Article:
Di Malta, G., Evans, C. and Cooper, M. (2019) Development and validation of the relational depth frequency scale. Psychother Research, 30 (2). pp. 213-227. ISSN: 1050-3307

https://doi.org/10.1080/10503307.2019.1585590

Article:
Di Malta, G orcid.org/0000-0002-6254-6522, Evans, C orcid.org/0000-0002-4197-4202 and Cooper, M (2019) Development and validation of the relational depth frequency scale. Psychother Res. pp. 1-15.

https://doi.org/10.1080/10503307.2019.1585590

Development and Validation of The Relational Depth Frequency Scale

Gina Di Malta
University of Roehampton
Chris Evans
University of Sheffield
Mick Cooper
University of Roehampton

Gina Di Malta
Department of Psychology, University of Roehampton, Holybourne Avenue, London SW15 4JD, gina.dimalta@roehampton.ac.uk

Chris Evans
Department of Psychology, University of Sheffield, Western Bank, Sheffield S10 2TN, chris@psyctc.org

Mick Cooper
Department of Psychology, University of Roehampton, Holybourne Avenue, London SW15 4JD, mick.cooper@roehampton.ac.uk

Reference

Di Malta, G., Evans, C., & Cooper, M. (2019): Development and validation of the relational depth frequency scale. Psychotherapy Research. DOI: 10.1080/10503307.2019.1585590

Acknowledgements

Thanks to Rosanne Knox, David Murphy, Sue Price, Shiri Gurvitz, Ariele Noble, Eddie Chauncy, Aisling Treanor Donnelly, Joel Vos, and Regina Pauli for their contributions to the development of the Relational Depth Frequency Scale.

Development and Validation of The Relational Depth Frequency Scale

Objective: The Relational Depth Frequency Scale (RDFS) was developed to measure the frequency of specific moments of profound contact and engagement in psychotherapy. **Method:** Following an initial process of item generation and rating, Three-Step Test Interviews were conducted with eight therapists and clients to further refine potential items. Sixteen relational depth items were then taken forward for psychometric assessment in an online sample of 336 therapists and 220 clients, each divided into separate "shortening" and "checking" subsamples. **Results:** Following psychometric scale shortening involving confirmatory factor analysis (CFA) and Rasch analysis, we formed a six item RDFS that could be used with both therapists and clients. The parameters of the shortened form replicated well in the independent checking subsamples with good internal consistency (Cronbach's $\alpha = .85$ and .93 in therapist and clients respectively), acceptable fit statistics in CFA and Rasch analysis, and moderate to high levels of convergent validity against the Working Alliance Inventory (Short Form, Revised) and Relational Depth Inventory (RDI-R2). **Conclusion:** As a brief self-report measure, the RDFS can be used to further assess the relationship between relational depth and therapeutic outcomes. Further research is needed to examine the validity of the RDFS in clinical settings.

Keywords: Relational depth, scale development, therapeutic alliance, therapist–patient relationship, measurement.

Clinical or Methodological Significance of this Article:

A self-report measure was developed to assess the frequency of moments of profound contact and engagement in the therapeutic relationship. Previous research suggests that the presence of such moments is associated with better therapeutic outcomes. This measure has the capacity to support further research into this phenomenon, and its potential role in facilitating improved therapeutic outcomes.

An extensive body of psychotherapy research suggests that the therapy relationship makes "substantial and consistent" contributions to outcomes, across different types of psychological treatment (Norcross & Lambert, 2018, p. 308). Although much of this evidence is correlational, more advanced statistical methods, such as cross-lagged panel models, demonstrate that improvements in the quality of the therapeutic relationship directly precede reductions in psychological distress (e.g., Falkenström, Kuria, Othieno, & Kumar, 2018; Zilcha-Mano, Dinger, McCarthy, & Barber, 2014). This suggests a causal relationship from former to latter. However, to date, little is known about the specific subcomponents of the therapy relationship that most strongly affect outcomes (Zilcha-Mano, 2017). This is an important area for further research.

Subcomponents of the therapy relationship can be organized into three types: qualities of the therapeutic interaction (e.g., cohesion in group therapy); therapist relational skills (e.g., empathy); and skills at focusing on, or attempting to manage, the therapy relationship itself (e.g., repairing alliance ruptures) (Castonguay et al., 2006). Of these, evidence is particularly robust that the qualities of the therapeutic interaction are associated with outcomes. Most specifically, this is for the *working alliance* (Bordin, 1979), with an average effect size from approximately 30,000 clients of $r = .28$, 95% CI [0.26–0.30] (Fluckiger, Del Re, Wampold, & Horvath, 2018). The working alliance has been defined as the, "quality and strength of the collaborative relationship between client and therapist" (Hovarth & Bedi, 2002, p.41). It has been conceptualized as having three components: agreement on the goals of therapy, consensus on the tasks of therapy, and the existence of a positive affective bond (Bordin, 1979).

A connected, but distinct, quality of the therapeutic interaction is the *real relationship*. This is conceptualized as a personal relationship characterized by mutual genuineness and a perception of the other, which befits the other (Gelso, 2004). The real relationship differs from the working alliance in that it is characterized by a co-constructed experience of mutual genuineness and "realism" in the relationship (Gelso, 2005). As predicted, the real relationship

has been found to make a contribution to treatment progress, above and beyond the working alliance (Fuertes et al., 2007), with a moderate-sized association to outcome, $r = .38$ (Gelso, Kivlighan & Markin, 2018).

Mutuality is a third quality of the therapeutic interaction that is both related to—but also distinct from—the working alliance and the real relationship. Mutuality has been operationalized as the bidirectional experiencing of Rogers's (1957) core conditions (empathy, congruence, and positive regard); with higher levels of mutuality associated with greater treatment progress (Murphy & Cramer, 2014). As with the real relationship, mutuality implies a co-openness across the therapy dyad. It also overlaps with the working alliance in its emphasis on collaboration. However, with mutuality, there is a particular emphasis on the sharedness of relational qualities across the therapeutic relationship. Closely related to mutuality, recent evidence suggests that "congruence" (i.e., alignment or agreement) between therapists' and clients' ratings of the therapeutic bond is associated with greater next session symptom improvement (Rubel, Barkalifa, Atzil-Slonim, Schmidt, & Lutz, 2018).

Within the field of person-centered and humanistic psychotherapies, researchers have proposed a further quality of the therapeutic interaction, *relational depth* (Mearns & Cooper, 2005, 2018; Knox, Murphy, Wiggins, & Cooper, 2013; Wiggins, Elliott, & Cooper, 2012). Relational depth is defined as, "a state of profound contact and engagement between two people in which each person is fully real with the other, and able to understand and value the other's experiences at a high level" (Mearns & Cooper, 2005, p. xii). Relational depth overlaps with the bond component of the working alliance, in that it involves a warm, personal connection between therapist and client. It also overlaps with the real relationship, in that both involve co-genuineness; and with mutuality, in that both involve a two-way, shared relationship. However, "relational depth" places particular emphasis on the level of interactive, inter-relational engagement between therapist and client (Mearns & Cooper, 2018). More significantly, while qualities such as the working alliance are conceptualized as "gradient" phenomena—more or

less present in the therapy relationship—relational depth has been conceptualized, for empirical purposes, as a "threshold" phenomenon (e.g., Cooper, 2005; Knox & Cooper, 2010): that is, present, or not present, during specific periods of therapy. Relational depth, then, refers to moments in therapy characterized by particularly intense contact and engagement. In this respect, it is synonymous with Stern's (2004) *moments of meeting*, characterized by the interpenetration of the therapist's and the client's minds, where both know and feel what the other knows and feels.

Qualitative research from across a range of client and therapist samples has indicated a high degree of commonality in how moments of relational depth are experienced (Cooper, 2013). This indicates that they are real and distinctive occurrences within the therapy relationship. Summarizing the empirical evidence, Cooper (2013) describes them as moments of connection and flow with another person; in which both individuals feel genuine, alive, and immersed in the relationship. Survey research indicates that nearly all therapists experience moments of relational depth with their clients, as do approximately 80% of clients (Leung, 2008). Wiggins, Elliott and Cooper (2012) found that relational depth was present in over a third of client–defined significant therapy events.

Mearns and Cooper (2005, 2018) argue that moments of relational depth can have an important therapeutic impact by enhancing clients' connections with—and acceptance of—their deepest, "existential" element of self; and by facilitating their capacity to form genuine connections with others. In support of this, qualitative research by Knox (2011) found that clients described these moments as having a significant positive impact. With respect to immediate effects, clients described moments of relational depth as "facilitative", "healing", and "changing"; and also described a positive impact on the therapeutic process itself. In the longer term, the most commonly reported impact was an increased connection with their own selves.

