from small sample sizes (De Prisco & Vieta, 2024).

Statistical Analyses

In the initial step, the Kaiser–Meyer–Olkin test and the Bartlett test of sphericity were performed to determine the suitability of the data to perform factor analysis. A confirmatory factor analysis (CFA) using a robust maximum likelihood estimator was carried out using the R package *lavaan* v0.6-18 (Rosseel, 2012) to examine the fit of the positively engaged, enmeshed, and stuck subscales of the CARE scales both individually and collectively (i.e., the three-factor model identified in the validation study). To evaluate the fit of the CFA model, the following criteria were applied: a comparative fit index (CFI) of .95 or higher, a Tucker Lewis index (TLI) of .95 or higher, a root mean square error of approximation (RMSEA) of .06 or lower, and a standardized root mean square residual (SRMR) of .08 or lower (Hu & Bentler, 1998; Kline & Little, 2023).

A multi-group CFA was conducted to assess the measurement invariance of the CARE scales across psychotherapy session formats (in-person versus remote, including video and telephone calls). Measurement invariance evaluates whether a construct is psychometrically equivalent across groups, ensuring that it retains the same structure and meaning in different contexts. Following established guidelines (Putnick & Bornstein, 2016), three levels of invariance were tested sequentially: (i) configural, (ii) metric, and (iii) scalar invariance. Configural invariance determines whether the factor structure (i.e., the relationships among items and factors) holds across groups. It assesses whether the same items load onto the same factors in both groups without imposing equality constraints, thereby confirming that the constructs are conceptualized similarly in both in-person and remote settings. Metric invariance (weak invariance) tests whether factor loadings are equivalent across groups, ensuring that the strength of the relationships between items and factors is consistent. This indicates that the construct is perceived similarly across groups. Scalar invariance (strong invariance) assesses whether item intercepts are equivalent across groups. Establishing scalar invariance ensures that observed mean differences reflect differences in the latent construct rather than group-specific response tendencies. Fit at each step was evaluated using changes in alternative fit indices (ΔCFI $\leq -.01$, ΔTLI $\leq -.01$, ΔRMSEA $\leq .015$, and ΔSRMR $\leq .03$ for metric invariance or $\leq .01$ for scalar invariance) and chi-square difference tests (Chen, 2007). These criteria balance sensitivity and robustness, addressing the limitations of chi-square difference tests, which can be overly sensitive in large samples. Residual invariance (strict invariance), which tests the equivalence of item-specific

error variances, was not assessed because it is not required for comparing latent means (Putnick & Bornstein, 2016; Vandenberg & Lance, 2000). To ensure sufficient power for the multi-group CFA (Kyriazos, 2018), data from the development and preliminary validation study ($N = 556$) were pooled with the present study sample ($N = 151$).

To assess the psychometric properties of the CARE scales, the following statistical analyzes were performed. The internal consistency for each scale was evaluated using Cronbach's α and McDonald's ω_{total} coefficients, and the average inter-item correlation, calculated with the R package *psych* v2.3.12 (Revelle, 2024). Standard error of measurement (SE_m) and difference (SE_d) were calculated to inform about the deviation of the score due to measurement error. The correlations between the CARE scales and (a) sociodemographic, clinical, and treatment variables, as well as (b) validated measures of specific elements of the therapeutic relationship and session outcomes, were calculated to establish criterion validity. The Benjamini-Hochberg method was applied to adjust the *p*-values (Benjamini & Hochberg, 1995). These adjusted *p*-values were then used to identify significant correlations. There were no missing data in the survey results, as the Qualtrics survey was designed to require responses to all questions, ensuring complete data collection.

Results

Preliminary Analyses

The Kaiser–Meyer–Olkin test (.78) and the Bartlett test of sphericity ($p < .001$) supported the suitability of the data for factor analysis.

Confirmatory Factor Analysis

Confirmatory factor analyses (CFAs) were performed on the positively engaged, enmeshed, and stuck subscales of the CARE scale individually and in combination in a three-factor model (see Table III). The results indicated that a single-factor model provided a good fit for the positively engaged and enmeshed factors, but was only partially satisfactory for the stuck factor. For the positively engaged factor: $\chi^2_{(5)} = 2.94$, CFI = 1.00; TLI = 1.02, RMSEA = .00 (90% CI [.00, .85]), and SRMR = .02. For the enmeshed factor: $\chi^2_{(5)} = 9.66$, CFI = .97; TLI = .94, RMSEA = .08 (90% CI [.00, .15]), and SRMR = .04. For the stuck factor: $\chi^2_{(5)} = 12.55$, CFI = .94; TLI = .88, RMSEA = .10 (90% CI [.03, .17]), and SRMR = .06. The three-factor model incorporating the three scales showed a good fit for the data: $\chi^2_{(87)} = 120.41$, CFI = .94; TLI = .93, RMSEA = .05 (90% CI [.03, .07]), and SRMR = .08.

We also tested a bi-factor model at the suggestion of a blind reviewer. While it showed excellent global fit ($\chi^2(5) = 83.82$, $p = .227$, CFI = .98; TLI = .98, RMSEA = .03 (90% CI [.00, .06]), and SRMR = .08), the explained common variance for the general factor was only 0.36, and its omega was 0.35—both suggesting the general factor primarily reflects shared variance rather than a distinct construct. Consequently, we regard the three-factor structure (Engaged, Enmeshed, and Stuck) as our primary focus, as it demonstrates clearer interpretability, stronger reliability, and better alignment with our theoretical framework.

Invariance testing with multigroup CFA

A multigroup CFA model was tested for measurement invariance between patients attending in-person (either face-to-face or on the couch) and remote (via video or telephone call) sessions.

Configural invariance. Configural invariance was tested to examine whether the same factor structure holds across both groups. The model exhibits a robust fit across both groups, indicating that the three-factor structure of the CARE scales is conceptually similar across session formats. The configural

Table III. Indices of fit for confirmatory factor analysis and internal reliability coefficients.

	Statistical index of fit			Practical index of fit			Internal reliability		
Model	χ^2	*df*	*p*	CFI	TLI	RMSEA [CI]	SRMR	α	ω
Positively engaged	2.94	5	.709	1.00	1.02	0.00 [0.00, 0.09]	0.02	0.76	0.79
Enmeshed	9.66	5	.086	0.97	0.94	0.08 [0.00, 0.15]	0.04	0.72	0.75
Stuck	12.55	5	.028	0.94	0.88	0.10 [0.03, 0.17]	0.06	0.70	0.71
Three-factor model	120.41	87	.010	0.94	0.93	0.05 [0.03, 0.07]	0.08	0.64	0.81

Note. α = Cronbach's alpha value; ω = McDonald's omega total; CFI = comparative fit index; *df* = degrees of freedom; RMSEA = root mean squared error of approximation; SRMR = standardized root mean square residual; TLI = Tucker-Lewis index; χ^2 = chi-square.

model fit indices were CFI= .992, TLI= .991, RMSEA= .030, and SRMR= .085. These values indicate that the pattern of item-factor loadings is uniform across the groups, satisfying the requirement for configural invariance.

