

RESEARCH ARTICLE

Session Reactions Scale-3: Initial psychometric evidence

TOMÁŠ ŘIHÁČEK ¹, ROBERT ELLIOTT ², JESSE OWEN³,
MICHAELA LADMANOVÁ ¹, JEREMY J. COLEMAN ⁴, & MATTEO BUGATTI ²

¹Faculty of Social Studies, Department of Psychology, Masaryk University, Brno, Czech Republic; ²Counselling Unit, School of Psychological Sciences and Health, University of Strathclyde, Glasgow, UK; ³Counseling Psychology Department, Morgridge College of Education, University of Denver, Denver, CO, USA & ⁴Department of Psychological Sciences, College of Science and Mathematics, Augusta University, Augusta, GA, USA

(Received 31 October 2022; revised 18 July 2023; accepted 24 July 2023)

ABSTRACT

Objective This study aimed to develop an updated brief self-report post-session measure, suitable for collecting systematic feedback on clients' session reactions in the context of measurement-based care (MBC). **Method:** The Session Reactions Scale-3 (SRS-3; 33 items) was developed by extending and adjusting the Revised Session Reactions Scale. In Study 1, the psychometric properties of the SRS-3 were tested on $N = 242$ clients. In Study 2, a brief version of the SRS-3 (SRS-3-B; 15 items) was developed using a combination of conceptual, empirical, and pragmatic criteria. In Study 3, the psychometric properties of the SRS-3-B were tested on a new sample of $N = 265$ clients. **Results:** Exploratory factor analysis supported the use of the SRS-3-B as a two-factor (helpful reactions, hindering reactions) or unidimensional (overall session evaluation) instrument. The SRS-3-B was meaningfully related to another process measure (Individual Therapy Process Questionnaire) both on the item and factor levels. **Conclusions:** The SRS-3-B is a reliable process measure to elicit rich and clinically meaningful feedback from clients within the MBC context and as a research instrument to assess the helpful and hindering aspects of therapy sessions.

Keywords: psychotherapy process; session rating; psychotherapy outcomes; client experience; routine outcome monitoring; measurement-based care

Clinical or methodological significance of this article: We developed brief and long versions of a pantheoretical self-report measure of client session reactions, that can be used both in research and in the context of routine outcome and process monitoring. The measures capture reactions that clients themselves are likely to consider important. They have satisfactory psychometric characteristics and can be interpreted at the item level and/or as composite scores.

As with any other health care profession, psychotherapy demands continuous assessment and feedback—aspects that have been recognized as essential to evidence-based practice (APA Presidential Task Force on Evidence-Based Practice, 2006; Boswell et al., 2023; Lambert et al., 2018). However, empirical evidence shows that psychotherapists' ability to obtain precise information

about their clients' progress or deterioration can be limited (Hannan et al., 2005; Hatfield et al., 2009). Measurement based care (MBC) practices have had a mixed impact on therapy outcomes (De Jong et al., 2021), which may in part be attributed to poor implementation of this practice in clinical settings (Miller et al., 2015). Another reason may be the fact that the existing MBC tools focus

Correspondence concerning this article should be addressed to Tomáš Řiháček, Faculty of Social Studies, Department of Psychology, Masaryk University, Joštova 10, 602 00 Brno, Czech Republic. Email: rihacek@fss.muni.cz

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

predominantly on symptom-based outcomes (Wampold, 2015). Although these tools are efficient in informing psychotherapists how well or poorly their clients are doing, they are typically not very informative in terms of suggesting remedies when the therapy has veered off track. Arguably, feedback on both the therapeutic process and outcome may provide information that can be more easily used to guide interventions compared to outcome data alone (Boswell et al., 2015).

Understanding the clients' perspective is key to effective psychotherapy (Bohart & Tallman, 1999; Duncan et al., 2004; Elliott & James, 1989), and psychotherapists must consider the breadth of clients' therapy experience and address it in a responsive manner (Constantino et al., 2020; Wu & Levitt, 2020). At the same time, some research evidence suggests that clients tend to defer to their therapists and do not always disclose their reactions to the therapeutic process (Farber, 2020; Rennie, 1994). Furthermore, clients and therapists often differ in the in-session moments they identify as significant or in the reasons for which they consider them significant (Elliott & Shapiro, 1992; Kivlighan & Arthur, 2000; Levitt & Piazza-Bonin, 2011; Llewellyn, 1988).

Thus, a systematic, comprehensive, and brief method to invite the client's voice to be more present in the therapeutic dialog is essential. Arguably, items on such a measure should represent a range of clients' reactions that clients themselves identify as important and are thus ready to recognize and rate after a session. Knowledge of what clients appreciate and dislike in psychotherapy has been best captured by qualitative studies and, more specifically, in clients' descriptions of significant events (see Timulak, 2010, for a review). However, in building a measure, we were not interested in the descriptions of these events per se. Rather, we focused on client-identified impacts of these events, as reported immediately after the session. These impacts represent immediate post-session "takeaways" that build up into longer-lasting therapeutic change (Greenberg, 1986; Orlinsky & Howard, 1986). There is a long tradition of studying the impacts of significant events (Elliott, 1985; Elliott et al., 1985; Ladmanová et al., 2022; Timulak, 2007). Terminologically, however, Reeker et al. (1996) argued that the word "impacts" implied the client to be a passive object of therapeutic interventions and preferred to use "reactions" instead, implying clients' agency in the therapeutic process. We treat the words impacts and reactions as synonyms in this text and we use them interchangeably depending on the study that is being discussed.

Instead of developing a new measure, we decided to refine an existing measure, Revised Session Reactions Scale (RSRS; Reeker et al., 1996) that, itself, was a revised version of the Session Impacts Scale (SIS; Elliott & Wexler, 1994). Initially, using cluster and content analysis, Elliott (1985; Elliott et al., 1985) developed an empirical classification of the immediate therapeutic impacts of significant events. This initial empirical taxonomy (Elliott, 1985) and the content analysis measure developed from it (Elliott et al., 1985) divided client reactions into two broad domains of helpful and hindering events. Based on this classification, Elliott and Wexler (1994) then developed the 16-item Session Impact Scale, which was later broadened into the 22-item Revised Session Reactions Scale by Reeker et al. (1996). To build upon their work, we utilized the RSRS as the basis of our new measure.

Measures designed for routine session-by-session use in a busy clinical environment must balance length with utility. Longer measures tend to have better psychometric properties but require more time to answer and may discourage clients and therapists from participating. Therefore, we adopted a different approach in which each construct is represented by a single carefully chosen item, trading off measurement precision for brevity and utility in routine clinical practice. This approach was inspired by the clinimetric tradition (Carrozzino et al., 2021) that prioritizes clinical usefulness over traditional psychometric criteria. The proponents of clinimetrics criticize traditional psychometric approaches for increasing redundancy by including multiple similar items in the pursuit of homogeneity. Instead, they argue that a scale should be composed of items that each possess incremental validity and provide distinct clinical information (Carrozzino et al., 2021). Furthermore, it should be composed of items that clinicians (and, in our case, clients) consider important (DeVet et al., 2003).

Although the SIS/RSRS items were originally developed to represent unique aspects (categories) of clients' session reactions, most studies used them as indices of more abstract, latent variables. Elliott and Wexler (1994) identified a three-factor SIS structure consisting of task impacts, relationship impacts, and hindering impacts. However, they demonstrated that a two-factor structure—with the first two factors merged into an overall helpfulness factor—also explained significant variability in session reactions. The same conclusions were reached about the RSRS structure (Reeker et al., 1996). This suggests that the SIS/RSRS can be meaningfully used both at the item level (yielding clients' scores for specific reactions) and at the latent-score level (yielding clients' global evaluation of the session).

Although two studies did not find strong connections between averaged session impacts and end of therapy outcomes (Elliott & Wexler, 1994; Kellett et al., 2007), session impacts predicted treatment outcomes (King et al., 2006) and treatment retention (Siqueland et al., 2004). Clients' SIS scores were also related to their attachment to the therapist (Janzen et al., 2008), working alliance (Fitzpatrick et al., 2005; King et al., 2006; Siqueland et al., 2004), and other measures of session quality (Elliott & Wexler, 1994; Fauth & Williams, 2005; Stiles et al., 1994). Furthermore, the level of agreement between client and therapist-identified significant events predicted outcome (Chui et al., 2020; Kivlighan & Arthur, 2000; Pivolusková et al., 2021). Collectively, these studies show that clients' session reactions have predictive potential and are worth monitoring during treatment.

