




Extended validation of the Relational Depth Frequency Scale: Retest reliability, divergent and criterion validity, and measurement invariance in UK- and US-stratified samples

Gina Di Malta¹  | Zhuang She²  | Brett Raymond-Barker³ | Mick Cooper³ 

¹School of Psychology and Counselling, Faculty of Arts and Social Sciences, The Open University, Milton Keynes, UK

²School of Psychology and Cognitive Science, East China Normal University, Shanghai, China

³School of Psychology, University of Roehampton, London, UK

Correspondence

Gina Di Malta, School of Psychology and Counselling, Faculty of Arts and Social Sciences, The Open University, Milton Keynes, UK.

Email: gina.dimalta@open.ac.uk

Abstract

Background: The Relational Depth Frequency Scale (RDFS) assesses moments of profound connection in psychotherapy, associated with therapeutic benefit. To date, the RDFS has not been tested for its retest reliability, divergent and criterion validity, and measurement invariance, nor has it been tested in stratified samples of psychotherapy patients.

Methods: Two stratified online samples of United Kingdom ($n = 514$) and United States ($n = 402$) psychotherapy patients filled out the RDFS, the Brief Social Desirability Scale (BSDS); and the Satisfaction with Therapy and Therapist Scale-revised (STTS-R). Two subsamples of patients (United Kingdom: $n = 50$ and United States: $n = 203$) filled out the RDFS again after 1 month.

Results: Reliability for the six-item RDFS were excellent in United Kingdom and United States samples (Cronbach's $\alpha = 0.91$ and 0.92 ; retest $r = 0.73$ and $r = 0.76$). Divergent ($r = 0.10$ and $r = 0.12$) and criterion validity ($r = 0.69$; and $r = 0.70$) were good. Full scalar invariance was established across countries, genders, and time.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2023 The Authors. *Journal of Clinical Psychology* published by Wiley Periodicals LLC.

Conclusion: This contributes important evidence to the validity of the RDFS. Future research should assess predictive validity against psychotherapy outcomes and replicate these analyses in diverse samples.

KEYWORDS

psychotherapy, psychotherapy satisfaction, relational depth, Relational Depth Frequency Scale, validation

1 | INTRODUCTION

The psychotherapy relationship is an essential common factor known to contribute to therapeutic outcomes (Norcross & Lambert, 2018; Wampold, 2021). There is extensive evidence that the psychotherapy relationship is associated with better engagement, less dropout, and increased satisfaction with psychotherapy and psychotherapists (Flückiger et al., 2018; Haan et al., 2014; Holdsworth et al., 2014; Keleher et al., 2019).

Relational depth (RD) is a unique quality of the therapeutic relationship 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). RD involves the presence of the Rogers (1957) conditions of empathy, warmth, and congruence, with both psychotherapist and patient active and open in the relationship (Mearns & Cooper, 2005/2018). RD is conceptualized as going beyond other relationship constructs in that it involves deep intimacy and complete mutual acceptance and openness (Cooper, 2005).

There is qualitative evidence that RD is associated with a range of therapeutic benefits, including increased satisfaction with psychotherapy. Fourteen counseling patients were interviewed about their experiences of RD (Knox, 2008). Most reported feeling real, whole, alive, and deeply understood during these moments (p. 184). They also said that the experience culminated in the sense of self-worth, self-acceptance, and empowerment. During moments of RD, patients perceived their therapists as warm, genuine, reliable, solid, experienced, and fearless, with a real desire to understand them (Knox, 2008; Knox & Cooper, 2011). Most patients felt that the therapist offered them something “over and beyond” what they would have expected from a professional relationship. Further evidence from 119 patients with single survey items by Leung (2008) suggested that, on average, patients rated moments of RD as being very important for therapeutic change. Patients also rated RD moments as strongly therapeutic and having an enduring impact. Recent research also indicates that levels of RD are associated with psychotherapy outcomes (Kim et al., 2020) and well-being (Di Malta, Bond, et al., [in press](#)).

As a result of these developments, counselor training in RD has already become integral to some psychotherapy and counseling courses across the United Kingdom and in the United States (e.g., Ray et al., 2021). However, further developments, and particularly outcome research, are needed; and this requires a reliable and well-validated instrument.

A first measure—the Relational Depth Inventory (RDI)—was developed to rate the presence of RD in a significant moment in psychotherapy (Wiggins et al., 2012). The RDI could be used to assess whether RD was present in a given moment as well as its intensity. The measure showed good convergent validity with the Working Alliance Inventory-Short Revised (WAI-SR, Hatcher & Gillaspay, 2006), a widely used short measure of the therapeutic alliance ($r = 0.72$, $n = 150$, $p < 0.01$). The scope of the RDI, however, was limited to a single moment in psychotherapy and did not capture RD across the overall psychotherapy.

