The Development of the Epistemic Trust Rating System (ETRS)

[1]Shimrit Fisher, [1]Timur Guralnik, [2]Peter Fonagy, [1]Sigal Zilcha-Mano

[1]The Department of Psychology, University of Haifa, Israel

[2] Research Department of Clinical, Educational and Health Psychology, University College London, London WC1E 7HB, United Kingdom

Running title: MEASURING EPISTEMIC TRUST IN PSYCHOTHERAPY

Acknowledgments: The research was supported by a grant from the Israeli Science Foundation (Grant no. 186/15).

Corresponding Author's Contact Information:

Shimrit Fisher, Ph.D. candidate

Department of Psychology, University of Haifa

Mount Carmel, Haifa 31905, Israel

Email: Shimrit.shahar@gmail.com

The Development of the Epistemic Trust Rating System (ETRS)

Keywords: Epistemic trust, Attachment, Psychotherapy, Psychometrics, measurement, therapeutic relationship

Abstract

Objective: The present study introduces and validates the Epistemic Trust Rating System (ETRS), an observer-based measure designed to assess epistemic trust (ET) within psychotherapy. ET in psychotherapy has gained much theoretical attention as a critical component in the therapeutic context, given its inherent link to social communication. However, its empirical validation remains pending, largely due to the absence of a refined instrument to gauge ET levels within the therapy environment. Therefore, this study aimed to translate ET's theoretical construct into tangible markers within the therapeutic context.

Method: One hundred eighteen patients enrolled in a randomized controlled trial received psychodynamic psychotherapy for depression. Incorporating top-down theoretical considerations with bottom-up empirical observations yielded an overall ETRS score accompanied by three distinct sub-scales, each assessing a singular ET element: one gauges the patient's propensity to share; another measures the degree to which 'we-mode' moments are achieved within the session; and the third evaluates the patient's receptiveness to learning.

Results: The findings demonstrate the psychometric robustness of the ETRS, with good internal consistency, interrater reliability (ICC(1,8)=.86-.90), and convergent (r=.23-.29) and discriminant validity (r=-.10).

Conclusion: The current study highlights the ETRS as a promising tool bridging theory and empirical exploration, enhancing our understanding of epistemic trust in psychotherapy.

Clinical and Methodological Significance of this Article: This study provides evidence of the reliability and validity of the ETRS, an observer-based measure of epistemic trust in therapeutic relationships. The ETRS demonstrates its potential to be a valuable asset for future research endeavors aimed at unraveling the significance of epistemic trust in the processes and outcomes of psychotherapy. Measuring a complex, multifaceted construct like epistemic trust narrows the existing gap between theoretical concepts and empirical assessment. This measure can equip therapists with a means to customize their approach to meet their patients' individual needs, potentially improving therapeutic outcomes.

In our busy social world, filled with so much information, people cannot take everything all at once. So, what strategies do we employ to navigate this information overload effectively? Epistemology, the philosophical study of how we gain knowledge (Fumerton, 2009), helps us understand this long-standing question. Central to this is the notion that our engagement with knowledge is influenced by our readiness to deem new information as both credible and pertinent. This ability, called epistemic trust (ET; Fonagy & Allison, 2014), mainly deals with how people first respond to new information (McCraw, 2015) and how much they trust others for information (Duschinsky & Foster, 2021; Wilholt, 2020). Even though this idea started in philosophy, many areas have looked into it from various perspectives. For instance, the field of educational sciences regards ET as integral to reasoning and learning from others' testimonies (Durkin & Shafto, 2016). Developmental psychology frames ET as a basic socio-emotional competence that aids in choosing between contradictory informants (Corriveau et al., 2009). In linguistics, ET is seen as a function of evaluating novel information rooted in verbal comprehension and aptitude (Clarke et al., 2018). Cognitive anthropology, by contrast, views ET as a social asset vital for secure knowledge exchange within a cultural context (Bergin, 2001). Despite nuances in definition across disciplines, the consensus suggests ET as a fundamental skill in human social adaptation to multifaceted challenges.

Conceptualizing ET in Psychotherapy

The idea of epistemic trust (ET) has become popular in many fields, but only recently has it gained much attention in psychotherapy (Fonagy et al., 2015; 2019). ET emerges as a critical component in the therapeutic context, given its inherent link to social communication. In particular, it involves partners (therapist and patient) collaboratively interpreting the patient's past and present experiences to foster more adaptive life approaches (Weiner & Borenstein,

2009). Thus, ET in psychotherapy is defined as the patient's propensity to perceive interpersonal cues from the therapist, the therapeutic relationship, and the therapy milieu as ensuring personal relevance. Such inclination makes subsequent communication feel actionable and generalizable to future situations (Fonagy & Allison, 2014; Fonagy et al., 2019). Building on this definition, further developments of this concept considered ET a potential key tool in therapy that enables patients to securely assimilate essential insights about themselves and others, facilitating adaptive social navigation (Fonagy et al., 2017). According to this paradigm, patients are more amenable to embracing new insights when therapists establish ET by presenting a coherent model to comprehend their emotions, cognitions, and actions—irrespective of the classes of therapeutic technique employed. The significance of patients' ET levels prior to therapy initiation, coupled with its potential as a catalyst for therapeutic transformation, has been emphasized (Fonagy & Allison, 2014).

Recent theoretical conceptualizations suggest that ET in psychotherapy, when viewed as a unified construct, is underpinned by three discernible components: sharing, we-mode, and learning (Fisher et al., 2022; 2023). The element of sharing is fundamentally a facet of social communication. It pertains to individuals' tendencies to either articulate or withhold their personal experiences with others during times of need. The decision matrix surrounding the act of sharing—specifically what to share, when to share, and with whom—often arises from a blend of tangible past experiences or envisaged future expectations. For instance, an individual might be inclined to share based on a past interaction ("Given my past experiences of sharing with him, he has consistently provided support") or predicated on prospective outcomes ("If I confide in him, he might offer assistance"). This inclination to disclose personal experiences, emotions, and

thoughts in pursuit of collaborative solutions facilitates a deeper understanding of self and interpersonal relationships (Mercier & Sperber, 2017).

