Article

Development and Initial Validation of the in-Session Patient Affective Reactions Questionnaire (SPARQ) and the Rift In-Session Questionnaire (RISQ)

Alberto Stefana [1,*], Joshua A. Langfus [2], Eduard Vieta [3], Paolo Fusar-Poli [1,4,5] and Eric A. Youngstrom [2,6]

1   Department of Brain and Behavioral Sciences, University of Pavia, 27100 Pavia, Italy;
    paolo.fusar-poli@unipv.it
2   Department of Psychology and Neuroscience, University of North Carolina at Chapel Hill,
    Chapel Hill, NC 27599, USA; langfus@unc.edu (J.A.L.); eay@unc.edu (E.A.Y.)
3   Bipolar and Depressive Disorders Unit, Hospital Clinic, IDIBAPS, CIBERSAM, University of Barcelona,
    08028 Barcelona, Catalonia, Spain; evieta@clinic.cat
4   OASIS Service, South London and Maudsley NHS Foundation Trust, London SE5 8AZ, UK
5   Early Psychosis: Interventions and Clinical-Detection (EPIC) Lab, Department of Psychosis Studies,
    Institute of Psychiatry, Psychology & Neuroscience, King's College London, London SE5 8AF, UK
6   Helping Give Away Psychological Science (HGAPS), 501c3, Chapel Hill, NC 27599, USA
*   Correspondence: alberto.stefana@gmail.com

Abstract: This article discusses the development and preliminary validation of a self-report inventory of the patient's perception of, and affective reaction to, their therapist during a psychotherapy session. First, we wrote a pool of 131 items, reviewed them based on subject matter experts' review, and then collected validation data from a clinical sample of adult patients in individual therapy (*N* = 701). We used exploratory factor analysis and item response theory graded response models to select items, confirmatory factor analysis (CFA) to test the factor structure, and *k*-fold cross-validation to verify model robustness. Multi-group CFA examined measurement invariance across patients with different diagnoses (unipolar depression, bipolar disorder, and neither of these). Three factors produced short scales retaining the strongest items. The in-Session Patient Affective Reactions Questionnaire (SPARQ) has a two-factor structure, yielding a four-item Negative affect scale and a four-item Positive affect scale. The Relationship In-Session Questionnaire (RISQ) is composed of four items from the third factor with dichotomized responses. Both scales showed excellent psychometric properties and evidence of metric invariance across the three diagnostic groups: unipolar depression, bipolar disorder, and neither of these. The SPARQ and the RISQ scale can be used in clinical or research settings, with particular value for capturing the patient's perspectives about their therapist and session-level emotional processes.

Keywords: therapeutic relationship; affective reaction; emotional reaction; in-session process; self-report measure

1. Introduction

Emotions are a central pillar of our human existence, and evolved as a biological mechanism for managing critical interpersonal interactions and other crucial life tasks [1]. Consequently, it is only natural that emotions are an integral component of both therapy-specific and nonspecific therapeutic processes [2,3] and play a substantial role in influencing outcomes at the level of individual sessions and the broader course of treatment [4–6].

Facilitating the patient to become aware of the emotions experienced during the session and to make constructive use of them is a key path to therapeutic change [7,8]. Among the range of emotions experienced in therapy, those directed toward the therapist hold particular significance, offering a valuable source of clinical information about their personality characteristics [9,10] and psychological functioning in the therapeutic relationship [11,12]

capable of informing more effective therapeutic interventions [13,14] and thus improving outcomes [15–20].

The challenge lies in allowing the therapists to effectively focus on the patient's emotions within the therapy session, supporting them in navigating, experiencing, accepting, and ultimately transforming these emotions [21–23]. Carefully and systematically assessing how the patient perceives, experiences, and reacts to the therapist during therapy sessions as part of the clinical work can facilitate the understanding of the nature of the patient's core intrapsychic problems and maladaptive schemas in interpersonal relationships [24,25]. However, the current landscape of therapeutic research and practice reveals a gap.

Despite numerous tools developed to assess emotions and emotional expression [6,26,27], only a few formally integrate attention to emotional processes within the unique dyadic context of the therapeutic relationship. Some work has directly examined emotional reactions and processes during the therapy encounter, using methods such as analysis of transcripts of sessions [28], microanalytic coding of video [29,30], or fundamental frequency of vocal expression [31], as well as long, clinician-reported measures [32,33]. Relatively less work has been undertaken to formally incorporate attention to dyadic affective processes into most psychotherapy modalities.

There currently does not exist a brief assessment tool that could be used to self-monitor affective reactions toward the clinician likely to influence the therapeutic process. This tool should be feasible to use in settings without access to more work on the part of the single therapist or more labor or analytically intensive methodologies used in research-oriented clinical settings. Existing self-report scales for emotion tend to focus on trait affect, or else are state measures that are decontextualized [34,35]. Other scales that measure constructs such as working alliance [36,37] only indirectly include affective content. An ideal tool would provide information about positive affect (e.g., liking, feeling understood, supported, and accepted) and negative affect (including both withdrawal-oriented emotions such as depression, shame, and anxiety, as well as approach/aggressive emotions such as anger and irritation). Additionally, they should be short enough to use repeatedly during a course of treatment. Embedding emotions in the context of in-session events is also likely to help different state emotion from more stable trait effects and reduce mood-congruent biases and halo effects [38].

Aim

The current study aimed to develop a patient-report measure of patterns of thought, feeling, and behavior activated and experienced in the therapeutic relationship that is clinically sophisticated, psychometrically valid, and easy enough to administer in real-world psychotherapeutic settings for both clinical and research purposes.

2. Materials and Methods

2.1. Procedures and Sample Characteristics

Eligibility criteria were being 18 years or older, fluent in English, and currently engaged in individual psychotherapy treatment for a diagnosed mental disorder. Participants meeting the study criteria were recruited via two online patient registers (i.e., ResearchMatch and Research for Me) from March through April 2022. ResearchMatch is a disease- and institution-neutral, United States national registry to recruit volunteers for clinical research [39] created by several academic institutions and funded in part by the National Institutes of Health (NIH) National Center for Advancing Translational Sciences (NCATS). Research for Me is a community of volunteers that serves as the central entry point for patients and community members interested in engaging with research at UNC; it was created by the North Carolina Translational & Clinical Sciences Institute (NC TraCS), the integrated hub of the NIH CTSA program at UNC-CH. Evidence indicates that participants recruited through online research platforms are consistent in their self-reported demographic and psychological information, and do not use deception when not financially incentivized [40].

Participants completed an anonymous online survey on Qualtrics, which lasted an average of 15 min.

A small pilot study tested the items on a convenience sample, debugging the Qualtrics programming and checking the wording and clarity of instructions. Participants were then recruited via ResearchMatch and Research for Me. Inclusion criteria were being 18 years or older, fluent in English, and in psychotherapy treatment for a mental disorder; exclusion criteria were deliberately kept minimal. These exclusions were having been declared legally incompetent or having a support administrator.

2.2. Item Generation

We followed best practices in scale development, starting with item generation [41,42]. We began with theoretical models motivating item content. Our affective models included the widely accepted positive affect and negative affect model of emotions [43]. We examined the item pool for the extended Positive Affect and Negative Affect Scales (PANAS-X) for candidate content [44]. We enriched it by also considering emotions related to the social/interpersonal dominance dimension [45], which separates emotions such as fear, guilt, and shame (strongly negative valence, but low dominance) from anger, social disgust, and contempt (also strongly negative, but high dominance). We accomplished this by reviewing and including exemplars of discrete emotions that might have distinct motivational properties, including high and low dominance negative emotions [35]. From an evolutionary perspective, these types of emotions served different functions, including fight versus flight in threatening situations [46], and shutting down to conserve resources when helpless [47,48]. We also considered models from the therapeutic process literature, covering affective, cognitive, and behavioral responses, drawing from the clinical-theoretical and empirical literature on transference [9,15,49–51] and related concepts [36,52,53].

Item generation combined inductive methods, looking at existing scales for discrete emotions [35], as well as therapy process scales cited above. We also used deductive methods, with experienced clinicians generating items reflecting affective features of good and challenging sessions. The initial item pool was reviewed for content validity by six experienced clinicians, three of whom primarily identify as cognitive-behavioral, two as eclectic, and one as psychodynamic. Six items were eliminated as redundant or poorly worded, and more than a dozen were reworded and reviewed again.

