Development and preliminary validation of the Clinician Affective REsponse (CARE) scale

Alberto Stefana,[1] Paolo Fusar-Poli,[1-3] Joshua A. Langfus,[4] Eduard Vieta,[5] Eric A. Youngstrom[4,6,7]

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

ABSTRACT

The present study reports on the development and validation of the clinician affective response (CARE) scale. The CARE scale was designed as a self-report measure of therapists' patterns of thoughts, feelings, and behaviors toward the patient during an individual psychotherapy session. An initial pool of 116 items was generated, and its quality was evaluated by subject matter experts. Validation data were gathered from licensed psychotherapists (n=554). We used exploratory factor analysis and item response theory-graded response modeling to select items, confirmatory factor analysis to test how well the factor structure fit the data, and k-fold cross-validation to ascertain the robustness of the model. Criterion validity was evaluated by correlating the scores of the scale with the characteristics of therapists, patients, and treatment. The selected model consists of 15 items and a 3-factor structure, which showed excellent model fit, good internal consistency, and evidence of criterion validity. The CARE scale, short and quick to complete, enables therapists to reflect on and recognize their inner experiences and quantify these experiences in ways conducive to statistical analysis and research. Furthermore, the monitoring of these affective reactions toward their patients can guide therapeutic interventions and inform clinical supervisors.

Key words: therapeutic relationship, therapist emotional reaction, psychotherapy process research, self-report measure, CARE scale.

Introduction

Psychotherapy can (re)activate a wide range of emotions not only in patients but also in therapists (Atzil-Slonim *et al.*, 2018; Kimerling *et al.*, 2000; Pope & Tabachnick, 1993). Among all the affective states and reactions experienced by a therapist in the here and now of the session, those felt towards the patients

are particularly important for two main reasons. Firstly, they are a complex and jointly created phenomenon that involves contributions from both members of the therapeutic dyad (Gelso & Hayes, 2007; Stefana, 2017). As a consequence, the overall affective reactions of the therapist are potentially informative about specific personality traits or pathologies (Stefana *et al.*, 2020; Tanzilli *et al.*, 2022) or mood states (Stefana, Fusar-Poli *et al.*, 2022; Stefana, Youngstrom *et al.*, 2022) of the patient. Secondly, countertransference reactions, *i.e.*, a subset of reactions in which unresolved conflicts of the therapist have been implicated (Gelso & Hayes, 2007), are modestly related to poorer therapy outcomes, while their successful management is associated with better outcomes (Hayes *et al.*, 2018).

Countertransference awareness appears to have a moderate effect on the relationship between time and treatment outcome (Abargil & Tishby, 2022). Furthermore, negative countertransference patterns are correlated with working alliances (Bhatia & Gelso, 2018) and positively associated with more ruptures and less resolution (Tishby & Wiseman, 2022). Conversely, positive countertransference patterns predict resolution when the clinician repeats positive patterns but, differently, predict ruptures when they try to repair previous negative patterns (Tishby & Wiseman, 2022).

The crucial importance of the emotions experienced during the session is further supported by a recent study showing that the emotions of the therapists before the session predict neither their empathy with the patient nor the quality of the session (Chui *et al.*, 2022). This is particularly important when considering that an empathetic and positive relationship is not only essential for facilitating change but also necessary for the successful execution of psychotherapeutic interventions (Hayes & Vinca, 2017; Peluso & Freund, 2018). This implies that a therapist's inner experience during sessions can predict certain negative outcomes.

Therefore, therapists need to regularly monitor their own affective reactions toward their patients, preferably using brief, validated, and reliable self-administered tools. However, although some measures have been developed to assess the therapist's overall reactions, most of them are not fully validated (Brody & Farber, 1996) or are quite long and investigate not a single session but a series of consecutive sessions (Betan *et al.*, 2005). Scale length is a key consideration given the complexity of the therapeutic process, which often requires the collection of multiple measures (Wampold & Flückiger, 2023). Furthermore, the two most frequently used scales, the therapist response questionnaire (TRQ) (Betan *et al.*, 2005) and the assessment of clinician's subjective experience (ACSE) (Pallagrosi *et al.*, 2014), require authorization from their developers for use. The ACSE, a measure of clinicians' subjective experience during their first interaction with a patient, consists of 46 items. It comprises five factorially derived subscales: tension, difficulty in attunement, engagement, disconfirmation, and impotence (Pallagrosi *et al.*, 2014). This scale was validated and has mostly been used in psychiatric contexts, subsequently being tested in psychotherapeutic settings (Tanzilli *et al.*, 2018). The TRQ, with its 79 items, evaluates psychotherapists' in-session reactions to patients over a typical period of 6-8 sessions. It encompasses a broad spectrum of cognitive, affective, and behavioral responses. The original English version identified an 8-factor structure: overwhelmed/disorganized, helpless/inadequate, positive, special/overinvolved, sexualized (describing sexual feelings towards the patient or experiences of sexual tension), disengaged, parental/protective, and criticized/mistreated (Betan *et al.*, 2005). Validation studies for the TRQ's translations have indi-

cated slightly varied factorial structures (Berg *et al.*, 2019; Tanzilli *et al.*, 2016).

The affective reaction of the therapist toward the patient during sessions warrants careful scrutiny also within clinical supervision (Shafranske & Falender, 2008), particularly when working with challenging patients, regardless of the psychotherapeutic approach or the theoretical orientation of the supervisors (Prasko *et al.*, 2022). For this purpose, there are supervisor-rated scales designed to evaluate the supervisee's handling of their reactions toward patients, such as the countertransference management scale (Pérez-Rojas *et al.*, 2017). However, the use of a self-report scale, either in combination with or as a substitute for supervisor-rated tools, could be valuable. Such scales would allow supervisors to effectively monitor a key aspect of the therapeutic process.