The Relational Depth Frequency Scale (RDFS) is a short unidimensional self-report measure developed to assess the frequency of RD over the course of psychotherapy (Di Malta, Evans, et al., 2020). The RDFS was

developed based on standard procedures of scale development (DeVellis & Thorpe, 2021) and cognitive pretesting methods with patients and psychotherapists (Di Malta, Cooper, et al., 2020). The RDFS was psychometrically tested using confirmatory factor analysis (CFA) and Rasch analyses in online samples of psychotherapists and patients in the United Kingdom (Di Malta, Evans, et al., 2020). It had a stable factor structure and showed excellent internal consistency (e.g., Cronbach's $\alpha = 0.93$ in a patient sample). It showed good convergent validity with the WAI-SR ($r = 0.71$) and the RDI ($r = 0.77$; Di Malta, Evans, et al., 2020). The validity of the instrument was later tested in clinical samples of patients and psychotherapists dyads in Poland, where its structure was replicated (Zarzycka et al., 2022). It showed good internal consistency (e.g., Cronbach's $\alpha = 0.81$ for patients), and good convergent validity with the Scale to Assess the Therapeutic Relationship (r ranging from 0.37 to 0.65; Zarzycka et al., 2022).

While the RDFS has undergone extensive assessments of reliability and validity, the measure has never been tested for retest reliability, divergent validity, and criterion validity. In addition, reliability and validity assessments have not been carried out in stratified samples of psychotherapy patients. Such stratified samples, representative of the general population, would mean that results can be generalized, irrespective of such characteristics as gender, age, and ethnicity of participants.

Acceptable retest reliability has ranged from 0.4 to 0.7 (Cicchetti, 1994; Fleiss, 1986). It is expected, when measuring the frequency of RD over the course of psychotherapy, that retest reliability should be highly consistent over time—provided the course of therapy is finished, as patients would report on the same frequency of RD over the course of psychotherapy (DeVellis & Thorpe, 2021). In ongoing psychotherapy, however, there may be differences at the two-time points as patients may have experienced more or less frequent RD over the remaining psychotherapy sessions, which could impact on their overall test score. In addition, if patients are still in psychotherapy at retest, a short time lapse between the time points is likely to show a higher correlation.

Criterion validity examines the scale's correlation with an external outcome, here satisfaction with psychotherapy and the psychotherapist (DeVellis & Thorpe, 2021). Satisfaction with treatment is an important patient-level outcome (e.g., Conte et al., 1995). Numerous studies show that it is key to engagement and crucial to positive improvement on symptoms (e.g., Bados et al., 2007; Moors & Zech, 2017; Tryon, 1990). Psychotherapists' characteristics associated with patient satisfaction also seem to align with those perceived in a moment of RD, including warm behaviors (e.g., Moors & Zech, 2017), unconditional positive regard (Roos & Werbart, 2013), being active and involved in the psychotherapy (Ackerman & Hilsenroth, 2003; Conte et al., 1995), and actively listening (Weger et al., 2014). In addition, patient satisfaction may be due to increased patient insight and depth in the first session (Tryon, 1990). With the existing evidence of associations of RD with outcomes and positive patient experiences, as well as its overlap with psychotherapist qualities, we predict that RDFS scores will correlate with the criterion of patient satisfaction in psychotherapy.

Divergent validity is demonstrated by a small correlation for two constructs that are not theoretically related (DeVellis & Thorpe, 2021). While RD frequency may be perceived as an experience that is socially desirable to achieve in psychotherapy, RD frequency scores should not show significant overlap with social desirability scores.

Another unsolved issue regarding the RDFS is its measurement invariance (MI). Two types of MI have been evidenced (C.-C. Chen et al., 2020): one is multiple-group MI—whether factor structure stays the same across groups (e.g., across countries, genders, or populations); a second is longitudinal invariance—whether the factor structure remains unchanged over time. These two types of MI are prerequisites for making an unbiased comparison of latent means across groups or time (Putnick & Bornstein, 2016). So far, only one study has attempted to establish the MI of the RDFS across clinical dyad populations of psychotherapy clients and therapists (Zarzycka et al., 2022). However, we still do not know whether the RDFS could be compared across countries, gender, and whether it could make meaningful mean comparisons when assessed over time in longitudinal studies. Addressing these issues would support the applications of the RDFS in different countries, subgroups, and future longitudinal surveys.

The aim of this study is to extend the validity of the RDFS in UK- and US-stratified samples. Based on prior literature, we hypothesize that:

- Retest reliability of the RDFS after 1 month is good with at least a moderate association ($r > 0.50$).

- The RDFS has good criterion validity with a moderate association with patient satisfaction with psychotherapy and psychotherapist (Satisfaction with Therapy and Therapist Scale-revised [STTS-R]; $r > 0.5$; $p < 0.001$).
- The RDFS has good divergent validity, with no association with social desirability (Brief Social Desirability Scale [BSDS]; $r < 0.2$; $p > 0.05$).
- The RDFS shows full scalar invariance across gender, countries, and over time, and has the same factor structure across countries, gender, and over time.

1.1 | METHODS

The research project was submitted for ethics consideration under the reference PSYC15/564 in the Department of Psychology and was approved under the procedures of the University of Roehampton's Ethics Committee.