Authorities recommend generating a large initial item pool, much more extensive than the intended scale [41,54]. The longest process scales with which we were familiar suggest an outer limit of around 40–45 simple items for scales intended to be repeated regularly during therapy (e.g., OQ [55]). Our item pool was thus more than triple the size of the longest we would consider feasible for a working instrument. The items were written in everyday language so that the questionnaire could be completed by people with different educational levels. Of note, items were written to sample from positive as well as negative affective domains, ensuring breadth of coverage and avoiding the pitfall of having all items only assessing negative constructs. However, items were written so that each was unipolar. For example, there were separate items for happy and sad affect, rather than one bipolar item ranging from sad to happy, and the response anchors ranged from "Not at all true" to "Very true" for each. These design decisions reflected the current understanding of the measurement of state affect (vs. trait affect or mood) [56], as well as recommendations about reducing cognitive load and improving response accuracy [41]. Items used a 5-point Likert-type scaling, and the instructions directed respondents to think about their most recent therapy session before responding, as our goal was to create a measure of the current process, not attitude towards treatment (as the target audience is people already in therapy), and not yet another measure of temperament or personality. The Flesch–Kincaid readability index was a 7.25, corresponding to a "7th grade reading level—fairly easy to read".

2.3. Additional Measures

The participants completed a 7-item demographic and clinical data form, which recorded their age, biological sex, the clinician's sex, the frequency of therapy sessions, the length of the course of treatment, and the patient's clinical diagnoses. Participants were asked to think about their most recent psychotherapy session, read a series of statements that people in psychotherapy might use to describe how they feel toward their therapist (e.g., "During my last therapy session, I felt happy to see my therapist"), and rate each of them on the extent to which each was true of the way they felt during that session. They were asked to respond using a five-point Likert scale: 0 = not at all true, 1 = a little true, 2 = somewhat true, 3 = a lot true, and 4 = very true. Higher scores indicated greater levels of affective reaction.

2.4. Statistical Analyses

Scale development followed best practices [44,57,58], and data analyses followed steps similar to prior work [59]. Specifically, the large pool of candidate items was evaluated using descriptive statistics (minimum, maximum, standard deviation, skewness, and kurtosis), checking suitability of individual items for inclusion in subsequent analyses. Those showing insufficient variability were dropped. The Kaiser–Meyer–Olkin test and the Bartlett test of sphericity verified the suitability of the data for factor analysis [60]. Parallel analysis using multiple factor retention methods was run using the R packages *paran* v1.5.2 (Dinno, 2018 [61]) and *EFAtools* v0.4.1 [62] to help determine how many factors might have enough related items to support interpretation. Subsequently, iterative exploratory factor analysis (EFA) using the R package *EFAtools* v0.4.1 [62] analyzed a matrix of the inter-item correlations using polychoric estimation and a PROMAX rotation to achieve simple structure. Items with factor loading less than 0.40 and those with more than 0.30 on two or more factors were removed [63,64], as the large item pool should allow selection of univocal items for major factors. This strategy facilitates unit-weighted scales, which would be easier to use in subsequent validation studies as well as in later clinical applications.

Item response theory (IRT) analyses were implemented in the R package *mirt* v1.36.1 [65] to estimate a Graded Response Model (GRM) for each scale identified by EFA. Item information and coverage were evaluated based on these models. Final item selection chose items with high information across a wide range of theta (θ) levels. IRT methods provide information about whether items are more helpful at low, medium, or high levels of a factor, as well as changes in scale reliability depending on the levels of the factor. In two instances, two items had similar factor loadings and theta levels, so a pool of clinical experts selected one for retention.

The fit of the final factor solution was tested by conducting CFAs using the R package *lavaan* v0.6-11 [66]. Multi-group confirmatory factor analysis examined measurement invariance of the scales across patients with different diagnoses (unipolar depression, bipolar disorder, neither of these). Furthermore, *k*-fold cross-validation using the R package *kfa* v0.2.0 [67] verified the robustness of our final model. We compared results using ML, MLR, and ULS estimators and also examined the statistical power in combination with the expected parameter changes and modification indices to look for model mis-specification [58].

Reliability statistics for the resulting scales were estimated using raw items and 1000 bootstrapped replications [57]. Correlations between questionnaire and patient demographic-clinical features, as well as treatment characteristics, offered preliminary information about the criterion validity of questionnaire scores. Sensitivity analyses re-ran the main models and reliability statistics after trimming the sample to eliminate extreme response times as a filter for online data collection.

3. Results

3.1. Sample Characteristics

The scale development sample consisted of 701 adults in psychotherapy for a mental health disorder. Most (80%, *n* = 564) were women. The most common age range was 18 to

29 years (40%, *n* = 282), followed by 30 to 39 years (19%, *n* = 131). Each participant had an average of 2.55 (*SD* = 1.53) DSM diagnoses at the diagnostic category level. Many were in psychotherapy over more than 24 months (48%, *n* = 335), typically at a frequency of two to four sessions per month (71%, *n* = 500). More than half of the participants had their most recent session less than one week prior to the study. Table 1 reports sample demographic and clinical characteristics.

Table 1. Demographics, Clinical, and Treatment Characteristics of Participating Patients (*N* = 701).

Demographics	% (*n*)
Biological sex	
Male	18% (126)
Female	80% (564)
I prefer not to say	2% (11)
Age (years)	
18–29	40% (282)
30–39	19% (131)
40–49	15% (104)
50–59	15% (105)
≥60	11% (79)
Clinical Characteristics	
Average number of diagnoses, *M (SD)*	2.55 (1.53)
Any anxiety disorder	75% (529)
Any (unipolar) depressive disorder	54% (378)
Any bipolar or related disorder	17% (117)
Any personality disorder	13% (93)
Any trauma- and stressor-related disorders	30% (209)
Treatment Characteristics	
In psychotherapy from	
0 to 3 months	18% (124)
4 to 6 months	10% (72)
7 to 12 months	12% (86)
13 to 24 months	12% (84)
>24 months	48% (335)
Session frequency	
≤1 per month	24% (171)
2 to 3 per month	35% (244)
1 per week	37% (256)
≥2 per week	4% (30)
Therapist's biological sex	
Female	77% (539)
Patient–Therapist biological sex match	
Same-sex	74% (521)

* Note: *N* sums to more than 701 because cases could have more than one diagnosis.

3.2. Preliminary Analyses

The Kaiser–Meyer–Olkin test (0.96) and the Bartlett test of sphericity (*p* < 0.001) verified the suitability of the data for factor analysis. Two items were deleted because they had polychoric correlations (with smoothing) of >0.90 with other items.

3.3. Item Pool Reduction—Iterative Exploratory Factor Analysis

Horn's [68] parallel analysis using 5000 iterations with simulated *N* = 701 patients and *k* = 129 items identified six factors with eigenvalues greater than one. A six-factor solution was also found when considering eigenvalues higher than those of the 99th percentile of the simulated eigenvalues using 10,000 iterations [69]. Finally, the Hull method [70] and comparison data [71] indicated four factors.

We conducted the first round of EFA extracting six factors, evaluating these for having adequate indicators (at least four items with strong loading) as well as conceptual coherence. Because we had a large initial item pool, we also eliminated items that cross-loaded without a clear dominant loading, to improve the interpretability of scales based on unit-weighted scores. After iterative EFA rounds, 68 items remained in contention, producing a three-dimensional factor structure. Items showed adequate to strong loadings on the respective factor: smallest loadings were 0.58, 0.33, and 0.44, respectively, for factors 1, 2, and 3; while median loadings were 0.74, 0.75, and 0.65, respectively). The items on factor 1 had relatively low endorsement rates compared to the others, creating positive skew at the item level. We considered whether this might be an artifactual "difficulty" factor. After considering the clinical coherence and ramifications of the item content, we opted to dichotomize this subset of items for subsequent IRT analyses and observed score interpretation, such that endorsing any but the lowest option was treated as a concerning, "yes" response. The items on the other two factors all showed acceptable item distributions and satisfied other guidelines for both factor analysis and graded response modeling.

3.4. Item Response Theory

Analyses for the Relationship In-Session Questionnaire (RISQ) used dichotomized items; Samejima's graded response model [72] evaluated the items for the in-Session Patient Affective Reactions Questionnaire (SPARQ) "Positive Affect" and "Negative Affect" factors. Table 2 reports the item discrimination and difficulty parameters of the final scales. Interestingly, the factors had different ranges of theta coverage, despite efforts to select items across a range of levels. The Positive Affect factor had reliability >0.80 from theta of −2.4 to +1.1, indicating that the items were informative and likely to be endorsed even at low levels of the latent variable. In contrast, the Negative Affect factor showed reliability >0.80 at theta ranging from +0.2 to +2.6, and the RISQ factor had reliability >0.80 between theta +0.9 and +3.4, indicating that these items had more information at high levels of the latent variable. Figure 1 shows the item characteristic curves and reliability for the scale scores.