A revision of the existing measures (RSRS and SIS) is warranted for two reasons. First, the original SIS items (Elliott & Wexler, 1994), as well as the items added in its revised version (Reeker et al., 1996), were developed from classifications that were based on single datasets and did not consider more robust qualitative meta-analytic findings about client-identified session impacts (Ladmanová et al., 2022; Timulak, 2007). Second, the number and length of items of the 22-item RSRS, appeared to be excessive to be used in the context of session-by-session routine measurement.

Aim of Study

This study aimed to develop and test two versions of the Session Reactions Scale-3, including a brief post-session self-report version based on the RSRS that would be (a) short enough to be used for routine process monitoring (i.e., no more than 15 items, so that the measure could be answered within two minutes), (b) trans-theoretical, capturing a broad range of session reactions that clients themselves consider important, and (c) clinically useful. After describing how we developed the measure, we introduce three studies. In Study 1, we tested the psychometric properties of the full version of the new measure, the Session Reactions Scale-3 (SRS-3). In Study 2, we developed the brief version of the SRS-3 (SRS-3-B) based on a combination of conceptual, empirical, and pragmatic criteria. In Study 3, we tested the psychometric properties of the SRS-3-B on a new sample. Although we primarily focused on the individual items, we also explored the factor structure. We did so because it is preferable to have a measure that is meaningful on both the item level and the latent factor level. The project was approved by the University of Denver IRB board (#1794574-3).

Developing the Full Version of the Session Reactions Scale-3

The Session Reactions Scale-3 (SRS-3) is based on the *Revised Session Reactions Scale* (RSRS; Reeker et al., 1996), a brief self-report scale assessing clients' immediate post-session reactions to their sessions, which itself is an extended version of the Session Impact Scale (SIS; Elliott & Wexler, 1994). The RSRS consists of 22 items. Fourteen items capture helpful reactions, while 8 items capture hindering reactions. All items are rated on a five-point Likert scale ranging from 1 = "not at all" to 5 = "very much". Each item contains a boldfaced heading (e.g., "Seeing things from another person's perspective"), followed by a more detailed description for each item (e.g., "As a result of this session, I have begun to see things (about myself or others) from another person's point of view, including that of my therapist."). See Elliott (2010) for full item wording. Reeker et al. (1996) reported excellent interitem reliability and good convergence with other measures of session quality (e.g., Session Evaluation Questionnaire smoothness and positivity) but not with weekly symptom change. Both the SIS and the RSRS also contain an open-ended item that allows clients to describe and rate any reaction not covered by the closed-ended items.

The SRS-3 was developed from the RSRS in several steps. First, we checked whether the two qualitative meta-analyses of client-reported session impacts contained any category not covered by the RSRS items (Ladmanová et al., 2022; Timulak, 2007). Based on this step, we added two helpful items, namely, "accepting oneself or the situation" and "improving one's skills or strategies", and three hindering items, namely, "disconnection from the therapist", "unwillingness to disclose," and "difficulties participating in therapy."

Second, we used an archival set of clients' responses to the RSRS open-ended item ("Please describe any other reactions you might have had to this session") from Elliott et al.'s (2016) study on an outpatient sample in emotion-focused therapy. The dataset contained 102 responses from 47 clients who each provided one to eight responses ($M = 2.2$, $SD = 1.6$). The data originated from sessions 3 to 41 ($M = 14.9$, $SD = 11.2$). The first and second authors coded these responses using the existing RSRS categories and identified two additional categories, namely, "emotional arousal" and "a sense of progress/accomplishment."

Third, we considered some of the existing RSRS items too complex and decided to split them into two. RSRS Item 15 ("distressed") was split into "feeling worse" and "being more bothered by one's

thoughts, feelings or memories” (the latter item emphasized the attitude toward one’s symptoms, rather than the worsening of symptoms as such); RSRS Item 19 (“misunderstood”) into “feeling misunderstood” and “feeling that the therapist’s approach does not fit the client”; and RSRS Item 21 (“distracted or confused”) into “distracted” and “confused.”

Fourth, we abbreviated the original RSRS items while maintaining their core meanings. A typical RSRS item had three or four lines of text, which made it less suitable for repeated use in routine care. Therefore, we simplified the items, highlighting the core meaning in each of them. Based on very low scores on many of the hindering items in previous studies, we also tried to improve the hindering items to make them less discouraging for clients to respond to. For instance, RSRS Item 2 (“Pressured or controlled. As a result of this session, I feel too much pressure is being put on me to confront something or to change; or I feel controlled or manipulated by my therapist, or pushed to do something I don’t want to do”) was reworded into “I feel uncomfortable doing what my therapist is suggesting for me to do.”

In accordance with our intention to cover as broad a range of clients’ session reactions as possible, we decided to keep the open-ended item (“Please describe and rate any other reactions you might have had to this session”). Open-ended text has several advantages. First, therapists might find the responses more directly connected to the session content, in their clients’ own words. Second, open-ended text can be utilized in research from natural language processing methods. Therefore, we assumed that providing clients with an opportunity to use free language may strengthen clients’ perceptions of the meaningfulness of the measure and the clinical utility of using this measure as a feedback tool. For the final list of SRS-3 items, see Supplement 1.

Study 1: Testing the Full Version of the Session Reactions Scale-3

Method

Sample

Clients. A sample of $N = 242$ clients (57.4% men and 42.6% women) participated in the study. Their ages ranged from 18 to 70 years ($M = 38.16$, $SD = 11.36$), and their races/ethnicities included White non-Hispanic (75.6%), African American (12.4%), Latino/a/x or Hispanic (2.9%), Asian American (2.9%), Indigenous (0.8%), and Other/Multiracial (5.4%). The clients’ highest level of education included bachelor’s degree (64.0%), master’s degree (19.4%), high school (14.0%), other

(2.1%), and doctoral degree (0.4%). At the time of taking the survey, they reported having attended only one session (8.3%), two to five sessions (31.8%), six to 10 sessions (30.6%), 11 to 15 sessions (9.9%), 16 to 20 sessions (9.5%), 21 to 50 sessions (4.1%), and more than 50 sessions (5.8%). Their most recent session took place within one week (29.3%), between one and two weeks (35.5%), between two and three weeks (24.4%), or between three weeks and four weeks (10.7%) of their participation in the study. Clients whose most recent session took place more than four weeks prior to their participation in the study were considered ineligible to participate. This interval was chosen as a tradeoff between the recency of the rated experience (the longer the interval, the less accessible the memory of the session experience would be) and the possibility to recruit a large enough sample (the shorter the interval, the lesser the chance of recruiting enough respondents).

Measures

Session Reactions Scale-3 (SRS-3). The SRS-3 is a self-report measure assessing clients’ immediate post-session reactions to their sessions. It consists of 32 items (18 helpful and 14 hindering). All items are rated on a five-point Likert scale ranging from 1 = “not at all” to 5 = “very much”. Furthermore, there was one open-ended item inviting clients to describe any additional reactions not captured by the closed-ended items (data obtained by this item were not analyzed in this study). The development of SRS-3 was described in the previous section.

The Individual Therapy Process Questionnaire (ITPQ). The ITPQ (Mander et al., 2015) is a 36-item measure developed to evaluate general change mechanisms in psychotherapy. Clients are asked about their experience of the last session and to respond to each item on a five-point scale. The measure was originally developed in German; Mander et al. (2015) published an English translation. We adopted their translation of the items, but with the authors’ consent, we improved the translation of the rating scale to 0 = “Not at all”, 1 = “Slightly”, 2 = “Somewhat”, 3 = “Pretty much”, and 4 = “Very much”, which worked better in our field testing than the translation suggested by the authors. Based on Grawe’s (2004) integrative theory of psychotherapy, it was intended to measure eight theoretical dimensions, including resource activation, problem actuation, mastery, clarification of meaning, emotional bond, goals and tasks, therapist interference, and patient fear. However, psychometric analysis of the original

German version revealed a six-factor structure of the measure, including the in-session impact (composed of resource activation, mastery, and clarification of meaning items), problem actuation (i.e., being highly emotionally involved in and affected by the session), confident collaboration (i.e., hope for change and trust in the therapeutic process), global alliance (composed of goals, tasks, and bond items of the working alliance), patient fear (i.e., withholding from the session), and therapist interference (i.e., perceiving the therapist as a hindering influence) factors (Mander et al., 2015). Except for problem actuation, all subscales demonstrated meaningful relationships with outcome scores in the original study. Using our study data, the reliability of the subscale scores was $\alpha = .85$ for in-session impact, $.56$ for problem actuation, $.73$ for confident collaboration, $.88$ for global alliance, $.84$ for patient fear, and $.84$ for therapist interference. We chose this measure because it comprehensively covered relational aspects (i.e., working alliance and collaboration), task-related aspects (i.e., in-session impacts and problem actuation), and hindering experiences (i.e., patient fear and therapist interference), which corresponds to the structure of the preceding measures (i.e., SIS and RSRS).