It is important to note that, in addition to self-assessment methods, clinicians' in-session emotional reactions can be evaluated using observer-rated scales. Such scales include the countertransference behavior inventory (Friedman & Gelso, 2000), Q-sort techniques applied to recordings of single therapy sessions, such as the Psychotherapy Process Q-set (Jones, 2000), and the core conflictual relationship theme method for the analysis of narratives of clinicians' interactions with their patients (Messina *et al.*, 2018).

It follows from the above that there remains a pressing need for freely usable, theoretically and psychometrically sound, brief self-reported assessment tools of the reactions of the therapist toward the patient during a single session that are currently available.

Overview of the present study

The objective of this study was to develop and test the psychometric properties of a self-report measure of the patterns of thoughts, feelings, and behaviors of therapists toward patients that is clinically sophisticated, psychometrically robust, and short enough to be used in research, training supervision, and real-world clinical settings. We analyzed i) items' descriptive characteristics and factor loadings; ii) confirmatory factor analysis (CFA); iii) internal consistency and score precision; iv) criterion validity. Although we did not have a specific hypothesis about the final model of the scale, we anticipated very small or no correlations between the scale/subscales and the sociodemographic characteristics of both patients and therapists. We also expected small to moderate correlations between some of the subscales and the diagnoses of personality disorder by patients, as well as their level of global functioning.

Methods

Participants

The development sample for the scale comprised 556 licensed psychotherapists, predominantly female (70%, n=392 for sex, n=390 for gender). The most represented age group was those aged 60 years and older (44%, n=244), followed by those aged 50 to 59 years (25%, n=138). In terms of therapeutic orientation, a plurality was psychodynamically oriented (35%, n=193), with cognitive-behavioral and eclectic approaches each accounting for 19% of the sample. About half of the respondents had ≥20 years of post-license clinical experience and spent ≥21 hours per week practicing therapy. Each of the 556 therapists reported about one single patient. Regarding the patients, 68% (n=376) were female;

55% (n=307) ranged between 18 and 39 years of age. Most patients (82%, n=454) had a diagnosis of psychiatric disorder and were rated mildly (28%, n=156) or moderately (25%, n=139) ill on the clinical global impressions (CGI) scale; the average global assessment of functioning (GAF) score was 65.0 (14.2). Many were in psychotherapy for more than 21 months (52%, n=289), typically at a frequency of one session per month (51%, n=285). Almost all the therapeutic dyads (92%, n=509) had the target session less than one week before filling out the study survey. Table 1 reports the demographics and professional characteristics of therapists, and Table 2 provides patient demographics and clinical and treatment characteristics.

Item generation

The development of the present scale followed best practice recommendations (DeVellis & Thorpe, 2022). This process included defining the construct to be measured, generating a large item pool using a mixed-methods strategy and involving representatives of the target population, conducting an expert review, administering the scale to an appropriately selected sample, and

running both traditional (test-level) and modern (item-level) statistical analyses. The generation of items relied on theoretical models and definitions that motivated the content of the item (DeVellis & Thorpe, 2022). Central to the design of our scale is the hierarchical model of affect (Tellegen *et al.*, 1999), a conceptual framework that describes individual differences in emotional experience. This model can be represented in terms of a 2-dimen-

Table 1. Therapist demographic and professional characteristics (N=556).

Demographic characteristics	% (n)
Biological sex	
Male	29 (160)
Female	70 (392)
Prefer not to say	1 (4)
Gender	
Male	29 (159)
Female	70 (390)
Non-binary/third gender	1 (7)
Age, years	
18-29	1 (5)
30-39	11 (62)
40-49	19 (107)
50-59	25 (138)
≥60	44 (244)
Professional characteristics/Professional background	
Psychologist	30 (168)
Psychiatrist	11 (63)
Social worker	20 (109)
Nurse	1 (5)
Licensed marriage/family therapist	11 (61)
Other	27 (150)
Theoretical orientation	
Psychodynamic/psychoanalytic	35 (193)
Cognitive-behavioral	19 (107)
Eclectic	19 (104)
Humanistic	9 (52)
EMDR	6 (34)
Other	12 (66)
Post-licensed experience	
1 to 9 years	20 (112)
10 to 19 years	31 (172)
≥20 years	49 (272)
Time spent practicing therapy	
1 to 10 hours per week	15 (83)
11 to 20 hours per week	31 (171)
≥21 hours per week	54 (302)

EMRD, eye movement desensitization and reprocessing.

Table 2. Patient demographic, clinical, and treatment characteristics (N=556).

	% (n)
Demographic characteristics	
Biological sex	
Female	68 (376)
Age, years	
18-29	27 (152)
30-39	28 (155)
40-49	20 (109)
50-59	15 (83)
≥60	10 (57)
Clinical characteristics	
Presence of any psychiatric mental disorder	82 (454)
Any anxiety disorder	45 (251)
Any (unipolar) depressive disorder	24 (135)
Any bipolar or related disorder	5 (29)
Any dissociative disorder	6 (32)
Any obsessive compulsive disorder	8 (42)
Any personality disorder	16 (89)
Cluster A	1 (2)
Cluster B	9 (52)
Cluster C	5 (26)
Any substance disorder	7 (41)
Any trauma- and stressor-related disorders	30 (165)
CGI	
Normal, not at all ill	28 (155)
Borderline mentally ill	13 (72)
Mildly ill	28 (156)
Moderately ill	25 (139)
Markedly/Severely ill	6 (34)
GAF, mean (SD)	65.0 (14.2)
Treatment characteristics	
Therapy length	
0 to 3 months	19 (106)
4 to 6 months	13 (75)
7 to 12 months	16 (86)
13 to 24 months	18 (100)
>24 months	34 (189)
Session frequency	
≤1 per month	8 (47)
2 to 3 per month	27 (148)
1 per week	51 (285)
≥2 per week	14 (76)
Session attendance	
In person	41 (229)
Telephone call	9 (47)
Video call	50 (280)
Patient-therapist biological same-sex match	69 (381)