Further evidence for an association between relational depth and therapeutic outcomes comes from research based on Wiggins's Relational Depth Inventory (RDI) (Wiggins, 2012,

2013; Wiggins, Elliott and Cooper, 2012). The RDI was developed to assess levels of relational depth in therapy events that clients identified as "important". Items were developed using thematic analysis and rated by three judges for their suitability. Through psychometric analysis, a revised 24-item measure—the Relational Depth Inventory (RDI-R)—was developed that showed good internal consistency and convergent validity with the Working Alliance Inventory–Short (revised) (WAI-SR) in a sample of 342 therapists and clients (Wiggins, *et al.*, 2012). Using hierarchical regression in a sample of 42 clients, Wiggins (2012) found that depth of relating accounted for 10 to 30 per cent of the variance in improvement, as assessed by the CORE-OM measure of psychological distress (Evans *et al*., 2002). Moreover, RDI-R scores predicted therapeutic outcome above and beyond scores on the Working Alliance Inventory (Short Form, WAI-SF–Revised, Hatcher & Gillaspy, 2006). These findings provided preliminary evidence that moments of relational depth may be an important, distinctive element of the therapy relationship that contribute to outcomes.

However, although the RDI-R is a reliable and valid means of assessing depth in the therapeutic interaction, its specific focus is on the intensity of relating in single events in therapy. This provides important insights into that one, particular experience, but does not capture the frequency of in-depth relating across the therapeutic relationship as a whole.

The Present Study

The overall purpose of this paper, therefore, was to develop a measure of the frequency of moments of relational depth in psychotherapy: complementing both the RDI-R and other gradient-based measures of the quality of the therapeutic interaction. We aimed to develop a measure that would be short enough for repeated use, but with acceptable reliability and a single factor structure. We also wanted it to have convergent validity, shown by strong, and statistically significant correlations, at the top end of Cohen's "medium" correlation effect sizes ($.5 \geqslant r \geqslant .3$), with working alliance inventory (WAI-SR, Hatcher & Gillaspie, 2006) scores, and within Cohen's "large" correlation effect sizes ($.8 \geqslant r \geqslant .5$) with RDI-R2 scores—the

latest development of the RDI-R (Wiggins, personal communications, 6th July 2018). Here, we expected the three measures to show conceptual overlap but measure different constructs, with the RDI-R2 being nearer in theory to the RDFS than the WAI-SR. Within an initial exploration of criterion validity, we expected that RDFS scores would show only weak relationships with gender. As described in the relational depth literature, there have been no *a priori* hypotheses on possible correlations with gender. Therefore, we believe it would be evidence of poor criterion validity (as per Widaman, Little, Preacher, & Sawalani, 2011) were there strong gender effects. In addition, despite small differences in clients' and therapists' experiences of relational depth (e.g. Knox, 2011), we sought a measure with items that could be used with both clients and therapists and showing similar reliability, if not necessarily similar scores, when used in both groups.

Method

The research project was submitted for ethics consideration under the reference PSYC 15/ 164 in the Department of Psychology and was approved under the procedures of the University of Roehampton's Ethics Committee on 20.05.15

This measure development and validation study involved two phases: an initial qualitative process of item generation and refinement, followed by a quantitative selection of items. The process began with the generation of items and their selection and refinement through expert rating, and *Three-Step Test Interviews* focusing on the creation of items that would address frequency of moments of relational depth.

Phase 1: Qualitative process of item generation and refinement

Generating and Refining Items. We followed methods of deductive scale development, an approach considered most appropriate where there exists some theory about a construct (DeVellis, 2016; Hinkin, Tracey, & Enz, 1997). The first and final authors, and a fellow researcher, generated items reflecting relational depth, as defined in the literature (Knox, *et al.*, 2013; Mearns & Cooper, 2005; Wiggins, *et al.*, 2012). This resulted in a pool of 45 items. In

addition, two external postdoctoral researchers who had conducted research on relational depth were invited to generate items, resulting in an additional 83 items, taking the total to 128 items. Items were developed to follow the stem, "Over the course of therapy with my therapist/client, there were moments where…". For instance, "I felt I was being understood beyond my words," and "We were deeply connected to one another." We used this stem to capture experiences of relational depth as a threshold phenomenon. That is, we wanted to know how frequently, over the course of therapy, individuals had experienced moments of intense connection and engagement, characterized by mutual genuineness, warmth, and understanding.

This item pool was refined through a process of expert rating (Hinkin, 2005). A questionnaire with the 128 items was posted using the online data collection software Qualtrics. Seven experts rated each item on a four-point Likert scale: (1) not at all, (2) a little, (3) moderately, and (4) very well, on DeVellis's (2016) three criteria: (a) how well it matches the target definition, (b) how well formulated it is for participants to fill in, and (c) how well, overall, it is suited to the measure. Experts included two doctoral researchers conducting studies in the field of relational depth, three postdoctoral researchers with published research on relational depth, the first author (a doctoral student in relational depth), and the final author (Professor of Counselling Psychology with multiple published research on the topic of relational depth). An average score was calculated for each item, and those under a cutpoint of 3.0 were excluded from further analysis. This left 50 items. These were reviewed by the research team prior to the subsequent stage of refinement, in order to attend to any poor structure and highly similar wording in the items. Two items containing multiple clauses were removed, as were 12 items which were nearly synonymous with other items in the scale. This resulted in a 36-item scale taken forward for Three-Step Test Interviews.

Three-Step Test Interview. The Three-Step Test Interview (TSTI) is a cognitive pretesting method that assesses the quality of self-completion questionnaires (Hak, van der Veer, & Jansen, 2004). It has been used in several scale development studies (e.g., Hak, van der Veer,

& Ommundsen, 2006), and is considered a productive methodological tool in the refinement of scales (Busse & Ferri, 2003; Jansen & Hak, 2005). The initial stage consists in the *concurrent think aloud* process, where a participant completes the scale while saying aloud what he or she is thinking. The second stage, the *focused interview,* is then used to elicit a participant's feedback on any gaps the researcher observed in the thought process and noted during the initial stage. The last stage—the *semi-structured interview*—is then used to elicit a participant's reflections, definitions of terms, opinions, and experiences of completing the scale. Questions at this stage included: "How was it for you to complete the scale?", "What do you understand is the meaning of this term?", and "Can you give any feedback or recommendations on how we can improve this scale?"

Following Hak *et al.*'s (2004) standard procedures, we conducted interviews with four therapists and four clients (five females and three males, mean age = 49 years old (range 29–90), two from mixed ethnic backgrounds and six white Caucasian), achieving data saturation with regard to problematic patterns of responses. Following interviews, the first author developed an analysis aimed at refining the scale structure and content based on patterns in the test-takers' responses, theory on relational depth, and theory in scale development. The scale was amended and items were removed based on the identification of five problem clusters interpreted from the data: (a) *Double-barrelled*: where the wording of the item had multiple clauses (n = 2 excluded); (b) *Confusions*: where the meaning of a word was understood differently by different participants or caused uncertainty (n = 6 excluded); (c) *Repetitions*: where the same wording had been used in another or more other items (n = 6 excluded); (d) *Redundant:* where the wording was redundant with the opening stem or response scale (n = 2 excluded); (e) *Mundane*: where items reflected mundane characteristics of the therapeutic relationship rather than more intense levels of therapeutic engagement (as shown by most participants answering readily and selecting highest frequency items on the scale, n = 4 excluded). Following the TSTI process, 16 items remained and were taken forward for shortening.

Phase 2: Scale Reduction and Psychometric Exploration

Following the initial qualitative selection of the 16 items, data from four samples of participants, identifying as therapists and as clients, were used to drive shortening of the measure. To avoid problems of capitalization on chance, the four samples were merged into a client and a therapist sample, which were then each separated into two random and completely independent subsamples. The first random client and therapist subsamples, termed *shortening subsamples* hereafter, were used to select items for a very short form suitable for repeated use. The other randomly selected client and therapist subsamples, the *checking subsamples*, were used to check that the psychometric parameters from the shortening analyses replicated to exclude capitalization on chance having driven the selection of the final items.

Sample size. Worthington & Whittaker (2006) suggested that *n* should be at least 50 when seeking a single factor. This fits with the earlier simulation work of Guadagnoli and Velicer (1988) which showed that sample size is not a linear function of items but of the number of factors. We sought a single factor hence we needed to power to detect a more complex population factor structure had this been present. The work of Guadagnoli and Velicer (1988) showed $n \geq 100$ is sufficient for robust detection of a two-factor solution so an online survey was designed to produce a smallest subsample size (across the four subsamples) of 100.