Session Evaluation Questionnaire (SEQ).

The SEQ (Stiles et al., 1994) is a client-rated measure that contains 21 pairs of bipolar adjectives. It assesses two dimensions of session evaluation (i.e., depth and smoothness) and two dimensions of post-session mood (i.e., positivity and arousal). Typically, SEQ is administered immediately after a session. Due to a variable period of time that might have elapsed between a session and participation in this study, we modified the instruction to ask clients how they felt “right after the most recent session” instead of how they felt “right now.” Each item is scored on a seven-point scale. Items 3, 5, 7, 9, 11, 12, 14, 17, 18, and 20 are reversed, and subscale scores are computed as the mean of constituent item ratings. The reliability of the subscale scores was high in the original study for all four subscales (Stiles et al., 1994). The SEQ scores were related to SIS scores (Elliott & Wexler, 1994; Stiles et al., 1994) but not to improvement (Stiles et al., 1990). Reliability was $\alpha = .78$ for depth, $.81$ for smoothness, $.84$ for positivity and $.16$ for arousal in our study. Because of low reliability, we did not use the arousal subscale. We selected this measure because it was repeatedly used in studies that validated the preceding measures (Elliott & Wexler, 1994; Reeker et al., 1996) and because it captured evaluative aspects not covered by the ITPQ.

Clients’ Experience of Therapy Scale (CETS).

The CETS (Levitt et al., 2021) is a 15-item self-report measure of clients’ overall experience with their therapy. Items are rated on a seven-point scale ranging from 1 = “Not at all” to 7 = “To a very great degree.” The CETS measures five dimensions, namely, pattern identification, disconnection/disengagement, therapist responsiveness, client agency, and transformative acceptance/safety, all of which are moderately to highly correlated. The total score correlated with alliance and outcome measures in the original study. We found reliability was $\alpha = .79$ for pattern identification, $.83$ for disconnection/disengagement, $.47$ for therapist responsiveness, $.56$ for client agency, and $.61$ for transformative acceptance/safety. We included this measure because it was, very much like the SRS-3 itself, developed based on clients’ qualitative reports of their psychotherapy experience.

Procedure

Client sample recruitment and data collection.

The clients’ sample data were collected on the Qualtrics online survey platform, and clients were recruited via the Amazon Mechanical Turk (MTurk) platform (Follmer et al., 2017). Each client was paid 1 USD for participating in the survey. In their review, Woo et al. (2015) collected evidence that Mturk is an effective crowdsourcing platform and that Mturk workers are motivated to provide high-quality, reliable responses. Moreover, it was found to be a source of reliable and valid data in the context of psychotherapy research (Tompkins, 2019).

The measures were administered in the same order as they are presented in the Measures section, except for the demographic questionnaire that was presented at the beginning of the survey and served to filter out participants who did not meet the inclusion criteria. The survey was framed with the following instruction: “In the following questions, we will be asking you to rate the experience of your most recent individual therapy session. Although you may find it impossible to disregard the previous sessions’ experience entirely, we ask you to consider the most recent session as much as possible.”

To be eligible for the study, participants had to (1) be at least 18 years old, (2) live in the US, (3) currently be in individual psychotherapy, and (4) have their most recent session no more than four weeks ago. Furthermore, they had to pass all three attention checks inserted among the questionnaire items. Of the 1759 people who responded to the survey, 1531 finished all measures and passed all three attention checks. However, a closer inspection of the responses

revealed a large proportion of rushed and careless responses. Therefore, we employed a rigorous multi-step procedure to identify trustworthy responses (see Supplement 2 for details). First, we established a minimum survey duration cutoff (9.5 min) and removed all responses below that cutoff. This cutoff was based on the analysis of an archival RSRS dataset showing that clients tend to either score high on helpful items and low on hindering items or vice versa. A lack of this differentiation was considered a sign of careless responding. Second, two independent raters inspected all remaining responses visually; we removed those responses categorized by both raters as careless.

Statistical analysis. The analysis was conducted in R version 4.1.1 (R Core Team, 2021). We explored the descriptive statistics of SRS-3 items and interitem correlations (Spearman's correlation coefficient was utilized after a skewed distribution was observed for some of the items). Since we aimed to measure experiential phenomena distinct from each item, we sought items with high interitem correlations ($>.70$) to consider their removal.

Furthermore, we proceeded with exploratory factor analysis (EFA, principal axis factoring method with oblique rotation, based on polychoric correlations) to explore the latent structure. Oblique rotation was chosen because there was no reason to assume orthogonal factors and correlated factors have been commonly considered a more plausible representation of reality (Browne, 2001). We used Horn's parallel analysis, Kaiser's criterion (eigenvalue > 1), and a scree plot, as well as the interpretability of the model, to find the appropriate number of factors to extract. EFA was conducted using the psych package (Revelle, 2021), and parallel analysis was conducted using the nFactors package (Raiche & Magis, 2020). To assess concurrent validity, we computed Pearson's correlations with the external measures (i.e., ITPQ, CETS, and SEQ) for each SRS-3 subscale.

We also computed Spearman's correlation with the external measures (i.e., ITPQ, CETS, and SEQ) for each SRS-3 item to assess concurrent validity on the item level. These coefficients were used in Study 2 to assist in selecting items for the SRS-3 brief version.

Results

All items had reasonable variability, and none showed serious floor or ceiling effects. Nine pairs of items had correlations of $r > .70$. Six of these pairs contained Item 20 (Confused). We concluded that

this item was nonspecific and decided to remove it from the scale (however, we kept the original item numbers in tables and plots for the sake of consistency with the data). The remaining three pairs of items included Items 4 (Lack of guidance) and 12 (Lack of support), Items 22 (Bothered) and 26 (Feeling less warm or more distant from the therapist), and Items 26 and 28 (Feeling distracted). We concluded that despite the high correlations, these items represented distinct categories of experience, and we retained these items. See Supplement 3 for SRS-3 item description.

To explore the SRS-3 factor structure, we conducted EFA. The Kaiser-Meyer-Olkin measure verified the sampling adequacy for the analysis; $KMO = .92$, and the KMO values for individual items ranged from .81 to .95. Bartlett's test of sphericity, $\chi^2(496) = 3896.97$, $p < .001$, indicated that correlations between items were sufficiently large for EFA. While Horn's parallel analysis suggested a four-factor solution, the scree plot suggested three factors, and only two factors had eigenvalues larger than 1. Therefore, we explored all three solutions. Both the four- and three-factor solutions contained multiple cross-loadings, and only the two-factor solution had a clean factor structure, with the first factor loading on all helpful items and the second factor loading on all hindering items (see Supplement 3). This solution corresponds to previous SIS/RSRS studies (Elliott & Wexler, 1994; Reeker et al., 1996) and is also clinically meaningful. The two factors were independent ($r = -.08$) and together explained 50% of the variance. The reliability was $\alpha = .91$ for the helpful-reactions subscale and $\alpha = .92$ for the hindering-reactions subscale.

The helpful-reactions subscale had large positive correlations with the ITPQ in-session impact ($r = .78$), confident collaboration ($r = .70$), and global alliance subscales ($r = .76$), CETS pattern identification ($r = .66$), therapist responsiveness ($r = .51$), and transformative acceptance ($r = .57$) subscales and had medium to large correlations ($r = .45-.54$) with all three SEQ subscales. The hindering-reactions subscale had very large correlations with the ITPQ patient fear ($r = .80$) and therapist interference ($r = .80$) subscales and the CETS disconnection/disengagement ($r = .84$) subscale and was moderately negatively related to SEQ depth ($r = -.48$) (see Supplement 4). The fact that both the helpful- and hindering-reactions subscales correlated with the ITPQ problems actuation subscale (although the relationship was much weaker in the case of hindering reactions; $r = .50$ and $.28$) is consistent with the assumption that problem actuation is an ingredient of the therapeutic

change yet may be a source of unpleasant in-session experiences.

Study 2: Developing a Brief Version of the Session Reactions Scale-3

Method

Samples

Therapists. A sample of $N = 111$ therapists participated in the study (57.7% women, 39.6% men, 0.9% genderfluid, and 1.8% nonbinary). Their ages ranged from 24 to 77 years ($M = 44.59$, $SD = 15.02$), and their races/ethnicities included White (74.8%), Other/Multiracial (15.3%), Asian American (3.6%), African American (2.7%), Latino/a/x or Hispanic (2.7%), and Indigenous (0.9%). Their highest level of education was PhD (64%), MA/MS (20.7%), PsyD (8.1%), Other (3.6%), BA/BS (2.7%), and not reported (0.9%). Most therapists were trained professionals (82.0%), and the remaining therapists were trainees (17.1%) or did not report their professional status (0.9%). The therapists' length of professional experience varied between 1 and 52 years ($M = 15.36$, $SD = 13.3$). The self-reported theoretical orientation was integrative (47.7%), cognitive/behavioral (17.1%), psychodynamic/psychoanalytic (13.5%), humanistic/experiential/existential (11.7%), systemic/family/postmodern (2.7%), and other (7.2%).