Table 2. Item Option Characteristics for the three factors based on IRT models.

|	Item Content	α	β1	β2	β3	β4
Factor 1	I felt disparaged or belittled by my therapist	3.05	1.46			
	I felt rejected by my therapist	4.73	1.28			
RISQ	I felt provoked or attacked by my therapist	2.31	1.84			
	I felt scared, uneasy, like my therapist might harm me	2.23	2.06			
Factor 2	I felt respected by my therapist	2.60	−2.07	−1.32	−0.73	0.19
SPARQ	I felt my therapist cared about me	2.98	−1.95	−1.04	−0.40	0.51
Positive	I felt happy to see my therapist	2.47	−1.83	−0.91	−0.25	0.61
Affect	I felt appreciated by my therapist	2.44	−1.29	−0.57	0.22	1.14
Factor 3	I felt worried my therapist couldn't help me	1.42	−0.51	0.68	1.54	2.32
	I felt afraid to spoke my mind, for fear of being judged, criticized, disliked by my therapist	2.67	0.34	1.26	1.77	2.32
SPARQ Negative Affect	I felt ashamed with my therapist about my fantasy, desires, mindset, behavior, or symptoms	1.73	0.37	1.32	2.12	2.95
	I felt shy, like I wanted to hide from my therapist or end the session early	1.96	0.54	1.48	2.09	2.79
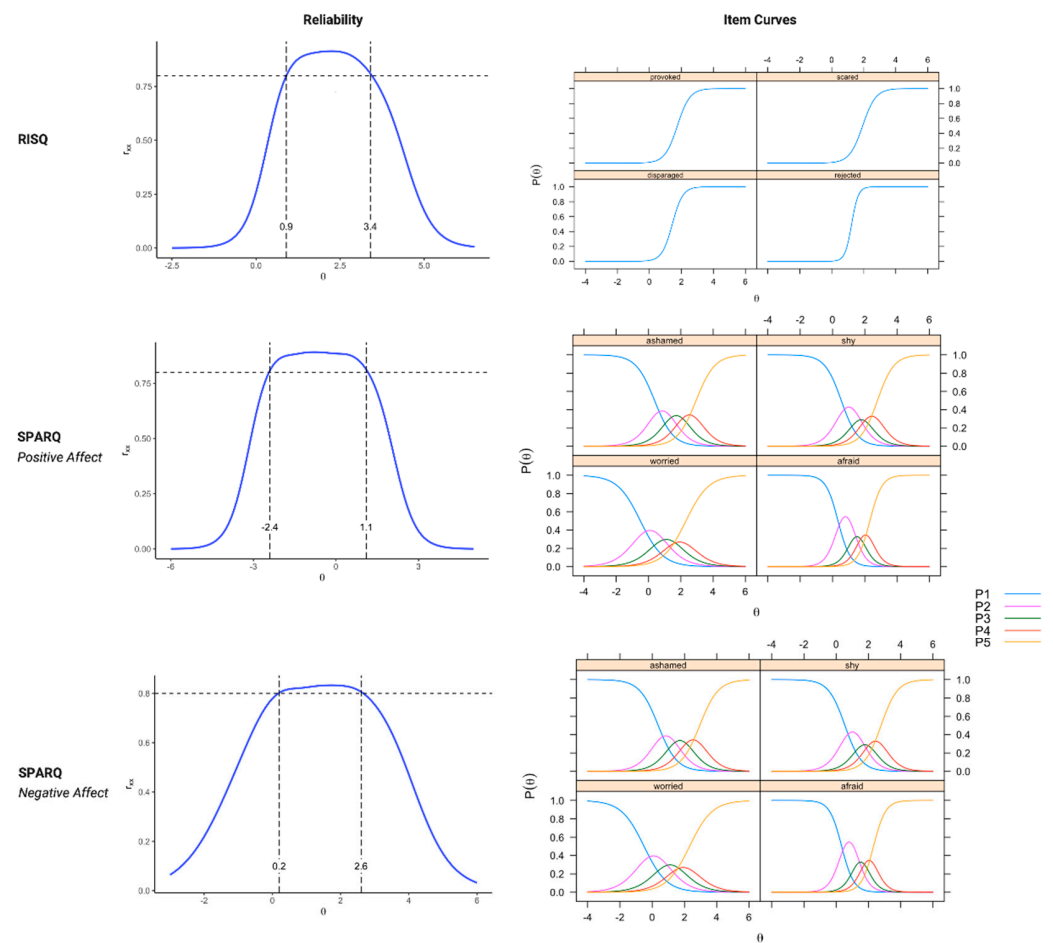
Figure 1. Item Option Characteristic Curves and Reliability for the Scale Scores. The curves on the right show the threshold where a patient's probability changes from a lower to the next higher option on the item. The reliability curves transform test information into a reliability estimate (between 0 and 1.0) and show how reliability changes at low (negative θ values), average (θ = 0), and high levels (positive θ) of the underlying factor.

3.5. Confirmatory Factor Analysis

K-fold cross-validations were performed with *k* = 3 to verify the robustness of our models. The two-factor model, named SPARQ, had the following fit indices: $X^2(df = 19) = 36.70$, CFI = 0.97, TLI = 0.96, RMSEA = 0.06 (90% CI [0.04, 0.08]), and SRMR = 0.05. The one-factor model, named RISQ, had fit indices above a satisfactory range: $X^2(df = 2) = 8.25$, CFI = 0.97, TLI = 0.93, RMSEA = 0.09 (90% CI [0.00, 0.19]), and SRMR = 0.03. A final set of models pooled the samples to provide a final set of parameter estimates (see Figure 2). The two-factor model of the SPARQ provided an excellent fit for the data: $X^2(df = 19) = 61.48$, CFI = 0.98, TLI = 0.97, RMSEA = 0.06 (90% CI [0.04, 0.07]), and SRMR = 0.05. Similarly, a good fit for the data was provided by the one-factor solution of the RISQ: $X^2(df = 2) = 14.32$, CFI = 0.98, TLI = 0.94, RMSEA = 0.09 (90% CI [0.05, 0.14]), and SRMR = 0.03. Statistical power was high for all expected parameter changes and modification indices; none of the tests indicated model mis-specification. Model fit remained good across estimators.

3.6. Invariance Testing with Multigroup CFA

To assess measurement invariance across patients with different diagnoses (unipolar depression, bipolar disorder, and neither of these), multigroup CFA models were fit for the SPARQ and the RISQ, respectively. For the SPARQ, a model with no equality constraints across groups showed good model fit ($X^2 = 133.28$, $df = 57$, $p < 0.001$, CFI = 0.98, TLI = 0.98, RMSEA = 0.08). Equating the loadings, item intercepts, and item thresholds did

not significantly harm model fit (ΔX^2 = 53.9, *df* = 56, *p* = 0.55), providing evidence of metric invariance across the three diagnostic groups for the SPARQ.
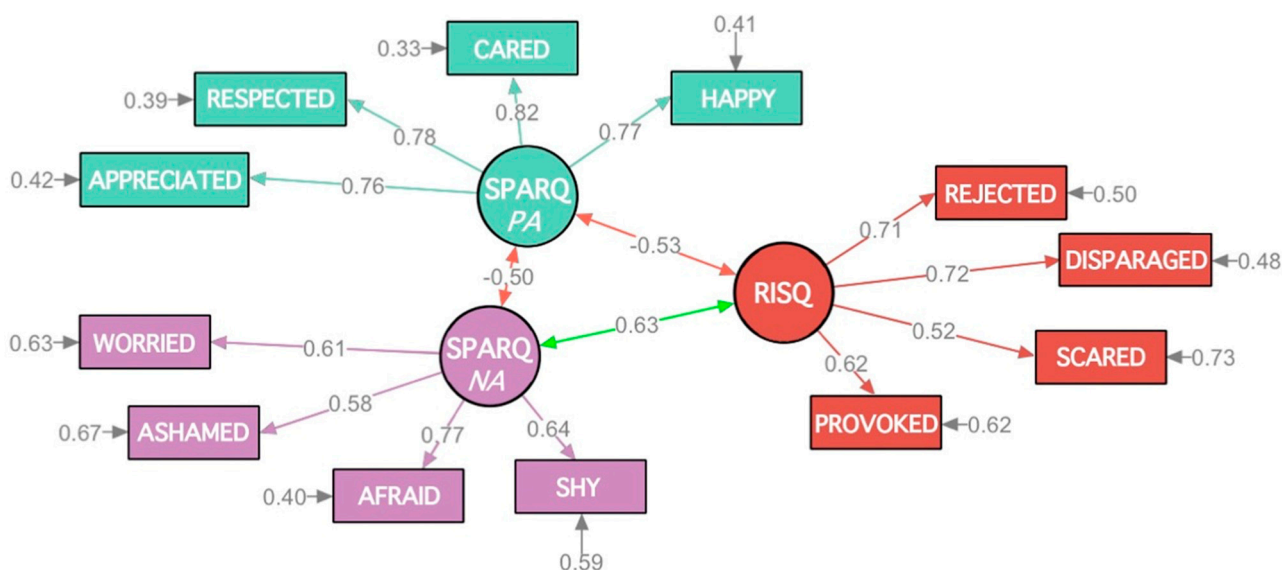

Figure 2. Measurement Model from Confirmatory Factor Analysis (*N* = 701) presenting a fully standardized solution using robust maximum likelihood estimation. Note: This Figure presents abbreviated item content for both the RISQ and the SPARQ items.