We sought a sample that would be as representative as possible of current UK therapist and client populations. Inclusion criteria for clients were aged 18 years old and above, and currently in therapy or having attended therapy in the past. For therapists, it was qualified, including both accredited and non-accredited therapists, or still in training, with a minimum of one year practice. Based on a response rate for a prior study with similar recruiting methods (Wiggins, Elliott, & Cooper, 2012), we estimated that the response rate would be between 15% and 43% and the non-completion rate around 50%. We recruited through four methods.

Firstly, the survey was posted on social media sites that were accessed by people with an interest in psychotherapy and counseling, and we sent a link via email to the research team's

contacts. The survey was reposted on social media sites and emailed at four time points over the course of four months. Participants recruited through this method were primarily members of the UK psychology and counseling community. Secondly, we approached counseling psychologist trainees on eight doctoral courses in the UK. Five courses' administrators agreed for their course to be approached and forwarded the survey to their trainees across the three years. For one training course, the survey was also sent to clinical psychologist trainees. We estimated that the survey was sent to 360 trainee psychologists, who could choose to participate as therapists or clients. In total, 49 (14%) trainees completed the survey. Thirdly, we approached psychotherapists and counsellors listed on the British Association for Counselling and Psychotherapy (BACP) register, across 25 cities in the United Kingdom. Emails were sent individually to 1,246 therapists, 139 (11%) completed the survey. Finally, we approached clients at a charity in central London. Emails were sent, individually, to 92 clients. Of these, eight (9%) completed the survey.

Survey questionnaire.

The survey form used the online data collection software Qualtrics and started with detailed information about the study and the consent statement, then the questionnaires in the following order.

Socio-demographics questionnaire. This included questions on gender, age, ethnicity, whether the respondent was a mental health professional, and whether they were participating as a client or a therapist. The questionnaire included additional questions on the type of profession, level of training, number of years' post qualification, therapeutic orientation, and number of sessions of therapy.

Relational Depth Frequency Scale **(RDFS).** As described above, the RDFS items were designed to measure the temporal frequency of moments of relational depth over the course of therapy. This may be in a single therapy session, over several sessions, or in the overall therapy. Temporal frequency can be defined as the number of occurrences of an event over a period of

time. A frequency scaling was chosen over an "amount" or "value" scaling because our focus was on relational depth as specific moments of interaction. Despite small differences in clients' and therapists' experiences of relational depth (Knox & Cooper, 2010), we opted to maintain identical item wording for therapist and client versions, with minor variations in the introductory paragraph. These read as follows:

> Below is a list of items representing experiences people might have in therapy. Please think of your relationship with your client [for therapist version]/therapist [for client version] and select how frequently you have experienced the moments described in each item. There is no right or wrong answer, individuals relate differently.

Participants were then told that each item followed the statement, "Over the course of therapy with my client/therapist, there were moments where…" and asked to rate each item on a 5-point Likert scale, where 1 = *not at all*, 2 = *only occasionally*, 3 = *sometimes*, 4 = *often*, and 5 = *most or all of the time*. Items included: "I felt a clarity of perception between us" and "I experienced a meeting that was beyond words".

The Relational Depth Inventory (revised 2) **(RDI-R2).** The RDI is a 64-item measure that was developed to assess the depth of relating in particular moments in therapy (Wiggins, 2012, 2013). The RDI-R is a shortened 24-item inventory that has demonstrated good internal consistency in a pilot sample of 189 therapists and 152 clients (Cronbach's alpha = .93; Wiggins, 2012). The RDI-R has showed a large correlation with the WAI-SR in a sample of 150 clients ($r = .72$, 95% CI .63 to .79, $p < .01$), suggesting initial convergent validity (Wiggins, et. al., 2012). The RDI-R2 is the latest development of the RDI and includes two items that were added following Wiggins et al.'s (2012) Rasch analysis to target a wider range of respondents (Wiggins, personal communications, 6th July 2018). The RDI-R2 begins with a question asking respondents to describe, in their own words, an important event experienced during a therapy session. The respondents are then asked to rate how accurately the 26 items fit this experience,

using a 5-point Likert scale (1 = *not at all*, 2 = *slightly*, 3 = *somewhat*, 4 = *very much*, 5 = *completely*). Example items are "I felt a profound connection between my therapist and me," and "I felt I had lost all sense of time".

The Working Alliance Inventory–Short Revised (WAI-SR). The WAI-SR is one of two short versions of the 36-item WAI, which has demonstrated the strongest psychometric properties (Falkenström, Granström, & Holmqvist, 2013). The WAI-SR has two versions: a 10-item therapist version and a 12-item client version, sharing a 5-point Likert scale, and both giving scores on Bordin's (1979) three alliance subscales (Hatcher & Gillaspie, 2006). The WAI-SR has demonstrated good internal consistency in a sample of 88 German outpatients and 243 inpatients (α = .90, 95% CI [.87–.93]; α = .93, 95% CI [.92–.94], respectively); and good convergent validity with the Helping Alliance Questionnaire (Pearson's r = .71 for correlation between total scores, n = 237, 95% CI [.64 to .77]) (Munder, Wilmers, Leonhart, Linster, & Barth, 2010).

Participants. A total of 751 individuals accessed the online survey. Three (0.5%) declined to consent and are omitted from all analyses. In total, 192 participants (25.6%) exited the survey before they finished completing the demographics questionnaire and RDFS. This left 556 (74%) participants who completed all items on the demographics questionnaire and the relational depth frequency items. All these were included in the internal reliability and scale shortening process. Of these, 436 completed the RDI-R2 and 434 (58% of the 751, 78% of the 556) completed the full online survey, including all items of the WAI-SR, and could be included in the explorations of convergent and divergent validity. The largest proportion of participants (120; 16% of the 751 and 22% of the 556) exited the survey when reaching the qualitative question on the RDI-R2.

Of the 556 participants, 455 (82%) were females, 100 (18%) were males, and one identified as "other" (<1%). The mean age was 46 years old. Participants were predominantly from a white Caucasian background (90%), other represented ethnicities were mixed ethnicity

(3%), Asian/Asian British (3%), Other (3%), and Black/Black Caribbean/Black British (1%). There were 476 (85%) mental health professionals, of which 336 (71%) participated as therapists and 138 (29%) participated as clients; 82 (15%) participants were laypersons participating as clients. None of the comparisons in the client subsample between mental health professionals and laypeople reached statistical significance. Mental health professionals included 295 counsellors (62%), 217 psychotherapists (46%), 66 counselling psychologists (14%), 21 clinical psychologists (4%), and 38 other (8%). Of these, 363 were qualified (76%) and 110 were in training (24%). The majority of therapist participants were qualified (83%), most had over 10 years-experience (57%), 34% had between five and ten years-experience, and 24% had between one to five years' experience.

Subsample split. The sample completing the survey as therapists was larger than that completing it as clients (n = 336 and 220 respectively). This provided sufficient data to meet our criterion of n ≥ 100 in each of the four (shortening/checking * therapists/clients) subsamples. As scale shortening can be distorted by capitalization on chance (see, for instance, Widaman, et al., 2011)—that is, produce a shortened form only sound for the particular sample—our method used a random split of the therapist and the client samples to provide two completely separate samples, one used for the shortening, the other to check for replication of the psychometric parameters of the chosen shortening. The splits of the therapist subsample, and of the client subsample, into the shortening (n = 168 and n = 110 respectively) and checking subsamples (n = 168 and n = 110 respectively) were done by pseudorandom number generation.

The sample characteristics for the shortening and checking subsamples of therapists and clients are summarized in Table 1 (in Appendix). The only statistically significant difference between the therapist and client subsamples was on age, with means of 48 and 43 years old respectively (SDs = 11, W = 46626, p < .0005). As at least one statistically significant difference is likely given the number of tests, this first random split of the samples was used in the analyses reported below.

Scale shortening methods. The objective was a short measure that would not be prohibitive for use in routine practice, and would show respectable psychometric properties, particularly of internal reliability and of criterion validity. There is a tension in shortening a scale between over-focusing to maximize reliability versus losing reliability and hence, inevitably, also losing validity (e.g., Widaman et al., 2011). Over-focusing on reliability will inevitably lead to selection of more similar items and hence lose breadth of coverage of the phenomenon being measured, lowering true validity. Our *a priori* criteria for shortening were: (a) retaining acceptable internal reliability (Cronbach's α); (b) reasonable fit to a one dimensional model, defined as acceptable fit in single factor CFA without large modification indices for correlated residuals suggesting a higher order factor structure *and* acceptable fit to a single factor Rasch model; (c) plausible (medium to low end large effect size) correlations with WAI and RDI scores within both the client and therapist samples; (d) small or non-significant effects of gender; and finally (e) psychometric properties that were not radically different in the client and therapist samples so that, though these are clearly different viewpoints on the phenomena of interest, it would be possible to use the same scale for both groups. Shortening was stopped when continuing led to deterioration in Cronbach's alpha and no improvement on any other parameters in the analyses.