1.1.1 | Participants

We collected data in two online samples via the participant recruitment site [Prolific.co](https://prolific.co). Recruitment was conducted with the aim of achieving a sample representative of UK and US populations across categories of age, gender, and ethnicity in accordance with national census data for each country, respectively (United Kingdom Census, 2011; United States Census, 2010). We used the most recent census data at the time of conducting the study. The inclusion criteria for participants were being at least 18 (United Kingdom) or 20 (United States) years of age, resident in the respective country, and having experienced a course of psychotherapy (four sessions or more). Participants' demographics are detailed in Table 1.

1.2 | Measures

1.2.1 | RDFS

The RDFS is a six-item unidimensional scale to assess the frequency of moments of RD over the course of therapy. Items include: "We were deeply connected to one another" and "We were immersed in the present moment" and

TABLE 1 Participant [retest] demographics.

Country of residence	UK sample (N = 514) [50]		US sample (N = 402) [203]	
	Female (N = 268) [32]	Male (N = 243) [18]	Female (N = 196) [100]	Male (N = 183) [92]
White (UK–2, US–15 [3])	234 [30]	210 [15]	136 [52]	133 [58]
Black (UK–0, US–1 [0])	8 [0]	9 [1]	24 [17]	19 [9]
Asian (UK–0, US–0 [0])	18 [1]	18 [2]	10 [8]	11 [11]
Mixed (UK–1, US–3 [2])	5 [1]	3 [0]	15 [12]	8 [4]
Other (UK–0, US–4 [2])	3 [0]	3 [0]	11 [11]	12 [9]

Note: Participants making nonbinary gender selections detailed beside ethnicity.

Abbreviation: SD, standard deviation.

are rated on a 5-point Likert scale where 1 = *not at all*, 2 = *only occasionally*, 3 = *sometimes*, 4 = *often*, 5 = *most or all of the time*. Each item follows the statement, "Over the course of therapy with my therapist, there were moments where...". There are no reverse items. The total score for the RDFS is summed and ranging from six to 30 with a higher score meaning more frequent RD, and no specified cut-off scores. The RDFS has shown excellent internal consistency ($\alpha > 0.90$) and validity in online samples of therapists and patients (Di Malta, Cooper, et al., 2020; Di Malta, Evans, et al., 2020), and in a clinical sample of psychotherapy dyads (Zarzycka et al., 2022). In this study, the RDFS' internal consistency was Cronbach's $\alpha > 0.90$ in the UK and US samples.

1.2.2 | STTS-R

The STTS-R (Oei & Green, 2008) is a 12-item, two-dimensional scale that assesses patients' levels of satisfaction with psychotherapy and psychotherapist. The two-factor structure accounts for over 60% of the variance and includes the Satisfaction with Therapy (SwT; six items) and the Satisfaction with Therapist (SwTt; six items) subscales. The revised form is a development of the original STTS where four original items were modified, three were removed, and three were added to better reflect patients' experience with psychotherapy and provide an improved measure of patient satisfaction (Oei & Green, 2008; Oei & Shuttlewood, 1999). Items are framed positively, for instance: "I am now able to deal more effectively with my problems" and "the therapist seemed to understand what I was thinking and feeling." Items are rated on a 5-point Likert scale ranging from 1 = *Strongly agree* to 5 = *Strongly disagree*. The scores are summed for all even number items to obtain the patient's level of SwT and for all odd-number items to obtain the patient's level of SwTt. The higher the score, the greater the level of patient satisfaction. The STTS-R has shown good psychometric properties (SwT $\alpha = 0.91$ and SwTt $\alpha = 0.96$) and a stable factor structure (Oei & Green, 2008). In this study, internal consistency for the SwT, SwTt, and STTS-R were excellent in both UK and US samples (Cronbach's $\alpha s > 0.90$).

1.2.3 | BSDS

The BSDS (Haghighat, 2007) is a four-item scale of social desirability developed for its brevity and practicality. The BSDS is the most commonly used social desirability scale across psychology and other disciplines. Items include "Would you smile at people every time you meet them?" and "Do you always practice what you preach?" Item 4: "Would you ever lie to people?" is reversed scored. Items are answered with "yes" or "no" with a score of 1 allocated to "yes" and 0 for "no." The total scores are summed. The cut-off score can be set from anything > 1 (more than one socially desirable answer) to > 2 (more than two socially desirable answers) to exclude people with a high tendency towards social desirability. The BSDS has shown good validity and moderate reliability, which are acceptable for its brevity (Cronbach's $\alpha = 0.60$; Turner et al., 2018). In this study, the internal consistency for the BSDS was Cronbach's $\alpha = 0.47$. This is poor but consistent with low past reliability, and reflects a limitation associated with the practicality of its brevity, and its reverse-scored item.

1.3 | Procedures

Participants were recruited on Prolific.co, an online platform for participant recruitment. We prescreened a pool of potential participants who had answered "Yes" to the Prolific.co prescreening item "Do you have—or have you had—a diagnosed, on-going mental health/illness/condition?" These individuals were then shown the survey advert, which invited participants who had had, or were having, psychotherapy. Sign-up was on a first come first served basis. Participants who responded to the advert on Prolific.co were redirected to the Qualtrics survey platform to

complete the study. Participants were asked to provide informed consent before completing the RDFS, STTS-R, BSDS, and answering demographic questions. Most participants signing up to the study went on to complete it.