Upon sharing an experience, the expectation is that the listener will be able to empathetically comprehend the conveyed subjective experience (Luyten et al., 2020). When this expectation is met, a dialogue ensues, hallmarked by mutual recognition, joint focus, and a thorough exploration of the topic. This phenomenon of mutual acknowledgment, where both parties perceive their understanding as aligned, is termed 'we-mode' (Fonagy et al., 2021). Functioning as a sophisticated cognitive state, we-mode establishes a conducive backdrop for the critique of existing perspectives or the genesis of new insights (Benjamin, 2004; Fonagy et al., 2019). Within the therapeutic environment, we-mode moments ideally manifest when a patient's candid disclosure aligns with a therapist's genuine empathic engagement and understanding. Such interactions, consequently, can be depicted as comprehensive dialogues wherein both therapist and patient work synergistically to derive meaning from occurrences, cognitive processes, and emotional states. This collaboration is evident in their adept alternation during discourse, sustained eye contact, the avoidance of prolonged awkward silences, consistent topic engagement, and the seamless flow of conversation with both parties contributing.

When patients and therapists jointly endeavor to derive meaning from events, thoughts, and emotions, the process of learning is facilitated. Conceptually, learning is characterized by an innate human propensity to be open to acquiring new knowledge from others concerning the physical and mental world, including knowledge about oneself or others (Csibra & Gergely, 2006). Within the therapeutic context, this openness towards learning manifests in various ways. An individual might, for instance, discern the presence of a learning opportunity yet struggle to pinpoint its nature or implications ("*Can you explain to me why this is happening to me?*"). They

might also be able to attribute their challenges to a particular behavioral pattern ("*I now see how comprehensive it is in my life*"), understand the prerequisites for its alteration, and apply this newfound understanding in distinct social situations, exemplified by: "*When I got angry, this time, I didn't avoid her. I told her how it felt. I was relieved*."

These three elements are interrelated, jointly constituting the multidimensional yet cohesive construct of ET. While each facet presents its unique insights, the composite ET construct that emerges from their synergy is undeniably of import. The propensity to view the therapeutic milieu as a space conducive to social *learning* might hinge on the therapeutic dyad's capability to cultivate and maintain a state of *we-mode*, which fundamentally relies on shared intentionality, the unreserved *sharing* of experiences, emotions, and cognitions. However, the empirical validation of these conceptualizations remains pending, largely due to the absence of a refined instrument to gauge ET levels within the therapy environment.

Operationalizing ET

While the triadic model has not been operationalized within the realm of psychotherapy, numerous studies have attempted to develop empirical frameworks to gauge ET across different life stages: infancy and early childhood (e.g., Corriveau et al., 2009; Egyed et al., 2013), and adulthood (Shafto et al., 2012; Schröder-Pfeifer et al., 2018; 2022). Predominantly, these investigations adopt a cognitive lens, emphasizing participants' capacity to discern differential accuracy when assimilating new information (Pasquini et al., 2007). For instance, Durkin & Shafto (2016) explored the role of informant reliability in augmenting children's and adults' conceptual knowledge and in recalibrating pre-existing misconceptions based on credible feedback. Similarly, Schröder-Pfeifer et al. (2018) probed the extent to which individuals

adjusted self-perceptions in light of input from expert committees. While these endeavors have employed experimental methodologies to assess ET, their applicability is limited in the therapeutic context. Recent efforts have centered on crafting self-report tools tailored to encapsulate respondents' subjective experiences (e.g., Campbell et al., 2021). However, to critically assess the proposed theoretical underpinnings of ET and its variance, a bespoke rating system for the therapeutic environment is imperative. Such a system could serve as a key adjunct to insights derived from self-report instruments, primarily illuminating the subjective realm of experiences. By leveraging this rating approach, we stand to access layers of understanding that may elude self-reporting, thus fostering a more subtle and holistic grasp of their patients' experiences (Fonagy et al., 2016).

The present study

The primary aim of this study is to develop and validate the Epistemic Trust Rating System (ETRS), an observer-based instrument designed to assess epistemic trust (ET) in psychotherapy sessions using the framework of the triadic model. Our methodology combined top-down theoretical approaches with bottom-up empirical observations, allowing us to identify expected ET markers and refine them through direct clinical observation. The video recordings used in this study were drawn from a larger randomized control trial (RCT) undertaken at the University of Haifa (Zilcha-Mano et al., 2018; 2021). We focused particularly on the fourth session of the supportive-expressive (SE) treatment protocol (Luborsky, 1984), which Leibovich et al. (2019) identified as crucial in influencing the therapy's future direction. To establish the scale's development systematically, we examined the convergent validity of the ETRS by investigating its relationship with two related constructs: Attachment and Working Alliance.

a) **Attachment**: The development of ET is theorized to occur within attachment relationships, suggesting that ET's evolution is likely influenced by different attachment styles (Luyten et al., 2021). Accordingly, we hypothesized a positive correlation between ETRS scores and secure attachment, as indicated by high scores on the secure dimension of the Clients Attachment to Therapist Scale (CATS; Mallinckrodt et al., 1995). In contrast, we expected negative correlations with insecure attachment styles, as reflected in the CATS preoccupied/merger and avoidant/fearful subscales, and the anxiety and avoidance dimensions of the Experience in Close Relationships-RS (ECR-RS; Fraley et al., 2011).
b) **Working Alliance**: Considering the mutual focus on the collaborative dynamics between patient and therapist in both ET and the therapeutic working alliance, we anticipated a positive correlation between ETRS scores and the working alliance. This relationship is to be measured using the Working Alliance Inventory (WAI; Horvath & Greenberg, 1989; Tracey & Kokotovic, 1989).