CGI, clinical global impressions; GAF, global assessment of functioning; SD, standard deviation.

sional space in which emotions can be plotted based on two primary dimensions: valence and arousal. We also incorporated emotions related to the social/interpersonal dimension of the dominance behavioral system (Johnson *et al.*, 2012). This system probes the nuances of how individuals assert themselves or take a backseat in social contexts. By assimilating this principle, our scale becomes adept at identifying emotions linked to an individual's social standing and interactions. Moreover, in acknowledgment of the diverse spectrum of emotional motivations, our scale prioritizes the recognition of discrete emotions, each with its unique driving force. For example, anger can catalyze confrontational behavior, while fear can lead to evasive actions. Such detailed emotional categorizations, inspired by Izard *et al.* (1993), equip our scale to grasp these subtle emotional nuances with precision. Lastly, our aspiration for this scale to be more than just a theoretical tool but also a practical instrument in the therapeutic and counseling arenas led us to integrate evidence-based therapeutic process models (Gelso, 2014; Wampold, 2015). These models act as guideposts, ensuring our scale's applicability and relevance in real-world clinical settings.

To encapsulate it, the genesis of our scale represents a harmonious melding of time-tested knowledge with innovative insights, culminating in a comprehensive tool tailored to gauge emotional states intricately intertwined with social interactions and personal motivations.

An initial item pool was generated by combining inductive and deductive approaches. The authors reviewed 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), selected items of interest, and edited them to make them appropriate for reflecting the affective, cognitive, and behavioral responses that a psychotherapist can experience during either a "good/not-so-difficult" or "difficult" session. In addition, the authors and two experienced psychotherapists (one with a psychological background and the other with a psychiatric one) generated novel items. The content validity of the initial item pool was reviewed by six experienced clinicians of different orientations (three cognitive-behavioral, two eclectic, and one psychodynamic). Five items were eliminated as redundant or not appropriate, and about twenty were reworded and reviewed for clarity.

Consistently with the empirical understanding of the measurement of feeling states (Russell & Carroll, 1999) and the recommendations for the reduction of cognitive load and the improvement of response accuracy (DeVellis & Thorpe, 2022), we wrote the items to sample from negative as well as positive affective domains, writing the anchors to avoid a bipolar format within items. Thus, items used a Likert scale on five levels: 0=not at all, 1=a little, 2=somewhat, 3=a lot, and 4=very much. Higher scores indicate greater levels of emotional activation.

Additional measures

Psychotherapists completed a 7-item demographic and professional data form, which recorded their age, biological sex, gender, professional background, theoretical orientation, post-license clinical experience, and average weekly time spent practicing therapy. They were asked to check their appointment book, identify the last adult patient (≥18 years old) they saw for an individual session, and complete a patient demographic and clinical data form and the questionnaire items. The former gathered information on the patient's age, sex, clinical diagnosis, and global functioning (using both the GAF and the CGI scales), as well as the

psychotherapeutic treatment's setting, session frequency, and length. Full instructions were provided for filling out the GAF and the CGI. The latter, as explained to participants in the instructions, contained a series of statements that describe how a therapist might feel toward a patient and rate each statement based on how true it was to them during the last session.

Procedures

All procedures and materials were reviewed and exempted by the Institutional Review Board of the University of North Carolina at Chapel Hill (UNC-CH) (IRB number 22-0356). A small pilot study verified the clarity and wording of both instructions and items and debugged Qualtrics programming. Psychotherapists (n=556) were then recruited through an email invitation that was distributed among members of mental health professional associations and registries from April through May 2022. The inclusion criterion was being a licensed psychotherapist; the sole exclusion criterion was not having at least one adult patient (≥18 years old) currently in psychotherapy treatment. Participants were instructed to review their appointment book and identify the most recent adult patient (aged 18 years or older) they had seen for an individual session. They were informed that the remainder of the survey would concentrate specifically on that patient and the therapeutic relationship established with them. Psychotherapists completed the online survey *via* Qualtrics, with an average duration of 20 minutes for completion. Each psychotherapist could participate once.

Sample size

According to the literature, a minimum sample size of 500 or more respondents is recommended for exploratory factor analysis (EFA) (Kyriazos, 2018). Similarly, for CFA, a sample size of up to 500 respondents is suggested, especially under challenging conditions such as low communalities and the presence of numerous weakly determined factors (Kyriazos, 2018). Consequently, our goal was to collect data from at least 500 therapists.

Statistical analyses

Data analyses followed the steps outlined in a previous work (Stefana *et al.*, 2023). Specifically, the initial candidate item pool was examined using descriptive statistics, and items exhibiting inadequate variability were dropped. All the items were evaluated on a 5-point Likert scale; however, we observed sparse data in the categories "a lot" and "very much." Statistically, such sparse data can impact the reliability of subsequent analyses. Clinically, distinguishing between any notable presence of a symptom, as indicated by "somewhat," *versus* its minimal or absent state is paramount. While there are nuanced differences between "somewhat," "a lot," and "very much," our primary clinical objective is to identify the presence of significant reactions. Given these considerations and to strengthen the robustness of our analysis, we combined "somewhat," a lot," and "very much" into a single category. The Kaiser-Meyer-Olkin test and the Bartlett test of sphericity assessed the data's suitability for factor analysis. Parallel analysis based on multiple factor retention criteria was performed using the R package *EFAtools* v0.4.3 to help determine the number of factors that might be tapped by enough items to support interpretation.