The study was readvertised to the same participants after 1 month. They were invited to take part in a retest survey that contained the RDFS only. Participants were instructed to rate the same course of psychotherapy. Recruitment for the retest happened in the same way as with the initial recruitment point: on a "first come first served" basis for the smaller UK sample and with the aim of achieving a larger subsample from US participants which remained broadly representative of the US population across categories of age, gender, and ethnicity in accordance with national census data (United States Census, 2010). In this regard, recruitment for the US retest sample gave priority to the retention of participants from minority demographic groupings. We aimed to recruit a UK retest subsample of 50 and a US retest subsample of 200.

Participants were paid the UK or US minimum wage equivalent of 5 min pro-rata for completing the study, with retested participants paid for a further 2 min pro-rata upon completing the RDFS again after a period of 1 month. UK participants received a total of £0.95, US participants received a total of £0.50.

The initial surveys were open to participants for up to 3 months to allow for representative samples to be collected. Retest surveys, which were advertised to participants 1 month after their initial survey completion, remained open for up to 3 weeks for the same reason.

1.4 | Analyses

Divergent, criterion validity, and retest reliability were assessed with Pearson correlations. Descriptive statistics, correlations, and tests of internal consistency (Cronbach's α and McDonald's ω) were performed with SPSS 21.0. Tests of factor structure and MI were performed with Mplus 8.3.

The comparative fit index (CFI), the root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR) were used to evaluate the model fit. RMSEA and SRMR values below 0.08, and CFI values above 0.90, are considered indicative of acceptable fit (Brown, 2015). There were no residual correlations in the CFA model.

For MI analyses, three models with increased constraints (configural, metric, and scalar invariance) were compared to test the MI of RDFS across countries (United Kingdom vs. United States), gender (men vs. women), and time (a month interval). Nonbinary participants were removed from the MI testing across gender due to the small sample size. We did not test strict equivalence because it is not a prerequisite for testing mean differences across groups or over time (Putnick & Bornstein, 2016). Multigroup and longitudinal invariance analyses were conducted with each successive step including more constraints. First, a baseline model was estimated without any constraints to examine whether the RDFS had equal factor structure across groups or over time (configural invariance); second, we tested whether the RDFS had equal factor loadings across groups or over time (metric invariance); third, we examined whether the RDFS had equal item intercepts across groups or over time (scalar invariance).

Invariance was assessed by measuring the change in fit compared to the previous level of MI: Δ CFI and Δ RMSEA values were used to evaluate the model fit with increased constraints. $|\Delta$ CFI < 0.01 and $|\Delta$ RMSEA < 0.015 indicates a good fit (F. F. Chen, 2007). All parameter estimates were obtained using the robust maximum likelihood estimation (MLR).

2 | RESULTS

2.1 | Descriptive statistics

Descriptive statistics for the RDFS items are shown in Table 2.

TABLE 2 Descriptive statistics for the six-item RDFS measure by country.

Country	Item	M	SD	Skewness	Kurtosis	RDFS questions
United Kingdom	RDFS 1	2.23	1.11	0.41	-0.86	1. We were deeply connected to one another.
	RDFS 2	2.88	1.09	-0.17	-0.77	2. We were immersed in the present moment.
	RDFS 3	2.62	1.17	0.17	-0.84	3. There was a deep understanding between us.
	RDFS 4	2.45	1.19	0.31	-0.84	4. It felt like a shared experience.
	RDFS 5	2.34	1.15	0.37	-0.86	5. I felt like we were totally in the moment together.
	RDFS 6	3.1	1.19	-0.19	-0.83	6. I felt we were completely open with each other.
United States	RDFS 1	2.88	1.05	-0.21	-0.57	1. We were deeply connected to one another.
	RDFS 2	3.41	0.97	-0.45	-0.17	2. We were immersed in the present moment.
	RDFS 3	3.26	1.08	-0.40	-0.46	3. There was a deep understanding between us.
	RDFS 4	3.03	1.18	-0.12	-0.83	4. It felt like a shared experience.
	RDFS 5	3.02	1.11	-0.13	-0.63	5. I felt like we were totally in the moment together.
	RDFS 6	3.64	1.13	-0.67	-0.34	6. I felt we were completely open with each other.

Abbreviations: RDFS, Relational Depth Frequency Scale; SD, standard deviation.

2.2 | Reliability

Internal consistency for the RDFS were Cronbach's $\alpha = 0.91$ ($\omega = 0.91$) in the UK sample ($n = 514$) and Cronbach's $\alpha = 0.92$ ($\omega = 0.92$) in the US sample ($n = 402$). The retest reliability for the RDFS was $r = 0.73$; 95% confidence interval (CI) [0.54, 0.87], $p < 0.001$ in the UK sample ($n = 50$) and $r = 0.76$; 95% CI [0.69, 0.82]; $p < 0.001$ in the US sample ($n = 203$).

