

CHAPTER 8

MEASURING THE RELATIONSHIP CONDITIONS IN PERSON-CENTRED AND EXPERIENTIAL PSYCHOTHERAPIES: PAST, PRESENT, AND FUTURE

ELIZABETH FREIRE & SOTI GRAFANAKI

Rogers' hypothesis of the 'necessary and sufficient conditions of therapeutic personality change' (1957) was formulated within the framework of logical positivism, which comprises the view that any construct to be investigated scientifically must be 'operationally defined' (Bickhard, 1992; Green, 1992). Rogers formulated his theory of the role of the therapeutic relationship in the promotion of personality change in a way that he believed would permit his theory to be tested with the traditional scientific methods of quantitative research. Therefore, for each of the relationship conditions, Rogers provided an operational definition that, according to his view, would allow an appropriate 'test' of the theory. These operational definitions of empathy, unconditional positive regard, and congruence, together with the overall positivist framework of Rogers' hypothesis, lured many researchers to the challenging task of developing instruments to measure these constructs. Hence, Rogers' hypothesis opened up an extraordinary new field of scientific exploration and throughout the next two decades the amount of research in the field of psychotherapy inspired by his formulation was unprecedented (Horvath, 1984; Patterson, 1984; Wyatt, 2001).

This chapter will review the instruments that have been developed over the years to investigate the therapist's facilitative conditions postulated by Rogers. There are a large number of instruments that were developed from within psychotherapeutic traditions other than person-centered or experiential and that have subscales that relate to more or less extent to some of the Rogers' relationship conditions (for instance, the 'Warmth' and 'Friendliness' subscales of the Vanderbilt Psychotherapy Process Scales, VPPS). However, considering the focus of this book and space limitations, we decided to restrict our review to instruments that were developed specifically from within the person-centered and experiential (PCEP) traditions.

We will present a brief description of each of these measures, alongside their historical context and psychometric properties. The main strengths and limitations of these measures will also be discussed. As a conclusion of the review, we will suggest future research directions for the investigation of the therapeutic relationship in PCE psychotherapies.

I. INSTRUMENTS THAT MADE HISTORY

Two instruments developed in the 60s played a fundamental role in the history of the person-centered and experiential psychotherapies: the *Barrett-Lennard Relationship Inventory* (BLRI) and the Scales for *Therapist Accurate Empathy, Nonpossessive Warmth, and Genuineness* (more known as the Truax Scales). Rogers himself was to some extent involved in the development of both measures, and most of the research that was developed out of Rogers' conditions theory used either one of these instruments (Barrett-Lennard, 1998).

1. Barrett-Lennard Relationship Inventory (BLRI)

In 1956, Barrett-Lennard was a graduate student at the Counseling Center of the University of Chicago looking for a topic for his doctoral thesis, when Rogers first circulated his theoretical formulation of the relationship conditions (one year before its publication). For his doctoral research, Barrett-Lennard decided to test Rogers' theory with actual clients in therapy (Barrett-Lennard, 1959). However, there were yet no measures of the therapist-to-client relationship conditions and then Barrett-Lennard had to 'invent them from the ground up' (Barrett-Lennard, 2002, p. 65).

Barrett-Lennard reasoned that the relationship 'as experienced by the client would be most crucially related to the outcome of therapy' (Barrett-Lennard, 2002, p. 67). Consequently, he decided to focus his instrument on the client's perceptions of the therapist's attitudes in the relationship, supplemented by the therapist's views of his/her own responses.

Description of the instrument

The BLRI comprises four subscales: 'Empathic Understanding', 'Level of Regard', 'Unconditionality', and 'Congruence'. Barrett-Lennard (1962) considered that the concept of Unconditional Positive Regard (UPR) could not be treated as a unitary dimension or single variable, and therefore he separated UPR into two distinct variables: 'Level of Regard' and 'Unconditionality'. In the initial version of the instrument, Barrett-Lennard (1962) had included a fifth variable called 'Willingness to be known' but the results for this variable were ambiguous and he decided to drop it from later versions of the inventory. However, some elements of this scale were absorbed into the Congruence dimension (Barrett-Lennard, 1978, 1986).

The BLRI is structured as a self-report questionnaire, with a six-point bipolar rating scale ranging from -3 ('NO, I strongly feel that it is not true') to +3 ('YES, I strongly feel that it is true'). The 64-item BLRI (Barrett-Lennard, 1978), the version most widely used today (Barrett-Lennard, 1998; 2003), contains 16 items (8 positively worded and 8 negatively worded) for each of the four sub-scales. Examples of items from the 64-item client form (Other-to-Self, or OS) are presented in the table below.

Clients are asked to mentally insert the name of the therapist in the underlined space in each item.