Two *a priori* decisions need further comment: the use of both CFA and Rasch analysis, and a particular point about the CFA. It is not common to use both CFA and Rasch analyses as they are often seen as rival methods. Other examples of using them appropriately for their complementarity include Cahill et al., (2012) and Mavranezouli, Brazier, Young, & Barkham, (2011) CFA treats item scores as arithmetic values that additively reflect the underlying dimension of interest matching how measures are generally scored using unit scores for levels of the items. Rasch analyses do not assume this and find the best fit of the data to a model in which the steps between response options on each item are thought to be reflections of the latent variable. This is a better fit to the psychological realities of questionnaire completion than CFA,

while the CFA model is a better fit to how questionnaires are scored than Rasch analysis. However, the Rasch analyses essentially check fit to the CFA assumption of a single underlying factor and of linearity: if the Rasch model shows markedly unequal steps between response levels for an item the CFA results, and conventional scoring, would be suspect. Our CFA is of the pure single factor model whereas many CFA analyses reported in the literature allow correlated item error terms to achieve high fit indices. Such analyses inevitably confound error and true variance so we did not allow these, preferring to accept that CFA fit indices might be less than ideal given the very short scale we sought, but knowing that interpretation of the model would be clear.

The process of shortening, starting with all 16 items, was an evaluation of misfit to the CFA and Rasch models with items removed which either showed most involvement in correlated residuals in the CFA (modification indices for correlated residuals over 10), or showed most uneven item response step locations and/or most differential item functioning between the therapist and client subsamples. When a final shortened scale was reached, the same Rasch and CFA analyses were conducted on the checking subsamples to assess the replicability of the parameters. As there are many statistics from each of the CFA and Rasch analyses of two separate shortening samples for each shortening step, these are not reported here. Full details are available on request.

Analyses of the full client and therapist samples and the total dataset were planned *a priori*, in order to achieve maximum precision of estimation subject to the checking analyses refuting that the choice of items had been achieved by capitalization on chance.

Most analyses were conducted using R, however, as a number of tests conventional in Rasch analyses are provided by the Winsteps program (Linacre, 2017) but not by the R packages, the complete sample data were analyzed using Winsteps to report these values. All scale development analyses were conducted using R version 3.4.4 (R Core Team, 2018). We used the ltm and lordif packages to compute Lord's test of differential item functioning

(between therapist and client subsamples) (Choi, 2016; Rizopoulos, 2006); and the lavaan package to explore the CFA of the data (Rosseel, 2012). In R, the Generalized Partial Credit Model (GPCM) was used initially without constraints to check for poor severity step separation.

Results

Psychometric Scale Shortening

The shortening steps are summarized in Table 2. Based on analyses of the shortening subsamples, items 1, 2 and 10 were first removed as they were involved in most of the high correlated error modification indices (MIs) in the CFA. Item 15 was removed because it showed strong differential item functioning (DIF) in the Rasch analyses. That is, the item behaved differently from the others in the analyses of the client as against the therapist subsamples.

From the 12-item stage, item 14 was removed as it was involved in most of the high correlated error MIs in the CFA, followed by items 5, 11, and 12 which showed poor fit in Rasch analysis and generally significantly correlated error terms with high MIs. Item 6 was then removed because of poor fit in Rasch analysis. Finally, item 3 was removed as it was involved in a high MI (with item 9). At this point further shortening would result in unacceptably low Cronbach alpha values.

Following this process, the final items were inspected by the authors to review their alignment with "relational depth", as conceptually described. As intended, most items depicted qualities of intense moments of interpersonal encounter, for instance "We were deeply connected to one another" and "I felt like we were totally in-the-moment together". One item, "it felt like a shared experience", described a less intense moment of encounter. However, we decided to retain it in the final scale as it remained consistent with descriptions of relational depth, to maintain a range of item "difficulties", and because the item would increase the face validity of the measure with a wider range of respondents. The final six items which constituted the RDFS were items 4, 7, 8, 9, 13, and 16 (marked in Table 2 with asterisks).

Table 2

Items Taken Forward for Psychometric Exploration and reasons for removal

Items	Reason for removal
01: I experienced an intense connection with him/her 02: I experienced a very profound engagement with her/him 10: I felt we connected on a human level 15: I felt we truly acknowledged each other at a very deep level	Involved in most correlated error terms with high MIs (MIs over 10.0) Marked DIF in client/therapist comparison
03: I experienced what felt like true mutuality	Involved (with item 9) in one high MI (21.3))
04: We were deeply connected to one another*	
05: I felt a clarity of perception between us	Poor fit in Rasch analysis
11: I experienced a deep sense of encounter 12: I experienced a meeting that was beyond words	Involved in most correlated error terms with high MIs (MIs over 10.0)
06: I felt intensely present with him/her	Poor fit in Rasch analysis
07: We were immersed in the present moment* 08: There was a deep understanding between us* 09: It felt like a shared experience* 13: I felt like we were totally in-the-moment together*	
14: I felt we were really close to each other	Involved in most correlated error terms with high MIs (over 10.0)
16: I felt we were completely open with each other*	

Note. * Indicates inclusion in final RDFS

Confirmatory Factor Analysis Fit Statistics

Confirmatory Factor Analysis fit statistics for a six item, unidimensional measure for all seven samples are given in Table 3. As can be seen here, CFA of the samples and the full dataset showed statistically significantly imperfect fit ($p < .05$) for all but the replication sample of clients ($p = .05$), RMSEA values indicated that the fit to the model is not perfect. However, all the CFI values were excellent, as were the SRMR values. The loadings for the six items in the seven samples are shown in table 4. The internal reliability and most other psychometric parameters in the checking sample were better than in the shortening sample, which rules out capitalization on chance. Analyses of the combined data have the tightest confidence intervals and greatest generalizability by virtue of the double sample size. All loadings were strong. There were some differences between the shortening and checking loadings but no consistent pattern for the loadings to be worse in the checking samples. There were stable differences between the

items in their loadings within samples and loadings in the client samples were higher than in the therapist samples.

Table 3
Confirmatory Factor Analysis Fit Statistics: All Subsamples and Full Dataset

Sample	n	Chisq	Df	P	CFI	RMSEA	SRMR
Exp. Th.	168	27.5	9	.001	.94	.11	.04
Conf. Th.	168	17.3	9	.045	.98	.07	.03
Exp. Cl.	110	18.2	9	.03	.98	.10	.03
Conf. Cl.	110	17.2	9	.05	.98	.09	.03
Th.	336	32.6	9	.0002	.97	.09	.03
Cl.	220	21.4	9	.01	.99	.08	.02
All	556	38.8	9	<.0001	.98	.08	.02

Note. Exp. Th. = Shortening therapist subsample, Conf. Th. = Checking therapist subsample, Exp. Cl. = Shortening client subsample, Exp. Th. = Checking client subsample

Table 4
Item loading table

	T(S)	T(c)	T(comb)	C(S)	C(C)	C(comb)	All
RDFST04	0.76	0.71	0.73	1.05	0.98	1.04	0.88
RDFST07	0.43	0.66	0.55	0.90	0.88	0.89	0.71
RDFST08	0.66	0.75	0.71	1.01	1.07	1.05	0.87
RDFST09	0.63	0.59	0.62	1.03	1.09	1.08	0.85
RDFST13	0.72	0.72	0.71	1.05	1.05	1.06	0.88
RDFST16	0.61	0.59	0.60	0.86	0.94	0.91	0.74

Internal consistency and Distribution Parameters for RDFS

For therapists, the internal consistency (Cronbach alpha) of the final, six item RDFS was .85 (95% CI .82–.88) and for clients it was .93 (95% CI .91–.94) (Table 5). The mean score (between a maximum of 30 and a minimum of 6) was 21.3 for therapists (95% CI 20.9–21.8) and 19.9 (95% CI 19.1–20.7) for clients. Internal consistency and distribution parameters for the RDFS in all samples and subsamples are summarized in Table 5.