2.3 | Divergent validity with social desirability

Pearson correlations between RDFS and the BSDS were used to assess the divergent validity of RDFS. The correlations were: $r = 0.10$; 95% CI [0.01, 0.19]; $p = 0.02$ and Cohen's $d = 0.20$ in the UK sample ($n = 514$); and $r = 0.12$; 95% CI [0.04, 0.19]; $p = 0.01$ and Cohen's $d = 0.24$ in the US sample ($n = 402$).

2.4 | Criterion validity with psychotherapy satisfaction

Pearson correlations between RDFS and STTS-R were used to assess the criterion validity of RDFS. The correlations were $r = 0.69$; 95% CI [0.64, 0.73]; $p < 0.0001$ in the UK sample; and $r = 0.70$; 95% CI [0.66, 0.73] $p < 0.0001$ in the US sample. Pearson correlations for overall satisfaction and the two subscales are shown in Table 3.

2.5 | Factor structure

The RDFS showed good model fit in both national samples (United Kingdom: CFI = 0.98, RMSEA = 0.08, SRMR = 0.02; United States: CFI = 0.98, RMSEA = 0.07, SRMR = 0.02). The CFA factor loadings are presented in

TABLE 3 Criterion validity of relational depth with psychotherapy satisfaction.

	UK sample, N = 514		US sample, N = 402	
	Pearson correlations <i>r</i>	95% CI	Pearson correlations <i>r</i>	95% CI
RDFS and patient satisfaction with psychotherapy (STTS-R)	0.69*	0.64, 0.73	0.70*	0.65, 0.73
RDFS and patient satisfaction with psychotherapy subscale (SWT)	0.64*	0.59, 0.68	0.66*	0.56, 0.68
RDFS and patient satisfaction with psychotherapist subscale (SWTt)	0.68*	0.64, 0.72	0.71*	0.66, 0.74

Abbreviations: CI, confidence interval; RDFS, Relational Depth Frequency Scale; STTS-R, Satisfaction with Therapy and Therapist Scale-revised; SwT, Satisfaction with Therapy; SwTt, Satisfaction with Therapist.

* $p < 0.0001$.

TABLE 4 Standardized factor loadings in the UK and US samples.

Items	Standardized factor loading (United Kingdom)	Standardized factor loading (United States)
Item 1	0.80***	0.82***
Item 2	0.74***	0.73***
Item 3	0.86***	0.86***
Item 4	0.82***	0.81***
Item 5	0.88***	0.87***
Item 6	0.65***	0.64***

*** $p < 0.001$.

Table 4 for the UK sample and US sample. In both samples, all items loaded highly and significantly on the target factor.

2.6 | Measurement invariance

As presented in Table 5, full scalar invariance was established across gender (in both UK and US samples), countries, and time as modes with increased constraints, and did not show a significant deterioration compared with the previous one (i.e., $|\Delta CFI| < 0.01$ and $|\Delta RMSEA| < 0.01$) in all tested MI analyses (see Table 3 for model comparisons). The results supported the full scalar invariance of RDFS across groups tested and time.

3 | DISCUSSION

These findings provide further evidence for the reliability and validity of the RDFS in stratified samples of psychotherapy patients. Tests of internal consistency, retest reliability, divergent validity with a measure of social desirability, and criterion validity with psychotherapy satisfaction were good. In addition, we established the full scalar invariance of the RDFS across gender, countries, and time. These results provide additional evidence for the reliability and validity of the RDFS in a Western context.

TABLE 5 Measurement Invariance across gender, countries, and time.

Subgroup comparison	χ^2	df	CFI	RMSEA	M comparison	Δ CFI	Δ RMSEA
Across gender							
For UK sample (N = 511)							
M1. Configural invariance	46.940	18	0.981	0.079	-	-	-
M2. Full metric invariance	58.263	23	0.977	0.077	M2-M1	-0.004	-0.002
M3. Full scalar invariance	62.768	28	0.977	0.070	M3-M2	0.000	-0.007
For US sample (N = 379)							
M4. Configural invariance	26.189	18	0.993	0.049	-	-	-
M5. Full metric invariance	36.251	23	0.989	0.055	M5-M4	-0.004	0.006
M6. Full scalar invariance	38.274	28	0.992	0.044	M6-M5	0.003	-0.011
Across countries (N: United Kingdom = 514; United States= 402)							
M7. Configural invariance	59.609	18	0.985	0.071	-	-	-
M8. Full metric invariance	69.393	23	0.984	0.066	M8-M7	-0.001	-0.005
M9. Full scalar invariance	100.147	28	0.975	0.075	M9-M8	-0.009	0.009
Across time for US sample (N = 203)							
M10. Configural invariance	74.883	47	0.982	0.054	-	-	-
M11. Full metric invariance	80.186	52	0.982	0.052	M11-M10	0.000	-0.002
M12. Full scalar invariance	82.550	57	0.983	0.047	M12-M11	0.001	-0.005

Note: All χ^2 estimates were significant at $p < 0.001$.

Abbreviations: CFI, comparative fit index; RMSEA, root mean square error of approximation.