SUBSCALE	ITEM
37. Level of Regard (+)	_____ is friendly and warm toward me.
33. Level of Regard (-)	_____ just tolerates me.
30. Empathic Understanding (+)	_____ realises what I mean even when I have difficulty in saying it.
58. Empathic Understanding (-)	_____’s response to me is usually so fixed and automatic that I don’t get through to him/her.
51. Unconditionality (+)	Whether thoughts and feelings I express are ‘good’ or ‘bad’ makes no difference to _____’s feeling toward me.
11. Unconditionality (-)	Depending on the way I am, _____ has a better (or worse) opinion of me sometimes than at other times.
12. Congruence (+)	I feel that _____ is real and genuine with me.
52. Congruence (-)	There are times when I feel that _____’s outward response to me is quite different from the way he/she feels underneath.

The items in the therapist’s form (‘Myself-to-the-Other’, or MO) are worded in the first person for therapists to describe their response to their clients. These items are equivalent to the items in the client’s form (Barrett-Lennard, 1986). However, this equivalence is not exact because that would make the items sound ‘unnatural’ (Barrett-Lennard, 2002, p.71). The following examples (see over) of the therapist’s form (MO) correspond to like-numbered items in the client’s form (OS) listed above.

Revisions

The first version (1962) of the BLRI consisted of 85 items, but since then the instrument has undergone a number of modifications, which have resulted in a considerable reduction in the number of items. These modifications have primarily been directed toward enhancing the wording of the items and reducing response bias by balancing positively and negatively stated items. However, the essential structure and rationale of the various versions are identical to the original one (Barrett-Lennard, 2002; 2003).

MEASURING THE RELATIONSHIP CONDITIONS IN PCEP

SUBSCALE	ITEM
37. Level of Regard (+)	I feel friendly and warm toward _____ .
33. Level of Regard (-)	I put up with _____ .
30. Empathic Understanding (+)	I can tell what _____ means, even when he/she has difficulty in saying it.
58. Empathic Understanding (-)	I often respond to _____ rather automatically, without taking in what he/she is experiencing.
51. Unconditionality (+)	Whether _____ is expressing 'good' thoughts and feelings, or 'bad' ones, does not affect the way I feel toward him/her.
11. Unconditionality (-)	Depending on _____'s actions, I have a better opinion of him/her sometimes than I do at other times.
12. Congruence (+)	I feel that I am genuinely myself with _____.
52. Congruence (-)	There are times when my outward response to _____ is quite different from the way I feel underneath.

The only substantial alteration of the BLRI has been the 40-item version, which has 10 items for each subscale (Barrett-Lennard, 2002). In this version, Barrett-Lennard not only dropped some items, but he also merged some others and reversed the positive/negative wording of a few items. These modifications were based more on 'experience and judgement' than on psychometric analysis of the items (Barrett-Lennard, 2002, p. 73).

Moreover, many distinct adaptations of the main 64-item and 40-item forms have been developed for particular uses or for specific populations. There are BLRI forms developed for students/teachers, children, groups, dyads, 'relational life space', supervisory relationships, nurse/patient, and doctor/patient relationships. Other further developments include an observer form (O-64) and a form for group members outside of therapy (OS-G-64) (Barrett-Lennard, 1984, 1998, 2002, 2003).

Other researchers have also added items to the original BLRI for the purposes of their own investigation. For instance, Lietaer (1976) added items related to 'directivity' in his Dutch-language translation, and Cramer (1986a) added an 'advice-given' scale to the BLRI in his study.

Validity

The initial items in the BLRI were derived from Rogers' (1957) paper and from the Relationship Q-Sort (Bown, 1954). The content of these items were revised following discussions with the staff members at the University of Chicago Counseling Center. According to Barrett-Lennard (1962), 'the preparation of items involved constant interaction between theory and operational expression and resulted in a continuous growth and progressive refinement of meaning relating to each concept' (p. 6). The construct validity of the BLRI is also supported by the formal content-validation procedure carried out to eliminate non-differential items. Five qualified judges (Rogers 'might have been' one of them¹) analyzed and checked carefully each item in order to eliminate items that did not express the variable they were designed to represent (Barrett-Lennard, 1978). The subscales were derived using a combination of item analysis and rational-theoretical considerations (Barrett-Lennard, 1959). Moreover, according to Barrett-Lennard, the considerable range of independent studies that have demonstrated an association between the BLRI and therapy outcome provides substantial evidence of 'predictive construct validity' (Barrett-Lennard, 1998, 2003)

Reliability

In an extensive review of the evidence on the BLRI, Gurman (1977) reported internal reliability data from 14 studies (five in actual therapy settings, four in therapy analogue settings, and five studies on other type of relationships, e.g., teachers, parents, and friends) with differing versions of the instrument. The mean internal consistency reliabilities across these 14 studies (four of them used alpha coefficients, the others used split-half reliabilities) were .91 for Level of Regard, .88 for Congruence, .84 for Empathy, .74 for Unconditionality and .91 for the total score. These results indicate that the 85-item and 64-item forms of the BLRI have high internal reliability.