Subsequently, iterative EFA was performed to analyze a polychoric matrix with pattern coefficients with PROMAX ro-

tation for simple structure. Items with factor loading <.40 and those with ≥.30 on multiple factors were removed (Boateng *et al.*, 2018). CFA using the R package *lavaan* v0.6-12 was used to check the unidimensionality of the latent variable for the identified factors separately, as the item response theory (IRT) requires unidimensional latent variables. Graded response model (GRM) IRT models were applied using the R package *mirt* v1.37.1 to evaluate the option characteristics of the scales identified through factor analysis. The selection of the final scale items involved picking items across the range of theta levels. A CFA tested the fit of the final factor model. *K*-fold cross-validation was then implemented using the R package *kfa* v0.2.1 to verify the robustness of the model. The results were compared using maximum likelihood, maximum likelihood with robust standard errors, and unweighted least squares estimators, and the model misspecification was checked by examining the statistical power in combination with the modification indices and expected parameter changes (Watkins, 2018). Reliability statistics for the final scale were computed using raw items and 1000 bootstrapped replications (Revelle & Condon, 2019). Correlations between subscales and therapist demographic-professional features, patient demographic-clinical features, and psychotherapy treatment characteristics provided preliminary data on the criterion validity of questionnaire scores.

Results

Factor analyses

In the preliminary item reduction, we kept only items with ≥5% of responses "somewhat to very much" to ensure that the items included described clinical phenomena that usually occur in an individual psychotherapy session (Devellis & Thorpe, 2021). After this procedure, 38 of the initial 116 items (*Supplementary Table 1*) remained.

The Kaiser-Meyer-Olkin test (.89) and the Bartlett test of sphericity (p<.001) verified the suitability of the data for factor analysis.

Parallel analysis with either EFA or principal component analysis suggested four factors, whereas using comparison data or Hull's methods found two factors.

The first round of EFA was conducted by extracting four factors and evaluating their indicators and conceptual coherence. Items with unique loadings of <.40 and those that cross-loaded without a clear dominant loading were deleted to improve scale interpretability. After the second EFA round, 21 items remained in contention, producing a 3-dimensional model. Items showed appreciable loadings on the corresponding factor: the smallest loadings for factors 1, 2, and 3 were respectively .51, .40, and .44; while median loadings were .80, .76, and .57). Factor 1 items showed relatively low endorsement rates that resulted in a slight negative skew, while items on factors 2 and 3 showed normal item distributions.

CFAs were conducted individually for the three factors. The results revealed that a one-factor model was a suitable fit for each factor, suggesting that each scale possessed a single latent variable, making them suitable for IRT analysis. *Supplementary Table 2* reports the fit indices for each model.

Samejima's GRM evaluated the items for each of the three factors: factor 1="positively engaged," factor 2="enmeshed," and factor 3="stuck." For each factor, five items that offered high information across a wide range of theta values were chosen. Table 3 provides the item discrimination and difficulty parameters of the final scale. The positively engaged factor had reliability >.80 from theta of -2.8 to -.6. In contrast, the enmeshed factor showed reliability >.80 at theta ranging from .1 to 1.9, and the stuck factor had reliability >.80 between theta .0 to +2.1. *Supplementary Figure 1* shows the item characteristic curve and reliability for the scale scores.

The 3-factor model showed an excellent fit for the data: $X^2(df=87)=133.43$, comparative fit index (CFI)=.98, Tucker-Lewis index (TLI)=.97, root mean square error of approximation (RMSEA)=.03 [90% confidence interval (CI) (.03, .04)], and standardized root mean square residual (SRMR)=.04. Furthermore, k-fold cross-validation with k=2 was performed to test the robustness of the model and provided the following fit indices: $X^2(df=87)=125.0$, CFI=.96, TLI=.96, RMSEA=.04 [90% CI (.03, .04)], and SRMR=.05. Lastly, a bifactor model with a general factor and the three specific factors was tested and demonstrated acceptable fit indices: $X^2(df=75)=216.51$, CFI=.93, TLI=.90, RMSEA=.06 [90% CI (.05, .07)], and SRMR=.11. *Supplementary Figure 2* shows the measurement models presenting the final sets of parameter estimates.

Table 3. Item option characteristics for the three factors based on item response theory models.

Scale	Item content	α	β1	β2
Factor 1: positively engaged	I really like them as a person.	2.89	-2.36	-1.38
	I think what we are working on was interesting.	1.77	-2.48	-1.73
	I feel overall comfortable and enjoy working with them.	1.96	-2.24	-1.71
	I feel enthusiastic about working together with them.	2.60	-2.07	-1.14
	I feel happy to see them.	2.81	-2.03	-.90
Factor 2: enmeshed	I wish I could give them the genuine love and care that they never received, needed, or deserved.	1.03	.12	1.60
	I do all I could for them to a greater extent than for my other patients.	1.94	1.29	2.16
	I was more concerned with their feelings, needs, and wishes than with other patients.	1.88	.97	1.95
	I feel tenderness towards them, more than I usually feel for my other patients.	1.85	.06	1.29
	I feel more protective of them than of most patients in my care.	3.14	.39	1.36
Factor 3: stuck	I feel frustrated in my efforts to help them.	4.17	.37	1.28
	I feel like I was incompetent or 'not good enough' to help them.	1.40	1.08	2.48
	I find it hard to step into their inner world.	1.27	.59	2.70
	I feel hopeless for them.	2.17	.93	1.97
	I feel annoyed, irritated, or angry with them.	2.31	1.05	2.08

α=discrimination parameter; β=difficulty parameter.

Internal consistency and score precision

The 3-factor Clinician Affective REsponse (CARE) scale (see the Appendix), consisting of 15 items, exhibited strong internal consistency: Cronbach's a=.77 and Mc Donald's w total=.83. Each of the subscales also demonstrated good internal consistency: positive engaged factor (k=5, a=.78, w=.78, average inter-item r=.41), enmeshed factor (k=5, a=.74, w=.75, average inter-item r=.36), and Stuck factor (k=5, a=.76, w=.77, average inter-item r=.39). It is worth noting that, given the brevity of the subscales (k=5), the average correlation offers a more reliable measure of internal consistency than a (Streiner *et al.*, 2015).