Metric invariance. Metric invariance was tested by constraining the factor loadings to be equal across groups. This step assesses whether the items contribute equally to the latent constructs in both groups. The model fit indices showed a minimal change: the CFI changed from .992 to .989 (ΔCFI= −.003), the TLI changed from .991 to .987 (ΔTLI= −.004), the RMSEA increased from .030 to .035 (ΔRMSEA= + .005), and the SRMR increased from .085 to .089 (ΔSRMR= −.006). According to established criteria (Cheung & Rensvold, 2002), Δ values indicate that metric invariance holds and the factor loadings are largely consistent across groups.

Scalar invariance. Scalar invariance was then tested by additionally constraining the item intercepts to be equal across groups. This step is crucial for comparing latent means between groups. The model fit indices exhibited minimal changes: the CFI changed from .989 to .988 (ΔCFI= −.001), the TLI remained the same at .987 (ΔTLI= .000), and the RMSEA decreased from .035 to .033 (ΔRMSEA = −.001). The SRMR remained at .089 (ΔSRMR= .000). Δ values are well within the acceptable range and thus indicate that scalar invariance holds.

Model comparisons. The chi-square difference test comparing the metric (weak) invariance model to the configural model showed a significant difference ($p < .001$), suggesting some differences in factor loadings across groups. However, Δ values were within acceptable limits, supporting metric invariance. The comparison between the scalar (strong) invariance model and the metric invariance model showed no significant difference ($p = .794$), further supporting scalar invariance.

Summary. Overall, the results indicate that the CARE scales exhibit configural, metric, and scalar invariance across in-person and remote session formats. This suggests that the scales measure the same constructs in the same way across these groups, thus allowing for meaningful comparisons of scores between patients attending different session formats.

Correlations between CARE scales

The positively engaged factor correlated $r = .16$ with the enmeshed factor and −.35 with the stuck factor, while the enmeshed and stuck factors correlated at $r = .23$.

Internal Consistency and Normal Distribution

Each of the three scales demonstrated acceptable internal consistency: positive engagement factor ($k = 5$, $\alpha = .76$, ω total = .78, average inter-item $r = .39$), enmeshed factor ($k = 5$, $\alpha = .72$, ω total = .72, average inter-item $r = .34$), and stuck factor ($k = 5$, $\alpha = .70$, ω total = .71, average inter-item $r = .31$). Skewness and kurtosis values were within normal limits: −1.46 and 1.47 for the positively engaged factor, 1.16 and .68 for the enmeshed factor, and 1.07 and .24 for the stuck factor.

Score Precision

The mean sum scores for the positively engaged, enmeshed, and stuck factors were 8.79 ($SD = 1.65$), 2.25 ($SD = 2.23$), and 1.46 ($SD = 1.69$), respectively. The percentages of the maximum possible scores were 88%, 22% and 15%, respectively. SE_m and SE_d were 4.12 and 5.83 for the positively engaged scale, 1.18 and 1.67 for the enmeshed scale, and .91 and 1.29 for the stuck scale.

Associations between Subscale Scores and Sociodemographic, Clinical, and Treatment Variables

As detailed in Tables IV and V, the CARE scale scores showed very weak correlations (Pearson r coefficients ranging from −.20 to .21) with all demographic data of patients and clinicians, as well as the clinical and treatment variables listed in Tables I and II.

Convergent Criterion Validity

Table IV presents all the correlation coefficients calculated for each CARE scale with the measures of therapeutic relationship and the measures of session outcome.

Therapeutic relationship measures. The positively engaged scale was strongly and positively correlated with the quality of the task ($r = .51$) and the bond ($r = .54$) dimensions of the working alliance (WAI-SR-T) and with the genuineness dimension

Table IV. Convergent and divergent validity correlations.

	Scale score	Engaged	Enmeshed	Stuck
Therapeutic relationship measures				
WAI-SR-T total score	67.22 (8.98)	.50***	−.00	−.45***
WAI-SR-T goal	20.85 (3.94)	.30***	−.06	−.37***
WAI-SR-T task	22.36 (3.24)	.51***	−.00	−.51***
WAI-SR-T bond	24.01 (3.02)	.54***	.08	−.32***
RRI-T total score	94.68 (10.77)	.53***	.27**	−.32***
RRI-T genuineness	47.74 (5.89)	.55***	.26**	−.31***
RRI-T realism	46.93 (5.51)	.44***	.25**	−.29***
ICB total amount	24.67 (3.84)	−.16	.32***	.28**
ICB positive	12.43 (2.69)	−.01	.38***	.23*
ICB negative	12.23 (1.82)	−.32***	.10	.26**
Tranference total amount	2.25 (1.13)	.14	.21*	.13
Tranference positive	2.63 (1.27)	.28**	.22*	−.03
Tranference negative	1.50 (.86)	−.01	.19*	.18*
SPARQ-C positive affect	12.46 (2.69)	.57**	.22*	−.24***
SPARQ-C negative affect	1.83 (2.24)	−.14	.17	.28**
PSQ (yes)[a]	20 (13%)	.00	.07	.22
Degree of tension	2.40 (1.05)	−.37	.34	.29
Extent issue addressed	3.60 (1.54)	.04	−.03	−.33
Degree of resolution	3.15 (1.42)	.16	−.08	−.42
Session outcome measures				
Insight overall	3.85 (.98)	.42***	.17	−.41***
Insight intellectual	3.90 (1.08)	.26**	.06	−.28**
Insight integrative	3.66 (1.10)	.47***	.11	−.45***
SES-3	12.54 (1.82)	.43***	.10	−.41***
SQ	4.01 (.84)	.45***	.17	−.40***
SPS	14.09 (3.64)	−.45***	−.06	.46***
SEQ depth	3.06 (.66)	.33***	.21*	−.24**
SEQ smoothness	4.19 (1.18)	.18	−.02	−.29**
SEQ positivity	4.37 (.90)	.41***	−.03	−.36***
SEQ arousal	2.84 (.94)	.12	.15	.02

[a]Number of observations and percentage on the whole sample.
*$p < .05$, **$p < .01$, ***$p < .001$, two-tailed.

($r = .55$) of the real relationship (RRI-T). Furthermore, the positively engaged scale was strongly correlated with the patient's safe and comfortable experience of the therapeutic relationship as perceived and rated by the clinician (SPARQ-C) ($r = .57$). Lastly, it was weakly and moderately correlated with negative countertransference (ICB) levels ($r = -.32$).