There are no psychometric data reported on the 40-item version, although the reliability is expected to be a little lower, given the reduced number of items. Thus, Barrett-Lennard recommends the use of the 64-item version 'where length is not a problem' (Barrett-Lennard, 2002, p. 74).

Intercorrelation of the BLRI subscales

Gurman (1977) reviewed 16 studies that reported intercorrelations among the BLRI subscales and concluded that (a) Empathy, Level of Regard, and Congruence present a moderate positive correlation, i.e., these dimensions 'appear to be relatively dependent'; (b) Unconditionality bears a very low (and in one case negative) correlation with the other dimensions (i.e., it is 'quite independent'); and (c) Empathy, Level of Regard, and Congruence are all either moderately or highly correlated to the total score (p. 510).

1. Barrett-Lennard, 2010, personal communication

Factor Analysis

Gurman (1977), after reviewing three studies that factor analysed the BLRI using item inter-correlation, concluded that the BLRI is ‘tapping dimensions that are quite consistent with Barrett-Lennard’s original work on the inventory’ (p. 513). However, he pointed out that more factor-analytic work on the BLRI on actual therapy settings should be undertaken. Almost ten years later, Cramer (1986b) factor analysed the original version of the BLRI and found that the first four factors accounted for 49.5% of the variance and reflected the four subscales postulated by the instrument. However, according to Cramer, half of the items did not “clearly distinguish the four factors”, and he concluded that “further refinement of this questionnaire was necessary to improve its factorial validity” (p. 126).

Further comments

The BLRI has been the most extensively used measure in PCE psychotherapy research. It has been considered the most suitable instrument to test Rogers’ theory of the relationship conditions since it taps into the client’s perceptions of the therapeutic relationship (Asay & Lambert, 2001; Gurman, 1977; Lockhart, 1984; Watson & Prosser, 2002). The BLRI has gained wide reputation and has been translated in many languages, including Arabic, Dutch, French, German, Greek, Italian, Japanese, Korean, Polish, Portuguese, Slovak, Spanish, and Swedish (Barrett-Lennard, 2002). The usage of the BLRI has expanded beyond the psychotherapeutic context to wider applications in other human service contexts and significant personal life relationships (e.g. family, friendship, work, and classroom relationships). An important strength of the BLRI is its extensive use in clinical settings and its validation primarily in actual counselling interactions, rather than analogue settings. However, the use of different forms, modifications of content and response format, and the use of isolated sub-scales (usually empathy) rather than the whole inventory by various researchers have posed significant challenges to its further empirical validation and systematic psychometric assessment (Ponterotto, & Furlong, 1985).

2. Scales for Therapist Accurate Empathy, Nonpossessive Warmth, and Genuineness

The *Scales for Therapist Accurate Empathy, Nonpossessive Warmth, and Genuineness* are observer-rated measures of the relationship conditions. They were developed by Truax and Carkhuff (1967) after they participated in a seminar with Rogers in the early part of 1957. Although the scales were developed ‘closely tied to Rogers’ statements’ on the relationship conditions (Truax & Carkhuff, 1967, p. 43), the identifying labels for the three therapist’s conditions were changed from Rogers’ original formulation. Empathy was changed to ‘Accurate Empathy’, which, according to Truax and Carkhuff ‘contains elements of the psychoanalytic view of moment-to-

moment diagnostic accuracy' (p. 43); Unconditional Positive Regard was changed to 'Nonpossessive Warmth (NW)' because the authors considered that "unconditionality of positive regard does not greatly contribute to outcome"; and Congruence was labelled 'Genuineness' (G), since they considered that "what seemed most related to client improvement was not simply a congruence between the therapist's organismic self and his behaviour or self-concept, but rather the absence of defensiveness or phoniness" (p. 43).

The scales were devised to be applied by trained independent raters, usually to samples of therapy tapes. The 'scoring unit' is typically two to five minutes of therapy interaction randomly extracted from the total therapy session (Bohart, Elliott, Greenberg, & Watson, 2002)

Description of the scales²

- A Tentative Scale for the Measurement of Accurate Empathy (AE)

This is a nine-point anchored rating scale. The range of this scale extends from a low point where the therapist has no awareness 'of even the most conspicuous of the client's feelings' (stage 1) through a middle stage of where the therapists responds 'to all of the client's more readily discernible feelings' (stage 5), to a high point where the therapist 'unerringly responds to the client's full range of feelings in their exact intensity' (stage 9).