Our approach, which does not rely on self-report, presents a unique opportunity to explore the congruence between objectively coded data and individuals' self-perception. By comparing the ETRS to self-report instruments, we were able to investigate the relationship between the coded information from session transcripts (as per the ETRS) and the patient's own subjective experience and self-awareness. To evaluate the discriminant validity of the ETRS, we examined its correlation with the Rosenberg Self-Esteem Scale (RSES; Rosenberg, 1979). Given that ET primarily assesses interpersonal dynamics, while the RSES focuses on intrapersonal aspects of self-esteem, we did not expect a significant correlation between the ETRS and RSES scores.

Method

Trial design and treatments

Patients were enlisted in response to advertisements offering free treatment at the psychotherapy research lab clinic (Zilcha-Mano et al., 2018; 2021). Participants who met the predefined inclusion and exclusion criteria (detailed below) were randomly assigned to either supportive-focused or supportive-expressive-focused treatment (Luborsky, 1984), each spanning 16 individual weekly sessions. Both operate within the framework of manualized psychodynamic psychotherapies. Assignment to the treatment arm was conducted by an outside institution not involved in the study. The study received approval from the institutional review board where data acquisition was undertaken. All participating patients and therapists signed consent forms agreeing to have their treatment sessions videotaped and for all quantitative and qualitative data obtained from the treatment to be used for research purposes, with personal information anonymized.

Participants

Patients

Data were collected from 118 patients, encompassing the entire RCT as well as the pilot phase. Technical failures in recording sessions led to the exclusion of two patients, rendering a final sample of 116. Participants were eligible for the study if they met the following criteria: (a) a diagnosis of Major Depressive Disorder (MDD) based on structured clinical interviews in accordance with the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, with scores above 14 on the 17-item Hamilton Rating Scale for Depression (HRSD; Hamilton, 1967) at two evaluation points, one week apart, and a diagnosis of MDD based on the Mini

International Neuropsychiatric Interview (MINI; Sheehan et al., 1998); (b) if the patients were on medication, their dosage must have been stable for at least 3 months before entering the study, and they had to be willing to maintain a stable dosage for the duration of the treatment; (c) patient's age must be between 18 and 65 years. (d) Hebrew language fluency; and (e) provision of written informed consent. Patients were also screened for the following exclusion criteria: (a) current risk of suicide or self-harm (HRSD suicide item > 2); (b) current substance abuse disorder; (c) current or past schizophrenia or psychosis, bipolar disorder, or severe eating disorder requiring medical monitoring; (d) history of organic mental disease; and (e) currently in psychotherapy. Table 1 provides a detailed overview of the demographic and clinical attributes of the participants. The sample comprises patients with an average age of 30.97 years (SD = 8.17) and an average educational attainment of 14.27 years (SD = 2.21). The majority of participants are female (59.9%) and employed (67.2%). Regarding income, 23% of the participants reported earnings above the average. In terms of marital status, a significant proportion are single (77.6%), followed by 17.2% who are married and 4.3% who are divorced. The majority of the sample identifies as Jewish (71.6%), with smaller percentages being Christian (1.7%), Muslim (7.8%), atheists or having other religious beliefs (9.5%). Clinically, 12.1% of the participants are currently on medication, 25% have a history of medication use, and 48.3% have previously undergone psychotherapy. Notably, a considerable portion of the sample (76.4%) has comorbidity with a personality disorder.

Table 1

Patients demographic and clinical characteristics

Therapists

The RCT and its pilot phase involved nine therapists, comprised of six women and three men, averaging 14.42 years of clinical experience. The therapists attended a 20-hour training workshop in supportive and expressive therapeutic techniques. Before proceeding to the trial phase, each therapist completed treatments for two pilot patients, one for each type, and was required to demonstrate adequate adherence to the treatment protocols before moving into the trial phase. Subsequent to the pilot phase, two therapists did not continue to the active phase of the RCT. One secured a full-time position elsewhere, while the other demonstrated low levels of adherence. Nevertheless, data from their pilot cases were retained for analysis. Each therapist received weekly group supervision from two supervisors, as well as weekly individual supervision from one of the supervisors, using videotaped sessions for feedback.

Research team

The development of the ETRS proceeded in two successive rounds involving two distinct teams. The first team was responsible for the construction of the rating system, while the second team rated the entire sample to derive ET scores. This dual approach, primarily influenced by logistical challenges due to rotating team memberships over the protracted development period (which spanned several years), inadvertently offered an opportunity. The rotation allowed us to evaluate if a fresh team could effectively learn the newly developed rating system and adhere to the manual, ensuring its usability extended beyond its original creators.

The **development team** comprised two Ph.D. candidates, one of whom is a licensed clinical psychologist and four undergraduate psychology students. They were instrumental in translating theoretical constructs into a tangible, measurable framework, leading to the formulation of a preliminary rating system template. Their efforts were overseen by two

additional experts: one, the pioneer of the ET concept and its application in psychotherapy, and the other, with a deep-rooted expertise in empirical research methodologies within psychotherapy.

The **rating team** retained the two aforementioned Ph.D. candidates but introduced a new contingent of six undergraduate psychology students for the rating process. To ensure rigor, these undergraduate team members underwent comprehensive training on the ET concept and its associated rating system. Once they exhibited acceptable adherence to the rating procedure, each rater was assigned approximately 25 sessions for independent evaluation.

Scale Development

The ETRS was designed to function as an observer-based measure for pinpointing ET within therapeutic encounters. Incorporating both top-down and bottom-up methodologies, this system sought to identify specific markers of ET evident during therapy sessions. Defined as verbal utterances or observable behaviors manifested by the patient during a session, these markers closely align with the triadic elements of ET. Specifically, any reference made by patients about disclosing emotions, thoughts, or personal experiences was categorized as a marker of 'sharing.' Expressions indicating fresh perceptions about oneself or others, stemming from interpersonal interactions, were identified as 'learning' markers. Additionally, the dynamics of the discourse between the patient and therapist were seen as potential indicative markers of the 'we-mode' or its absence. Both the development and the rating teams of the ETRS worked following a systematic step-wise process.