Table 5
Reliability and Distribution Parameters for RDFS in All Samples and Subsamples

Sample	n	Mean (95% CI)	SD (95% CI)	Range	Cronbach alpha (95% CI)
Exp. Th.	168	21.2 (20.6 to 21.8)	4.2 (3.8 to 4.7)	8 to 30	.83 (.78 to .88)
Conf. Th.	168	21.5 (20.8 to 22.1)	4.3 (3.9 to 4.9)	7 to 30	.87 (.84 to .91)
Exp. Cl.	110	18.9 (17.8 to 20.1)	6.2 (5.6 to 6.9)	6 to 30	.92 (.90 to .95)
Conf. Cl.	110	20.9 (19.7 to 22.1)	6.3 (5.6 to 7.0)	6 to 30	.93 (.91 to .96)
Th.	336	21.3 (20.9 to 21.8)	4.2 (3.9 to 4.6)	6 to 30	.85 (.82 to .88)
Cl.	220	19.9 (19.1 to 20.7)	6.3 (5.8 to 6.8)	7 to 30	.93 (.91 to .94)
All	556	20.8 (20.4 to 21.2)	5.2 (4.9 to 5.5)	6 to 30	.90 (.88 to .91)

Note. Exp. Th. = Shortening therapist subsample, Conf. Th. = Checking therapist subsample, Exp. Cl. = Shortening client subsample, Exp. Th. = Checking client subsample

Rasch Analysis

Winsteps standardized residuals had mean 0.00 and standard deviation 0.97 (marked deviations from zero and one indicate misfit to the Rasch model). The log-likelihood test of fit gave $\chi^2 = 5949.1$ with approximately 6005 ±6 *df*, $p = .69$, suggesting the fit to the model is not statistically significantly imperfect. Similarly, the global Root-Mean-Square Residual (RMSR) was .61 with expected value: .62.

In the light of these fit indicators, the specific Rasch statistics for the scale can be considered. For the items in the full dataset the separation based on the "worst case: real" analysis in Winsteps was 8.08 and the "model" analysis was 8.32, these values equate to item reliability values of .98 and .99. The "model" analysis assumes that all misfit was down to random sampling vagaries in this dataset and that the model will replicate in all future use of the measure. The likely generalizable value is going to lie between the two. For person separation, the "real" separation was 2.60 and "model" separation was 2.89 equating to person reliabilities of .87 and .89 respectively. In addition, visual inspection of the item response curves from the Rasch analyses showed all the response steps fairly evenly spaced and none out of their logical order.

Criterion validity

Convergent validity. Inspection of scatterplots with linear regression lines showed no marked deviation from linear relationships between the RDFS scores and any of the WAI-SR and RDI-R2 scores. The convergent validity correlations with the WAI-SR and RDI-R2 scores in the client and therapist samples are shown in Tables 6 and 7 (in Appendix), respectively. For clients, Pearson correlations with the RDFS were .71 for the total WAI-SR scale (.40 for WAI-SR Goals, .69 for WAI-SR Tasks, and .76 for WAI-SR Bond), and .77 for the RDI-R2. All correlations are significant at $p < 4*10^{-9}$. For therapists, Pearson correlations with the RDFS were .52 for the total WAI-SR scale (.40 for WAI-SR Goals, .49 for WAI-SR Tasks, and .46 for WAI-SR Bond), and .56 for the RDI-R2. All correlations were significant with $p < 0.001$.

Effects of Gender. This effect was small and not statistically significant and was of a different direction in the client and the therapist subsamples, with a higher mean for the women than the men in the therapists' subsample and *vice versa* in the clients' (Table 8).

Table 8
Criterion Validity: Gender Effects on RDFS Scores

Sample	Female mean	Female SD	Male mean	Male SD	Cohen's *d* (95% CI)
Therapists	21.5	4.3	20.7	3.9	0.18 (-0.05 to 0.45)
Clients	19.8	6.4	20.4	6.0	-0.11 (-0.43 to 0.21)
All	20.8	5.3	20.6	4.8	0.04 (-0.17 to 0.26)

Effect of Therapist vs. Client Status of Participant. As suggested by the means in Table 8, and shown in the boxplot in Figure 1, there was a small but statistically significant tendency for therapists to give higher RDFS scores than clients (Mann-Whitney test $W = 40{,}809$, $p = .03$, Cohen's $d = 0.28$, 95% CI [0.11–0.46].

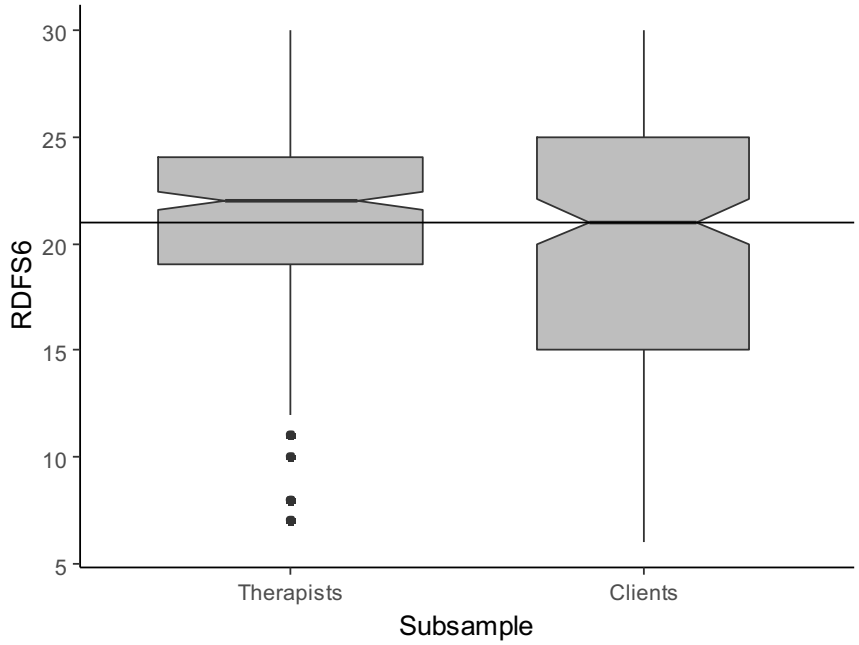

Figure 1. Boxplot of RDFS scores: clients vs. therapists

Discussion

This study demonstrated the development of a brief self-report measure of relational depth, which can complement measures of the quality of the therapeutic interaction by focusing on in-depth relating as a "threshold" phenomenon. Initial processes of item refinement, followed by quantitative scale shortening led to a six item RDFS whose psychometric parameters

replicated remarkably well in an independent checking subsample. This supported that it had not been capitalizing on chance sampling variations. The measure had good internal consistency considering its brevity. Rasch analyses showed both excellent fit between data and model, excellent reliability, even spacing of response stepping points, and no differential item functioning across the client and therapist groups of participants. The statistical fit to a strict single factor model was by no means perfect. This finding was unsurprising with only six items and five response levels in the scale, implying the data are not continuous nor likely to show correlations unaffected by the short ordinal scale. On balance, these findings support the use of the RDFS as a reliable measure of the frequency of moments of profound contact and engagement in psychotherapy.

With regards to criterion validity, the RDFS showed a strong association with the RDI-R2 in both the client and therapist samples, with correlations that were within the predicted range. This suggests that there is a strong relationship between the frequency with which individuals experience moments of relational depth in therapy (RDFS), and the depth of relating they ascribe to a single self-identified important moment in therapy (RDI-R2; Wiggins 2012).

Associations between RDFS and WAI-SR scores were overall lower, but showing similar differences between client and therapist-completed scales. Correlations for all three therapist-completed subscales, and the client-completed goal subscale were within the predicted range, and the therapist-completed total scale was just above it. This association is comparable to the convergence of the Real Relationship Inventory with the WAI-SR in a therapist sample (r = .47; Gelso et al, 2005), and suggests that for therapists, there is a moderate association between the quality of the working alliance and the frequency with which moments of relational depth are experienced. It may also support the hypothesis that relational depth occurs more often when there is already a good working alliance.

As with the RDI-R2, correlations of the RDFS with the client-completed WAI-SR total scale, and the Tasks and Bond subscale were considerably larger and outside the predicted

range. This suggests that, for clients, there could be more overlap between the quality of the alliance (in terms of bond and task agreement), and the frequency with which moments of relational depth are experienced. This can also support that clients are likely to need the safety of an even stronger alliance in order to experience relational depth in therapy.

While the RDFS was designed to be usable for both therapists and clients—and this was supported in our results—psychometric parameters were, nonetheless, not expected to be the same in therapists and clients. Higher internal reliability in the client sample may have contributed—through "attenuation" (meaning that correlations between two less than perfectly reliable measures of correlated phenomena will always be lower than the correlations between those phenomena had they been measured with high reliability) —to higher correlations for clients with the RDI-R2 and WAI-SR (Fisher, 2014). Another plausible explanation of this difference is that therapists, by their training and experience, have a more nuanced understanding of the therapy relationship, and are better able to discern different components within it. This interpretation is supported by the similar trend in correlations between the RDI-R2 and the WAI-SR in our client and therapist samples. Further research is needed to explain these differences: particularly data collected in clinical settings from pairs of therapists and clients in ongoing therapies (e.g., Rubel et al., 2018).