Reliability tests corroborated prior results. The excellent internal consistency of > 0.9 in both samples was comparable to internal consistency from prior studies in online samples (Di Malta, Evans, et al., 2020) and in clinical samples (Zarzycka et al., 2022). Retest reliability after 1 month was above 0.7 in both samples. Although we expected this measurement to be consistent over time, this moderately strong correlation is above the recommended acceptable minimum suggested by Nunally and Bernstein (1978) of 0.7, and close to excellent, however, by the standards of Fleiss (1986). The moderate correlation may have reflected that some patients had not ended psychotherapy at these timepoints and, therefore, may have experienced more or less frequent RD during the month between the two timepoints, which would have changed their score at retest. This would have an impact on retest correlations. Future research may investigate retest reliability of the RDFS in patients who have finished their course of therapy.

As hypothesized, correlations with social desirability were < 0.2 . However, these associations were significant (unsurprising, given the large sample size), with a small but marginally meaningful effect size ($d > 0.20$). This suggests there may be a small social desirability bias when patients use the RDFS (Cohen, 1992). This is not unexpected as RD is likely to be perceived as an experience that is socially desirable to achieve in psychotherapy. However, correlations were close to zero and substantially smaller than correlations that assessed convergent validity, which suggests these can constitute evidence of acceptable divergent validity. Such evidence of divergent validity with a social desirability measure is an important step to demonstrate the validity of the RDFS. In effect, it suggests that the RDFS's reliability is not mostly due to a response bias (DeVellis & Thorpe, 2021). This finding, however, must be

interpreted with caution. It will also need replication with a measure of social desirability that has shown better internal consistency.

As predicted, RD frequency was associated with both psychotherapy satisfaction and satisfaction with the psychotherapist. This supports prior evidence which suggest that psychotherapists' characteristics associated with patient satisfaction are closely comparable with those perceived by patients in moments of RD (e.g., Ackerman & Hilsenroth, 2003; Conte et al., 1995; Knox et al., 2012; Roos & Werbart, 2013; Weger et al., 2014). This finding constitutes the first empirical evidence for the association of RD with satisfaction with psychotherapy and the psychotherapist.

Scalar invariance results showed full invariance across gender, countries, and time. These results suggested that all RDFS items means were statistically equivalent across these populations and over time, and participants from the tested populations and in their repeated measures generally answered the same to the RDFS items. Thus, it is appropriate to make RDFS mean comparisons between men/women and the United Kingdom/United States, as well as over time. Moreover, the results indicated a stable factor structure across groups and measurement time points. The findings also support the generalizability of the RDFS's unidimensional factor across countries, gender, and over time.

The study was limited in that it relied on online samples of patients who remembered past courses of psychotherapy. It is possible that this could have influenced the way in which patients responded to questionnaires. It may be useful to test these associations in clinical settings. In addition, the retest period was not consistent across endings of therapy, which may have made the retest less consistent. Gender invariance analyses were limited to two genders: men and women, due to the other category having a small sample size. Similarly, we could not test MI across ethnicity groups due to small sample sizes. MI was only tested across two Western countries, which limits evidence of MI outside of a Western context. Future research needs to address the MI of the RDFS across other gender categories, non-Western countries, and ethnicity groups.

Overall, this study supports the use of the RDFS in research. The RDFS can also be used as a process measure in clinical practice. Future research should examine the predictive validity of the RDFS on different psychotherapeutic outcomes and continue to assess the validity of the RDFS in diverse samples. Assessing the predictive validity of the RDFS on symptomatic outcomes can support the implementation of RD training with the view of improving psychotherapy services (Ray et al., 2021).

3.1 | Implications for practice

This study extends the evidence for the reliability and validity of the RDFS—a tool to assess the frequency of moments of profound connection in psychotherapy. The RDFS can be used in practice as a process measure to assess moments of deep contact over the course of treatment. Monitoring such moments may be important for practice as they are associated with increased psychotherapy satisfaction and satisfaction with the psychotherapist. In addition, RD moments have been associated with enhanced psychotherapy benefits and psychological well-being. The RDFS can also be used in future research to study the impact of RD on psychotherapy outcomes. Establishing such links can support the development of psychotherapy.

AUTHOR CONTRIBUTIONS

Gina Di Malta formulated the conceptualization of study aims and goals, carried out the funding acquisition, formal analysis, project administration, writing of the manuscript, and reviewing and editing the work. Zhuang She carried out the formal analysis and was involved in writing the manuscript and reviewing and editing the work. Brett Raymond-Barker was involved in the data curation, project administration, and writing the manuscript. Mick Cooper supervised the project, was involved in writing the manuscript, and reviewing and editing the work.

ACKNOWLEDGMENTS

Thanks to all patients who participated in this study. This work was supported by the Strategic Research Investment Fund (SRIF) at the Open University.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data and material used in this study are available on the Open Science Framework repository. They can be accessed upon reasonable request to the corresponding author.

ETHICS STATEMENT

The study received ethical approval from the University of Roehampton's Ethics Committee. All patients gave informed consent to participate in this study. This study was not preregistered.