For the RISQ, the baseline model showed good fit (X^2 = 17.28, *df* = 6, *p* = 0.19, CFI = 0.99, TLI= 0.97, RMSEA = 0.09). As with the SPARQ, equating the item loadings, intercepts, and thresholds did not significantly harm model fit (ΔX^2 = 3.01, *df* = 4, *p*= 0.56), providing evidence of metric invariance across the three diagnostic groups for the RISQ.

3.7. Internal Consistency and Score Precision

Table 3 presents the scale descriptive statistics, reliability estimates, and standard errors. The internal consistencies of the final scales were good [54,57]: RISQ (*k* = 4, McDonald's omega = 0.74, Cronbach's alpha = 0.75, average inter-item *r* = 0.43), Positive Affect (*k* = 4, omega = 0.86, alpha = 0.86, average inter-item *r* = 0.61), and Negative Affect (*k* = 4, omega = 75, alpha = 0.74, average inter-item *r* = 0.41). The mean scores on the SPARQ "Positive Affect" and "Negative Affect" scales were, respectively, 10.45 (*SD* = 4.16) and 3.03 (*SD* = 3.11). The mean score on the dichotomized RISQ scale was 0.36 (*SD* = 0.87).

For measures of individual precision, we also included the standard error of the measure (*SEm*) and standard error of the difference (*SEd*), along with critical values corresponding to the reliable change index (RCI) propounded by Jacobson and colleagues (e.g., [73]). The 90% value is 1.65 times the *SEd*, and the 95% is 1.96 times the *SEd*. These provide thresholds as being 90% confident that the patient change between the two evaluations was likely to reflect "real" change rather than measurement error. Jacobson stipulated this as a first condition for his two-part nomothetic definition of "clinically significant change". The second aspect, transitioning past a benchmark based on clinical and or non-clinical reference distributions, is less applicable here: It is not clear what it would mean conceptually to have a "nonclinical reference group" for scores on a scale measuring emotional reactions during therapy sessions. However, it is feasible to define one of the benchmarks based on the clinical distribution. We estimated the 5th percentile for the Positive Affect score, marking where the score would be concerningly low compared to the clinical distribution, and the 95th percentile for the Negative Affect scale, above which the score would be concerningly high based on this nomothetic comparison.

We also included estimates of minimally important difference (MID), using the *d* of 0.5 operational definition [54]. This estimate of MID is more liberal than Jacobson's

RCI-type methods, but it aligns with patient subjective experiences across a broad swathe of constructs and outcome measures [74].

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

	RISQ	SPARQ	
		Positive Affect	Negative Affect
Descriptive statistics			
Potential Range	0 to 4	0 to 16	0 to 16
Observed Range	0 to 4	0 to 16	0 to 16
Mean, *SD*	0.36 (0.87)	10.45 (4.16)	3.03 (3.11)
POMP, *SD*	9.00 (21.75)	65.31 (26.00)	18.94 (19.44)
Skew	2.67	−0.58	1.33
Kurtosis	6.75	−0.53	1.57
Standard Error of Measurement (SE_m)	0.44	1.56	1.55
Standard Error of Difference (SE_d)	0.63	2.20	2.50
Internal consistency reliability			
X^2/df	7.16	3.24	
CFI	0.98	0.98	
TLI	0.94	0.97	
RMSEA	0.09	0.06	
SRMR	0.03	0.05	
Average inter-item *r*	0.43	0.61	0.41
Alpha	0.75	0.86	0.74
Omega total	0.74	0.86	0.75
Clinical change benchmarks			
90% Critical Change	1.02	3.63	3.63
95% Critical Change	1.21	4.31	4.61
Minimally important difference	0.44	2.08	1.55
Jacobson benchmark threshold (5% tail)	--	LB: 2.30	UB: 9.13
Scale correlations			
SPARQ—Positive Affect	−0.45 *	1	
SPARQ—Negative Affect	0.49 *	−0.40 *	1

Note: SPARQ = in-Session Patient Affective Reactions Questionnaire; LB = Lower Bound; POMP = Percentage of Maximum Possible; RISQ = Relationship In-Session Questionnaire; UB = Upper Bound. RISQ uses dichotomized answers as "Not at all true" versus the other options. Minimally important difference was operationally defined as $d = 0.5$. * $p < 0.0005$, two-tailed.

3.8. Criterion Validity

The criterion validity of the SPARQ was examined in relation to patient demographic and clinical features, as well as treatment characteristics (see Table 4). All the correlations were small, with none greater than 0.20. However, as expected, the strongest correlations with factors describing negative attitudes were observed with personality disorder. More specifically, small correlations were observed between any personality disorder and the Negative Affect scale of the SPARQ ($r = 0.17$, $p < 0.001$), and between cluster B personality disorder and the RISQ ($r = 0.20$, $p < 0.001$).

3.9. Sensitivity Analyses

Sensitivity analyses trimmed the cases to eliminate the fastest and slowest completion times, a standard check for online surveys [40]. Dropping these 36 cases left a sample with $N = 665$. IRT analyses and reliability coefficients and the CFAs all produced results that were identical or changed only at the second decimal place. Model fit was essentially identical: the three-factor model had a robust CFI of 0.995, TLI = 0.994, RMSEA = 0.030 (0.017 to 0.042 90% CI), and SRMR = 0.043, with all factor loadings large and similar to the untrimmed sample. The criterion correlations showed the same substantive results (all available upon request as an *R* notebook). The pattern of response times was highly positively skewed—the fastest completion times were still close to the median, whereas there were

some cases with extremely long responses, often an artifact of not clicking past the "Thank you" at the end. The lack of ultra-fast responders is consistent with ResearchMatch being a register of people volunteering to help with research and not expecting compensation.

Table 4. Criterion Validity Correlations with Patient Diagnoses, Demographics, and Objective Therapy Characteristics.

Criterion Variable	RISQ	SPARQ	
		Positive Affect	Negative Affect
Age	−0.04	0.05	−0.15 ***
Sex	−0.06	0.04	−0.06
Average # of diagnoses	0.04	−0.07	0.17 ***
Any anxiety disorder	−0.10 *	−0.00	0.03
Any bipolar disorder	0.05	−0.05	0.03
Any depressive disorder	−0.14 ***	0.06	0.02
Any personality disorder	0.20 ***	−0.11 *	0.17 ***
Cluster A PD	0.08 *	−0.08	0.08 *
Cluster B PD	0.19 ***	−0.12 **	0.13 **
Cluster C PD	0.07	−0.08	0.09 *
Any trauma- and stressor-related disorder	−0.02	0.04	0.05
Therapy length	−0.11 *	0.13 **	−0.05
Session frequency	0.03	0.10 *	0.06
Therapist's sex	−0.13 **	0.10 *	−0.07
Patient–Therapist sex match	−0.10 *	0.09 *	−0.12 **

Note: Coefficients are point-biserial correlations for dichotomized variables, point-biserial correlations for dummy-coded categorical variables, Spearman correlations for ordinal variables, and Pearson correlations for continuous variables. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

4. Discussion

The goal of this article was to develop and rigorously evaluate a freely available, short self-reported questionnaire assessing the patient's perceptions and affective reactions to the therapist after a session (see final scales in Appendix A). Although feedback from patients is subject to biases and distortions [75], it can also be a valuable measure of in-session experience. These sorts of affective processes contribute both to positive outcomes [30] and premature dropout (e.g., [31]). Patient-report also could offer a helpful contrasting source of information, counterbalancing potential bias in therapist relationship and process ratings—a considerable source of error in therapists' ratings of patient emotional experiences and insight, accounting for 30% of the total variance in scores after accounting for perceived emotional intelligence [76]. Gathering the patient's perspective about their affective reactions would offer therapists more information ("objective data", from the patient's point of view), as well as guiding opportunities to disprove negative interpersonal expectations, enhance insight, and reinforce alliance, and ultimately outcome. The literature on routine outcome monitoring in psychotherapy [77,78] indicates that focusing on affective reactions experienced by the patient toward the therapist [79] may be especially effective for those patients who are not doing well in therapy. Furthermore, psychological assessment itself can be a therapeutic intervention when combined with personalized feedback, able to produce positive clinically meaningful effects, especially on treatment processes [80]. The SPARQ represents a further step toward a measurement feedback system that uses valid, reliable, and standardized measures to improve mental health outcomes [81].