While correlations for the client group were large, the largest correlation for the bond component suggests that 42.2% of the variance in clients' RDFS scores was not accounted for by their WAI-SR Bond scores. This supports the argument that the RDFS is measuring a separable dimension of the quality of the therapeutic interaction in even the most correlated of the RDI-R2 and WAI-SR scores. This also suggests that the RDFS can be used to complement the RDI-R2's emphasis and focus: the RDFS assesses frequency of relational depth over an episode of therapy, as opposed to the depth of specific moments. In addition, while the RDI-R2 functions to elicit vivid memories of such moments, the RDFS may minimize participant fatigue

in that it has only six items. As such, it may be more appropriate for regular use in clinical practice and in research designs requiring repeated assessment.

Finally, we had sought and found a measure showing a low effect of gender on mean scores. As described in our nomological net, there were no strong *a priori* hypotheses in the literature on effects of gender. Thus, a strong gender effect would show poor criterion validity (Widaman et al., 2011). Therefore, this finding supports initial evidence in the exploration of criterion validity.

The study had several limitations. While a range of participants and samples were targeted, our estimated response rates were only 9–14%. In addition, we did not control for who responded to the survey, such that this may have been influenced by a self-selecting bias. That is, the title of the project called for participation in a study on the quality of the therapeutic relationship, and this could have attracted participants who were relationally-oriented. In addition, our results may have been influenced by a social desirability bias, as therapists may have attempted to portray their therapeutic relationship in positive terms. However, this was controlled for when looking at reliability in the client sample, as clients would have less reason to present themselves in this way. Another limitation of this study was that 82% of participants were females, and fewer than 10% of participants came from an ethnic minority background. For our therapist sample, these proportions are not unusual in the psychotherapy workforce in the UK and the US (American Psychological Association, 2015; Orlinsky & Rønnestad, 2005); but future research would benefit from the validation of the RDFS in more balanced and culturally diverse samples.

Another important limitation of our study was that approximately 62% of client participants were also mental health professionals. This meant that responses to the items may not have been representative of a true population of psychotherapy clients. The RDFS, therefore, requires validation with a client sample that is more truly representative of psychotherapy users. A study carried out in a clinical setting would also enable more control over its procedures and

would allow for the clustering of therapist and client data. Furthermore, data collection at multiple time points in a clinical setting would enable further exploration of the sensitivity to change of the RDFS in therapeutic processes over time. Further research is also needed to test the construct validity of the RDFS against other measures, in particular the Real Relationship Inventory (Kelley et al., 2010), and an indicator of mutuality using the Barrett-Lennard Relationship Inventory (Barrett-Lennard, 2015). Finally, the RDFS needs formal testing of divergent validity, for instance with a measure of social desirability.

With initial validity and reliability in an online sample, the RDFS—a measure of a unique component of the therapy relationship, moments of relational depth—can be used to further understand the processes and mechanisms by which particular qualities of the therapeutic interaction can lead to improved treatment outcomes. As a brief measure, the RDFS has the potential for future use in cross-lagged panel studies and other methodologically advanced designs to evaluate the direction of causality in any identified associations. In such a study, it will be essential to assess the sensitivity of the RDFS to change over the course of treatment, as well as to use other relationship measures, such as the WAI-SR and BLRI, to examine whether moments of relational depth make a unique contribution to outcomes. Such research will help to identify whether the greatest contribution to outcomes comes from the general quality of the therapeutic interaction; or by the frequency of particularly intense moments of therapeutic engagement; or, indeed, by the depth of relating in single moments in therapy, as assessed by the RDI-R2 (Wiggins, 2012). With respect to these latter two possibilities, moments of relational depth may function like "significant therapy events" (Elliott & Shapiro, 1992), or akin to episodes of "sudden change" (Tang & DeRubeis, 1999), making qualitatively distinct contributions to therapeutic improvement. A measure of the frequency of relational depth could also be used to extend existing research on mutuality of the core conditions, or congruence of bond associations, across the therapeutic relationship and its association to outcomes (e.g., Rubel et al., 2018; Murphy & Cramer, 2014).

If relational depth is found to make a distinctive contribution to outcomes, this may have important implications for practice. Mearns and Cooper (2018) identified a number of strategies by which therapists may be trained to deepen their levels of relating. This includes the development of practical skills such as "holistic listening", and self-reflective work on identifying one's own "chronic strategies of disconnection" (Cooper & Knox, 2018). It also includes a broader developmental agenda, such as developing the capacity to draw on "existential touchstones". There is initial evidence suggesting that therapists can learn to relate at depth and undergo relational depth training if they have a capacity or a desire for it (Tangen & Cashwell, 2016).

The RDFS measure is licensed under the Creative Commons Attribution-NoDerivatives 4.0 International licence. It should be used cautiously until larger and more diverse and representative referential samples are collected. Nevertheless, as Zilcha-Mano (2017) states, "more exciting work [on the alliance–outcome relationship] awaits us yet…exploring new perspectives on the complexity of therapeutic change" (p.322). Our hope is that the RDFS will contribute to this process.

References

American Psychological Association. (2015). *Demographics of the US psychology workforce: Findings from the American Community Survey*. Washington, DC: APA. Retrieved from: http://www.apa.org/workforce/publications/13-demographics/index.aspx

Barrett-Lennard, G. T. (2015). *The Relationship Inventory: A Complete Resource and Guide*. London: John Wiley.

Bordin, E. (1979). The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy*, 16(3), 252-260. doi. 10.1037/h0085885

Busse, A., & Ferri, R. B. (2003). Methodological reflections on a three-step-design combining observation, stimulated recall and interview. *Zentralblatt für Didaktik der Mathematik*, *35*(6), 257-264. doi: 10.1007/BF02656690

Cahill, J., Stiles, W. B., Barkham, M., Hardy, G. E., Stone, G., Agnew-Davies, R., & Unsworth, G. (2012). Two short forms of the Agnew Relationship Measure: The ARM-5 and ARM-12. *Psychotherapy Research*, *22*(3), 241-255.

Castonguay, L. G., Constantino, M. J., & Holtforth, M. G. (2006). The working alliance: Where are we and where should we go?. *Psychotherapy: Theory, Research, Practice, Training*, *43*(3), 271. doi: 10.1037/0033-3204.43.3.271

Choi, S. W. (2016). lordif: Logistic Ordinal Regression Differential Item Functioning using IRT. R package version 0.3-3. (With contributions from Gibbons, L. E. and Crane, P. K.): ftp://200.236.31.3/CRAN/web/packages/lordif/lordif.pdf

Cooper, M. (2005). Therapists' experiences of relational depth: A qualitative interview study. Counselling and Psychotherapy Research, 5(2), 87-95. doi: 10.1080/17441690500211130

Cooper, M. (2013). Experiencing relational depth in therapy: What we know so far. In R. Knox, D. Murphy, S. Wiggins & M. Cooper (Eds.), *Relational depth: New perspectives and developments* (pp. 62-76). Basingstoke: Palgrave.

Cooper, M., & Knox, R. (2018). Therapists' self-reported chronic strategies of disconnection in everyday life and in counselling and psychotherapy: an exploratory study. *British Journal of Guidance & Counselling, 46*(2), 185-200. doi: 10.1080/03069885.2017.1343457

DeVellis, R. F. (2016). *Scale development: Theory and applications* (Vol. 26). Sage publications.

Elliott, R., & Shapiro, D. A. (1992). Client and therapist as analysts of significant events. In S. G. Toukmanian & D. L. Rennie (Eds.), Psychotherapy Process Research: Paradigmatic and Narrative Approaches (pp. 163-186). London: Sage.

Evans, C., Connell, J., Barkham, M., Margison, F., McGrath, G., Mellor-Clark, J., & Audin, K. (2002). Towards a standardised brief outcome measure: Psychometric properties and utility of the CORE–OM. *The British Journal of Psychiatry, 180*(1), 51-60. doi:10.1192/bjp.180.1.51

Falkenström F., Granström F., Holmqvist R. (2013). Therapeutic alliance predicts symptomatic improvement session by session. *Journal of Counseling Psychology*, 60, 317-328. doi. 10.1037/a0032258

Falkenström, F., Kuria, M., Othieno, C., & Kumar, M. (2018). Working alliance predicts symptomatic improvement in public hospital-delivered psychotherapy in Nairobi, Kenya. Journal of consulting and clinical psychology. doi: 10.1192/S2056474000001938

Fisher, C. R. (2014) A Pedagogic Demonstration of Attenuation of Correlation Due to Measurement Error, Spreadsheets in Education (eJSiE): Vol. 7: Iss. 1, Article 4. Available at: http://epublications.bond.edu.au/ejsie/vol7/iss1/4

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

Fuertes, J. N., Mislowack, A., Brown, S., Gur-Arie, S., Wilkinson, S., & Gelso, C. J. (2007). Correlates of the real relationship in psychotherapy: A study of dyads. *Psychotherapy Research*, *17*(4), 423-430. doi: 10.1080/10503300600789189

Gelso, C. J. (2004, June). A theory of real relationship in psychotherapy. In *International Conference of the Society for Psychotherapy Research, Rome*.