The development process

Step 1: Top-Down Approach – Translating ET Elements. The goal of this first step was to translate ET elements into observable markers. This initial phase aimed to bridge theoretical conceptualizations of ET elements with potential discernible markers manifested in patients' speech and behavior. To this end, the development team delineated prototypical examples anticipated to surface within patients' discourse, as illustrated in Text Box 1. Subsequent consultation with the two supervisors ensured the accurate and valid translation of theoretical constructs into tangible markers within the therapeutic context.

Text box 1:

Observational markers of ET in therapy sessions

Step 2: Bottom-Up Approach – Formulation of the ETRS. Armed with the postulated markers, team members scrutinized the fourth videotaped session of fifteen patients, seeking these markers and achieving consensus on their relevance to ET. The rationale for selecting the fourth session stemmed from its key role in the treatment manual (Book, 1998). In this specific session, therapists are guided to articulate the patient's unspoken formulation overtly, delivering insights that could gauge the patient's openness to novel, previously non-shared knowledge. Each member independently undertook a careful examination of the sessions, identifying pronounced ET manifestations and formulating initial assessments based on their evaluations. A dynamic interplay between the theoretical framework and actual session data followed, exemplifying the inductive nature of the approach. Regular team meetings facilitated a unified recognition of the relevant markers. Points of contention among raters were thoroughly discussed until consensus

was achieved. This iterative process was continued until saturation was reached, aligning with established research practice (Hennink & Kaiser, 2022; Levitt et al., 2017).

Step 3: Formulation and Refinement of the Rating Manual. During this phase, an initial rating manual was devised. It was subsequently piloted, leading to revisions that incorporated a consolidated rating sheet with illustrative examples spanning different levels of rating. Iterative modifications were implemented to address subtle nuances in the rating process, culminating in a finalized protocol. As the protocol took shape, markers of ET were systematically classified. They were structured to represent a continuum for each distinct element, calibrated on a 5-point Likert scale, with endpoints ranging from 0 (indicating very low or absent ET) to 4 (indicating very high or fully present ET).

The Rating Process

Step 1: Orientation and Reflection. Initially, the rating team familiarized themselves with the theoretical underpinnings of ET and its proposed three elements as they manifest in psychotherapy. Weekly discussion groups were initiated to review and reflect upon team members' interpretations and address their concerns and beliefs informed by relevant scholarly work (Levitt et al., 2018; Hill, 2012).

Step 2: Rating Practice Sessions. The team engaged in both collective and individual rating exercises using a subset of pilot sessions guided by the newly formulated rating manual. They were tasked with identifying and rating instances of ET within these sessions. These practice sessions served to acquaint team members with the practical application of the rating system and work towards uniformity in their assessments.

Step 3: Rating Calibration and Consensus Building. The fourth session in each treatment was transcribed verbatim and included non-verbal communication elements (fillers ('umm'), sighs, silences, simultaneous speech, humming, crying, etc.). The transcripts were then rated by two independent raters. The team convened weekly to address any drifts or variations in their ratings, aiming to rectify these through collaborative discussion. This cyclical process of rating, dialogue, and reflection went on until the raters achieved substantial agreement, supporting the consistency and trustworthiness of the ET assessments.

The systematic rating procedure yielded an overall ETRS score accompanied by three distinct sub-scales. Each sub-scale specifically assesses a singular ET element: one gauges the extent of patient sharing, another measures the degree to which we-mode moments are achieved within the session, and the third evaluates the patient's receptiveness to learning. These scores provide an individual profile of ET in psychotherapy tailored to each patient. The composite ET score for each patient was derived by averaging the values from the three sub-scales. Elevated ETRS total scores signify enhanced levels of epistemic trust. For the purposes of evaluating reliability, convergent, and discriminant validity, we analyzed these aggregated scores.

Measures

Epistemic Trust Rating System (ETRS)

The ETRS scale evaluates the manifestation of a patient's epistemic trust during therapeutic sessions. Utilizing verbatim transcriptions of the sessions, trained raters employed a 5-point Likert scale to assess dimensions of sharing, 'we-mode,' and learning. The ETRS total score was computed by averaging the scores of these three elements.

Convergent validity

Experience in Close Relationships - Relationship Structures (ECR-RS; Fraley et al., 2011). An adapted 18-item self-report measure, originally designed to assess romantic attachment relationships, was utilized. In this adaptation, patients were instructed to respond based on their perceptions of their relationship with their therapist. Responses are indicated on a 7-point Likert scale, ranging from 0 (strongly disagree) to 6 (strongly agree). The measure yields scores for two dimensions: anxious attachment and avoidant attachment. In the present study, the reliability of the two scales was .79 for the anxiety dimension and .83 for the avoidant dimension.

Client's Attachment to the Therapist Scale (CATS; Mallinckrodt et al., 1995). A 36-item self-report measure rooted in attachment theory gauges clients' perceptions of their therapeutic relationships. Participants provide their responses about their interpersonal feelings, thoughts, and experiences on a 6-point Likert scale, spanning from 1 (disagree) to 6 (strongly agree). The scale produces scores for three distinct dimensions: secure, preoccupied/merger, and avoidant/fearful attachment. For this study, reliability was .84 for the secure dimension, .84 for the preoccupied/merger dimension, and .86 for the avoidant/fearful dimension.

Working Alliance Inventory-Short Form (WAI-S; Tracey & Kokotovic, 1989). A 12-item self-report instrument measures the therapeutic alliance based on Bordin's (1979) working alliance theory. Respondents rate items on a 7-point Likert scale, from 1 (never) to 7 (always). The measure produces a cumulative general alliance score, with higher scores indicative of a more robust working alliance. While internal consistency estimates for the client version typically range between .90-.98 (Hanson et al., 2002), the reliability coefficient in the current study was .93.

Discriminant validity

Rosenberg Self Esteem Scale (RSES; Rosenberg, 1979). A 10-item self-report tool assesses subjective self-evaluation. The instrument includes five positively phrased and five negatively phrased items. Responses are captured on a 4-point Likert-type scale, ranging from 1 (strongly agree) to 4 (strongly disagree). Higher scores denote greater global self-esteem. Past research has reported Cronbach's alpha reliability coefficients between .72 and .90 (Gray-Little et al., 1997; Robins et al., 2001). In the current study, the reliability coefficient was .87.