The mean scores on positively engaged, enmeshed, and stuck factors were, respectively, 9.07 [standard deviation (SD)=1.75], 2.37 (SD=2.38), and 1.74 (SD=2.15). *Supplementary Table 3* includes descriptive statistics, reliability estimates, and standard errors of the final subscales. The scale score correlations aligned with the factor correlations in the 3-factor models (refer to *Supplementary Figure 1*). The most notable association was a moderate negative correlation between stuck and positively engaged scores (r=-.38, p<.0001). This was followed by a small positive correlation between stuck and enmeshed scores (r=.27, p<.0001), and a smaller positive correlation between enmeshed and positively engaged scores (r=.11, p<.001).

Table 4 reports the reliable change index's critical values (Wise, 2004) as measures of individual precision. Furthermore, a nomothetic benchmark based on the clinical distribution has been estimated for each factor score. The 5th percentile was calculated for the positively engaged score, indicating which score would be alarmingly low compared to the population distribution, and the 95th percentile for the enmeshed and stuck scores, above which the score would be alarmingly high. Estimates of minimally important differences were also included as a rule of thumb for what might often be perceived as a meaningful shift in affective tone by the therapist regarding the patient.

Criterion validity

The criterion validity of the CARE scale was assessed by associations with the demographic and professional characteristics of the therapist listed in Table 1, as well as the demographic, clinical, and treatment characteristics of the patient listed in Table 2.

As expected, the positively engaged subscale negatively correlated with the patient's diagnosis of cluster B personality disorder (r=-.17, p<.001) and positively with the patient's better global functioning (r=.17, p<.001 for the GAF).

The enmeshed subscale correlated with the patient's age (r=-.13, p<.01), as well as session frequency (r=.12, p<.01) and therapy duration (r=.14, p<.001).

Lastly, the stuck subscale was associated with patient's global functioning (r=-.25, p<.001 for GAF score; r=.19, p<.001 for CGI score), diagnosis of cluster C personality disorder (r=-.19,

p<.001), or any trauma- and stressor-related disorder (r=-.12, p<.05). Furthermore, it was associated with therapy duration (r=.15, p<.001) and session frequency (r=.11, p<.001) (Table 5).

Discussion

The present study aimed to develop, test, and gather initial validity data for a short and freely available self-reported scale assessing the patterns of affective, cognitive, and behavioral responses psychotherapists have to their patients during a session. We began with an initial item pool larger than the intended final length of the inventory to guarantee adequate coverage of the constructs and guard against poor internal consistency, and we kept items with adequate endorsement and variability to ensure that content reflected clinical experiences that take place in a psychotherapy session reasonably frequently (DeVellis & Thorpe, 2022). Then, we iteratively refined the item set by examining item descriptive characteristics, clear factor loadings, values of retained dimensions, and clinical meaningfulness (DeVellis & Thorpe, 2022). Factor analyses indicated that the scale measures 3-dimensional factors (positively engaged, enmeshed, and stuck) that are theoretically sound and supported by theoretical and empirical evidence, clinically meaningful, and internally consistent and reliable (Betan *et al.*, 2005; Colli *et al.*, 2014; Tanzilli *et al.*, 2016).

The three scales have different features, suggesting potentially differentiated roles in the context of therapy research. The Positively Engaged factor includes items that describe feelings of appreciating the patient as a person, of contentment in seeing them, of being comfortable and enthusiastic about working with them, and of being interested in what they are working on. It delineates a positive experience (from the therapist's perspective) of the therapeutic relationship and process, which appears to be characterized by empathic attunement and positive alliance, potentially regardless of the current therapy outcomes. The Positive Engagement scale is the "easiest" in an IRT sense, meaning that it is quite usual to get high scores after most therapy sessions. The average was very close to the top of the possible range of scores (*M*~90% of the maximum possible), although there was still wide variation in scores in our sample. Findings suggest that typical sessions have strong positive engagement from the therapist (total scores <6 occurred in only 5% of cases). A high level of positive engagement for the therapist may offer the dyad a foundation for more challenging work. Positive engagement is likely to correlate highly with working alliance, and might be protective against rupture, although these hypotheses need to be tested in future work.

The relational component of treatment engagement has predominantly been examined through the lens of the therapeutic alliance. Recent meta-analyses have established a strong positive correlation between the patient-therapist alliance (as evaluated by patients, observers, and therapists) and treatment outcomes

Table 4. Descriptive statistics, internal consistency reliability, precision, and inter-scale correlations.

Clinical change benchmarks	Factor 1 Positively engaged	Factor 2 Enmeshed	Factor 3 Stuck
90% critical change	1.96	2.66	2.41
95% critical change	2.33	3.16	2.86
Minimally important difference (MID)	.88	1.19	1.08
Jacobson benchmark threshold (5% tail)	<5.64	>7.03	>5.95

A minimally important difference was operationally defined as d=.5 (Norman *et al.*, 2003; Streiner *et al.*, 2015).

(Flückiger *et al.*, 2018). These findings also emphasize the critical influence of therapists (*i.e.*, therapist effect) in this dynamic (Del Re *et al.*, 2021). This data suggests the need for more research focusing on affective engagement within the therapeutic process. Effective psychotherapy requires not only proactive patient engagement and contribution (Pope & Vasquez, 2016) but also active involvement from therapists (Stefana, 2017). Such reciprocal engagement is essential for achieving meaningful therapeutic results.

In contrast, the enmeshed scale and the stuck scale showed a lower central tendency in their score distributions, with means around 20% of the maximum possible range. The content of both scales was less frequent as a response to the session. The enmeshed factor contains items that describe the desire to give them love, activation to do for them, concern for their feelings/needs/wishes, tenderness, and tendency to protect them, which were considerably greater as compared to those for their other patients. Given that this dimension describes a pattern of responses that deviate from those the therapist typically has toward patients and does not relate to any diagnosis or clinical features of the patient, it is possible to hypothesize that the enmeshed factor describes what is more properly called countertransference (Hayes *et al.*, 2018).