We started with an item pool much larger than the intended final length of the scales, aiming to ensure good coverage of the constructs. We reduced the item pool iteratively using a combination of examining item characteristics, clear univocal loadings, adequate indicators for retained dimensions, and clinical cohesiveness of the set [41,82]. Factor analyses converged on a three-factor solution for the SPARQ that was theoretically coherent, clinically meaningful, and had very good internal reliability and consistency. The Positive Affect factor includes items indicating the patient's perception of being cared for, appreciated, respected, and guided by the therapist. It delineates a secure and comfortable (from the patient's perspective) experience of the therapeutic relationship, which appears

characterized by trust, affective attunement, and positive working alliance. The Negative Affect factor contained items describing feelings of shyness and shame with the therapist, fear of speaking openly, worry of not being helped, and a sense of personal failure due to their need for help from the therapist. An additional factor, the RISQ, has items describing the patient's tendency to feel disparaged, belittled, rejected, and attacked.

The dimensions of the SPARQ and the RISQ closely reflect the emotional configurations emerging in psychotherapeutic clinical practice [11,13,83,84] and allow therapists and researchers to identify patients' affective reactions toward their therapist, measure varying levels of them across sessions, and/or assess their relationship to session and treatment outcome. As such, the SPARQ is likely to prove useful for transference work [79], determining ways in which patients interact with their therapists, and increasing the therapist's understanding of the types and amount of emotional reactions. The identified dimensions likely reflect a mixture of the patient's own interpersonal dynamics, partially elicited by the therapist and therapeutic setting, and the interaction of patient and therapist in-session attitudes and behaviors.

The medium-size correlations among the SPARQ and the RISQ indicate on the one hand that these are distinct yet related dimensions, and on the other that was possible for a patient to feel cared for by the therapist even when they felt ashamed, afraid to open up with their therapist, or worried that their therapist could not help them, as well as when they are disappointed due to feeling criticized, attacked, or rejected by him/her.

This study included a preliminary investigation of the new scales' criterion validity by examining the associations between patients' affects toward their therapist and diagnosis of mental disorders. We found that patients' in-session affective patterns were not arbitrary but tended to relate to specific diagnoses in clinically meaningful and predictable ways. Consistent with results from previous studies [9,10,85], personality disorders were related to the negative dimensions of the therapeutic relationship. These results suggest that therapists treating a patient with a personality disorder, notably cluster B personality disorders, can expect some occurrence of negative attitudes and behavior against them. By being aware of this situation, the therapist may be able to provide prompt and effective therapeutic intervention, which, among other things, can help decrease premature discontinuation (which is a particularly high risk in patients who have a personality disorder [86]). At the same time, these associations involved only small to moderate amounts of the reliable variance in the scales, indicating that the scales likely measure variations in affective tone in sessions rather than being driven by depression-distortion or personality biases [38]. Put another way, although the scores may be influenced by patient traits, they are not a redundant measure of symptoms.

The different scales that emerged from the analyses have distinct features, and these suggest somewhat varying roles in the context of therapy and treatment research. The Positive Affect scale is the "easiest", meaning that it would be typical to have high scores after most therapy sessions. It is worth noting that the average was still towards the middle of the possible range of scores (M ~60% of the maximum possible), and there were few at the floor or ceiling. This suggests that typical (or good?) sessions may involve some challenging work, and the goal should not be to aim for "perfect" scores (unlike other consumer rating situations, such as Uber, Yelp, or course evaluations at most institutions). When thinking about Jacobson-style normative benchmarks, we debated whether a goal should be to exceed a high bar (e.g., 95% percentile compared to the reference distribution). Having the patient feel substantially less positive about the session and therapist seems more clearly problematic. Scores <4 occurred in only 5% of cases. In contrast, the Negative Affect scale had a lower distribution of scores, with a mean closer to 20% of the maximum possible range. Being closer to the floor, a benchmark pegged to a location in the 5th percentile is impossible (which is quite common in clinical scales in wide use [87]); 24% of cases had an observed score of zero. For the Negative Affect scale, having session ratings above the 95th percentile seems clearly concerning, corresponding to raw scores of 12+.

The RISQ items were rarely endorsed, as evident from item means and the pronounced shift in the region where the items were informative in the IRT analyses. Yet these also showed the most significant correlations with therapy features and patient diagnoses, underscoring their clinical relevance. More work can evaluate whether these are best used in the original Likert format, which would maximize the variance, but have a strongly skewed distribution, versus dichotomizing as "present at all" versus absent, which reduces variability but is frequently performed with clinical items (e.g., [88]). A third approach would be to use it as a checklist where the patient simply checked the box, and endorsing any of them would be a warning flag, given the combination of rarity and severity. Even that most liberal definition only occurred in 25% of cases. Used as a suite, the current data with the three scales suggest that it could be worth checking on the therapy process if Positive Scores fall below 4, Negative Scores rise above 10, and any of the RISQ items are endorsed. These offer provisional operational definitions for investigation in new samples.

The availability of tools such as these that are feasible to use in clinical settings opens up an important set of questions about how best to incorporate these into ongoing treatment. A related consideration would be who is the intended audience for the scores. The focus on the client (trying to improve the therapy process from their own perspective), the therapist, or the clinical supervisor each involves a different interpretive frame and set of goals, and perhaps different ethical considerations.

Strengths, Limitations, and Future Directions

There were several strengths associated with this study. The development of a self-report questionnaire is a novel contribution to the operationalization of patients' affective reactions to the therapist, which represents a key component of the therapeutic relationship. The SPARQ has excellent psychometric properties, can be completed in less than three minutes, and is easy to score. Moreover, the development itself followed best practices and utilized a combination of traditional and modern test theories. Finally, a large clinical sample was used.

There were also several limitations that should be addressed, including both technical and conceptual issues. The first is the exclusive reliance on the patient as the sole informant. Patients' perceptions do greatly matter, but they present only one piece of a complex system. The second limitation concerns the possible bias in patients' self-reporting their own affective, cognitive, and behavioral reactions. Traditionally, reports on the patient's emotional responses toward the therapists are obtained via clinicians or external observers/raters. However, the same issues hold true for measuring countertransference (i.e., the therapist's affective, cognitive, and behavioral responses toward the patient), but there is a body of literature that provides support for using clinicians' ratings of countertransference [18,89–91]. Patient ratings of their own affective responses make particular sense when considering the clinical importance of assessing the subjective emotional experience of the patient. Third, the psychometrics of the 15 extracted items should be confirmed in a sample where they are not embedded in the larger original item pool, checking that performance is similar without context effects. These tend to be small with scales that are homogeneous and have strong factor loadings, as is the case here, but they remain worth corroborating. Although we used *k*-fold cross validation and a large patient sample, our CFA models were still based on the same clinical sample as the exploratory analyses. It will be important to replicate with patients ascertained in a different way or for additional clinical issues to address generalizability. Fourth, now that a reduced item set has been identified, systematic exploration of its dependability, retest stability, and sensitivity to treatment effects will be an important next step in validation [57].

Future research using the SPARQ should examine affective states and processes from multiple perspectives to assess its validity and correlates and understand how self-reported affective reaction relates to therapists' perceptions of this phenomenon. Future studies should also investigate how this measure relates to the process and outcome of therapy, as well as to other components of the therapy relationship, especially countertransference,

working alliance, and real relationships. Finally, longitudinal research will add to our how these processes unfold over the course of psychotherapy and predict different trajectories. In sum, a major next step would be to examine multivariate models for combining information from different therapy processes and outcome measures, different informants (e.g., therapist and external observer), patient's personal family history/demographic/clinical characteristics, and therapist's personal characteristics and in-session attitudes and behaviors (including the therapeutic interventions) to examine incremental validity and to develop decision support algorithms and optimal sequences.