Data Analysis

The internal consistencies of the three rating scales of the ETRS are reported as Cronbach's alpha reliability coefficients. The ETRS's interrater reliability was evaluated using the ICC(1,8) (Shrout & Fleiss, 1979). Given that every session in the sample was double-rated and exhibited satisfactory agreement (ICC(1,8) ≥ .75, according to Cicchetti (1994)), the average ETRS score was employed in the validity analyses. Convergent validity for the ETRS was assessed via Pearson correlations with scores from the CATS, ECR-RS, and WAI-S. Discriminant validity was further explored using Pearson correlation with RSES scores. We applied Cohen's (1988) conventions to gauge the magnitude of effect size. Specifically, a correlation coefficient of .10 is suggestive of a "weak" association, .30 is deemed indicative of a "medium" association, and .50 or larger is interpreted as a "large" effect.

Results

Descriptive Data and Internal Consistency

Through adopting a systematic approach to scale development, the team worked from theoretical concepts to evolve an observable measure of ET as it manifests in therapy sessions. The ETRS

has markers of sharing, we-mode, and learning, along with their varying manifestations and corresponding ratings for each level. These are presented in Table S1 (available in the online Supplemental Material). Descriptive statistics and interrater reliability for the ETRS total score and its three constituent elements are displayed in Table 2. The distribution of ETRS ratings within our sample approximated a normal distribution. Further, Table 3 provides correlation coefficients between the three ETRS elements, revealing associations ranging from weak to medium in strength (Cohen, 1988). This indicates that while the three elements have some degree of convergence, they represent unique facets of ET.

Table 2

Descriptive Statistics and Interrater Reliability of the ETRS

Table 3

Correlations among the three elements of the ETRS

Convergent and Discriminant Validity of the ETRS

Convergent validity assessments of the ETRS in relation to the CATS, ECR-R, and WAI-S lent partial support to our stated hypotheses, as illustrated in Table 4. Specifically, the ETRS total score exhibited a significant positive correlation with the secure attachment measure from the CATS. This suggests that individuals with higher epistemic trust are more likely to report a more secure attachment to their therapists. Moreover, as anticipated, there was a significant negative correlation between the ETRS total score and the avoidant attachment measure on the ECR-RS. This relationship suggests that individuals exhibiting pronounced avoidant tendencies with therapists tend to manifest lower trust in the therapeutic relationship. Additionally, consistent

with our convergent validity expectations, a positive and significant association was noted between the ETRS total score and the WAI-S total score. However, contrary to our hypotheses, no significant relationships emerged between the ECR-RS anxious attachment and CATS avoidant/fearful attachment measures.

In terms of discriminant validity and in alignment with our hypotheses, the correlation between the ETRS total score and the RSES score was both low and non-significant, underscoring the premise that epistemic trust and self-esteem are distinctly separate constructs.

Secondary Results: Associations of ETRS Elements

Beyond the primary findings relating to the ETRS total score, this section considers the relationships observed between individual ETRS elements and several self-report metrics, including attachment dimensions, therapeutic alliance, and self-esteem (see Table 4). Notably, significant positive associations emerged between the CATS secure dimension and the sharing and learning elements of the ETRS. The CATS preoccupied/merger dimension displayed a significant positive correlation with the ETRS's learning element but not with its sharing or we-mode elements. Additionally, all three ETRS elements—sharing, learning, and we-mode—were found to negatively correlate with the ECR-RS avoidant attachment dimension. A noteworthy observation is the significant positive correlation between both the WAI-S and the CATS preoccupied/merger subscale with the ETRS learning element. These findings imply that individuals demonstrating a more preoccupied attachment style with their therapist or those fostering stronger therapeutic alliances appear to derive enhanced learning from the therapeutic engagement. Correlations outside of those described here were non-significant.

Table 4

Pearson Correlations Between the ETRS and Measures to Determine Convergent and Discriminant Validity

Sensitivity Analysis

Due to the Covid-19 pandemic, thirteen patients received psychotherapy via an online platform. We repeated our analyses, excluding these thirteen patients, and found that the findings and interrater reliability remain similar to those reported in the original analysis. The results appear in Supplemental Tables S3-S5.

Discussion

The present study is pioneering in its presentation of the development and evaluation of the observer-based Epistemic Trust Rating System (ETRS) within psychotherapy. The ETRS encapsulates the patient's receptivity towards viewing interpersonal cues from the therapist, the therapeutic relationship, and the therapeutic context as ensuring personal relevance and generalizability. Crafted meticulously, the ETRS amalgamates a total score reflecting the patient's overarching epistemic trust, harmoniously blending its three distinct components. By synergizing the patient's propensity to share, their capacity to cultivate we-mode moments, and their receptivity to introspection and external insights, the total score provides a holistic assessment of their epistemic trust. The preliminary evaluation of the instrument's psychometric robustness is encouraging. Our results underscore the premise that evaluators, despite their individual perspectives, can reach a consensus on the ETRS, which is corroborated by its commendable inter-rater reliability. The observed discriminant and convergent validity of the measure resonates harmoniously with both theoretical anticipations and the conceptual framework.

Consistent with our predictions, data reveal that pronounced levels of secure attachment to therapists positively correlate with heightened epistemic trust. This implies that patients, when they perceive their therapists as emotionally attuned and validating, are predisposed to foster trust. Consequently, they become more amenable in therapy, even when confronted with insights (such as therapist interpretations) that may challenge their pre-existing convictions (Fonagy & Allison, 2014). Similarly, our findings indicate that pronounced levels of avoidant attachment towards the therapist inversely correlate with epistemic trust. A plausible interpretation suggests that entrenched insecurities in attachment, potentially stemming from apprehensions of rejection, feelings of detachment, and a cynical perspective on relationships, could impede individuals' capacity to trust external communication, encompassing interactions with their therapists.

Corroborating our predictions, the analyses elucidated that a fortified working alliance positively correlates with enhanced epistemic trust. This alludes to the notion that patients, when they synergize collaboratively with therapists, not only align on therapeutic objectives and form a robust emotional rapport but also cultivate a conducive environment fostering interpersonal learning within the therapy.