The stuck factor consisted of items that describe the therapist feeling difficulty stepping into the patient's inner world, frustration in their efforts to help, incompetence, hopelessness, and annoyance towards the patient. These responses also appeared to be associated with more challenging cases or interactions. For both the enmeshed and stuck scales, even moderate elevations should be given attention, as they might indicate a negative process where the therapist experiences a loss of agency and reflective capacity during their interaction with the patient (Moltu *et al.*, 2010; Werbart *et al.*, 2022).

The low correlations between the CARE's three factors showed that these are three distinct yet related dimensions. Furthermore, they indicate that it is possible for a therapist to be positively engaged with a patient while deeply enmeshed in the relationship, or even while feeling stuck in it.

We evaluated the CARE scales' criterion validity by examining their association with patients' diagnoses of mental disorders and global functioning. We found that clinicians' patterns of responses related to specific psychiatric diagnoses in plausible and clinically coherent ways. In line with findings from previous studies (Betan *et al.*, 2005; Colli *et al.*, 2014; Tanzilli *et al.*, 2016), cluster B disorders were negatively associated with the therapists' positive engagement in the relationship, whereas cluster C disor-

Table 5. Criterion validity correlations or mean differences across patient diagnoses, demographics, and objective therapy characteristics.

Criterion variable	Factor 1 Positively engaged	Factor 2 Enmeshed	Factor 3 Stuck
Therapist age	.08	-.08	-.01
Therapist sex[a] (female higher)	.08	.01	-.04
Therapist gender[b] (female higher)	.09	.01	-.07
Professional background	.54	.88	1.57
Years of post-licensed experience	.01	-.06	-.02
Weekly time spent practicing therapy	-.05	-.06	-.03
Patient age	-.07	-.13**	.02
Patient biological sex (female higher)	-.02	.02	-.10
Any psychiatric disorder	-.02	-.04	-.05
Any anxiety disorder	-.01	.00	-.06
Any bipolar disorder	-.07	-.02	.08
Any depressive disorder	.05	.09	.02
Any dissociative disorder	.00	-.04	.04
Any obsessive compulsive disorder	-.02	-.03	.02
Any personality disorder	-.17***	.03	.18***
Cluster A	-.05	.00	-.05
Cluster B	-.17	.03	.09
Cluster C	-.08**	.00	.19***
Any substance disorder	-.00	.08	.07
Any trauma- and stressor-related disorder	.09	.03	-.12*
CGI	-.05	.07	.19***
GAF	.17***	-.07	-.25***
Therapy length (months, ordinal; see prior table)	.06	.14***	.15***
Session frequency (ordinal, see prior table)	.01	.12**	.11**
Session attendance (in person, telephone, video)	2.88	.73	2.41
Patient-therapist sex match (yes/no)	.03	.00	-.06

CGI, clinical global impressions; GAF, global assessment of functioning.
[a]Respondents who did not disclose their sex (*i.e.*, 1%) were excluded from this analysis; [b]respondents who reported their gender as non-binary/third gender (*i.e.*, 1%) were excluded from this analysis; *p<.05; **p<.01; ***p<.001, 2-tailed.
Coefficients are point-biserial correlations for dichotomized variables, analysis of variance F-value for the categorical variables "Professional background" and "Session attendance," Spearman correlations for ordinal variables, and Pearson correlations for continuous variables.

ders were negatively associated with feelings of being stuck and hopeless. Additionally, the reaction of feeling stuck was associated with dealing with a patient suffering from trauma- or stressor-related disorder, consistent with literature addressing clinical work with traumatized people (Lindy & Lindy, 2004). Furthermore, we found that when dealing with lower global functioning patients, therapists' responses were characterized by stronger (negative) feelings of being stuck and weaker (positive) feelings of being engaged with the patient, which is again consistent with prior work (Colli *et al.*, 2014).

The enmeshed and stuck scales were also significantly correlated to both therapy length and session frequency. A recent systematic review of the dose-response effect in routine psychotherapies, *i.e.*, the relationship between the dose (primarily length and frequency) of treatment and the subsequent probability of clinical improvement, reported a curvilinear relationship between the number of sessions and the odds of positive response to treatment, illustrating a trend of smaller and non-significant gains with extended treatment (Robinson *et al.*, 2020). It is plausible that as the duration of treatment lengthens, especially when carried out with a different frequency of sessions, therapists might sometimes feel stuck or enmeshed. More speculatively, higher enmeshment might sometimes lead to continuing the treatment beyond a point of diminishing returns in patient improvement.

The three dimensions of the CARE scale closely reflect fundamental dimensions of therapists' affective responses emerging in clinical encounters (Prasko *et al.*, 2022; Stefana *et al.*, 2020; Stefana, Fusar-Poli *et al.*, 2022). Using brief scales such as these could help allow therapists to identify patterns of emotional responses toward and ways in which they interact with their patients. The different content coverage across the three scales is likely to increase therapists' recognition of the types and amounts of their in-session reactions. The CARE scale might be useful in monitoring varying levels of these dimensions across sessions and/or assessing their associations with session and treatment outcomes. Furthermore, in alignment with recent proposals to integrate emotional competency and these components in clinical research and practice competency movement concerns (Gonsalvez *et al.*, 2020; Price *et al.*, 2017), the CARE scale could be a valuable option. The CARE scale could be instrumental in addressing emotional competency components, as many have proposed the integration of these components in clinical research and practice. Recent studies highlight the profound impact of therapists' countertransference management competencies (Hayes *et al.*, 2018) and, more generally, foundational and functional competencies (Dimmick *et al.*, 2023) on client outcomes. This highlights the relevance of a scale like the CARE scale in advancing emotional competency assessment and linking it to real-world clinical outcomes. The scale dimensions likely reflect not just a linear, objective response to the person, personality, and clinical features of the patient, but a more complex mixture derived also from the therapist's own idiosyncratic reactions, shaped by factors like training, orientation, and personal features, and the interactions of patient and therapist during the therapeutic encounter.