5. Conclusions

The patient's experience and perceptions of their psychotherapist must be accurately identified ("diagnosed") and discussed with the patient in a form and at a time that suits them. This article details the development and validation of two new brief self-report measures of the patient's affective reactions toward their psychotherapist. Both the SPARQ and the RISQ show excellent psychometric properties and are short and easy for patients to complete on their own. The results support the potential usefulness of these scales in assessing the patient's affective responses during therapy, and they provide initial evidence that these measures are appropriate for research and clinical use in individual psychotherapy settings. By enabling patients to rate their own affective reactions toward their therapist on a carefully developed, normed questionnaire, we turn patients' emotional experiences into quantifiable dimensions that can be analyzed, used to guide clinical interventions, and employed as indices of clinical change. These questionnaires may also be a useful tool in clinical supervision for psychotherapy trainees.

Appendix A

Appendix A.1. in-Session Patient Affective Reactions Questionnaire (SPARQ)

The following is a series of statements that people in psychotherapy might use to describe how they feel toward their psychotherapist. Think about your last psychotherapy session and remember some details from it. Then read each statement and indicate how you felt during that session. Select the response that corresponds with your answer with

"0" being not at all and "4" being very much. Do not worry if your responses appear to be inconsistent, as people often experience mixed and conflicting feelings.

Item nr.		Not at All	A Little	Somewhat	A Lot	Very Much
1	I felt happy to see my therapist.	0	1	2	3	4
2	I felt ashamed with my therapist about my fantasy, desires, mindset, behavior, or symptoms.	0	1	2	3	4
3	I felt worried my therapist couldn't help me.	0	1	2	3	4
4	I felt shy, like I wanted to hide from my therapist or end the session early.	0	1	2	3	4
5	I felt afraid to spoke my mind, for fear of being judged, criticized, disliked by my therapist.	0	1	2	3	4
6	I felt my therapist cared about me.	0	1	2	3	4
7	I felt respected by my therapist.	0	1	2	3	4
8	I felt appreciated by my therapist.	0	1	2	3	4

Positive Affect items: 1, 6, 7, and 8. Negative Affect items: 2, 3, 4, and 5.

Appendix A.2. Rift In-Session Questionnaire (RISQ)

The following are five statements that people in psychotherapy might use to describe how they feel toward their therapist. Think about your last psychotherapy session. Then read the following statements and rate each of them to indicate if it is true of the way you felt during that session. Select the response that corresponds with your answer by placing a cross on "Yes" to indicate that you experienced that feeling (irrespective of its intensity) during the session, or by placing a cross on "No" to indicate that you did not have that feeling.

Item nr.			
1	I felt provoked or attacked by my therapist.	No	Yes
2	I felt scared, uneasy, like my therapist might harm me.	No	Yes
3	I felt rejected by my therapist.	No	Yes
4	I felt disparaged or belittled by my therapist.	No	Yes

References

1. Robinson, M.D.; Watkins, E.; Harmon-Jones, E. (Eds.) *Handbook of Cognition and Emotion*; Guilford Press: New York, NY, USA, 2013; ISBN 9781462509997.
2. Faustino, B. Minding My Brain: Fourteen Neuroscience-based Principles to Enhance Psychotherapy Responsiveness. *Clin. Psychol. Psychoth* **2022**, *29*, 1254–1275. [CrossRef]
3. Lane, R.D.; Subic-Wrana, C.; Greenberg, L.; Yovel, I. The Role of Enhanced Emotional Awareness in Promoting Change across Psychotherapy Modalities. *J. Psychother. Integr.* **2022**, *32*, 131–150. [CrossRef]
4. Hayes, J.A.; Gelso, C.J.; Goldberg, S.; Kivlighan, D.M. Countertransference Management and Effective Psychotherapy: Meta-Analytic Findings. *Psychotherapy* **2018**, *55*, 496–507. [CrossRef]
5. Pascual-Leone, A. How Clients "Change Emotion with Emotion": A Programme of Research on Emotional Processing. *Psychother. Res.* **2018**, *28*, 165–182. [CrossRef]
6. Peluso, P.R.; Freund, R.R. Therapist and Client Emotional Expression and Psychotherapy Outcomes: A Meta-Analysis. *Psychotherapy* **2018**, *55*, 461–472. [CrossRef] [PubMed]
7. Greenberg, L.S. Emotions, the Great Captains of Our Lives: Their Role in the Process of Change in Psychotherapy. *Am. Psychol.* **2012**, *67*, 697–707. [CrossRef] [PubMed]
8. Stefana, A. Erotic Transference. *Br. J. Psychother.* **2017**, *33*, 505–513. [CrossRef]
9. Bradley, R.; Heim, A.K.; Westen, D. Transference Patterns in the Psychotherapy of Personality Disorders: Empirical Investigation. *Br. J. Psychiatry* **2005**, *186*, 342–349. [CrossRef] [PubMed]
10. Tanzilli, A.; Colli, A.; Gualco, I.; Lingiardi, V. Patient Personality and Relational Patterns in Psychotherapy: Factor Structure, Reliability, and Validity of the Psychotherapy Relationship Questionnaire. *J. Personal. Assess.* **2018**, *100*, 96–106. [CrossRef]
11. Gabbard, G.O. *Psychodynamic Psychiatry in Clinical Practice*, 5th ed.; American Psychiatric Publishing: Arlington, VA, USA, 2014.

12. McWilliams, N. *Psychoanalytic Diagnosis: Understanding Personality Structure in the Clinical Process*; The Guilford Press: New York, NY, USA, 2011.

13. Lingiardi, V.; McWilliams, N. *Psychodynamic Diagnostic Manual: PDM-2*; Guilford Press: New York, NY, USA, 2017.

14. Clarkin, J.F.; Caligor, E.; Sowislo, J.F. Transference-Focused Psychotherapy for Levels of Personality Pathology Severity. In *Gabbard's Textbook of Psychotherapeutic Treatments*; Crisp, H., Gabbard, G.O., Eds.; American Psychiatric Publishing: Washington, DC, USA, 2022.

15. Høglend, P. Exploration of the Patient-Therapist Relationship in Psychotherapy. *AJP* **2014**, *171*, 1056–1066. [CrossRef]

16. Markin, R.D. Toward a Common Identity for Relationally Oriented Clinicians: A Place to Hang One's Hat. *Psychotherapy* **2014**, *51*, 327–333. [CrossRef]

17. Ulberg, R.; Hummelen, B.; Hersoug, A.G.; Midgley, N.; Høglend, P.A.; Dahl, H.-S.J. The First Experimental Study of Transference Work–in Teenagers (FEST–IT): A Multicentre, Observer- and Patient-Blind, Randomised Controlled Component Study. *BMC Psychiatry* **2021**, *21*, 106. [CrossRef]

18. Bhatia, A.; Gelso, C.J. Therapists' Perspective on the Therapeutic Relationship: Examining a Tripartite Model. *Couns. Psychol. Q.* **2018**, *31*, 271–293. [CrossRef]

19. Høglend, P.; Hagtvet, K. Change Mechanisms in Psychotherapy: Both Improved Insight and Improved Affective Awareness Are Necessary. *J. Consult. Clin. Psychol.* **2019**, *87*, 332–344. [CrossRef]

20. Markin, R.D.; McCarthy, K.S.; Barber, J.P. Transference, Countertransference, Emotional Expression, and Session Quality over the Course of Supportive Expressive Therapy: The Raters' Perspective. *Psychother. Res.* **2013**, *23*, 152–168. [CrossRef] [PubMed]

21. Greenberg, L.S. Theory of Functioning in Emotion-Focused Therapy. In *Clinical Handbook of Emotion-Focused Therapy*; Greenberg, L.S., Goldman, R.N., Eds.; American Psychological Association: Washington, DC, USA, 2019; pp. 37–59, ISBN 9781433829772.

22. Greenberg, L.S.; Goldman, R.N. Theory of Practice of Emotion-Focused Therapy. In *Clinical Handbook of Emotion-Focused Therapy*; Greenberg, L.S., Goldman, R.N., Eds.; American Psychological Association: Washington, DC, USA, 2019; pp. 61–89, ISBN 9781433829772.