The stuck scale was strongly and negatively correlated with the task dimension ($r = -.51$) of the working alliance (WAI-SR-T), while it was only moderately correlated with the goal ($r = -.37$) and bond ($r = -.32$) dimensions. Furthermore, it was moderately and negatively correlated with the real relationship (RRI-T) ($r = -.32$) and weakly with the patient's positive experience of the relationship as perceived and rated by the clinician (SPARQ-C) ($r = -.24$). In contrast, the stuck scale was weakly and positively correlated with the levels of countertransference (ICB), both positive ($r = .23$) and negative ($r = .26$), the patient's negative experience of the relationship ($r = .28$), and the amount of negative transference ($r = .18$).

The enmeshed scale was moderately correlated with the amount of positive countertransference (ICB) ($r = .38$) and only weakly associated with the real relationship dimensions (RRI-T) ($r = .27$) and the amount of patient's positive emotions toward the therapist and the therapeutic relationship—in the form of positive transference and SPARQ-C positive affect as perceived and rated by the clinician (both $r = .22$).

It is important to note that each scale retains a significant amount of reliable specific variance ($=\omega - r^2$), indicating that there is reliable variance unique to each scale, beyond the variance shared with other measures.

Session outcome measures. The positively engaged scale was moderately correlated with five measures of session outcome (Insight integrative, SES-3, SQ, SPS, and SEQ Positivity) (*r*s range = .41–.45), while it was weakly and negatively correlated with the degree of tension between the patient and the clinician during the session (PSQ) ($r = -.37$). The direction of the correlations was in line with

Table V. Discriminant validity correlations.

	Positively engaged	Enmeshed	Stuck
Therapist age	.16*	−.11	−.16
Therapist gender (Man higher)	−.06	.03	.12
Professional background	.02	.04	−.08
Psychotherapeutic approach	.09	.11	.11
Years of post-licensed experience	.06	−.07	−.05
Weekly Time spent practicing therapy	.03	−.08	−.05
Patient age	.16	−.20*	−.04
Patient gender	−.04	−.13	.12
Patient ethnicity	.08	.23	−.03
Any psychiatric disorder	.08	.09	−.04
Any anxiety disorder	.04	.03	−.16
Any trauma- and stressor-related disorders	.06	.14	.03
Any (unipolar) depressive disorder			
Any neurodevelopmental disorder	.02	.09	.04
Any personality disorder	.12	.14	.11
Any eating disorder	.14	.23	−.03
Any bipolar or related disorder	−.01	−.05	.15
Any other disorder	−.14	.02	.02
CGI	.06	.15	.15
GAF	−.00	−.13	−.19*
Therapy length (months, ordinal; see prior table)	.16	.16	.20*
Session frequency (ordinal; see prior table)	.09	.21***	.02
Session attendance (in person, telephone, video)	−.09	.05	.14
Location (categorical; see prior table)	−.03	.00	−.07

Note. Coefficients are point-biserial correlations for dichotomized variables, Spearman correlations for ordinal variables, and Pearson correlations for continuous variables.
*$p < .05$, **$p < .01$, ***$p < .001$, two-tailed.

expectations, indicating a positive correlation between the clinician's level of positive engagement in the therapeutic relationship and positive session outcomes.

The stuck scale showed a pattern of correlations similar to that of the positively engaged scale but in the opposite direction. The main differences were moderate correlations with the real relationship ($r = -.32$) and between the stuck scale and the problem addressed ($r = -.33$) and problem resolved ($r = -.42$) scales of the PSQ. These negative correlations align with theoretical expectations.

The enmeshed scale was very weakly correlated with all measures of session outcome (*r*s range = .03–.21), except one: the PSQ highest tension ($r = -.34$), which evaluated the highest degree of conflict, misunderstanding, or tension experienced in the therapeutic relationship during the session.

Discussion

The present study examined the structure, reliability, and validity of a new brief tool intended to gather potentially clinically useful information on the patterns of affective, cognitive, and behavioral responses experienced by mental health clinicians towards their patients during individual psychological and psychotherapeutic sessions. Clinician self-report feedback, despite its inherent biases and distortions (Bhatia & Gelso, 2018; Ciuk et al., 2015), is an essential indicator of their subjective experience during psychological interventions. Clinicians' awareness of their affective states in the relationship, as reflected in such feedback, plays a crucial role in enhancing treatment outcomes (Abargil & Tishby, 2022; Hayes et al., 2018).

Factor analyses confirmed that the CARE scales measure three-dimensional factors—namely positively engaged, enmeshed, and stuck—that are theoretically sound, clinically meaningful, and supported by empirical evidence (Betan et al., 2005; Colli et al., 2014; Tanzilli et al., 2016). These scales reflect different emotional patterns and have specific features and correlations with measures of other elements of the therapeutic relationship, as well as of session outcome, suggesting potentially differentiated roles in the context of therapy research.

Importantly, our findings show that the CARE scales work similarly across both in-person and remote sessions, meaning the fundamental structure and item meanings are consistent between the two groups. In other words, the tool reliably measures the same concepts regardless of the session format. As a result, the CARE scales are a valuable and reliable tool for examining and comparing therapists' inner dynamics in both teletherapy and traditional in-person psychotherapy. This result is especially significant given the rapid growth of teletherapy from a supplementary service to a standard treatment option in recent decades (Pierce et al., 2020, 2021), a trend that has been further accelerated by the COVID-19 pandemic (Stefana, Youngstrom, Chen, et al., 2020; Stefana, Youngstrom, Hopwood, et al., 2020).

The positively engaged factor encompasses items reflecting feelings of appreciation for the patient, contentment in their presence, comfort and enthusiasm in working with them, and interest in their progress. This factor represents the therapist's positive experience of the therapeutic relationship and process, characterized by empathic attunement and a strong positive alliance, potentially independent of the therapy outcomes. The positive engagement scale tends to yield high scores frequently. In line with the preliminary validation study (Stefana, Fusar-Poli, Langfus, et al., 2024), the average score was near the

upper limit (88% of the maximum possible), although there was still variability within the sample. The results also indicate that typical sessions are characterized by strong positive engagement from clinicians.

Regarding convergent validity, positive engagement is strongly correlated with the bond dimension of the alliance and the clinician's perception that the patient experiences the therapeutic relationship as secure and comfortable. Furthermore, positive engagement during the session was moderately correlated with the clinician's positive mood post-session. Evidence of discriminant validity was demonstrated by weak or moderate correlations of the positive engagement scale with the demographic and professional variables of the therapists, as well as with the demographic, clinical, and treatment variables of the patients. These correlations between elements of the therapeutic relationship, especially the alliance, are consistent with the existing clinical-theoretical and empirical literature (Stefana, 2017; Tanzilli et al., 2018).