Experts. To cluster SRS-3 items to create a short form of the SRS3, we created a purposeful sample of $N = 11$ psychologists, seven PhD-level and four PhD candidates, who also had experience as both psychotherapy clients and therapists (six of them are coauthors of this article, and the remaining five were their colleagues). Seven were men and four were women, predominantly White (one Asian and one Biracial-White/Asian), aged from 26 to 71 ($M = 37.6$, $SD = 12.9$), with one to 49 years of professional experience ($M = 11.5$, $SD = 13.4$). Their self-identified theoretical orientation was integrative ($n = 5$), systemic ($n = 2$), humanistic/experiential ($n = 1$), dynamic ($n = 1$), cognitive/behavioral ($n = 1$), and relational-cultural ($n = 1$).

Procedure

Therapist sample. The data in the therapists' sample were collected in Qualtrics. Therapists were recruited via professional listservs of the Society for Psychotherapy Research (SPR), Society for the Exploration of Psychotherapy Integration (SEPI), and Division 12 (Society of Clinical Psychology), Division 17 (Society of Counseling Psychology),

and Division 29 (Society for the Advancement of Psychotherapy) of the American Psychological Association. Apart from demographic questions (gender identity, age, race/ethnicity, highest degree, professional status, years of professional experience, and theoretical orientation), therapists were administered the SRS-3 items with the following instruction: "Please rate the extent to which you, as a clinician, consider obtaining a client's session-by-session responses on each of these items as vital feedback for your further work with this client." They rated each item on a five-point Likert scale ranging from 1 = "not at all" to 5 = "very much". One hundred thirty-four therapists opened the survey, but 23 of them did not answer to any items beyond the demographic questions.

We then computed the means and standard deviations of each item's perceived clinical utility, as well as the correlation between the perceived utility and the years or therapists' professional experience.

Expert sample. We asked the expert sample to categorize the SRS-3 items based on their conceptual similarity. They were instructed to use as many or as few piles as desired but to keep only one level of hierarchy (i.e., not to create subcategories). This produced 11 independent classifications. We then used these classifications to create a co-occurrence matrix: the number in each cell represented how many times the two items were placed in the same category. This co-occurrence matrix then served as the input for hierarchical clustering, using Ward's method and Euclidean distance as the metric. The level of cluster resolution was determined based on team discussion: we aimed to include as many distinct types of reactions as possible while eliminating conceptual overlaps. A similar procedure was used by Elliott (1985) and Tracey et al. (2003).

Item selection process. We used a combination of conceptual, empirical, and pragmatic criteria to select items for the brief version. Conceptually, we strove to represent diverse types of reactions. For this purpose, we used the conceptual clusters obtained from the expert sample ($N = 11$; see Supplement 5) as a basis for item selection, with the intention of retaining at least one item from each cluster. Empirically, we examined the concurrent validity of the items (i.e., items' correlations with external measures obtained from the client sample in Study 1, $N = 242$; see Supplement 4) to see which items best represent established psychotherapy process constructs. Furthermore, we used existing research where relevant (see the Results section). Pragmatically, we strove to retain items

that were perceived as clinically useful, using the clinical utility ratings obtained from the therapist sample ($N = 111$; see Supplement 4). We also considered the likelihood that a client would be willing to answer the item frankly in the context of routine monitoring. Finally, we sought to keep the measure short enough so that it could be completed within two minutes. Therefore, we decided to limit the number of items in the brief version to 15. Instead of relying on one source of data, the decisions made about retaining and/or modifying the items were based multiple data sources, team discussion, and were also influenced by the authors' clinical experience. A similar multi-perspective procedure was used in the development of the CORE-10 measure (Barkham et al., 2013).

Results

The hierarchical clustering of the similarity ratings provided by the expert sample identified 11 clusters of conceptually similar items (see Supplement 5). Item correlations with other measures are provided in Supplement 4. We used this information as a basis for selecting items for the brief version. Except for Cluster 2, all clusters are represented by one or more items in the brief version. In two cases, we decided to create new, composite items instead of selecting one of the original items. For instance, we combined items representing new insight/awareness about the self or others into a single item. Since Cluster 2 (Clients' disengagement) items were conceptually represented by the emotional bond item, we did not include this cluster in the brief version. Supplement 6 shows the rationale for item selection, and Supplement 7 contains the wording of the final brief version (SRS-3-B).

Study 3: Testing the Brief Version of the Session Reactions Scale-3

Method

Sample

Clients. A sample of $N = 265$ clients (61.1% women, 36.6% men, and 2.3% nonbinary) participated in the study. Their ages ranged from 18 to 78 years ($M = 41.68$, $SD = 13.44$), and their ethnicities included White Non-Hispanic (76.2%), African American (9.8%), Other/Multiracial (7.5%), Latino/a/x or Hispanic (3.0%), Asian American (1.5%), Indigenous (1.5%), and Native Hawaiian (0.4%). Clients' highest level of education included High School (47.5%), Bachelor's Degree (29.8%), Other (11.0%), Master's Degree (8.7%), and

Doctoral Degree (3.0%). At the time of taking the survey, they reported having attended only one session (3.4%), two to five sessions (23.0%), six to 10 sessions (22.3%), 11 to 15 sessions (13.6%), 16 to 20 sessions (11.3%), 21 to 50 sessions (9.8%), and more than 50 sessions (16.6%). Their most recent session took place within one week ago (44.5%), between one and two weeks ago (35.8%), between two and three weeks ago (11.3%), and between three weeks and four weeks ago (8.3%). Clients whose most recent session took place more than four weeks ago were deemed ineligible to participate.

Procedure

Recruitment and data collection. The clients' sample data were collected on the Qualtrics online survey platform, and clients were recruited via the CloudResearch platform (Chandler et al., 2019) and via social media (Facebook and Reddit). Clients were administered the SRS-3-B and ITPQ measures (the reliability of the ITPQ subscale scores was $\alpha = .94$ for in-session impact, .69 for problem actualization, .90 for confident collaboration, .93 for global alliance, .84 for patient fear, and .68 for therapist interference in Study 2). The current study applied the same eligibility criteria as in Study 1.

Of the total of 2983 people who accessed the survey on the CloudResearch platform, 2511 were ineligible for the study, another 186 missed one or more attention checks, 36 responded too quickly (we set the minimal duration cutoff to 240 s; this value was based on the cutoff used in Study 1, see Supplement 2, reflecting the lower number of items in Study 2), and three were removed based on manual rating of response patterns (see Supplement 2). The CloudResearch platform thus contributed 247 responses; these respondents were paid for participation. Another thirty people responded to the survey based on a social media advertisement; 12 of them were removed based on the duration cutoff. The social media pathway thus contributed 18 responses. Respondents in the social media pathway were not paid.

Statistical analysis. Statistical analyses followed the analytic plan employed in Study 1. We conducted descriptive analyses of the SRS-3-B items and inter-item correlations, applied EFA to explore the factor structure, and examined the concurrent validity by computing Spearman's correlation with the ITPQ subscales for each SRS-3 item (as well as Pearson's correlations for SRS-3-B subscales).

Table I. SRS-3-B item description and exploratory factor analysis (Study 3).

Item	Item description	Item description			Two-factor EFA			One-factor EFA	
		<i>M</i>	<i>SD</i>	<i>Mdn</i>	F1	F2	Com.	F1	Com.
1	Realized something new	3.12	1.14	3	.79	.26	.48	.55	.30
2	Understood, supported, or reassured	4.03	1.05	4	.61	-.13	.48	.69	.47
3	More distanced	2.42	1.13	2	.47	.38	.18	.16	.03
4	Empowered, hopeful, or positive about self	3.27	1.18	3	.71	-.17	.66	.81	.66
5	Stuck, blocked, or unable to progress	1.96	1.18	2	-.17	.71	.66	-.68	.46
6	Positively or hopeful about another person	3.08	1.25	3	.63		.46	.67	.45
7	Clearer about problems/goals	3.57	1.10	4	.73	-.12	.64	.79	.63
8	Uncomfortable	1.73	1.04	1	.13	.70	.42	-.39	.15
9	Improved skills/learned strategies	3.45	1.19	4	.78		.65	.78	.60
10	Feel worse	1.34	0.83	1		.88	.84	-.69	.47
11	Personally invested	3.77	1.12	4	.62	-.18	.53	.73	.54
12	Lack of direction or guidance	1.51	0.96	1	-.21	.65	.61	-.67	.45
13	Relieved or less burdened	3.21	1.20	3	.71	-.15	.63	.79	.62
14	Accepted self or situation	3.63	1.03	4	.81		.62	.74	.54
Eigenvalue					6.46	1.39		6.37	
% variance explained					36%	20%		46%	
Cronbach's alpha					.88	.77		.89	

Note. F1 = Helpful reactions, F2 = hindering reactions. Factor loadings over .40 appear in bold, loadings less than .10 were removed. Correlation between the factors in the two-factor solution was $r = -.52$. For full item wording see the Supplement 6.