Gelso, C. J., Kivlighan, D. M. J., & Markin, R. D. (2018). The real relationship and its role in psychotherapy outcome: A meta-analysis. *Psychotherapy, 55*(4), 434-445. doi: 10.1037/pst0000183

Gelso, C. J., Kelley, F. A., Fuertes, J. N., Marmarosh, C., Holmes, S. E., Costa, C., & Hancock, G. R. (2005). Measuring the Real Relationship in Psychotherapy: Initial Validation of the Therapist Form. *Journal of Counseling Psychology*, *52*(4), 640-649. doi. 10.1037/0022-0167.52.4.640

Guadagnoli, E., & Velicer, W. F. (1988). Relation to sample size to the stability of component patterns. *Psychological Bulletin*, *103*(2), 265-275. doi: 10.1037/0033-2909.103.2.265

Hak, T., Van der Veer, K., & Jansen, H. (2004). *The Three-Step Test-Interview (TSTI): An observational instrument for pretesting self-completion questionnaires*. Rotterdam: Erasmus Research Institute of Management. doi: 10.18148/srm/2008.v2i3.1669

Hak, T., van der Veer, K., & Ommundsen, R. (2006). An Application of the Three-Step Test Interview (TSTI): A validation study of the Dutch and Norwegian versions of the 'Illegal Aliens Scale'. *International Journal of Social Research Methodology*, 9(3), 215-227. doi. 10.1080/13645570600656264

Hatcher, R. L., & Gillaspy, J. A. (2006). Development and validation of a revised short version of the Working Alliance Inventory. *Psychotherapy Research*, *16*(1), 12-25. doi: 10.1080/10503300500352500

Hermida, R. (2015). The problem of allowing correlated errors in structural equation modeling: Concerns and considerations. Computational Methods in Social Sciences, 3, 5–17. DOI: 10.1177/2056305115610141

Hinkin, T. R. (2005). Scale development principles and practices. In R. A. Swanson & E. F. Holton III (eds.) *Research in organizations: Foundations and methods of inquiry*, (pp. 161-179). San Francisco: Berrett-Koehler.

Hinkin, T. R., Tracey, J. B., & Enz, C. A. (1997). *Scale construction: Developing reliable and valid measurement instruments* [Electronic version]. Retrieved 22nd April 2018 from Cornell University, School of Hotel Administration site: http://scholarship.sha.cornell.edu/articles/613

Horvarth, A. O., & Bedi, R. P. (2002). The alliance. In J. C. Norcross (Ed.), *Psychotherapy Relationships that Work: Therapist Contributions and Responsiveness to Patients* (pp. 37-69). Oxford: Oxford University Press.

Jansen, H., & Hak, T. (2005). The productivity of the three-step test-interview (TSTI) compared to an expert review of a self-administered questionnaire on alcohol consumption. *Journal of Official Statistics*, *21*(1), 103-120. DOI: 10.18148/srm/2008.v2i3.1669

Kelley, F. A., Gelso, C. J., Fuertes, J. N., Marmarosh, C., & Lanier, S. H. (2010). The Real Relationship Inventory: Development and psychometric investigation of the client form. *Psychotherapy: Theory, Research, Practice, Training*, *47*(4), 540. DOI: 10.1037/a0022082

Knox, R. (2011). *Clients' experiences of relational depth.* (PhD dissertation), Glasgow: University of Strathclyde.

Knox, R., & Cooper, M. (2010). Relationship qualities that are associated with moments of relational depth: The client's perspective. *Person-Centered and Experiential Psychotherapies*, *9*(3), 236-256. doi. 10.1080/14779757.2010.9689069

Knox, R., Murphy, D., Wiggins, S., & Cooper, M. (Eds.). (2013). *Relational depth: New perspectives and developments*. Basingstoke: Palgrave.

Leung, J. (2008). *A quantitative online study exploring the factors associated with the experience and perception of relational depth.* (DPsych. dissertation), Glasgow: University of Strathclyde.

Linacre, J.M. (2017) A user's guide to WINSTEPS MINISTEP Rasch-model computer programs. Version 4.0.0. Winsteps.com. ISBN 0-941938-03-4.

Mavranezouli, I., Brazier, J. E., Young, T. A., & Barkham, M. (2011). Using Rasch analysis to form plausible health states amenable to valuation: the development of CORE-6D from a measure of common mental health problems (CORE-OM). *Quality of Life Research*, *20*(3), 321-333. Retrieved from: https://link.springer.com/article/10.1007/s11136-010-9768-4

Mearns, D., & Cooper, M. (2018). *Working at Relational Depth in Counselling and Psychotherapy* (2nd ed.). London: Sage.

Mearns, D., & Cooper, M. (2005). *Working at Relational Depth in Counselling and Psychotherapy*. London: Sage.

Munder, T., Wilmers, F., Leonhart, R., Linster, H. W., & Barth, J. (2010). Working Alliance Inventory-Short Revised (WAI-SR): psychometric properties in outpatients and inpatients. *Clinical Psychology & Psychotherapy*, *17*(3), 231-239. doi. 10.1002/cpp.658

Murphy, D., & Cramer, D. (2014). Mutuality of Rogers's therapeutic conditions and treatment progress in the first three psychotherapy sessions. *Psychotherapy Research*, *24*(6), 651-661. doi: 10.1080/10503307.2013.874051

Norcross, J. C. & Lambert, M. J. (2018). Psychotherapy relationships that work III. *Psychotherapy*, *55* (4), 303-315. doi: 10.1037/pst0000193.

Orlinsky, D. E., Rønnestad, M. H., & Collaborative Research Network of the Society for Psychotherapy Research. (2005). How psychotherapists develop: A study of therapeutic work and professional growth. Washington, DC, US: American Psychological Association. doi: 10.1037/11157-000

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rizopoulos, D. (2006). ltm: An R package for Latent Variable Modelling and Item Response Theory Analyses, *Journal of Statistical Software*, *17*(5), 1-25. doi: 10.18637/jss.v017.i05

Rogers, C.R. (1957). The necessary and sufficient conditions of therapeutic personality change. *Journal of Consulting Psychology*, *21*, 95-103.doi: 10.1037/h0045357

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(2), 1-36. doi: 10.18637/jss.v048.i02

Rubel, J. A., Bar-Kalifa, E., Atzil-Slonim, D., Schmidt, S., & Lutz, W. (2018). Congruence of therapeutic bond perceptions and its relation to treatment outcome: Within-and between-dyad effects.

Stern, D. N. (2004). *The Present Moment in Psychotherapy and Everyday Life*. New York: W. W. Norton.

Tang, T. Z., & DeRubeis, R. J. (1999). Sudden gains and critical sessions in cognitive-behavioral therapy for depression. Journal of Consulting and Clinical Psychology, 67(6), 894-904: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.488.1236&rep=rep1&type=pdf

Tangen, J. L., & Cashwell, C. S. (2016). Touchstones of connection: A concept mapping study of counselor factors that contribute to relational depth. *The Journal of Humanistic Counseling*, *55*(1), 20-36.