Strengths, limitations, and future directions

The present study has several strengths. The development of a short self-report tool is an important contribution to the operationalization of therapists' emotional, cognitive, and behavioral responses to the patient during a specific session, which constitutes a key element of the therapeutic relationship (San & Arranz,

2023). The CARE scale showed excellent psychometric properties in a large sample of practicing psychotherapists, drawn from various therapeutic orientations and settings. The suite of the three brief scales requires less than 2 minutes to complete, and it is easy to score. Moreover, we followed scale development procedural best practices and combined classical and modern test theories. These, combined with the sample employed, improve the likely external validity of the scales and supporting results.

However, there were also some limitations, including both conceptual and technical issues. First, therapists were the sole informants. This might constitute a possible bias in therapists' self-reporting of their own affective attitudes, thoughts, and behaviors. However, growing evidence provides support for using clinicians' ratings of their responses toward the patient (Betan *et al.*, 2005; Bhatia & Gelso, 2018). Second, psychometric properties should be re-assessed in a sample where scale items are not embedded in the larger original item pool to check for possible context effects. However, it should be mentioned that usually, these effects tend to be small in scales with homogeneous factor structure and strong subscale loadings, as is our case. Although we had a good sample size for scale development (Boateng *et al.*, 2018) and used the *k*-fold cross-validation method to evaluate the final model, the CFA was conducted on the same sample as the EFA. Third, now that a short scale has been developed, systematic exploration of the dependability of its subscales, test-retest stability, and sensitivity to treatment effects represents important next steps in validation (Revelle & Condon, 2019).

In future research, we will conduct a more comprehensive criterion validity analysis, exploring the relationships between our scales and other established measures of other elements of the therapeutic relationship, especially countertransference, real relationship, and working alliance. This would also allow us to evaluate the final model of the scale in a different sample, furthering our understanding of its robustness. Longitudinal research should investigate the use of the CARE scale as a self-report tool able to measure the weekly changes in the therapists' emotional responses toward a patient (which would represent a further step toward a measurement feedback system of therapists' in-session experiences). Lastly, prospective studies could add to our understanding of how therapists' subjective experiences unfold over the course of psychotherapy and influence different trajectories.

Conclusions

The therapist's pattern of responses toward the patient is a jointly created phenomenon that involves contributions from both members of the therapeutic couple; therefore, it can be a source of information about the patient. Recognition, processing, and management of these emotional responses do not just help inform therapists about their patient's personality functioning and interpersonal style (Stefana *et al.*, 2020; Stefana, Fusar-Poli *et al.*, 2022), but they can also improve therapy outcomes (Hayes *et al.*, 2018). It follows that it is crucial for therapists to accurately and systematically self-assess their inner experiences, perceptions of their patients, and competence in managing them. The CARE scale is short and quick to complete, making it easily implementable in clinical practice and supervision, as well as in research settings. By enabling therapists to self-rate a series of "common" affective reactions toward their patients on a normed scale measure with known clinical correlates, the CARE may help therapists reflect on and recognize their inner experiences and quantify these therapist experiences in ways that are amenable to

statistical analysis and research, as well as potentially used to guide therapeutic interventions, inform clinical supervisors, and serve as potential "process" or mediating variables in basic research on psychotherapy.

The Clinician Affective REsponse (CARE) scale is available online as Appendix at: https://www.researchinpsychotherapy.org/rpsy/article/view/736

References

Abargil, M., & Tishby, O. (2022). Countertransference awareness and treatment outcome. *Journal of Counseling Psychology*, *69*(5), 667-677. doi: 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. doi: 10.1037/cou0000250.

Berg, J., Lundh, L. G., & Falkenstrom, F. (2019). Countertransference in Swedish psychotherapists: Testing the factor structure of the Therapist response questionnaire. *Research in Psychotherapy : Psychopathology, Process, and Outcome*, *22*(1), 331. doi: 10.4081/ripppo.2019.331.

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. doi: 10.1176/appi.ajp.162.5.890.

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

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: a primer. *Frontiers in Public Health*, *6*, 149. doi: 10.3389/fpubh.2018.00149.

Brody, E. M., & Farber, B. A. (1996). The effects of therapist experience and patient diagnosis on countertransference. *Psychotherapy: Theory, Research, Practice, Training*, *33*(3), 372-380. doi: 10.1037/0033-3204.33.3.372.

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. doi: 10.1037/pst0000442.

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. doi: 10.1176/appi.ajp.2013.13020224.

Del Re, A. C., Flückiger, C., Horvath, A. O., & Wampold, B. E. (2021). Examining therapist effects in the alliance–outcome relationship: a multilevel meta-analysis. *Journal of Consulting and Clinical Psychology*, *89*(5), 371-378. doi: 10.1037/ccp0000637.

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. doi: 10.1037/tep0000416.

Flückiger, C., Del Re, A. C., Wampold, B. E., & Horvath, A. O. (2018). The alliance in adult psychotherapy: a meta-analytic synthesis. *Psychotherapy*, *55*(4), 316-340. doi: 10.1037/pst0000172.

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

Gelso, C. (2014). A tripartite model of the therapeutic relationship: theory, research, and practice. *Psychotherapy Research*, *24*(2), 2. doi: 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., 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), 640-649. doi: 10.1037/0022-0167.52.4.640.

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. doi: 10.1037/h0101779.

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. doi: 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. doi: 10.1037/0000034-006.

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

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. doi: 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. doi: 10.1037/a0027503.

Jones, E. E. (2000). *Therapeutic action: a guide to psychoanalytic therapy*. Rowman & Littlefield.

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. doi: 10.1016/S1077-7229(00)80089-9.

Kyriazos, T. A. (2018). Applied psychometrics: sample size and sample power considerations in factor analysis (EFA, CFA) and SEM in general. *Psychology*, *09*(08), 08. doi: 10.4236/psych.2018.98126.