Lastly, scales were moderately correlated with measures of session quality and patient emotional-intellectual understanding of the material discussed in the session. Given that therapists' overall in-session reactions become an obstacle mainly when it has a negative valence (Racker, 2018), our results align with a meta-analysis indicating that psychotherapists' ability to effectively manage countertransference, defined as "internal and external reactions in which unresolved conflicts of the therapist, usually but not always unconscious, are implicated," appears crucial for achieving positive therapeutic results (Hayes et al., 2018). Notably, the correlation between the positive engagement scale and session outcome measures was equal to or greater than the correlations reported for therapist-rated session outcomes with working alliance (Bhatia & Gelso, 2018; Zilcha-Mano et al., 2016), real relationship (Bhatia & Gelso, 2018; Gelso et al., 2018), and countertransference (Bhatia & Gelso, 2018). Importantly, the skewness values (skews < 3) observed for the three CARE scales suggest that the coefficients are unlikely to be biased, aligning with established guidelines that indicate such skewness levels do not typically distort statistical estimates (Chou & Bentler, 1995).

On the contrary, the enmeshed scale had a lower central tendency in its score distributions, with means around 20% of the maximum possible range. The items—reflecting a desire to give love, a drive to act for the patient, concern for their feelings, tenderness, and a tendency to protect them, significantly more than for other patients—were less commonly endorsed in response to sessions. Evidence for discriminant validity of the enmeshed scale was shown and convergent validity was supported by a statistically significant correlation with positive countertransference. As hypothesized in the preliminary validation study of the CARE scales (Stefana, Fusar-Poli, Langfus, et al., 2024), the enmeshed factor appears to represent aspects of what is more accurately described as positive countertransference (Gelso & Kline, 2022).

The stuck scale reflects the clinician's feelings of difficulty entering the patient's inner world, frustration in their efforts to help, incompetence, hopelessness, and annoyance towards the patient. Similar to the enmeshed scale, the stuck scale exhibited a lower central tendency in its score distributions, with means around 20% of the maximum possible range. Discriminant validity for the stuck scale was demonstrated, and convergent validity was supported by a moderate correlation with the task dimension of the working alliance. Regarding convergent validity, this scale was moderately and negatively correlated with measures of session quality outcome and the resolution of conflicts, misunderstandings, or tensions between the clinician and the patient during the session. Contrary to our expectations, the stuck scale was not associated with the occurrence of alliance ruptures. This result might be due to the therapist-reported nature of the data and might be different when considering patients' reported alliance ruptures. The stuck scale was also moderately and negatively correlated with the degree of integrative insight of the patient. This correlation pattern aligns with the findings from a study that employed the Assessment of Clinicians' Subjective Experience's (Pallagrosi et al., 2014) difficulty in attunement scale (which includes items related to difficulties in establishing emotional contact, being empathic, understanding the patient's experience, and communicating with the patient). That study found a correlation between the difficulty in attunement scale and working alliance and session outcome measures in adult psychotherapy settings (Tanzilli et al., 2018). The stuck scale thus appears to indicate a negative process in which the therapist experiences a loss of agency and reflective capacity during their in-session work with the patient (Moltu et al., 2010; Werbart et al., 2022).

The three dimensions of the CARE tool closely mirror the core aspects of clinicians' affective responses that arise during clinical interactions (Prasko et al., 2022; Stefana, Bulgari, et al., 2020). The use of brief scales, such as the CARE scales, helps therapists identify patterns of emotional reactions and ways of interaction in session. The varied content across the three scales is likely to enhance the awareness of the types and extents of clinicians' in-session reactions. The CARE scales may be beneficial for monitoring fluctuations in these dimensions across sessions and evaluating their relationships with session and treatment outcomes. Furthermore, in line with recent suggestions to incorporate emotional competency into clinical research

and practice (Gonsalvez et al., 2020; Price et al., 2017), the CARE scales could serve as a valuable tool. This scale can contribute to the evaluation of emotional competency, an area that has been proposed for integration into clinical practice. Recent research underscores the significant impact of countertransference management skills of therapists (Hayes et al., 2018) and, more broadly, foundational and functional competencies (Dimmick et al., 2023) on patient outcomes. This emphasizes the importance of a tool such as the CARE scales to advance the assessment of emotional competency and linking it to real-world clinical outcomes.

Strengths, Limitations, and Future Directions

This study has several strengths. Validation of a concise self-report tool contributes significantly to operationalization of the emotional, cognitive, and behavioral responses of clinicians towards patients during therapy sessions, which constitutes an important component of the therapy relationship (San & Arranz, 2023). The CARE scales exhibited strong psychometric properties, corroborating the findings from its development and preliminary validation, and fully establishing its validity. The set of three brief scales takes less than 2 min to complete and is straightforward to score.

However, some limitations must be considered when interpreting the findings. First, the data are based solely on the perspectives of the clinicians. This includes a potential bias in self-reporting one's own affective reactions compared to observer ratings of verbal and nonverbal behaviors or physiological recordings (Ciuk et al., 2015), which can capture aspects of emotional reactions beyond the awareness of respondents. Despite this, growing evidence supports using clinicians' ratings of their responses to patients (Betan et al., 2005; Bhatia & Gelso, 2018). Furthermore, self-report measures provide valuable insights into clinicians' perceptions and internal experiences of their patients and the therapeutic relationship during sessions. Second, the study lacks measures of clinicians' mental state awareness (i.e., mentalization), ability to identify and verbalize emotional states (i.e., alexithymia), and emotion regulation strategies. These factors are crucial to collect valid and reliable data on emotional reactions in session and are related to the first limitation, as integrating self-report questionnaires with observer-rated methods would be beneficial.

Future research using the CARE scales should explore affective states and processes from multiple perspectives to further validate the tool, assess correlations, and understand how the scores relate to patients' perceptions and their affective reactions towards clinicians. A comparison with other inventories assessing the overall emotional reactions of therapists toward patients—such as the *Therapist Response Questionnaire* (Betan & Westen, 2009) and the *Feeling Word Checklist* (Holmqvist & Armelius, 1994)—would also be important. Longitudinal studies are particularly needed to investigate the predictive validity of the CARE scales in terms of patient outcomes. Additionally, researchers should examine the CARE scales as a self-report tool to measure weekly changes in the emotional responses of therapists to patients, which would represent a significant step towards a measurement feedback system for the experiences in sessions of therapists. Finally, prospective studies could improve our understanding of how subjective experiences of therapists evolve throughout psychotherapy and influence various trajectories and outcomes.