Results

See Table I for item descriptions. All items had reasonable variation, and none of them evidenced a severe floor or ceiling effect. Two pairs of items had a correlation of $r > .70$, namely, Items 10 (Feeling worse) and 5 (Stuck, blocked, or unable to progress) and Items 10 (Feeling worse) and 12 (Lack of direction or guidance). We concluded that despite the high correlation, these items represented distinct categories of experience and were retained.

To explore the SRS-3-B factor structure, we conducted EFA. The Kaiser-Meyer-Olkin measure verified the sampling adequacy for the analysis; $KMO = .92$, and the KMO values for individual items ranged from .76 to .95. Bartlett's test of sphericity, $\chi^2(91) = 1623.17, p < .001$, indicated that correlations between items were sufficiently large for EFA. Horn's parallel, scree plot, and Kaiser's rule all agreed on two factors. The two-factor solution had a nearly clean structure with one cross-loading: Item 3 (distanced from certain feelings, thoughts, or memories) had a similarly strong loading on both the helpful- and the hindering-reactions factor (see Table I). The helpful- and hindering-reactions subscales, based on bolded items, had good to acceptable reliability ($\alpha = .88$ and $.77$, respectively). The helpful-reactions subscale's reliability would increase to $\alpha = .90$ if Item 3 was dropped.

Unlike the full SRS-3 version, the two factors were relatively strongly negatively correlated ($r = -.52$). Therefore, we also explored a unidimensional solution. All items had a reasonably high loading,

except for Item 3 (distanced from certain feelings, thoughts, or memories, see Table I). The reliability of the total scale was $\alpha = .89$ (.90 if Item 3 was dropped).

Table II shows the correlations between SRS-3-B items and the ITPQ subscales. Except for Item 3, all items show the expected direction of relationships: the helpful-reactions items had moderate to large positive correlations (.35-.71) with the positive ITPQ subscales (i.e., in-session impact, confident collaboration, global alliance, and problem actuation); most also had small to medium negative correlations with the negative ITPQ subscales (i.e., patient fear and therapist interference). Items 2 (Understood, supported, or reassured), 7 (Clearer about problems/goals), and 11 (Personally invested) were most strongly related to the ITPQ global alliance subscale. Items 1, 4, 6, 9, 13, and 14 represent specific micro-outcomes and, as such, were well related to the ITPQ in-session impact subscale. Most hindering items had small to medium positive relations to the ITPQ patient fear and therapist interference subscales. Item 3 had negligible relationships across all ITPQ subscales.

Discussion

This study aimed to develop a brief pantheoretical post-session self-report measure suitable for use in the context of routine process and outcome monitoring. We first created a longer measure (32 items) that captured a broad scope of clients' session reactions.

Table II. Correlations between SRS-3-B items and the ITPQ (Study 3).

Item	ITPQ subscales					
	In-session impact	Confident collaboration	Global alliance	Patient fear	Therapist interference	Problem actuation
1 Realized something new	.56	.51	.51	-.01	.01	.42
2 Understood, supported, or reassured	.55	.53	.66	-.20	-.24	.39
3 More distanced	.16	.13	.17	.12	.06	.12
4 Empowered, hopeful, or positive about self	.69	.61	.54	-.34	-.10	.39
5 Stuck, blocked, or unable to progress	-.42	-.42	-.38	.48	.30	-.22
6 Positively or hopeful about another person	.58	.52	.46	-.16	-.02	.35
7 Clearer about problems/goals	.67	.64	.61	-.26	-.16	.39
8 Uncomfortable	-.25	-.25	-.26	.35	.20	-.07
9 Improved skills/learned strategies	.65	.63	.55	-.22	-.16	.40
10 Feel worse	-.38	-.40	-.37	.36	.27	-.16
11 Personally invested	.61	.65	.61	-.26	-.21	.47
12 Lack of direction or guidance	-.45	-.45	-.45	.35	.31	-.29
13 Relieved or less burdened	.71	.65	.60	-.31	-.15	.47
14 Accepted self or situation	.67	.68	.60	-.25	-.19	.40
<i>SRS-3-B subscales</i>						
Helpful reactions	.86	.82	.78	-.29	-.23	.55
Hindering reactions	-.50	-.52	-.51	.51	.48	-.24
Total score	.76	.70	.65	-.11	-.05	.52

Note. Correlations $>.30$ (i.e., medium and higher) are in bold.

This was then shortened to a brief version (15 items) balancing clinical utility and brevity. The measure was developed in a bottom-up manner to represent session reactions that clients reported as significant in existing qualitative studies. Although we had no *a priori* intention to represent specific theoretical constructs, both the full and the brief versions of the measure cover aspects that have been recognized as key ingredients in the therapeutic process, including the therapeutic alliance (Flückiger et al., 2020; Horvath et al., 2011), various immediate gains (such as empowerment, new skills/strategies, awareness/insight, acceptance, and relief), the lack of fit between the therapists' approach and the clients' needs or expectations (Swift et al., 2018), and non-improvement/deterioration (Cuijpers et al., 2018).

We designed each item to represent a stand-alone, phenomenologically distinguishable category of clients' session reactions. Therefore, the items were primarily selected based on conceptual similarity rather than on mere co-occurrence (as would be the case in the factor-analytic approach). Although the lack of multiple items per construct did not allow us to model the measurement error when working with the measure on the item level, the items can be considered to possess high face validity. Since they have been derived from qualitative studies on client-reported session reactions, they represent

phenomena that are subjectively important and easy to reflect on and rate from the clients' perspective. This stands in contrast to process measures such as the Session Evaluation Questionnaire (Stiles et al., 1994), which consists of abstract items that do not directly translate to distinct session reactions. Furthermore, the breadth of reactions covered by the full and brief version provides potentially richer information than other widely used post-session measures such as the Session Rating Scale (Duncan et al., 2003).

Nevertheless, exploratory factor analyses showed that both long and brief versions of the measure have a meaningful factor structure composed of two factors, namely, overall helpfulness and hinderingness of the session, with acceptable to excellent reliability. Within the brief version, items can be summed to derive a total score that serves as an overall index of session evaluation, as described in Siqueland et al.'s (2004) study. Thus, besides obtaining feedback on the item-by-item level, the SRS-3 and SRS-3-B can be also employed in empirical studies as a measure of perceived session quality. However, clinicians and researchers should bear in mind that the perceived helpfulness and hinderingness is based on clients' subjective evaluation, which does not necessarily correspond to the role these experiences play in the treatment process. For

instance, some session events can be perceived as unpleasant, yet helpful in the long run (Marren et al., 2022).

Only one item (SRS-3-B Item 3) seemed to break this pattern: “I feel more distanced from certain feelings, thoughts, or memories”. While this item loaded only on the hindering subscale in the long version, it had relatively strong positive loadings on both the helpful- and hindering-reactions subscales in the brief version. Although the classic factor-analytic approach would require us to drop this item, the clinimetric perspective (Carrozzino et al., 2021; DeVet et al., 2003) underlying the development of this measure supported its retention. Furthermore, in a network modeling study based on RSRS data that also contained this item, feeling distanced from one’s feelings, thoughts, or memories was one of the most central nodes and positively influenced other helpful reactions, such as the therapeutic relationship and relief (Řiháček et al., 2023). The complex role of this item within the factor structure may reflect the fact that distancing from one’s experiencing can be interpreted both as experiential avoidance (Cookson et al., 2020) and as a disentanglement of the self from cognitive or emotional content, such as in cognitive defusion (Hayes et al., 1999) and mindfulness exercises (Dunn et al., 2013). Its role may also differ in various phases of the treatment (Řiháček et al., 2023). Although it might be useful to have an item that can distinguish these two aspects of distancing oneself from the experience, we suspect that it is seldom in the clients’ capacity to make such a distinction. Therefore, we find it more feasible to keep the item in its current form and leave therapists to interpret responses based on context.