Latts, M. G. (1997). *A revision and validation of the countertransference factors inventory*. Dissertation Abstracts International: Section B: the Sciences and Engineering, 58(1-B), 0455.

Lindy, J. D., & Lindy, D. C. (2004). Countertransference and disaster psychiatry: from Buffalo Creek to 9/11. *Psychiatric Clinics of North America*, *27*(3), 571-587. doi: 10.1016/j.psc.2004.03.009.

Meehan, K. B., Levy, K. N., & Clarkin, J. F. (2012). Construct validity of a measure of affective communication in psy-

chotherapy. *Psychoanalytic Psychology*, *29*(2), 145-165. doi: 10.1037/a0027450.

Messina, I., Solina, C., Arduin, A., Frangioni, V., Sambin, M., & Gelso, C. (2018). Origins of countertransference and core conflictual relationship theme of a psychotherapist in training as emerging in clinical supervision. *Psychotherapy*, *55*(3), 222-230. doi: 10.1037/pst0000148.

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. doi: 10.1080/10503300903470610.

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. doi: 10.1037/0033-3204.32.4.669.

Pallagrosi, M., Fonzi, L., Picardi, A., & Biondi, M. (2014). Assessing clinician's subjective experience during interaction with patients. *Psychopathology*, *47*(2), 111-118. doi: 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. doi: 10.1037/pst0000165.

Pérez-Rojas, A. E., Palma, B., Bhatia, A., Jackson, J., Norwood, E., Hayes, J. A., & Gelso, C. J. (2017). The development and initial validation of the countertransference management scale. *Psychotherapy*, *54*(3), 307-319. doi: 10.1037/pst0000126.

Pope, K. S., & Tabachnick, B. G. (1993). Therapists' anger, hate, fear, and sexual feelings: national survey of therapist responses, client characteristics, critical events, formal complaints, and training. *Professional Psychology: Research and Practice*, *24*(2), 142-152. doi: 10.1037/0735-7028.24.2.142.

Pope, K. S., & Vasquez, M. J. T. (2016). *Ethics in Psychotherapy and counseling: a practical guide*. John Wiley & Sons.

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. doi: 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. doi: 10.1037/tep0000133.

Revelle, W., & Condon, D. M. (2019). Reliability from α to ω: a tutorial. *Psychological Assessment*, *31*(12), 1395-1411. doi: 10.1037/pas0000754.

Robinson, L., Delgadillo, J., & Kellett, S. (2020). The dose-response effect in routinely delivered psychological therapies: a systematic review. *Psychotherapy Research*, *30*(1), 79-96. doi: 10.1080/10503307.2019.1566676.

Russell, J. A., & Carroll, J. M. (1999). On the bipolarity of positive and negative affect. *Psychological Bulletin*, *125*(1), 3-30. doi: 10.1037/0033-2909.125.1.3.

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

Shafranske, E. P., & Falender, C. A. (2008). Supervision addressing personal factors and countertransference. In C. A. Falender & E. P. Shafranske (Eds.), *Casebook for clinical supervision: a competency-based approach*. (pp. 97-120). American Psychological Association. doi: 10.1037/11792-005.

Stefana, A. (2017). *History of countertransference: From Freud to the British object relations school*. Routledge. doi: 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. doi: 10.1002/cpp.2455.

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), 22. doi: 10.3390/ijerph192215403.

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), 5156. https://doi.org/10.3390/jcm12155156.

Stefana, A., Youngstrom, E. A., & Vieta, E. (2022). Empirical support for the use and further study of the countertransference construct in the clinical care of patients with bipolar disorder. *Bipolar Disorders*, *24*(1), 84-85. doi: 10.1111/bdi.3153.

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.

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. doi: 10.1037/per0000146.

Tanzilli, A., Majorana, M., Fonzi, L., Pallagrosi, M., Picardi, A., Fornari, M. A. C. de', 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. doi: 10.4081/ripppo.2018.327.

Tanzilli, A., Trentini, C., Grecucci, A., Carone, N., Ciacchella, C., Lai, C., Sabogal-Rueda, M. D., & Lingiardi, V. (2022). Therapist reactions to patient personality: a pilot study of clinicians' emotional and neural responses using three clinical vignettes from in treatment series. *Frontiers in Human Neuroscience*, *16*, 1037486. doi: 10.3389/fnhum.2022.1037486.

Tellegen, A., Watson, D., & Clark, L. A. (1999). On the dimensional and hierarchical structure of affect. *Psychological Science*, *10*(4), 297-303. doi: 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), 29-44. doi: 10.1080/10503307.2020.1862934.

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

Wampold, B. E., & Flückiger, C. (2023). The alliance in mental health care: conceptualization, evidence and clinical applications. *World Psychiatry*, *22*(1), 25-41. doi: 10.1002/wps.21035.

OPEN ACCESS

Watkins, M. W. (2018). Exploratory factor analysis: a guide to best practice. *Journal of Black Psychology*, *44*(3), 219-246. doi: 10.1177/0095798418771807.

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. doi: 10.1080/09515070.2020.1863186.

Wise, E. A. (2004). Methods for analyzing psychotherapy outcomes: a review of clinical significance, reliable change, and recommendations for future directions. *Journal of Personality Assessment*, *82*(1), 50-59. doi: 10.1207/s15327752jpa 8201_10.

Online supplementary material:

Supplementary Table 1. Distribution of responses for each item on the initial pool.

Supplementary Table 2. Fit indices confirming the unidimensionality of the three subscales from confirmatory factor analyses.

Supplementary Table 3. Descriptive statistics, internal consistency reliability, precision, and inter-scale correlations.

Supplementary Figure 1. Reliability and item characteristic curves for the scale scores.

Supplementary Figure 2. Measurement models from confirmatory factor analysis (N=556) presenting a fully standardized solution using robust maximum likelihood estimation.