23. Stefana, A.; Bulgari, V.; Youngstrom, E.A.; Dakanalis, A.; Bordin, C.; Hopwood, C.J. Patient Personality and Psychotherapist Reactions in Individual Psychotherapy Setting: A Systematic Review. *Clin. Psychol. Psychother.* **2020**, *27*, 697–713. [CrossRef] [PubMed]

24. Stefana, A. *History of Countertransference: From Freud to the British Object Relations School*, 1st ed.; Routledge: London, UK, 2017; ISBN 9781315445601. [CrossRef]

25. Stefana, A.; Hinshelwood, R.D.; Borensztejn, C.L. Racker and Heimann on Countertransference: Similarities and Differences. *Psychoanal. Q.* **2021**, *90*, 105–137. [CrossRef]

26. Norcross, J.C.; Lambert, M.J. *Psychotherapy Relationships That Work, Volume 1: Evidence-Based Therapist Contributions*, 3rd ed.; Oxford University Press: Oxford, UK, 2019.

27. Stefana, A.; Fusar-Poli, P.; Gnisci, C.; Vieta, E.; Youngstrom, E.A. Clinicians' Emotional Reactions toward Patients with Depressive Symptoms in Mood Disorders: A Narrative Scoping Review of Empirical Research. *Int. J. Environ. Res. Public Health* **2022**, *19*, 15403. [CrossRef]

28. Luborsky, L.; Crits-Christoph, P. *Understanding Transference: The Core Conflictual Relationship Theme Method*; Subsequent Edition; American Psychological Association: Washington, DC, USA, 1998; ISBN 9781557984531.

29. Bänninger-Huber, E.; Salvenauer, S. Different Types of Laughter and Their Function for Emotion Regulation in Dyadic Interactions. *Curr. Psychol.* **2022**. [CrossRef]

30. Benecke, C.; Peham, D.; Bänninger-Huber, E. Nonverbal Relationship Regulation in Psychotherapy. *Psychother. Res.* **2005**, *15*, 81–90. [CrossRef]

31. Wieder, G.; Fischer, M.S.; Einsle, F.; Baucom, D.H.; Hahlweg, K.; Wittchen, H.-U.; Weusthoff, S. Fundamental Frequency during Cognitive Preparation and Its Impact on Therapy Outcome for Panic Disorder with Agoraphobia. *Behav. Res. Ther.* **2020**, *135*, 103728. [CrossRef]

32. Multon, K.D.; Patton, M.J.; Kivlighan, D.M. Development of the Missouri Identifying Transference Scale. *J. Couns. Psychol.* **1996**, *43*, 243–252. [CrossRef]

33. Zittel Conklin, C.; Westen, D. *The Countertransference Questionnaire*; Emory University Departments of Psychology and Psychiatry and Behavioral Sciences: Atlanta, GE, USA, 2003.

34. McLaughlin, A.A.; Keller, S.M.; Feeny, N.C.; Youngstrom, E.A.; Zoellner, L.A. Patterns of Therapeutic Alliance: Rupture–Repair Episodes in Prolonged Exposure for Posttraumatic Stress Disorder. *J. Consult. Clin. Psychol.* **2014**, *82*, 112–121. [CrossRef]

35. Izard, C.E.; Libero, D.Z.; Putnam, P.; Haynes, O.M. Stability of Emotion Experiences and Their Relations to Traits of Personality. *J. Personal. Soc. Psychol.* **1993**, *64*, 847–860. [CrossRef] [PubMed]

36. Wampold, B.E.; Flückiger, C. The Alliance in Mental Health Care: Conceptualization, Evidence and Clinical Applications. *World Psychiatry* **2023**, *22*, 25–41. [CrossRef] [PubMed]

37. Flückiger, C.; Del Re, A.C.; Wampold, B.E.; Horvath, A.O. The Alliance in Adult Psychotherapy: A Meta-Analytic Synthesis. *Psychotherapy* **2018**, *55*, 316–340. [CrossRef]

38. Faul, L.; LaBar, K.S. Mood-Congruent Memory Revisited. *Psychol. Rev.* **2022**. [CrossRef]

39. Harris, P.A.; Scott, K.W.; Lebo, L.; Hassan, N.; Lightner, C.; Pulley, J. ResearchMatch: A National Registry to Recruit Volunteers for Clinical Research. *Acad. Med.* **2012**, *87*, 66–73. [CrossRef]

40. Chandler, J.; Shapiro, D. Conducting Clinical Research Using Crowdsourced Convenience Samples. *Annu. Rev. Clin. Psychol.* **2016**, *12*, 53–81. [CrossRef]
41. DeVellis, R.F.; Thorpe, C.T. *Scale Development: Theory and Applications*, 5th ed.; Sage: Thousand Oaks, CA, USA, 2022.
42. McCoach, D.B.; Gable, R.K.; Madura, J.P. *Instrument Development in the Affective Domain: School and Corporate Applications*, 3rd ed.; Springer Science + Business Media: New York, NY, USA, 2013; ISBN 978-1-4614-7134-9.
43. Tellegen, A.; Watson, D.; Clark, L.A. On the Dimensional and Hierarchical Structure of Affect. *Psychol. Sci.* **1999**, *10*, 297–303. [CrossRef]
44. Clark, L.A.; Watson, D. Constructing Validity: New Developments in Creating Objective Measuring Instruments. *Psychol. Assess.* **2019**, *31*, 1412–1427. [CrossRef]
45. Johnson, S.L.; Leedom, L.J.; Muhtadie, L. The Dominance Behavioral System and Psychopathology: Evidence from Self-Report, Observational, and Biological Studies. *Psychol. Bull.* **2012**, *138*, 692–743. [CrossRef]
46. Gray, J.A.; McNaughton, N. The Neuropsychology of Anxiety: Reprise. In *Perspectives in Anxiety, Panic and Fear*; Hope, D.A., Ed.; University of Nebraska Press: Lincoln, NE, USA, 1996; Volume 43, pp. 61–134.
47. Gilbert, P.; Allan, S.; Brough, S.; Melley, S.; Miles, J.N.V. Relationship of Anhedonia and Anxiety to Social Rank, Defeat and Entrapment. *J. Affect. Disord.* **2002**, *71*, 141–151. [CrossRef] [PubMed]
48. Plutchik, R. *Emotion: A Psychoevolutionary Synthesis*; Harper & Row: Manhattan, NY, USA, 1980.
49. Eagle, M.N. *Core Concepts in Contemporary Psychoanalysis: Clinical, Research Evidence and Conceptual Critiques*, 1st ed.; Routledge: Abingdon-on-Thames, UK, 2017; ISBN 9781315142111.
50. Ulberg, R.; Amlo, S.; Critchfield, K.L.; Marble, A.; Høglend, P. Transference Interventions and the Process between Therapist and Patient. *Psychotherapy* **2014**, *51*, 258–269. [CrossRef]
51. Racker, H. *Transference and Countertransference*, 1st ed.; Routledge: Abingdon-on-Thames, UK, 2018; ISBN 9780429484209.
52. Gelso, C.J.; Kline, K.V. The Sister Concepts of the Working Alliance and the Real Relationship: On Their Development, Rupture, and Repair. *Res. Psychother. Psychopathol. Process Outcome* **2019**, *22*, 373. [CrossRef] [PubMed]
53. Gelso, C.J.; Kivlighan, D.M.; Markin, R.D. The Real Relationship. In *Psychotherapy Relationships that Work*; Oxford University Press: Oxford, UK, 2019; pp. 351–378. ISBN 9780190843953.
54. Streiner, D.L.; Norman, G.R.; Cairney, J. *Health Measurement Scales: A Practical Guide to Their Development and Use*, 5th ed.; Oxford University Press: Oxford, UK, 2015; ISBN 9780199685219.
55. Lambert, M.J.; Burlingame, G.M.; Umphress, V.; Hansen, N.B.; Vermeersch, D.A.; Clouse, G.C.; Yanchar, S.C. The Reliability and Validity of the Outcome Questionnaire. *Clin. Psychol. Psychother.* **1996**, *3*, 249–258. [CrossRef]
56. Russell, J.A.; Carroll, J.M. On the Bipolarity of Positive and Negative Affect. *Psychol. Bull.* **1999**, *125*, 3–30. [CrossRef]
57. Revelle, W.; Condon, D.M. Reliability from α to ω: A Tutorial. *Psychol. Assess.* **2019**, *31*, 1395–1411. [CrossRef]
58. Sellbom, M.; Tellegen, A. Factor Analysis in Psychological Assessment Research: Common Pitfalls and Recommendations. *Psychol. Assess.* **2019**, *31*, 1428–1441. [CrossRef] [PubMed]
59. Youngstrom, E.A.; Van Meter, A.; Frazier, T.W.; Youngstrom, J.K.; Findling, R.L. Developing and Validating Short Forms of the Parent General Behavior Inventory Mania and Depression Scales for Rating Youth Mood Symptoms. *J. Clin. Child Adolesc. Psychol.* **2020**, *49*, 162–177. [CrossRef]
60. Hair, J.F., Jr.; Black, W.C.; Babin, B.J.; Anderson, R.A. *Multivariate Data Analysis*, 8th ed.; Cenage: Boston, MA, USA, 2023.
61. Dinno, A. Paran: Horn's Test of Principal Components/Factors. 2018. Available online: https://cran.r-project.org/web/packages/paran/paran.pdf (accessed on 2 August 2023).
62. Steiner, M.; Grieder, S. EFAtools: An R Package with Fast and Flexible Implementations of Exploratory Factor Analysis Tools. *J. Open Source Softw.* **2020**, *5*, 2521. [CrossRef]
63. Nunnally, J.C.; Bernstein, I.H. Psychometric Theory. In *McGraw-Hill Series in Psychology*, 3rd ed.; McGraw-Hill: New York, NY, USA, 1994; ISBN 9780070478497.
64. Raykov, T.; Marcoulides, G.A. *Introduction to Psychometric Theory*; Routledge: Abingdon-on-Thames, UK, 2011; ISBN 9781136900037.
65. Chalmers, R.P. Mirt: A Multidimensional Item Response Theory Package for the *R* Environment. *J. Stat. Soft.* **2012**, *48*. [CrossRef]
66. Rosseel, Y. Lavaan: An *R* Package for Structural Equation Modeling. *J. Stat. Soft.* **2012**, *48*, 1–36. [CrossRef]
67. Nickodem, K.; Halpin, P. Kfa: K-Fold Cross Validation for Factor Analysis. 2022. Available online: https://cran.r-project.org/web/packages/kfa/kfa.pdf (accessed on 2 August 2023).
68. Horn, J.L. A Rationale and Test for the Number of Factors in Factor Analysis. *Psychometrika* **1965**, *30*, 179–185. [CrossRef] [PubMed]
69. Glorfeld, L.W. An Improvement on Horn's Parallel Analysis Methodology for Selecting the Correct Number of Factors to Retain. *Educ. Psychol. Meas.* **1995**, *55*, 377–393. [CrossRef]
70. Lorenzo-Seva, U.; Timmerman, M.E.; Kiers, H.A.L. The Hull Method for Selecting the Number of Common Factors. *Multivar. Behav. Res.* **2011**, *46*, 340–364. [CrossRef] [PubMed]
71. Ruscio, J.; Roche, B. Determining the Number of Factors to Retain in an Exploratory Factor Analysis Using Comparison Data of Known Factorial Structure. *Psychol. Assess.* **2012**, *24*, 282–292. [CrossRef]
72. Samejima, F. The General Graded Response Model. In *Handbook of Polytomous Item Response Theory Models*; Nering, M.L., Ostini, R., Eds.; Routledge: Abingon-on-Thames, UK, 2010; pp. 77–107.