Conclusion

Our findings endorse the utility of the CARE scales in both clinical and research contexts, particularly to evaluate the subjective affective reactions of clinicians toward their patients and affective processes at the session level. The CARE scales are brief and easy to complete, making them easily implementable in clinical practice, supervision, and research environments. By allowing therapists to self-rate a set of common affective reactions toward their patients on a standardized scale with established correlates, the CARE scales can help therapists (a) reflect and recognize their internal experiences and (b) quantify these experiences for statistical analysis and research. Furthermore, the CARE scales can potentially guide therapeutic interventions, inform clinical supervisors.

Acknowledgements

The first author used ChatGPT o1 to edit the language (grammar, syntax, clarity, and readability) of the manuscript.

Funding

This study has received funding from the European Union's Horizon 2020; European Research Council research and innovation program under the Marie Sklodowska-Curie grant agreement No 101030608.

Disclosure Statement

EAY received royalties from the American Psychological Association and Guilford Press and consulted about psychological assessment with Signant Health. He is the co-founder and Executive Director of Helping Give Away Psychological Science, a 501c3 in the United States, where he serves as an unpaid volunteer. EV has received grants and served as consultant, advisor, or continuing medical education (CME) speaker for the following entities outside the submitted work: AB-Biotics, AbbVie, Adamed, Angelini, Biogen, Biohaven, Boehringer-Ingelheim, Celon Pharma, Compass, Dainippon Sumitomo Pharma, Ethypharm, Ferrer, Gedeon Richter, GH Research, Glaxo-Smith Kline, HMNC, Idorsia, Johnson & Johnson, Lundbeck, Medincell, Merck, Newron, Novartis, Orion Corporation, Organon, Otsuka, Roche, Rovi, Sage, Sanofi-Aventis, Sunovion, Takeda, and Viatris. AS and PF-P have nothing to disclose.

ORCID

Alberto Stefana http://orcid.org/0000-0002-4807-7184
Eduard Vieta http://orcid.org/0000-0002-0548-0053
Paolo Fusar-Poli http://orcid.org/0000-0003-3582-6788
Eric A. Youngstrom http://orcid.org/0000-0003-2251-6860

References

Abargil, M., & Tishby, O. (2022). Countertransference awareness and treatment outcome. *Journal of Counseling Psychology*, *69*(5), 667–677. https://doi.org/10.1037/cou0000620

Atzil-Slonim, D., Bar-Kalifa, E., Fisher, H., Peri, T., Lutz, W., Rubel, J., & Rafaeli, E. (2018). Emotional congruence between clients and therapists and its effect on treatment outcome. *Journal of Counseling Psychology*, *65*(1), 51–64. https://doi.org/10.1037/cou0000250

Barkham, M., De Jong, K., Delgadillo, J., & Lutz, W. (2023). Routine outcome monitoring (ROM) and feedback: Research review and recommendations. *Psychotherapy research : journal of the Society for Psychotherapy Research*, *33*(7), 841–855. https://doi.org/10.1080/10503307.2023.2181114.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Betan, E. J., Heim, A. K., Zittel Conklin, C., & Westen, D. (2005). Countertransference phenomena and personality pathology in clinical practice: An empirical investigation. *American Journal of Psychiatry*, *162*(5), 890–898. https://doi.org/10.1176/appi.ajp.162.5.890

Betan, E. J., & Westen, D. (2009). Countertransference and personality pathology: Development and clinical application of the countertransference questionnaire. In R. A. Levy & J. S. Ablon (Eds.), *Handbook of evidence-based psychodynamic psychotherapy* (pp. 179–198). Humana Press. https://doi.org/10.1007/978-1-59745-444-5_8

Bhatia, A., & Gelso, C. J. (2018). Therapists' perspective on the therapeutic relationship: Examining a tripartite model. *Counselling Psychology Quarterly*, *31*(3), 271–293. https://doi.org/10.1080/09515070.2017.1302409

Breivik, R., Wilberg, T., Evensen, J., Røssberg, J. I., Dahl, H. S. J., & Pedersen, G. (2020). Countertransference feelings and personality disorders: A psychometric evaluation of a brief version of the Feeling Word Checklist (FWC-BV). *BMC Psychiatry*, *20*(1), 141. https://doi.org/10.1186/s12888-020-02556-6

Brody, E. M., & Farber, B. A. (1996). The effects of therapist experience and patient diagnosis on countertransference. *Psychotherapy: Theory, Research, Practice, Training*, *33*(3), Article 3. https://doi.org/10.1037/0033-3204.33.3.372

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed., pp. xvii, 462). The Guilford Press.

Busner, J., & Targum, S. D. (2007). The clinical global impressions scale: Applying a research tool in clinical practice. *Psychiatry (Edgmont)*, *4*(7), Article 7.

Chen, F. F. (2007). Sensitivity of goodness of Fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5

Chou, C.-P., & Bentler, P. M. (1995). Estimates and tests in structural equation modeling. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 37–55). Sage Publications, Inc.

Chui, H., Li, X., & Luk, S. (2022). Therapist emotion and emotional change with clients: Effects on perceived empathy and session quality. *Psychotherapy*, *59*(4), 594–605. https://doi.org/10.1037/pst0000442

Ciuk, D., Troy, A. K., & Jones, M. C. (2015). Measuring emotion: Self-reports vs. physiological indicators. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2595359

Colli, A., Tanzilli, A., Dimaggio, G., & Lingiardi, V. (2014). Patient personality and therapist response: An empirical investigation. *American Journal of Psychiatry*, *171*(1), 102–108. https://doi.org/10.1176/appi.ajp.2013.13020224

Cologon, J., Schweitzer, R. D., King, R., & Nolte, T. (2017). Therapist reflective functioning, therapist attachment style and therapist effectiveness. *Administration and Policy in Mental Health*, *44*(5), 614–625. https://doi.org/10.1007/s10488-017-0790-5

De Prisco, M., & Vieta, E. (2024). The never-ending problem: Sample size matters. *European Neuropsychopharmacology: The Journal of the European College of Neuropsychopharmacology*, *79*, 17–18. https://doi.org/10.1016/j.euroneuro.2023.10.002

DeVellis, R. F., & Thorpe, C. T. (2022). *Scale development: Theory and applications* (5th ed.). Sage.