The SRS-3 and SRS-3-B items showed meaningful correlations with other process measures. Items that capture various in-session gains, such as new awareness/insight, empowerment, and improved skills, correlated most strongly with the ITPQ in-session impacts subscale. One item that was designed to capture the emotional quality of the therapeutic relationship was most related to the ITPQ global alliance subscale. The hindering-reaction items were correlated with the ITPQ patient fear and therapist interference subscales. Although clients tend to defer to their therapists and may refrain from disclosing their dissatisfaction with the therapist or the process (Farber, 2020; Rennie, 1994), SRS-3 and SRS-3-B seem capable of detecting these phenomena. Interestingly, all helpful SRS-3-B items were positively related to the ITPQ problem actuation subscale. Although previous studies failed to find a relationship between problem actuation and outcome (Mander et al., 2015), this finding supports

the importance of problem actuation as a psychotherapy process phenomenon.

The fact that the brief version emphasizes helpful reactions over hindering ones, can be seen as a limitation. In case the assessment of a wide array of reactions, including hindering ones, is a priority, the long version can be used — the SRS-3 represents a significant improvement on the RSRS because of its greater breadth and briefer, more streamlined items. Alternatively, lower scores on the helpful reaction items can be used as indices of suboptimal session experiences and can be subsequently explored with the client. For instance, a lower score on SRS-3-B Item 2 (“I feel understood, supported, or reassured by my therapist”) can be used as a signal of an alliance rupture. This is in line with the clinical guidelines for using another process measure, the Session Rating Scale (Duncan et al., 2003). However, clinicians should exert caution in interpreting clients’ item scores normatively. For example, it may be perfectly fine for a client to score “Not at All” on SRS-3-B Item 6 (“I feel more positively or hopeful about another person(s)”) if the session did not focus on such a topic.

Limitations

The data for this study were collected online via two crowdsourcing platforms. In Study 1, the quality of a large portion of the clients’ data was problematic. Despite the existing evidence supporting the use of the MTurk platform for research in industrial and organizational psychology (Woo et al., 2015) and psychotherapy (Tompkins, 2019), we cannot recommend this platform for study designs similar to the present study. Although we employed a rigorous multistep and multi-rater procedure to filter out untrustworthy responses, detecting careless responses is a nontrivial task that requires researchers to make arbitrary decisions and involves subjective judgment (Gottfried et al., 2022). The study relied on self-reports, and some associations in the data may thus result from self-report biases. As with any online survey, it was not possible to ensure that all respondents were actual psychotherapy clients. On the other hand, clients did not have to worry about the therapists’ reaction to the scores and could thus provide genuine answers. Furthermore, the data were collected up to four weeks after the clients’ most recent therapeutic session. While this allowed clients to provide a thoughtful reflection on the session experience, some aspect of the experience could have been more difficult to access for the clients and their responses might thus not capture the immediate post-session reactions.

For most external measures’ subscales, the reliability was comparable to that in the original

studies. However, the ITPQ problem actuation subscale, the arousal SEQ subscale, and several of the CETTS subscales have less than optimal reliability in our study. We attribute this to the fact that the data were collected up to four weeks after the most recent session.

When designing items for the brief version of the measure, we strove to maximize the coverage of the measure with as few items as possible. Therefore, in some cases, we created items that were a combination of several items from the long version. While this allowed us to create a compact and versatile measure, the brief version is no longer a subset of the SRS-3 items, and data collected by the two versions cannot be compared directly. Furthermore, the absence of data on treatment outcomes did not allow us to test the measure's predictive validity. More specifically and following the clinimetric tradition, it would be also desirable to test the measure's incremental validity in predicting psychotherapy outcomes both on the item level (i.e., each item's added value) and the scale level (i.e., in addition to other constructs such as the working alliance).

Conclusions and Future Directions

We developed a brief pantheoretical self-report measure of clients' session reactions that can be used both for routine process monitoring and for research purposes. Each item represents a stand-alone category of clients' reactions; therefore, in the context of clinical practice, the measure can be interpreted at the item level. Clients should be encouraged to differentiate among the items and provide responses that reflect their genuine session experiences. Nevertheless, our study also supported using either two composite scores (i.e., session helpfulness and hinderingness) or a single total score representing the overall session evaluation. A notable strength of the measure lies in the fact that it was derived from qualitative research on clients' reactions to significant events in psychotherapy sessions. Therefore, it contains categories that clients themselves are likely to consider important and, thus, can easily rate from their own perspective. While the SRS-3-B may be generally more suitable for the use in routine process monitoring, the full version is recommended in situations where a more in-depth assessment of negative reactions is desired (e.g., with challenging client populations such as those with borderline processes).

This study provided initial psychometric evidence for both versions of the measure, though more research is needed to elucidate their properties. First, the measure's ability to predict outcomes

(i.e., predictive validity) should be determined in future studies, both in terms of the total score and the incremental validity of each item. It is also possible that responses provided immediately after sessions differ from those obtained several hours or even days later, since clients' evaluation may change as they have time to reflect. Relatedly, reactions measured in different intervals may have different predictive power. Second, the measure's predictive validity above and beyond that of the working alliance should be investigated to see if adding information about a broad range of clients' session experiences explains additional outcome variance. Third, the clinical utility of the measure must be tested in a session-by-session routine monitoring scenario. Besides other things, this will also depend on the session-by-session variation of the ratings. Fourth, norms or "critical values" can be determined empirically that would serve as warning signals for clinicians. Fifth, a mixed-method study can be conducted to explore how frankly clients respond to the scale when they expect their therapist to review their responses.

In the clinical context, the measure is meant to serve as a means of initiating a discussion with a client and to receive rich and clinically meaningful feedback. While the measure itself does not allow clinicians to make connections between specific sessions events and clients' session reactions, these can be explored in a follow-up therapeutic conversation. In this sense, this measure can become a part of MBC (Zhu et al., 2021) or feedback-informed treatment (Prescott et al., 2017). Further research is needed to determine whether using this kind of process-focused feedback has the potential to increase the effectiveness of therapy.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Funding

This study was supported by the Czech Ministry of Education, Youth, and Sports under Grant MUNI/A/1446/2022 and by J. William Fulbright Commission in the Czech Republic. Jesse Owen was awarded funds to support this work by the John Templeton Foundation and SonderMind, Inc.

Data Availability Statement

The data will be made available by the first author upon reasonable request.

Supplemental Data

Supplemental data for this article can be accessed online at <https://doi.org/10.1080/10503307.2023.2241983>.