While the overarching construct of epistemic trust is encapsulated optimally by the total score, an in-depth examination of its individual elements can potentially unveil clinically pertinent nuances. As illuminated by the present study's data, the three subscales, although correlating with the total score, might offer insights into distinct and subtle patterns. A salient observation emerged from a meticulous review of the results: a notable positive correlation exists between pronounced levels of preoccupied/merger attachment to the therapist and an amplified learning element of ET. A conceivable post hoc rationale resonates with findings by Mallinckrodt et al. (2017) that delineate the pseudo-secure attachment style. Patients manifesting

this style typically exhibit elevated scores in both secure attachment and preoccupied/merger attachment dimensions. Given that our assessment was centered on the initial phase of therapy, specifically the fourth session, these patterns may bear semblance to those evocative of secure attachment styles. Notably, during the nascent stages of treatment, both styles might manifest overlapping attributes such as the propensity for rapport-building, candid self-disclosure, and a favorable perception of the therapist. Nonetheless, an inherent over-reliance on the therapist by these patients might translate into a heightened susceptibility to external influences, culminating in challenges in discerning credible from spurious information. As therapy progresses, we anticipate these individuals might struggle with leveraging the therapeutic relationship as a steadfast platform for exploration. This may eventually cascade into a decelerating correlation with epistemic trust.

Implications and Future Directions

The ETRS, being a valid and reliable rating system, offers promising opportunities for both research and clinical practice. One potential area of exploration lies in examining the role of ET as a prognosticator of therapeutic outcomes. By gauging interpersonal variations in ET at the commencement of therapy, researchers can explore its associations with outcomes like symptom alleviation, enhancements in attachment security, and augmented social support, among others. Such exploration would elucidate theoretical propositions suggesting that patients endowed with high epistemic trust may accrue more pronounced therapeutic benefits compared to their low ET counterparts (Fonagy et al., 2018). Another promising research direction involves exploring ET as a mechanism of change (Luyten et al., 2020). By assessing shifts in ET across diverse points during therapy, researchers could pinpoint its key components. This, in turn, can reveal efficacious strategies to convey intent in knowledge communication, especially when addressing

patients with diminished ET (Fisher et al., 2023). Consequently, epistemic trust and its associated learning process could emerge as central tenets of efficacious psychotherapeutic interventions. In tandem with this trajectory, we advocate for subsequent studies to consider ETRS as a standalone outcome metric, spotlighting the transformative capacity of epistemic trust within the therapeutic milieu. Undertaking such studies could augment our grasp of psychotherapeutic dynamics and refine clinical practices by designing interventions that fortify epistemic trust, thereby enhancing therapeutic outcomes.

These research endeavors are poised to yield profound clinical insights, reshaping our comprehension of therapeutic mechanisms and informing the design of future therapeutic protocols and therapist training curricula. This encompasses potential therapeutic aims, methodologies, and efficacy benchmarks. Drawing from the ET theoretical framework, an authentic therapeutic ambition could be to aid patients in discerning conducive environments or relationships where insights about oneself and others prove invaluable. To fulfill such aims, therapists are encouraged to craft case formulations underpinned by the ET paradigm, adeptly pinpointing the hallmarks of sharing, we-mode, and learning (Fisher et al., 2023). Additionally, the composite score can function as a succinct tool, offering valuable glimpses into a patient's overarching trajectory at various therapeutic stages.

Limitations

This study introduced the ETRS and provided an initial, promising validation. However, it is important to interpret these preliminary findings with caution, particularly regarding their generalizability and robustness. The study's success in consistently and validly inferring levels of epistemic trust (ET) from therapy session transcripts underscores the tangible nature of this phenomenon as reflected in verbal communication. This finding not only highlights the

robustness of ET as a measurable construct but also opens up exciting possibilities for future research. However, incorporating non-verbal aspects of patient-therapist interactions could provide deeper insights into the construct of epistemic trust, enhancing our understanding of this complex dynamic.

Additionally, advancing the validity and applicability of the ETRS across a more extensive range of contexts necessitates research with larger, more varied samples and the inclusion of longitudinal assessments. For instance, it is essential to further validate the ETRS in the context of constructs such as mentalization and personality disorders and to expand the sample to include clinical groups beyond MDD. It would also be advantageous for diverse research groups globally to replicate our study, reinforcing the findings' robustness and extending the ETRS's relevance across varying cultural and ethnic settings.

A notable aspect of our sample is that it encompassed all intent-to-treat RCT patients, which meant the inclusion of individuals randomized to two distinct psychotherapeutic protocols. During the ETRS's development, no differentiation was made between these two conditions. This approach is potentially underpinned by the theoretical proposition that ET is ubiquitous across treatments, and thus, its presence within the therapeutic relationship is anticipated, irrespective of the specific therapeutic method (Fonagy et al., 2019). Encouragingly, preliminary testing supported this rationale, as results were consistent even when the two conditions were evaluated separately. Lastly, while an observer-based measure offers distinct advantages, it also demands significant resources and time. Yet, coupling such a rating system with data from ET self-report instruments might yield a holistic view of ET's role throughout therapy. This integrated perspective can enrich therapeutic assessments and interventions, deepening our comprehension of a patient's epistemic trust.

Finally, the present study focuses on the psychometric properties of the ETRS and leaves many open questions for future research on the associations between ET and related constructs. Future studies could explore how state-like changes in ET occur in response to specific therapeutic techniques or during moments of rupture and repair in the therapeutic alliance. For instance, it is hypothesized that employing alliance-strengthening techniques (Eubanks et al., 2023) in the fourth session could enhance ET by the eighth session, potentially leading to improvements in depressive symptoms. This hypothesis is based on the premise that fortifying the therapeutic alliance may positively influence the 'we-mode' element of ET, thereby accelerating learning and sharing processes (Fisher et al., 2023). Additionally, changes in ET levels during treatment might correlate with a decrease in both the frequency and severity of ruptures in the therapeutic alliance. Investigating these dynamics could shed light on ET's role as a mechanism of therapeutic change.