Dimmick, A. A., Callahan, J. L., & Cox, R. J. (2023). Further validation of competency assessment: The practicum evaluation form. *Training and Education in Professional Psychology*, *17*(3), 269–276. https://doi.org/10.1037/tep0000416

Friedman, S. M., & Gelso, C. J. (2000). The development of the inventory of countertransference behavior. *Journal of Clinical Psychology*, *56*(9), 1221–1235. https://doi.org/10.1002/1097-4679(200009)56:9<1221::AID-JCLP8>3.0.CO;2-W

Fuertes, J. N., Gelso, C. J., Owen, J. J., & Cheng, D. (2013). Real relationship, working alliance, transference/countertransference and outcome in time-limited counseling and psychotherapy. *Counselling Psychology Quarterly*, *26*(3–4), 294–312. https://doi.org/10.1080/09515070.2013.845548

Gelso, C. J. (2014). A tripartite model of the therapeutic relationship: Theory, research, and practice. *Psychotherapy Research*, *24*(2), Article 2. https://doi.org/10.1080/10503307.2013.845920

Gelso, C. J., & Hayes, J. A. (2007). *Countertransference and the therapist's inner experience: Perils and possibilities*. Lawrence Erlbaum Associates Publishers.

Gelso, C. J., Hill, C. E., & Kivlighan, D. M. (1991). Transference, insight, and the counselor's intentions during a counseling hour. *Journal of Counseling & Development*, *69*(5), 428–433. https://doi.org/10.1002/j.1556-6676.1991.tb01539.x

Gelso, C. J., Kelley, F. A., Fuertes, J. N., Marmarosh, C., Holmes, S. E., Costa, C., & Hancock, G. R. (2005). Measuring the real relationship in psychotherapy: Initial validation of the therapist form. *Journal of Counseling Psychology*, *52*(4), Article 4. https://doi.org/10.1037/0022-0167.52.4.640

Gelso, C. J., Kivlighan, D. M., & Markin, R. D. (2018). The real relationship and its role in psychotherapy outcome: A meta-analysis. *Psychotherapy*, *55*(4), 434–444. https://doi.org/10.1037/pst0000183

Gelso, C. J., & Kline, K. V. (2022). Some directions for research and theory on countertransference. *Psychotherapy Research*, *32*(1), 46–51. https://doi.org/10.1080/10503307.2021.1968529

Gonsalvez, C. J., Deane, F. P., Terry, J., & Hoong Gooi, C. (2020). Anatomy of psychology practitioner competence: Clustering approach reveals the hierarchical organization of competencies. *Clinical Psychology: Science and Practice*, *27*(4), e12347. https://doi.org/10.1037/h0101779.

Graff, H., & Luborsky, L. (1977). Long-term trends in transference and resistance: A report on a quantitative-analytic method applied to four psychoanalyses. *Journal of the American Psychoanalytic Association*, *25*(2), 471–490. https://doi.org/10.1177/000306517702500210

Hair, J. F., Jr., Black, W. C., Babin, B. J., & & Anderson, R. A. (2023). *Multivariate data analysis* (8th ed.). Cenage.

Hatcher, R. L., Lindqvist, K., & Falkenström, F. (2020). Psychometric evaluation of the working alliance inventory-therapist version: Current and new short forms. *Psychotherapy Research: Journal of the Society for Psychotherapy Research*, *30*(6), 706–717. https://doi.org/10.1080/10503307.2019.1677964

Hayes, J. A., Gelso, C. J., Goldberg, S., & Kivlighan, D. M. (2018). Countertransference management and effective psychotherapy: Meta-analytic findings. *Psychotherapy*, *55*(4), 496–507. https://doi.org/10.1037/pst0000189

Hayes, J. A., & Vinca, M. (2017). Therapist presence, absence, and extraordinary presence. In L. G. Castonguay & C. E. Hill (Eds.), *How and why are some therapists better than others?: Understanding therapist effects* (pp. 85–99). American Psychological Association. https://doi.org/10.1037/0000034-006

Holmqvist, R., & Armelius, B. A. (1994). Emotional reactions to psychiatric patients. Analysis of a feeling checklist. *Acta Psychiatrica Scandinavica*, *90*(3), 204–209. https://doi.org/10.1111/j.1600-0447.1994.tb01578.x

Horvath, A. O., & Greenberg, L. S. (1989). Development and validation of the working alliance inventory. *Journal of Counseling Psychology*, *36*(2), 223–233. https://doi.org/10.1037/0022-0167.36.2.223

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*(4), 424–453. https://doi.org/10.1037/1082-989X.3.4.424

Izard, C. E., Libero, D. Z., Putnam, P., & Haynes, O. M. (1993). Stability of emotion experiences and their relations to traits of personality. *Journal of Personality and Social Psychology*, *64*(5), 847–860. https://doi.org/10.1037/0022-3514.64.5.847

Johnson, S. L., Leedom, L. J., & Muhtadie, L. (2012). The dominance behavioral system and psychopathology: Evidence from self-report, observational, and biological studies. *Psychological Bulletin*, *138*(4), 692–743. https://doi.org/10.1037/a0027503

Jones, S. H., Thornicroft, G., Coffey, M., & Dunn, G. (1995). A brief mental health outcome scale-reliability and validity of the global assessment of functioning (GAF). *The British Journal of Psychiatry: The Journal of Mental Science*, *166*(5), 654–659. https://doi.org/10.1192/bjp.166.5.654

Kimerling, R. E., Zeiss, A. M., & Zeiss, R. A. (2000). Therapist emotional responses to patients: Building a learning-based language. *Cognitive and Behavioral Practice*, *7*(3), 312–321. https://doi.org/10.1016/S1077-7229(00)80089-9

Kline, R. B., & Little, T. D. (2023). *Principles and practice of structural equation modeling* (5th ed.). The Guilford Press.

Kolden, G. G. (1991). The generic model of psychotherapy: An empirical investigation of patterns of process and outcome relationships. *Psychotherapy Research*, *1*(1), 62–73. https://doi.org/10.1080/10503309112331334071

Kolden, G. G., & Howard, K. I. (1992). An empirical test of the generic model of psychotherapy. *Journal of Psychotherapy Practice and Research*, *1*(3), 225–236.

Kyriazos, T. A. (2018). Applied psychometrics: The 3-faced construct validation method, a routine for evaluating a factor structure. *Psychology (Savannah, GA)*, *9*(8), Article 8. https://doi.org/10.4236/psych.2018.98117

Latts, M. G. (1997). *A revision and validation of the countertransference factors inventory* (Vol. 58, Issues 1-B, p. 0455). ProQuest Information & Learning.

Lent, R. W., Hoffman, M. A., Hill, C. E., Treistman, D., Mount, M., & Singley, D. (2006). Client-specific counselor self-efficacy in novice counselors: Relation to perceptions of session quality. *Journal of Counseling Psychology*, *53*(4), 453–463. https://doi.org/10.1037/0022-0167.53.4.453.