Widaman, K. F., Little, T. D., Preacher, K. J., & Sawalani, G. M. (2011). On creating and using short forms of scales in secondary research. In K. H. Trzesniewski, M. B. Donnellan, & R. E. Lucas (Eds.), *Secondary data analysis: An introduction for psychologists* (pp. 39-61). Washington, DC, US: APA. doi: 10.1037/12350-003

Wiggins, S. (2012). *Development and testing of a measure of relational depth in counselling and psychotherapy.* (PhD thesis), University of Strathclyde, Glasgow. Retrieved 24/4/2018 from http://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.576407

Wiggins, S., Elliott, R., & Cooper, M. (2012). The prevalence and characteristics of relational depth events in psychotherapy. *Psychotherapy Research*, *22*(2), 139-158. doi. 10.1080/10503307.2011.629635

Worthington, R. L., & Whittaker, T. A. (2006). Scale development research: A content analysis and recommendations for best practices. *The Counseling Psychologist*, *34*(6), 806-838. doi: 10.1177/0011000006288127

Zilcha-Mano, S. (2017). Is the alliance really therapeutic? Revisiting this question in light of recent methodological advances. *American Psychologist, 72*(4), 311-325. doi: 10.1037/a0040435

Zilcha-Mano, S., Dinger, U., McCarthy, K. S., & Barber, J. P. (2014). Does alliance predict symptoms throughout treatment, or is it the other way around? Journal of Consulting and Clinical Psychology, 82(6), 931-935. doi: 10.1037/a0035141

Table 1

Sample and Subsample Characteristics

	All	Therapists	Clients		Shortening and checking subsamples			
					Shortening		Checking	
	N = 556	N = 336	Mental health professionals N = 138	Laypersons N=82	Therapists N=168	Clients N=110	Therapists N=168	Clients N=110
Age (mean, SD)	46 (12)	48 (11)	43 (11)	43 (13)	48 (11)	45 (12)	49 (11)	42 (12)
Gender (N, %)								
Female	455 (82%)	274 (82%)	111 (80%)	70 (85%)	133 (79%)	86 (78%)	141 (84%)	95 (86%)
Male	100 (18%)	62 (18%)	27 (20%)	11 (13%)	35 (21%)	23 (21%)	27 (16%)	15 (14%)
Other	1 (0%)	0	0	1 (1%)	0	1 (1%)	0	0
Ethnicity								
White	502 (90%)	305 (91%)	124 (90%)	73 (89%)	151 (90%)	100 (91%)	154 (92%)	97 (88%)
Black/African/Caribbean	6 (1%)	4 (1%)	2 (1%)	0	2 (1%)	1 (1%)	2 (1%)	1 (1%)
Mixed ethnicity	16 (3%)	7 (2%)	7 (5%)	2 (2%)	4 (2%)	3 (3%)	3 (2%)	6 (5%)
Asian/Asian British	18 (3%)	15 (4%)	1 (1%)	2 (2%)	7 (4%)	3 (3%)	8 (5%)	0
Other	14 (3%)	5 (1%)	4 (3%)	5 (6%)	4 (2%)	3 (3%)	1 (1%)	6 (5%)
Profession*								
Counsellor	295 (62%)	216 (64%)	79 (57%)	-	109 (65%)	45 (41%)	107 (64%)	34 (31%)
Psychotherapist	217 (46%)	154 (46%)	63 (46%)	-	82 (49%)	27 (25%)	72 (43%)	36 (33%)
Clinical Psychologist	21 (4%)	15 (4%)	6 (4%)	-	8 (5%)	5 (5%)	7 (4%)	1 (1%)
Counselling Psychologist	66 (14%)	45 (13%)	21 (15%)	-	20 (12%)	9 (8%)	25 (15%)	12 (11%)
Other	38 (8%)	27 (8%)	11 (8%)	-	15 (9%)	5 (5%)	12 (7%)	6 (5%)
Level								
In training	110 (24%)	57 (17%)	53 (38%)	-	29 (17%)	25 (37%)	28 (17%)	28 (40%)
Qualified	363 (76%)	278 (83%)	85 (62%)	-	138 (83%)	43 (63%)	140 (83%)	42 (60%)
Years Post-qualification								
Less than one year	31 (8%)	24 (9%)	7 (8%)	-	11 (8%)	3 (7%)	13 (9%)	4 (10%)
1-5 years	129 (35%)	94 (34%)	35 (41%)	-	44 (32%)	17 (40%)	50 (36%)	18 (43%)
5-10 years	84 (23%)	68 (24%)	16 (19%)	-	24 (25%)	8 (19%)	34 (24%)	8 (19%)
10-20 years	79 (22%)	59 (21%)	20 (24%)	-	29 (21%)	12 (28%)	30 (21%)	8 (19%)
Over 20 years	40 (11%)	33 (12%)	7 (8%)	-	20 (14%)	3 (7%)	13 (9%)	4 (10%)
Duration of therapy								
Less than 6 sessions		33 (10%)	4 (3%)	3 (4%)	12 (7%)	4 (4%)	21 (12%)	3 (3%)
6-24 sessions		107 (32%)	25 (18%)	9 (11%)	51 (30%)	16 (15%)	56 (33%)	18 (16%)
Over 24 sessions		133 (40%)	53 (38%)	27 (33%)	75 (45%)	35 (32%)	58 (35%)	45 (41%)

Therapy has ended	63 (19%)	56 (41%)	43 (52%)	30 (18%)	55 (50%)	33 (20%)	44 (40%)
Therapeutic orientation*							
Cognitive-Behavioural	56 (17%)	5 (4%)	17 (21%)		14 (13%)		8 (7%)
Psychodynamic	71 (21%)	20 (14%)	9 (11%)		16 (15%)		13 (12%)
Person-centred	144 (43%)	24 (17%)	14 (17%)		20 (18%)		18 (16%)
Integrative	167 (50%)	44 (32%)	15 (18%)		28 (25%)		31 (28%)
Psychoanalytic	16 (5%)	16 (12%)	6 (7%)		11 (10%)		11 (10%)
I don't know	-	2 (1%)	11 (13%)		7 (6%)		6 (5%)
Existential	75 (22%)	-	-				
Systemic	14 (4%)	-	-				
Other	54 (16%)	27 (20%)	10 (12%)		14 (13%)		23 (21%)

Note. *Participants could select more than one option

Table 6

Convergent Validity in Client Subsample against WAI-SR and RDI-R2

	RDFS	WAI-SR	WAI-SR Goal	WAI-SR Task	WAI-SR Bond	RDI
RDFS	1.00	0.71 (0.61 to 0.79)	0.4 (0.28 to 0.51)	0.69 (0.6 to 0.76)	0.76 (0.69 to 0.82)	0.77 (0.7 to 0.83)
WAI-SR	0.68 (0.59 to 0.77)	1.00	0.83 (0.78 to 0.86)	0.9 (0.86 to 0.92)	0.85 (0.81 to 0.89)	0.67 (0.57 to 0.75)
WAI-SR Goal	0.39 (0.26 to 0.52)	0.84 (0.79 to 0.88)	1.00	0.6 (0.5 to 0.68)	0.48 (0.38 to 0.59)	0.41 (0.27 to 0.53)
WAI-SR Task	0.66 (0.56 to 0.75)	0.86 (0.8 to 0.9)	0.57 (0.45 to 0.67)	1.00	0.74 (0.66 to 0.8)	0.68 (0.58 to 0.76)
WAI-SR Bond	0.74 (0.66 to 0.8)	0.82 (0.76 to 0.87)	0.46 (0.34 to 0.58)	0.7 (0.62 to 0.77)	1.00	0.66 (0.54 to 0.74)
RDI	0.75 (0.66 to 0.82)	0.63 (0.52 to 0.73)	0.39 (0.25 to 0.52)	0.66 (0.55 to 0.75)	0.63 (0.51 to 0.72)	1.00

Note. Figures above leading diagonal are Pearson correlation, figures below are Spearman coefficients.

Table 7

Convergent Validity in Therapist Subsample against WAI-SR and RDI-R2

	RDFS	WAI-SR	WAI-SR Goal	WAI-SR Task	WAI-SR Bond	RDI
RDFS	1.00	0.5 (0.41 to 0.58)	0.4 (0.29 to 0.5)	0.47 (0.37 to 0.55)	0.43 (0.33 to 0.52)	0.56 (0.48 to 0.63)
WAI-SR	0.5 (0.41 to 0.59)	1.00	0.92 (0.9 to 0.93)	0.91 (0.89 to 0.93)	0.73 (0.65 to 0.8)	0.34 (0.23 to 0.44)
WAI-SR Goal	0.42 (0.32 to 0.5)	0.91 (0.88 to 0.93)	1.00	0.82 (0.78 to 0.86)	0.47 (0.37 to 0.57)	0.23 (0.11 to 0.35)
WAI-SR Task	0.46 (0.36 to 0.54)	0.89 (0.85 to 0.91)	0.79 (0.73 to 0.83)	1.00	0.49 (0.37 to 0.59)	0.31 (0.19 to 0.42)
WAI-SR Bond	0.41 (0.3 to 0.49)	0.7 (0.64 to 0.76)	0.47 (0.39 to 0.56)	0.43 (0.34 to 0.52)	1.00	0.35 (0.23 to 0.46)
RDI	0.56 (0.48 to 0.64)	0.34 (0.23 to 0.44)	0.26 (0.14 to 0.37)	0.31 (0.2 to 0.43)	0.3 (0.17 to 0.41)	1.00

Note. Figures above the leading diagonal are Pearson correlation coefficients, figures below are Spearman coefficients.