Conclusions

In summarizing, the ETRS represents the first observer-based assessment of ET within psychotherapy. Emerging from its initial development and validation stages, the ETRS demonstrates its potential as a promising evaluative tool. Its strength is evident in the consistency among evaluators in ratings and its established relationships with other theoretically pertinent constructs. Thus, the ETRS holds the potential to serve as a much-anticipated bridge, connecting the extensive theoretical literature with rigorous empirical exploration in this domain.

References

Benjamin, J. (2004). Beyond doer and done to: An intersubjective view of thirdness. *The Psychoanalytic Quarterly, 73*(1), 5-46. https://doi.org/10.1002/j.2167-4086.2004.tb00151.x

Bergin, L. A. (2001). The role of truth when communicating knowledge across epistemic differences. *Social Epistemology*, *15*(4), 367-378.

Book, H. E. (1998). *How to practice brief psychodynamic therapy: The core conflictual relationship theme method*. Washington, DC: American Psychological Association Press.

Bordin, E. S. (1979). The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy: Theory, Research, & Practice*, 16, 252–260. http://dx.doi.org/10.1037/h0085885

Bowlby, J. (1969). Attachment and loss: volume I: attachment. In *Attachment and Loss: Volume I: Attachment* (pp. 1-401). The Hogarth Press and the Institute of Psycho-Analysis.

Campbell, C., Tanzer, M., Saunders, R., Booker, T., Allison, E., Li, E., . . . Fonagy, P. (2021). Development and validation of a self-report measure of epistemic trust. *PLoS One, 16(4),* e0250264. doi: 10.1371/journal.pone.0250264

Central Bureau of Statistics. (23 March 2016). Population group. https://www.cbs.gov.il/en/Pages/All-Terms.aspx?k=Religion

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6,* 284–290.

Clarke, A., Meredith, P. J., Rose, T. A., & Daubney, M. (2018). A role for epistemic trust in speech-language pathology: A tutorial paper. *Journal of Communication Disorders, 72,* 54-63.

Cohen, J. (1988). Statistical power analysis for the behavioural sciences. 2nd edn. Hillsdale, NJ: Lawrence Erlbaum.

Corriveau, K. H., Harris, P. L., Meins, E., Fernyhough, C., Arnott, B., Elliott, L., Liddle, B., Hearn, A., Vittorini L., & De Rosnay, M. (2009). Young children's trust in their mother's claims: Longitudinal links with attachment security in infancy. *Child Development, 80*(3), 750-761. http://dx.doi.org/10.1111/j.1467-8624.2009.01295.x

Csibra, G., & Gergely, G. (2006). Social learning and social cognition: The case for pedagogy. *Processes of Change in Brain and Cognitive Development. Attention and Performance XXI, 21*, 249-274.

Durkin, K., & Shafto, P. (2016). Epistemic trust and education: effects of informant reliability on student learning of decimal concepts. *Child Development, 87*(1), 154-164.

Duschinsky, R., & Foster, S. (2021). *Mentalizing and Epistemic Trust: The Work of Peter Fonagy and Colleagues at the Anna Freud Centre*. Oxford University Press.

Egyed, K., Király, I., & Gergely, G. (2013). Communicating shared knowledge in infancy. *Psychological Science, 24*(7), 1348–1353. https:// doi.org/10.1177/0956797612471952.

Eubanks, C. F., Samstag, L. W. E., & Muran, J. (2023). *Rupture and repair in psychotherapy: A critical process for change* (pp. vii-338). American Psychological Association.

Fisher, S., Fonagy, P., Wiseman, H., & Zilcha-Mano, S. (2023). I see you as recognizing me; therefore, I trust you: Operationalizing epistemic trust in psychotherapy. *Psychotherapy*. Advance online publication. https://doi.org/10.1037/pst0000501

Fisher, S., Zilcha-Mano, S., & Fonagy, P. (2022). Conceptualizing Epistemic Trust in Psychotherapy: A Triadic Model, *Psychotherapy Bulletin*, *57*(2), 29-35.

Fonagy, P., & Allison, E. (2014). The role of mentalizing and epistemic trust in the therapeutic relationship. *Psychotherapy, 51*(3), 372–380. http://dx.doi.org/10.1037/a0036505

Fonagy, P., Campbell, C., Constantinou, M., Higgitt, A., Allison, E., & Luyten, P. (2021). Culture and psychopathology: An attempt at reconsidering the role of social learning. *Development and Psychopathology,* 1-16. https://doi.org/10.1017/S0954579421000092

Fonagy, P., Luyten, P., & Allison, E. (2015). Epistemic petrification and the restoration of epistemic trust: A new conceptualization of borderline personality disorder and its psychosocial treatment. *Journal of Personality Disorders*, *29*(5), 575-609. http://dx.doi.org/10.1521/pedi.2015.29.5.575

Fonagy, P., Luyten, P., Allison, E., & Campbell, C. (2017). What we have changed our minds about: Part 2. Borderline personality disorder, epistemic trust and the developmental significance of social communication. *Borderline Personality Disorder and Emotion Dysregulation*, *4*(1), 9. http://dx.doi.org/10.1186/s40479-017-0062-8

Fonagy, P., Luyten, P., Allison, E., & Campbell, C. (2018). *Reconciling psychoanalytic ideas with attachment theory*. Guilford Press.

Fonagy, P., Luyten, P., Allison, E., & Campbell, C. (2019). Mentalizing, epistemic trust and the phenomenology of psychotherapy. *Psychopathology*, *52*(2), 94-103. http://dx.doi.org/10.1159/000501526

Fonagy, P., Luyten, P., Moulton-Perkins, A., Lee, Y. W., Warren, F., Howard, S., ... & Lowyck, B. (2016). Development and validation of a self-report measure of mentalizing: The reflective functioning questionnaire. *PloS one, 11*(7), e0158678.