Markin, R. D., McCarthy, K. S., & Barber, J. P. (2013). Transference, countertransference, emotional expression, and session quality over the course of supportive expressive therapy: The raters' perspective. *Psychotherapy Research*, *23*(2), 152–168. https://doi.org/10.1080/10503307.2012.747013

Meehan, K. B., Levy, K. N., & Clarkin, J. F. (2012). Construct validity of a measure of affective communication in psychotherapy. *Psychoanalytic Psychology*, *29*(2), 145–165. https://doi.org/10.1037/a0027450

Mohr, J. J., Gelso, C. J., & Hill, C. E. (2005). Client and counselor trainee attachment as predictors of session evaluation and countertransference behavior in first counseling sessions. *Journal of Counseling Psychology*, *52*(3), 298–309. https://doi.org/10.1037/0022-0167.52.3.298

Moltu, C., Binder, P.-E., & Nielsen, G. H. (2010). Commitment under pressure: Experienced therapists' inner work during difficult therapeutic impasses. *Psychotherapy Research: Journal of the Society for Psychotherapy Research*, *20*(3), 309–320. https://doi.org/10.1080/10503300903470610

Multon, K. D., Patton, M. J., & Kivlighan, D. M. (1996). Development of the Missouri identifying transference scale. *Journal of Counseling Psychology*, *43*(3), 243–252. https://doi.org/10.1037/0022-0167.43.3.243

Munder, T., Wilmers, F., Leonhart, R., Linster, H. W., & Barth, J. (2010). Working alliance inventory-short revised (WAI-SR): psychometric properties in outpatients and inpatients. *Clinical*

Psychology and Psychotherapy, *17*(3), 231–239. https://doi.org/10.1002/cpp.658.

Muran, J. C., Safran, J. D., Gorman, B. S., Samstag, L. W., Eubanks-Carter, C., & Winston, A. (2009). The relationship of early alliance ruptures and their resolution to process and outcome in three time-limited psychotherapies for personality disorders. *Psychotherapy: Theory, Research, Practice, Training*, *46*(2), Article 2. https://doi.org/10.1037/a0016085

Najavits, L. M., Griffin, M. L., Frank, A., Weiss, R. D., Liese, B. S., Thompson, H., Nakayama, E., Siqueland, L., Daley, D., & Onken, L. S. (1995). Therapists' emotional reactions to substance abusers: A new questionnaire and initial findings. *Psychotherapy: Theory, Research, Practice, Training*, *32*(4), 669–677. https://doi.org/10.1037/0033-3204.32.4.669

Orlinsky, D. E., & Howard, K. I. (1966). *Psychotherapy session report, form P and form T*. Institute of Juvenile Research.

Pallagrosi, M., Fonzi, L., Picardi, A., & Biondi, M. (2014). Assessing clinician's subjective experience during interaction with patients. *Psychopathology*, *47*(2), 111–118. https://doi.org/10.1159/000351589

Peluso, P. R., & Freund, R. R. (2018). Therapist and client emotional expression and psychotherapy outcomes: A meta-analysis. *Psychotherapy*, *55*(4), 461–472. https://doi.org/10.1037/pst0000165

Pierce, B. S., Perrin, P. B., & McDonald, S. D. (2020). Demographic, organizational, and clinical practice predictors of U.S. psychologists' use of telepsychology. *Professional Psychology: Research and Practice*, *51*(2), 184–193. https://doi.org/10.1037/pro0000267

Pierce, B. S., Perrin, P. B., Tyler, C. M., McKee, G. B., & Watson, J. D. (2021). The COVID-19 telepsychology revolution: A national study of pandemic-based changes in U.S. mental health care delivery. *American Psychologist*, *76*(1), 14–25. https://doi.org/10.1037/amp0000722

Prasko, J., Ociskova, M., Vanek, J., Burkauskas, J., Slepecky, M., Bite, I., Krone, I., Sollar, T., & Juskiene, A. (2022). Managing transference and countertransference in cognitive behavioral supervision: Theoretical framework and clinical application. *Psychology Research and Behavior Management*, *15*, 2129–2155. https://doi.org/10.2147/PRBM.S369294

Price, S. D., Callahan, J. L., & Cox, R. J. (2017). Psychometric investigation of competency benchmarks. *Training and Education in Professional Psychology*, *11*(3), 128–139. https://doi.org/10.1037/tep0000133

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review: DR*, *41*, 71–90. https://doi.org/10.1016/j.dr.2016.06.004

Racker, H. (2018). *Transference and countertransference* (1st ed.). Routledge. https://doi.org/10.4324/9780429484209

Revelle, W. (2024). *Psych: Procedures for psychological, psychometric, and personality research. R package version 2.4.6.26*, https://personality-project.org/r/psych/

Rosseel, Y. (2012). lavaan: An *R* package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02.

Samejima, F. (2010). The general graded response model. In M. L. Nering & R. Ostini (Eds.), *Handbook of polytomous item response theory models* (pp. 77–107). Routledge.

Samstag, L. W., Batchelder, S. T., Muran, J. C., Safran, J. D., & Winston, A. (1998). Early identification of treatment failures in short-term psychotherapy. An assessment of therapeutic alliance and interpersonal behavior. *The Journal of Psychotherapy Practice and Research*, *7*(2), 126–143.

San, L., & Arranz, B. (2023). Effective patient-physician communication in the era of neuropsychopharmacology. *European Neuropsychopharmacology*, *71*, 1–2. https://doi.org/10.1016/j.euroneuro.2023.02.007

Stefana, A. (2017). *History of countertransference: From Freud to the British object relations school*. Routledge. https://doi.org/10.4324/9781315445601

Stefana, A., Bulgari, V., Youngstrom, E. A., Dakanalis, A., Bordin, C., & Hopwood, C. J. (2020). Patient personality and psychotherapist reactions in individual psychotherapy setting: A systematic review. *Clinical Psychology & Psychotherapy*, *27*(5), 697–713. https://doi.org/10.1002/cpp.2455

Stefana, A., Damiani, S., Granziol, U., Provenzani, U., Solmi, M., Youngstrom, E. A., & Fusar-Poli, P. (2025). Psychological, psychiatric, and behavioral sciences measurement scales: Best practice guidelines for their development and validation. *Frontiers in Psychology*, *15*, 1494261. https://doi.org/10.3389/fpsyg.2024.1494261

Stefana, A., Fusar-Poli, P., Gnisci, C., Vieta, E., & Youngstrom, E. A. (2022). Clinicians' emotional reactions toward patients with depressive symptoms in mood disorders: A narrative scoping review of empirical research. *International Journal of Environmental Research and Public Health*, *19*(22), Article 22. https://doi.org/10.3390/ijerph192215403

Stefana, A., Fusar-Poli, P., Langfus, J., Vieta, E., & Youngstrom, E. (2024). Development and preliminary validation of the Clinician Affective REsponse (CARE) scale. *Research in Psychotherapy: Psychopathology, Process and Outcome*, *27*(1), 736. https://doi.org/10.4081/ripppo.2024.736.