ORCID

Gina Di Malta  <https://orcid.org/0000-0002-6254-6522>

Zhuang She  <http://orcid.org/0000-0001-8702-837X>

Mick Cooper  <http://orcid.org/0000-0003-1492-2260>

PEER REVIEW

The peer review history for this article is available at <https://www.webofscience.com/api/gateway/wos/peer-review/10.1002/jclp.23525>.

REFERENCES

- Ackerman, S. J., & Hilsenroth, M. J. (2003). A review of therapist characteristics and techniques positively impacting the therapeutic alliance. *Clinical Psychology Review, 23*(1), 1–33. [https://doi.org/10.1016/S0272-7358\(02\)00146-0](https://doi.org/10.1016/S0272-7358(02)00146-0)
- Bados, A., Balaguer, G., & Saldaña, C. (2007). The efficacy of cognitive-behavioral therapy and the problem of drop-out. *Journal of Clinical Psychology, 63*(6), 585–592. <https://doi.org/10.1002/jclp.20368>
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). Guilford Press.
- Chen, C.-C., Lau, J. M., Richardson, G. B., & Dai, C.-L. (2020). Measurement invariance testing in counseling. *Journal of Professional Counseling: Practice, Theory & Research, 47*(2), 89–104. <https://doi.org/10.1080/15566382.2020.1795806>
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 14*(3), 464–504. <https://doi.org/10.1080/10705510701301834>
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*(4), 284–290.
- Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159.
- Conte, H. R., Ratto, R., Clutz, K., & Karasu, T. B. (1995). Determinants of outpatients' satisfaction with therapists: Relation to outcome. *The Journal of Psychotherapy Practice and Research, 4*(1), 43–51. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3330383/pdf/43.pdf>
- Cooper, M. (2005). Therapists' experiences of relational depth: A qualitative interview study. *Counselling and Psychotherapy Research, 5*(2), 87–95. <https://doi.org/10.1080/17441690500211130>
- Fleiss, J. L. (1986). Reliability of measurement, *The design and analysis of clinical experiments*. John Wiley and sons.
- Flückiger, C., Del Re, A. C., Wampold, B. E., & Horvath, A. O. (2018). The alliance in adult psychotherapy: A meta-analytic synthesis. *Psychotherapy, 55*(4), 316–340. <https://doi.org/10.1037/pst0000172>
- Haan, A. M., Boon, A. E., Jong, J. T. M., Geluk, C. A. L., & Vermeiren, R. R. M. (2014). Therapeutic relationship and dropout in youth mental health care with ethnic minority children and adolescents. *Clinical Psychologist, 18*(1), 1–9. <https://doi.org/10.1111/cp.12030>
- Haghighat, R. (2007). The development of the Brief Social Desirability Scale (BSDS). *Europe's Journal of Psychology, 3*(4), 10–5964.