Fraley, R. C., Heffernan, M. E., Vicary, A. M., & Brumbaugh, C. C. (2011). The experiences in close relationships—Relationship Structures Questionnaire: A method for assessing

attachment orientations across relationships. *Psychological Assessment, 23*(3), 615–625. https://doi.org/10.1037/a0022898

Fumerton, R. (2009). *Epistemology*. John Wiley & Sons.

Gray-Little, B., Williams, V. S., & Hancock, T. D. (1997). An item response theory analysis of the Rosenberg Self-Esteem Scale. *Personality and Social Psychology Bulletin*, *23*(5), 443-451.

Hamilton, M. (1967). Development of a rating scale for primary depressive illness. *The British Journal of Social and Clinical Psychology, 6*(4), 278– 296. https://doi.org/10.1111/j.2044-8260.1967.tb00530.x

Hanson, W. E., Curry, K. T., & Bandalos, D. L. (2002). Reliability generalization of working alliance inventory scale scores. *Educational and Psychological Measurement*, *62*(4), 659-673.

Hennink, M., & Kaiser, B. N. (2022). Sample sizes for saturation in qualitative research: A systematic review of empirical tests. *Social science & medicine*, *292*, 114523.

Horvath, A. O., & Greenberg, L. S. (1989). Development and validation of the Working Alliance Inventory. *Journal of Counseling Psychology, 36* ,223–233. http://dx.doi.org/10.1037/0022-0167.36.2.223

Krieger, J., & Crahan, M. E. (2001). The Oxford companion to politics of the world. Oxford University Press.

Leibovich, L., Front, O., McCarthy, K. S., & Zilcha-Mano, S. (2019). How do supportive techniques bring about therapeutic change: The role of therapeutic alliance as a potential mediator. *Psychotherapy,* Advance online publication. http://dx.doi.org/10.1037/pst0000253

Levitt, H. M. (2018). How to conduct a qualitative meta-analysis: Tailoring methods to enhance methodological integrity. *Psychotherapy Research, 28*(3), 367-378.

Levitt, H. M., Pomerville, A., Surace, F. I., & Grabowski, L. M. (2017). Metamethod study of qualitative psychotherapy research on clients' experiences: Review and recommendations. Journal of Counseling Psychology, 64(6), 626.

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

Luyten, P., Campbell, C., & Fonagy, P. (2021). Rethinking the relationship between attachment and personality disorder. *Current opinion in psychology*, *37*, 109-113.

Luyten, P., Campbell, C., Allison, E., & Fonagy, P. (2020). The mentalizing approach to psychopathology: State of the art and future directions. *Annual Review of Clinical Psychology*, *16*. http://dx.doi.org/10.1146/annurev-clinpsy-071919-015355

Mallinckrodt, B., Anderson, M., Choi, G., Levy, K., Petrowski, K., Sauer, E., Tishby, O., and Wiseman, H. (2017). Pseudosecure vs. individuated-secure client attachment to therapist: Implications for therapy process and outcome. *Psychotherapy Research, 27*(6), 677-691. DOI: 10.1080/10503307.2016.1152411

Mallinckrodt, B., Gantt, D. L., & Coble, H. M. (1995). Attachment patterns in the psychotherapy relationship: Development of the Client Attachment to Therapist Scale. Journal of Counseling Psychology, 42(3), 307.

McCraw, B. W. (2015). The nature of epistemic trust. *Social Epistemology*, *29*(4), 413-430.

Mercier, H., & Sperber, D. (2017). *The enigma of reason: A new theory of human understanding*. Allen Lane.

Pasquini, E. S., Corriveau, K. H., Koenig, M., & Harris, P. L. (2007). Preschoolers monitor the relative accuracy of informants. *Developmental Psychology, 43*(5), 1216.

Robins, R. W., Hendin, H. M., & Trzesniewski, K. H. (2001). Measuring global self-esteem: Construct validation of a single-item measure and the Rosenberg Self-Esteem Scale. *Personality and Social Psychology Bulletin, 27*(2), 151-161.

Rosenberg, M. (1979). *Conceiving the Self*. Basic Books.

Schröder-Pfeifer, P., Georg, A. K., Talia, A., Volkert, J., Ditzen, B., & Taubner, S. (2022). The Epistemic Trust Assessment—An experimental measure of epistemic trust. *Psychoanalytic Psychology*, *39*(1), 50.

Schröder-Pfeifer, P., Talia, A., Volkert, J., & Taubner, S. (2018). Developing an assessment of epistemic trust: a research protocol. *Research in Psychotherapy: Psychopathology, Process and Outcome, 21*(3), 123-131. http://dx.doi.org/10.4081/ripppo.2018.330

Shafto, P., Eaves, B., Navarro, D. J., & Perfors, A. (2012). Epistemic trust: Modeling children's reasoning about others' knowledge and intent. *Developmental Science, 15*(3), 436-447.

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., & Dunbar, G. C. (1998). The mini international neuropsychiatric interview (MINI): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *The Journal of Clinical Psychiatry, 59*(20), 22–33.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin, 86*(2), 420.

Tracey, T. J., & Kokotovic, A. M. (1989). Factor structure of the Working Alliance Inventory. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, *1*(3), 207-210. http://dx.doi.org/10.1037/1040-3590.1.3.207

Weiner, I. B., & Bornstein, R. F. (2009). *Principles of psychotherapy: Promoting evidence-based psychodynamic practice*. John Wiley & Sons.

Wilholt, T. (2020). Epistemic trust in science. *The British Journal for the Philosophy of Science*, *64*(2), 233-253 doi: http://bjps.oxfordjournals.org/

Zilcha-Mano, S., Dolev, T., Leibovich, L., & Barber, J. P. (2018). Identifying the most suitable treatment for depression based on patients' attachment: Study protocol for a randomized controlled trial of supportive-expressive vs. supportive treatments. *BMC Psychiatry*, *18*(1), 362.

Zilcha-Mano, S., Goldstein, P., Dolev-Amit, T., Ben David-Sela, T., & Barber, J. P. (2021). A randomized controlled trial for identifying the most suitable treatment for depression based on patients' attachment orientation. *Journal of Consulting and Clinical Psychology*, *89*(12), 985.