73.   Jacobson, N.S.; Truax, P. Clinical Significance: A Statistical Approach to Defining Meaningful Change in Psychotherapy Research. *J. Consult. Clin. Psychol.* **1991**, *59*, 12–19. [CrossRef]

74.   Norman, G.R.; Sloan, J.A.; Wyrwich, K.W. Interpretation of Changes in Health-Related Quality of Life: The Remarkable Universality of Half a Standard Deviation. *Med. Care* **2003**, *41*, 582–592. [CrossRef]

75.   Ægisdóttir, S.; White, M.J.; Spengler, P.M.; Maugherman, A.S.; Anderson, L.A.; Cook, R.S.; Nichols, C.N.; Lampropoulos, G.K.; Walker, B.S.; Cohen, G.; et al. The Meta-Analysis of Clinical Judgment Project: Fifty-Six Years of Accumulated Research on Clinical Versus Statistical Prediction. *Couns. Psychol.* **2006**, *34*, 341–382. [CrossRef]

76.   Markin, R.D.; Kivlighan, D.M. Bias in Psychotherapist Ratings of Client Transference and Insight. *Psychother. Theory Res. Pract. Train.* **2007**, *44*, 300–315. [CrossRef] [PubMed]

77.   Carlier, I.V.E.; Meuldijk, D.; Van Vliet, I.M.; Van Fenema, E.; Van der Wee, N.J.A.; Zitman, F.G. Routine Outcome Monitoring and Feedback on Physical or Mental Health Status: Evidence and Theory: Feedback on Physical or Mental Health Status. *J. Eval. Clin. Pract.* **2012**, *18*, 104–110. [CrossRef] [PubMed]

78.   Lambert, M.J.; Whipple, J.L.; Kleinstäuber, M. Collecting and Delivering Progress Feedback: A Meta-Analysis of Routine Outcome Monitoring. *Psychotherapy* **2018**, *55*, 520–537. [CrossRef] [PubMed]

79.   Høglend, P.; Gabbard, G.O. When Is Transference Work Useful in Psychodynamic Psychotherapy? A Review of Empirical Research. In *Psychodynamic Psychotherapy Research*; Levy, R.A., Ablon, J.S., Kächele, H., Eds.; Humana Press: Totowa, NJ, USA, 2012; pp. 449–467; ISBN 9781607617914.

80.   Poston, J.M.; Hanson, W.E. Meta-Analysis of Psychological Assessment as a Therapeutic Intervention. *Psychol. Assess.* **2010**, *22*, 203–212. [CrossRef] [PubMed]

81.   Bickman, L. A Measurement Feedback System (MFS) Is Necessary to Improve Mental Health Outcomes. *J. Am. Acad. Child Adolesc. Psychiatry* **2008**, *47*, 1114–1119. [CrossRef]

82.   MacKenzie, S.B.; Podsakoff, P.M.; Podsakoff, N.P. Construct Measurement and Validation Procedures in MIS and Behavioral Research: Integrating New and Existing Techniques. *MIS Q.* **2011**, *35*, 293–334. [CrossRef]

83.   Colli, A.; Tanzilli, A.; Gualco, I.; Lingiardi, V. Empirically Derived Relational Pattern Prototypes in the Treatment of Personality Disorders. *Psychopathology* **2016**, *49*, 364–373. [CrossRef]

84.   Prasko, J.; Ociskova, M.; Vanek, J.; Burkauskas, J.; Slepecky, M.; Bite, I.; Krone, I.; Sollar, T.; Juskiene, A. Managing Transference and Countertransference in Cognitive Behavioral Supervision: Theoretical Framework and Clinical Application. *Psychol. Res. Behav. Manag.* **2022**, *15*, 2129–2155. [CrossRef]

85.   Clarkin, J.F.; Yeomans, F.E.; Kernberg, O.F. *Psychotherapy for Borderline Personality*; Wiley: Hoboken, NJ, USA, 1999.

86.   Swift, J.K.; Greenberg, R.P. Premature Discontinuation in Adult Psychotherapy: A Meta-Analysis. *J. Consult. Clin. Psychol.* **2012**, *80*, 547–559. [CrossRef] [PubMed]

87.   Youngstrom, E.A.; Van Meter, A.; Frazier, T.W.; Hunsley, J.; Prinstein, M.J.; Ong, M.-L.; Youngstrom, J.K. Evidence-Based Assessment as an Integrative Model for Applying Psychological Science to Guide the Voyage of Treatment. *Clin. Psychol. Sci. Pract.* **2017**, *24*, 331–363. [CrossRef]

88.   Depue, R.A.; Slater, J.F.; Wolfstetter-Kausch, H.; Klein, D.; Goplerud, E.; Farr, D. A Behavioral Paradigm for Identifying Persons at Risk for Bipolar Depressive Disorder: A Conceptual Framework and Five Validation Studies. *J. Abnorm. Psychol.* **1981**, *90*, 381–437. [CrossRef] [PubMed]

89.   Betan, E.J.; Heim, A.K.; Zittel Conklin, C.; Westen, D. Countertransference Phenomena and Personality Pathology in Clinical Practice: An Empirical Investigation. *Am. J. Psychiatry* **2005**, *162*, 890–898. [CrossRef] [PubMed]

90.   Colli, A.; Tanzilli, A.; Dimaggio, G.; Lingiardi, V. Patient Personality and Therapist Response: An Empirical Investigation. *Am. J. Psychiatry* **2014**, *171*, 102–108. [CrossRef]

91.   Hayes, J.A.; Riker, J.; Ingram, K. Countertransference Behavior and Management in Brief Counseling: A Field Study. *Psychother. Res.* **1997**, *7*, 145–153. [CrossRef]