Stefana, A., Fusar-Poli, P., Vieta, E., & Youngstrom, E. A. (2024). Assessing the patient's affective perception of their psychotherapist: Validation of the in-session patient affective reactions questionnaire. *Frontiers in Psychiatry*, *15*, 1346760. https://doi.org/10.3389/fpsyt.2024.1346760

Stefana, A., & Hill, C. E. (under review). Session evaluation scale: Psychometric evaluation and development of short versions.

Stefana, A., Langfus, J. A., Vieta, E., Fusar-Poli, P., & Youngstrom, E. A. (2023). Development and initial validation of the in-session patient affective reactions questionnaire (SPARQ) and the rift In-session questionnaire (RISQ). *Journal of Clinical Medicine*, *12*(15), Article 15. https://doi.org/10.3390/jcm12155156

Stefana, A., Youngstrom, E. A., Chen, J., Hinshaw, S., Maxwell, V., Michalak, E., & Vieta, E. (2020). The COVID-19 pandemic is a crisis and opportunity for bipolar disorder. *Bipolar Disorders*, *22*(6), 641–643. https://doi.org/10.1111/bdi.12949

Stefana, A., Youngstrom, E. A., Hopwood, C. J., & Dakanalis, A. (2020). The COVID-19 pandemic brings a second wave of social isolation and disrupted services. *European Archives of Psychiatry and Clinical Neuroscience*, *270*(6), 785–786. https://doi.org/10.1007/s00406-020-01137-8

Stiles, W. B. (1980). Measurement of the impact of psychotherapy sessions. *Journal of Consulting and Clinical Psychology*, *48*(2), 176–185. https://doi.org/10.1037/0022-006X.48.2.176

Stiles, W. B. (2002). *Session evaluation questionnaire: Structure and use*. https://wbstiles.net/session_evaluation_questionnaire.htm

Stiles, W. B., Reynolds, S., Hardy, G. E., Rees, A., Barkham, M., & Shapiro, D. A. (1994). Evaluation and description of psychotherapy sessions by clients using the session evaluation questionnaire and the session impacts scale. *Journal of Counseling Psychology*, *41*(2), 175–185. https://doi.org/10.1037/0022-0167.41.2.175

Streiner, D. L., Norman, G. R., & Cairney, J. (2015). *Health measurement scales: A practical guide to their development and use* (5th ed.). Oxford University Press.

Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2019). *Using multivariate statistics* (7th ed.). Pearson. https://catalog.lib.unc.edu/catalog/UNCb11408088

Tanzilli, A., Colli, A., Del Corno, F., & Lingiardi, V. (2016). Factor structure, reliability, and validity of the therapist response questionnaire. *Personality Disorders: Theory. Research, and Treatment*, *7*(2), 147–158. https://doi.org/10.1037/per0000146

Tanzilli, A., Majorana, M., Fonzi, L., Pallagrosi, M., Picardi, A., Fornari, M. A. C. d., Biondi, M., & Lingiardi, V. (2018). Relational variables in short-term psychodynamic psychotherapy: An effectiveness study. *Research in Psychotherapy: Psychopathology, Process, and Outcome*, *21*(3), 327. https://doi.org/10.4081/ripppo.2018.327

Tellegen, A., Watson, D., & Clark, L. A. (1999a). Further support for a hierarchical model of affect: Reply to green and salovey. *Psychological Science*, *10*(4), 307–309. https://doi.org/10.1111/1467-9280.00159

Tellegen, A., Watson, D., & Clark, L. A. (1999b). On the dimensional and hierarchical structure of affect. *Psychological Science*, *10*(4), 297–303. https://doi.org/10.1111/1467-9280.00157

Tishby, O., & Wiseman, H. (2022). Countertransference types and their relation to rupture and repair in the alliance. *Psychotherapy Research*, *32*(1), 16–31. https://doi.org/10.1080/10503307.2020.1862934

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*(1), 4–70. https://doi.org/10.1177/109442810031002

Wampold, B. E. (2015). How important are the common factors in psychotherapy? An update. *World Psychiatry*, *14*(3), 270–277. https://doi.org/10.1002/wps.20238

Wampold, B. E., & Flückiger, C. (2023). The alliance in mental health care: conceptualization, evidence and clinical applications. *World psychiatry: official journal of the World Psychiatric Association (WPA)*, *22*(1), 25–41. https://doi.org/10.1002/wps.21035.

Werbart, A., Gråke, E., & Klingborg, F. (2022). Deadlock in psychotherapy: A phenomenological study of eight psychodynamic therapists' experiences. *Counselling Psychology Quarterly*, *35*(4), 744–762. https://doi.org/10.1080/09515070.2020.1863186

Zilcha-Mano, S., Muran, J. C., Hungr, C., Eubanks, C. F., Safran, J. D., & Winston, A. (2016). The relationship between alliance and outcome: Analysis of a two-person perspective on alliance and session outcome. *Journal of Consulting and Clinical Psychology*, *84*(6), 484–496. https://doi.org/10.1037/ccp0000058

Appendix

Clinician Affective REsponse (CARE) Scales

The following are 15 statements that psychotherapists might use to describe how they feel toward their patients. Think about your last session with that specific patient, then read each statement and rate it on the extent to which it is true of the way how you felt during that session. Select the response that corresponds with your answer placing a cross on the appropriate number. Do not worry if your responses appear to be inconsistent, as people often experience mixed and conflicting feelings.

	During the last therapy session, how much did	Not at all	A little	Somewhat to very much
1	I really like this patient as a person.	0	1	2
2	I do all I could for this patient to a greater extent than for my other patients.	0	1	2
3	I feel frustrated in my efforts to help this patient.	0	1	2
4	I wish I could give this patient the genuine love and care that they never received, need, or deserve.	0	1	2
5	I think what we are working on was interesting.	0	1	2
6	I feel like I was incompetent or "not good enough" to help this patient.	0	1	2
7	I feel tenderness towards this patient, more than I usually feel for my other patients.	0	1	2
8	I feel hopeless for this patient.	0	1	2
9	I feel annoyed, irritated, or angry with this patient.	0	1	2
10	I feel more protective of this patient than of most patients in my care.	0	1	2
11	I feel enthusiastic about working together with this patient.	0	1	2
12	I find it hard to step into their inner world.	0	1	2
13	I was more concerned with this patient's feelings, needs, and wishes than with other patients.	0	1	2
14	I feel happy to see this patient.	0	1	2
15	I feel overall comfortable and enjoy working with this patient.	0	1	2

Positively engaged: 1, 5, 11, 14, and 15. Enmeshed: 2, 4, 7, 10, and 13. Stuck: 3, 6, 8, 9, and 12.