ORCID

Tomáš Řiháček  <http://orcid.org/0000-0001-5893-9289>

Robert Elliott  <http://orcid.org/0000-0002-3527-3397>

Michaela Ladmanová  <http://orcid.org/0000-0001-9809-6044>

Jeremy J. Coleman  <http://orcid.org/0000-0002-8026-2101>

Matteo Bugatti  <http://orcid.org/0000-0001-5640-0992>

References

- APA Presidential Task Force on Evidence-Based Practice. (2006). Evidence-based practice in psychology. *American Psychologist*, 61(4), 271–285. <https://doi.org/10.1037/0003-066X.61.4.271>
- Barkham, M., Bewick, B., Mullin, T., Gilbody, S., Connell, J., Cahill, J., Mellor-Clark, J., Richards, D., Unsworth, G., & Evans, C. (2013). The CORE-10: A short measure of psychological distress for routine use in the psychological therapies. *Counselling and Psychotherapy Research*, 13(1), 3–13. <https://doi.org/10.1080/14733145.2012.729069>
- Bohart, A. C., & Tallman, K. L. (1999). *How clients make therapy work: The process of active self-healing*. American Psychological Association.
- Boswell, J. F., Hepner, K. A., Lysell, K., Rothrock, N. E., Bott, N., Childs, A. W., Douglas, S., Owings-Fonner, N., Wright, C. V., Stephens, K. A., Bard, D. E., Aajmain, S., & Bobbitt, B. L. (2023). The need for a measurement-based care professional practice guideline. *Psychotherapy*, 60(1), 1–16. Advance online publication. <https://doi.org/10.1037/pst0000439>
- Boswell, J. F., Kraus, D. R., Miller, S. D., & Lambert, M. J. (2015). Implementing routine outcome monitoring in clinical practice: Benefits, challenges, and solutions. *Psychotherapy Research*, 25(1), 6–19. <https://doi.org/10.1080/10503307.2013.817696>
- Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research*, 36(1), 111–150. https://doi.org/10.1207/S15327906MBR3601_05
- Carrozzino, D., Patierno, C., Guidi, J., Berrocal Montiel, C., Cao, J., Charlson, M. E., Christensen, K. S., Concato, J., De Las Cuevas, C., De Leon, J., Eöry, A., Fleck, M. P., Furukawa, T. A., Horwitz, R. I., Nierenberg, A. A., Rafanelli, C., Wang, H., Wise, T. N., Sonino, N., & Fava, G. A. (2021). Clinimetric criteria for patient-reported outcome measures. *Psychotherapy and Psychosomatics*, 90(4), 222–232. <https://doi.org/10.1159/000516599>
- Chandler, J., Rosenzweig, C., Moss, A. J., Robinson, J., & Litman, L. (2019). Online panels in social science research: Expanding sampling methods beyond Mechanical Turk. *Behavior Research Methods*, 51(5), 2022–2038. <https://doi.org/10.3758/s13428-019-01273-7>
- Chui, H., Palma, B., Jackson, J. L., & Hill, C. E. (2020). Therapist–client agreement on helpful and wished-for experiences in psychotherapy: Associations with outcome. *Journal of Counseling Psychology*, 67(3), 349–360. <https://doi.org/10.1037/cou0000393>
- Constantino, M. J., Coyne, A. E., & Muir, H. J. (2020). Evidence-based therapist responsiveness to disruptive clinical process. *Cognitive and Behavioral Practice*, 27(4), 405–416. <https://doi.org/10.1016/j.cbpra.2020.01.003>
- Cookson, C., Luzon, O., Newland, J., & Kingston, J. (2020). Examining the role of cognitive fusion and experiential avoidance in predicting anxiety and depression. *Psychology and Psychotherapy: Theory, Research and Practice*, 93(3), 456–473. <https://doi.org/10.1111/papt.12233>
- Cuijpers, P., Reijnders, M., Karyotaki, E., de Wit, L., & Ebert, D. D. (2018). Negative effects of psychotherapies for adult depression: A meta-analysis of deterioration rates. *Journal of Affective Disorders*, 239, 138–145. <https://doi.org/10.1016/j.jad.2018.05.050>
- De Jong, K., Conijn, J. M., Gallagher, R. A. V., Reshetnikova, A. S., Heij, M., & Lutz, M. C. (2021). Using progress feedback to improve outcomes and reduce drop-out, treatment duration, and deterioration: A multilevel meta-analysis. *Clinical Psychology Review*, 85, 102002. <https://doi.org/10.1016/j.cpr.2021.102002>
- DeVet, H. C. W., Terwee, C. B., & Bouter, L. M. (2003). Clinimetrics and psychometrics: Two sides of the same coin. *Journal of Clinical Epidemiology*, 56(12), 1146–1147. <https://doi.org/10.1016/j.jclinepi.2003.08.010>
- Duncan, B. L., Miller, S. D., & Sparks, J. A. (2004). *The heroic client: A revolutionary way to improve effectiveness through client-directed, outcome-informed therapy*. Jossey-Bass.
- Duncan, B. L., Miller, S. D., Sparks, J. A., Claud, D. A., Reynolds, L. R., Brown, J., & Johnson, L. D. (2003). The Session Rating Scale: Preliminary psychometric properties of a “working” alliance measure. *Journal of Brief Therapy*, 3(1), 3–12.
- Dunn, R., Callahan, J. L., & Swift, J. K. (2013). Mindfulness as a transtheoretical clinical process. *Psychotherapy*, 50(3), 312–315. <https://doi.org/10.1037/a0032153>
- Elliott, R. (1985). Helpful and nonhelpful events in brief counseling interviews: An empirical taxonomy. *Journal of Counseling Psychology*, 32(3), 307–322. <https://doi.org/10.1037/0022-0167.32.3.307>
- Elliott, R. (2010). *Revised Session Reactions Scale*. <http://www.eft-scotland.org/wp-content/uploads/2021/12/Revised-Session-Reactions-Scale-C-v3.doc>
- Elliott, R., & James, E. (1989). Varieties of client experience in psychotherapy: An analysis of the literature. *Clinical Psychology Review*, 9(4), 443–467. [https://doi.org/10.1016/0272-7358\(89\)90003-2](https://doi.org/10.1016/0272-7358(89)90003-2)
- Elliott, R., James, E., Reimschuessel, C., Cislo, D., & Sack, N. (1985). Significant events and the analysis of immediate therapeutic impacts. *Psychotherapy: Theory, Research, Practice, Training*, 22(3), 620–630. <https://doi.org/10.1037/h0085548>
- Elliott, R., & Shapiro, D. A. (1992). Clients and therapists as analysts of significant events. In S. R. Toukmanian & D. L. Rennie (Eds.), *Psychotherapy process research: Theory-guided and phenomenological research strategies* (pp. 163–186). Sage.
- Elliott, R., Wagner, J., Sales, C. M. D., Rodgers, B., Alves, P., & Café, M. J. (2016). Psychometrics of the Personal Questionnaire: A client-generated outcome measure. *Psychological Assessment*, 28(3), 263–278. <https://doi.org/10.1037/pas0000174>
- Elliott, R., & Wexler, M. M. (1994). Measuring the impact of sessions in process-experiential therapy of depression: The session impacts scale. *Journal of Counseling Psychology*, 41(2), 166–174. <https://doi.org/10.1037/0022-0167.41.2.166>
- Farber, B. A. (2020). Disclosure, concealment, and dishonesty in psychotherapy: A clinically focused review. *Journal of Clinical Psychology*, 76(2), 251–257. <https://doi.org/10.1002/jclp.22891>