- Hatcher, R. L., & Gillaspay, J. A. (2006). Development and validation of a revised short version of the Working Alliance Inventory. *Psychotherapy Research*, 16(1), 12–25. <https://doi.org/10.1080/10503300500352500>
- Holdsworth, E., Bowen, E., Brown, S., & Howat, D. (2014). Client engagement in psychotherapeutic treatment and associations with client characteristics, therapist characteristics, and treatment factors. *Clinical Psychology Review*, 34(5), 428–450. <https://doi.org/10.1016/j.cpr.2014.06.004>
- Keleher, B., Oakman, J. M., Capobianco, K., & Mittelstaedt, W. H. (2019). Basic psychological needs satisfaction, working alliance, and early termination in psychotherapy. *Counselling Psychology Quarterly*, 32(1), 64–80. <https://doi.org/10.1080/09515070.2017.1367271>
- Kim, J., Joseph, S., & Price, S. (2020). The positive psychology of relational depth and its association with unconditional positive self-regard and authenticity. *Person-Centered & Experiential Psychotherapies*, 19(1), 12–21. <https://doi.org/10.1080/14779757.2020.1717983>
- Knox, R. (2008). Clients' experiences of relational depth in person-centred counselling. *Counselling and Psychotherapy Research*, 8(3), 182–188. https://www.researchgate.net/publication/233445567_Clients%27_experiences_of_relational_depth_in_person-centred_counselling
- Knox, R., & Cooper, M. (2011). A state of readiness: An exploration of the client's role in meeting at relational depth. *Journal of Humanistic Psychology*, 51(1), 61–81. <https://doi.org/10.1177/0022167810361687>
- Knox, R., Murphy, D., Wiggins, S., & Cooper, M. (Eds.). (2012). *Relational depth: New perspectives and developments*. Macmillan International Higher Education.
- Leung, J. (2008). *A quantitative online study exploring the factors associated with the experience and perception of relational depth* [Unpublished dissertation DPsych]. University of Strathclyde.
- Di Malta, G., Bond, J., Raymond-Barker, B., Moller, N., & Cooper, M. (in press). The impact of relational depth, emotional intimacy, and relationship satisfaction on psychological wellbeing in close relationships in the community. *Journal of Happiness Studies*.
- Di Malta, G., Cooper, M., Vos, J., & van der Veer, K. (2020). An application of the Three-Step Test-Interview (TSTI) in the validation of the Relational Depth Frequency Scale. *Journal of Humanistic Psychology*. <https://doi.org/10.1177/0022167820962626>
- Di Malta, G., Evans, C., & Cooper, M. (2020). Development and validation of the relational depth frequency scale. *Psychotherapy Research*, 30, 213–227. <https://doi.org/10.1080/10503307.2019.1585590>
- Mearns, D., & Cooper, M. (2005/2018). *Working at relational depth in counselling and psychotherapy*. Sage.
- Moors, F., & Zech, E. (2017). The effects of psychotherapist's and clients' interpersonal behaviors during a first simulated session: A lab study investigating client satisfaction. *Frontiers in Psychology*, 8, 1868. <https://doi.org/10.3389/fpsyg.2017.01868>
- Norcross, J. C., & Lambert, M. J. (2018). Psychotherapy relationships that work III. *Psychotherapy*, 55(4), 303–315. <https://doi.org/10.1037/pst0000193>
- Nunnally, J. C., & Bernstein, I. (1978). *Psychometric theory*. McGraw.
- Oei, T. P. S., & Green, A. L. (2008). The Satisfaction With Therapy and Therapist Scale-revised (STTS-R) for group psychotherapy: Psychometric properties and confirmatory factor analysis. *Professional Psychology: Research and Practice*, 39(4), 435–442. <https://doi.org/10.1037/0735-7028.39.4.435>
- Oei, T. P. S., & Shuttlewood, G. J. (1999). Development of a Satisfaction with Therapy and Therapist Scale. *The Australian and New Zealand Journal of Psychiatry*, 33(5), 748–753. <https://doi.org/10.1080/j.1440-1614.1999.00628.x>
- Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, 41, 71–90. <https://doi.org/10.1016/j.dr.2016.06.004>
- Ray, D. C., Lankford, C. T., Malacara, A. B., Woehler, E., & McCullough, R. (2021). Exploring counselor experiences of training in relational depth: An interpretative phenomenological inquiry. *Journal of Counseling & Development*, 99(1), 84–95. <https://doi.org/10.1002/jcad.12356>
- Rogers, C. R. (1957). The necessary and sufficient conditions of therapeutic personality change. *Journal of Consulting Psychology*, 21(2), 95.
- Roos, J., & Werbart, A. (2013). Therapist and relationship factors influencing dropout from individual psychotherapy: A literature review. *Psychotherapy Research*, 23(4), 394–418. <https://doi.org/10.1080/10503307.2013.775528>
- Tryon, G. S. (1990). Session depth and smoothness in relation to the concept of engagement in counseling. *Journal of Counseling Psychology*, 37(3), 248–253. <https://doi.org/10.1037/0022-0167.37.3.248>
- Turner, M. J., Slater, M. J., Dixon, J., & Miller, A. (2018). Test-retest reliability of the irrational performance beliefs inventory. *European Journal of Sport Science*, 18(1), 123–129. <https://doi.org/10.1080/17461391.2017.1411527>
- United Kingdom Census. (2011). *Male and female usual resident populations by single year of age—Table Numbers PP02UK and PP03UK*. <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/2011censuspopulationestimatesbysingleyearofageandsexforlocalauthoritiesintheunitedkingdom>

- United States Census. (2010). *Age and sex composition* (Brief Number C2010BR-03). <https://www.census.gov/population/www/cen2010/glance/2010CensusBriefs.html>
- United States Census. (2010). *Overview of race and Hispanic origin* (Brief Number C2010BR-02). <https://www.census.gov/population/www/cen2010/glance/2010CensusBriefs.html>
- DeVellis, R. F., & Thorpe, C. T. (2021). *Scale development: Theory and applications*. Sage.
- Wampold, B. E. (2021). Healing in a social context: The importance of clinician and patient relationship. *Frontiers in Pain Research*, 2, 684768. <https://doi.org/10.3389/fpain.2021.684768>
- Weger, Jr., H., Castle Bell, G., Minei, E. M., & Robinson, M. C. (2014). The relative effectiveness of active listening in initial interactions. *International Journal of Listening*, 28(1), 13–31. <https://doi.org/10.1080/10904018.2013.813234>
- Wiggins, S., Elliott, R., & Cooper, M. (2012). The prevalence and characteristics of relational depth events in psychotherapy. *Psychotherapy Research*, 22(2), 139–158. <https://doi.org/10.1080/10503307.2011.629635>
- Zarzycka, B., Jankowski, T., Szostek, D., Di Malta, G., & Cooper, M. (2022). Relational depth from the perspective of the psychotherapy dyad: Psychometric properties of the Relational Depth Frequency Scale. *Psychotherapy Research: Journal of the Society for Psychotherapy Research*, 32, 910–921. <https://doi.org/10.1080/10503307.2022.2038803>

How to cite this article: Di Malta, G., She, Z., Raymond-Barker, B., & Cooper, M. (2023). Extended validation of the Relational Depth Frequency Scale: Retest reliability, divergent and criterion validity, and measurement invariance in UK- and US-stratified samples. *Journal of Clinical Psychology*, 1–13. <https://doi.org/10.1002/jclp.23525>