- Fauth, J., & Williams, E. N. (2005). The in-session self-awareness of therapist-trainees: Hindering or helpful? *Journal of Counseling Psychology, 52*(3), 443–447. <https://doi.org/10.1037/0022-0167.52.3.443>
- Fitzpatrick, M. R., Iwakabe, S., & Stalikas, A. (2005). Perspective divergence in the working alliance. *Psychotherapy Research, 15*(1–2), 69–80. <https://doi.org/10.1080/10503300512331327056>
- Flückiger, C., Rubel, J., Del Re, A. C., Horvath, A. O., Wampold, B. E., Crits-Christoph, P., Atzil-Slonim, D., Compare, A., Falkenström, F., Ekeblad, A., Errázuriz, P., Fisher, H., Hoffart, A., Huppert, J. D., Kivity, Y., Kumar, M., Lutz, W., Muran, J. C., Strunk, D. R., ... Barber, J. P. (2020). The reciprocal relationship between alliance and early treatment symptoms: A two-stage individual participant data meta-analysis. *Journal of Consulting and Clinical Psychology, 88*(9), 829–843. <https://doi.org/10.1037/ccp0000594>
- Follmer, D. J., Sperling, R. A., & Suen, H. K. (2017). The role of MTurk in education research: Advantages, issues, and future directions. *Educational Researcher, 46*(6), 329–334. <https://doi.org/10.3102/0013189X17725519>
- Gottfried, J., Ježek, S., Králová, M., & Řiháček, T. (2022). Autocorrelation screening: A potentially efficient method for detecting repetitive response patterns in questionnaire data. *Practical Assessment, Research and Evaluation, 27*, Article 2. <https://doi.org/10.7275/vyx-bt24>
- Grawe, K. (2004). *Psychological therapy*. Hogrefe & Huber Publishers.
- Greenberg, L. S. (1986). Change process research. *Journal of Consulting and Clinical Psychology, 54*(1), 4–9. <https://doi.org/10.1037/0022-006X.54.1.4>
- Hannan, C., Lambert, M. J., Harmon, C., Nielsen, S. L., Smart, D. W., Shimokawa, K., & Sutton, S. W. (2005). A lab test and algorithms for identifying clients at risk for treatment failure. *Journal of Clinical Psychology, 61*(2), 155–163. <https://doi.org/10.1002/jclp.20108>
- Hatfield, D., McCullough, L., Frantz, S. H. B., Krieger, K., Hatfi, D., McCullough, L., Frantz, S. H. B., & Krieger, K. (2009). Do we know when our clients get worse? An investigation of therapists' ability to detect negative client change. *Clinical Psychology & Psychotherapy, 17*(1), 25–32. <https://doi.org/10.1002/cpp.656>
- Hayes, S. C., Strosahl, K. D., & Wilson, K. G. (1999). *Acceptance and commitment: An experiential approach to behavior change*. The Guilford Press.
- Horvath, A. O., Del Re, A. C., Flückiger, C., & Symonds, D. (2011). Alliance in individual psychotherapy. *Psychotherapy, 48*(1), 9–16. <https://doi.org/10.1037/a0022186>
- Janzen, J., Fitzpatrick, M., & Drapeau, M. (2008). Processes involved in client-nominated relationship building incidents: Client attachment, attachment to therapist, and session impact. *Psychotherapy: Theory, Research, Practice, Training, 45*(3), 377–390. <https://doi.org/10.1037/a0013310>
- Kellett, S., Clarke, S., & Matthews, L. (2007). Session impact and outcome in group psychoeducative cognitive behavioural therapy. *Behavioural and Cognitive Psychotherapy, 35*(3), 335–342. <https://doi.org/10.1017/S135246580600347X>
- King, R., Bambling, M., Reid, W., & Thomas, I. (2006). Telephone and online counselling for young people: A naturalistic comparison of session outcome, session impact and therapeutic alliance. *Counselling and Psychotherapy Research, 6*(3), 175–181. <https://doi.org/10.1080/14733140600874084>
- Kivlighan, D. M., & Arthur, E. G. (2000). Convergence in client and counselor recall of important session events. *Journal of Counseling Psychology, 47*(1), 79–84. <https://doi.org/10.1037/0022-0167.47.1.79>
- Ladmanová, M., Řiháček, T., & Timulak, L. (2022). Client-identified impacts of helpful and hindering events in psychotherapy: A qualitative meta-analysis. *Psychotherapy Research, 32*(6), 723–735. <https://doi.org/10.1080/10503307.2021.2003885>
- Lambert, M. J., Whipple, J. L., & Kleinstäuber, M. (2018). Collecting and delivering progress feedback: A meta-analysis of routine outcome monitoring. *Psychotherapy, 55*(4), 520–537. <https://doi.org/10.1037/pst0000167>
- Levitt, H. M., Grabowski, L. M., Minami, T., & Morrill, Z. (2021). *The Clients' Experience of Therapy Scale (CETS): Assessing outcome by centralizing clients' view of therapy quality* [Manuscript under review].
- Levitt, H. M., & Piazza-Bonin, E. (2011). Therapists' and clients' significant experiences underlying psychotherapy discourse. *Psychotherapy Research, 21*(1), 70–85. <https://doi.org/10.1080/10503307.2010.518634>
- Llewellyn, S. P. (1988). Psychological therapy as viewed by clients and therapists. *British Journal of Clinical Psychology, 27*(3), 223–237. <https://doi.org/10.1111/j.2044-8260.1988.tb00779.x>
- Mander, J., Schlarb, A., Teufel, M., Keller, F., Hautzinger, M., Zipfel, S., Wittorf, A., & Sammet, I. (2015). The Individual Therapy Process Questionnaire: Development and validation of a revised measure to evaluate general change mechanisms in psychotherapy. *Clinical Psychology & Psychotherapy, 22*(4), 328–345. <https://doi.org/10.1002/cpp.1892>
- Marren, C., Mikoška, P., O'Brien, S., & Timulak, L. (2022). A qualitative meta-analysis of the clients' experiences of emotion-focused therapy. *Clinical Psychology & Psychotherapy, 29*(5), 1611–1625. <https://doi.org/10.1002/cpp.2745>
- Miller, S. D., Hubble, M. A., Chow, D., & Seidel, J. (2015). Beyond measures and monitoring: Realizing the potential of feedback-informed treatment. *Psychotherapy, 52*(4), 449–457. <https://doi.org/10.1037/pst0000031>
- Orlinsky, D. E., & Howard, K. I. (1986). The relation of process to outcome in psychotherapy. In S. L. Garfield & A. E. Bergin (Eds.), *Handbook of psychotherapy and behavior change* (3rd ed., pp. 311–381). Wiley.
- Pivolusková, H., Řiháček, T., Čevelíček, M., & Ukropová, L. (2021). Are client- and therapist-identified significant events related to outcome? A systematic review. *Counselling Quarterly, 34*(1), 1–20. <https://doi.org/10.1080/09515070.2019.1642851>
- Prescott, D. S., Maeschalck, C. L., & Miller, S. D. (2017). *Feedback-informed treatment in clinical practice: Reaching for excellence*. American Psychological Association.
- Raiche, G., & Magis, D. (2020). *nFactors: An R package for parallel analysis and non graphical solutions to the Cattell scree test*. <https://cran.r-project.org/package=nFactors>
- R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.r-project.org/>
- Reeker, J., Elliott, R., & Ensing, D. (1996). *Measuring session effects in process-experiential and cognitive-behavioral therapies of PTSD: The Revised Session Reaction Scale*. Paper Presented at Meeting of the Society for Psychotherapy Research, Amelia Island, Florida. <http://www.eft-scotland.org/wp-content/uploads/2021/12/Reeker-Elliott-Ensing-1996.doc>
- Rennie, D. L. (1994). Clients' deference in psychotherapy. *Journal of Counseling Psychology, 41*(4), 427–437. <https://doi.org/10.1037/0022-0167.41.4.427>
- Revelle, W. (2021). *psych: Procedures for personality and psychological research*. Northwestern University. <http://cran.r-project.org/package=psych>
- Řiháček, T., Elliott, R., Owen, J., & Ladmanová, M. (2023). *The structure and dynamics of client session reactions: A longitudinal network analysis* [Manuscript submitted for review].
- Siqueland, L., Crits-Christoph, P., Barber, J. P., Connolly Gibbons, M. B., Gallop, R., Griffin, M., Frank, A., Thase, M. E., Luborsky, L., & Liese, B. (2004). What aspects of

- treatment matter to the patient in the treatment of cocaine dependence? *Journal of Substance Abuse Treatment*, 27(2), 169–178. <https://doi.org/10.1016/j.jsat.2004.06.007>
- Stiles, W. B., Reynolds, S., Hardy, G. E., Rees, A., Barkham, M., & Shapiro, D. A. (1994). Evaluation and description of psychotherapy sessions by clients using the session evaluation questionnaire and the session impacts scale. *Journal of Counseling Psychology*, 41(2), 175–185. <https://doi.org/10.1037/0022-0167.41.2.175>
- Stiles, W. B., Shapiro, D. A., & Firth-Cozens, J. A. (1990). Correlations of session evaluations with treatment outcome. *British Journal of Clinical Psychology*, 29(1), 13–21. <https://doi.org/10.1111/j.2044-8260.1990.tb00844.x>
- Swift, J. K., Callahan, J. L., Cooper, M., & Parkin, S. R. (2018). The impact of accommodating client preference in psychotherapy: A meta-analysis. *Journal of Clinical Psychology*, 74(11), 1924–1937. <https://doi.org/10.1002/jclp.22680>
- Timulak, L. (2007). Identifying core categories of client-identified impact of helpful events in psychotherapy: A qualitative meta-analysis. *Psychotherapy Research*, 17(3), 305–314. <https://doi.org/10.1080/10503300600608116>
- Timulak, L. (2010). Significant events in psychotherapy: An update of research findings. *Psychology and Psychotherapy: Theory, Research and Practice*, 83(4), 421–447. <https://doi.org/10.1348/147608310X499404>
- Tompkins, K. A. (2019). *A comparison of Mechanical Turk with traditional recruitment settings for conducting psychotherapy research* [Doctoral dissertation]. University of Alaska Anchorage. ProQuest Dissertations and Theses Global.
- Tracey, T. J. G. G., Lichtenberg, J. W., Goodyear, R. K., Claiborn, C. D., & Wampold, B. E. (2003). Concept mapping of therapeutic common factors. *Psychotherapy Research*, 13(4), 401–413. <https://doi.org/10.1093/ptr/kpg041>
- Wampold, B. E. (2015). Routine outcome monitoring: Coming of age—With the usual developmental challenges. *Psychotherapy*, 52(4), 458–462. <https://doi.org/10.1037/pst0000037>
- Woo, S. E., Keith, M., & Thornton, M. A. (2015). Amazon Mechanical Turk for industrial and organizational psychology: Advantages, challenges, and practical recommendations. *Industrial and Organizational Psychology*, 8(2), 171–179. <https://doi.org/10.1017/iop.2015.21>
- Wu, M. B., & Levitt, H. M. (2020). A qualitative meta-analytic review of the therapist responsiveness literature: Guidelines for practice and training. *Journal of Contemporary Psychotherapy*, 50(3), 161–175. <https://doi.org/10.1007/s10879-020-09450-y>
- Zhu, M., Hong, R. H., Yan, T., Yang, X., Want, X., Liu, J., Murphy, J. K., Michalak, E. E., Wang, Z., Yatham, L. N., Chen, J., & Lam, R. W. (2021). The efficacy of measurement-based care for depressive disorders: A systematic review and meta-analysis of randomized controlled trials. *The Journal of Clinical Psychiatry*, 82(5), 21r14034. <https://doi.org/10.4088/JCP.21r14034>