- Tentative Scale for the measurement of Non-Possessive Warmth (NW)

This is a five-point anchored rating scale. In the lowest level of the scale, the therapist is 'actively offering advice or giving clear negative regard' (stage 1). At the mid-point, the therapist shows a 'positive caring . . . but it is a semipossessive caring' (stage 3), and at the top level the therapist conveys a prizing of the client uncontaminated by 'evaluations of his behaviour or his thoughts' (stage 5).

- A Tentative Scale for the measurement of Therapist Genuineness of Self-Congruence (G)

This is a five-point anchored rating scale ranging from a very low level where 'there is explicit evidence of a very considerable discrepancy between what he [the therapist] says and what he experiences' (stage 1) through a mid-point where 'the therapist is implicitly either defensive or professional, although there is no explicit evidence' (stage 3), to a top level where 'the therapist is freely and deeply himself in the relationship' and 'his verbalizations match his inner experiences' (stage 5).

2. The Scales for Therapist Accurate Empathy, Nonpossessive Warmth, and Genuine were published in its integrity in Kiesler (1973) at pages 428–431.

Carkhuff revision

The main important revision of these scales was developed by Carkhuff (1969). He shortened the original Accurate Empathy scale to 5 points with the aim of increasing its reliability. The Carkhuff's empathy scale was named *Empathic Understanding in Interpersonal Process Scale* (EU) and it is considered a 'truncated version' of the original Accurate Empathy scale (Engram & Vandergoot, 1978, p. 349). According to Engram and Vandergoot (1978), the overall correlation between the AE and Carkuff's EU scales is very high ($r = .89$). Carkhuff (1969) also added other scales to reflect more active therapy strategies. The Carkhuff *Scales for Assessing Facilitative Interpersonal Counseling* are: *Empathy, Respect, Concreteness, Genuineness and Self-Disclosure, Confrontation, and Immediacy*.

Validity

The construct validity of the Truax and Carkuff scales has been object of much controversy. The only published study presented by Truax (1972) as evidence of the construct validity of these scales was a study undertaken by Shapiro (1968), on which the ratings on the scales were correlated with observer-ratings on a 7-point 'semantic differential scale' of 18 variables that included understanding-not understanding, accepting-rejecting, genuine-false, and good-bad. The 'Accurate Empathy' ratings correlated significantly ($r = .67$) with the understanding-not understanding variable. However, Rappaport and Chinsky (1972) pointed out that the understanding- not understanding ratings correlated even higher with ratings of therapist warmth ($r = .87$) and genuineness ($r = .73$). Moreover, 'Accurate Empathy' correlated higher ($r = .71$) with the variable good-bad than it did with understanding- not understanding. These results are discussed by Rappaport and Chinsky as evidence that the Accurate Empathy ratings represent a 'more general therapist quality' than accurate empathy (p. 401). Truax (1972) also cited positive correlations between Accurate Empathy and various measures of therapeutic outcome as support for the construct validity of this scale. Chinsky and Rappaport (1970), however, contended that correlational data between outcome and a given variable does not provide evidence of the construct validity of that variable.

Furthermore, the discriminant validity of the Accurate Empathy scale was heavily challenged by the results of a study that compared ratings of tapes with and without client statements (Truax, 1966). In this study, samples of tape-recorded therapeutic sessions were edited so as to remove client's statements. Ratings were then made on both sets of tapes (i.e., the tapes with therapists' and clients' statements, and the edited tapes containing only the therapists' statements). No significant differences were found in the ratings of accurate empathy and nonpossessive warmth between the edited and unedited tapes. Moreover, these two sets of Accurate Empathy ratings were highly correlated ($r = .68$).

Truax (1972) presented these findings to argue that therapist's responses can be easily rated independent of the client's responses. According to Truax, the 'Accurate Empathy' raters are specifically trained to listen as much as possible only to the therapist responses and to not be influenced by the patient content' (p. 398). However, the fact that the accuracy of a therapist's empathy could be determined in absence of client statements seems unreasonable: 'How can one assess the accuracy of a therapist's empathy unless there is someone to whom the therapist is responding?' (Chinsky & Rappaport, 1970, p. 380). According to Chinsky and Rappaport (1970), the most parsimonious explanation for these findings is that 'raters are responding to some quality of the therapist (perhaps voice quality, tone, inflection, or language style) or that more general therapist trait is being measured not implied in the definition of Accurate Empathy' (p. 380).

Caracena and Vicory (1969) also questioned the construct validity of the AE scale after they found that ratings on AE were significantly correlated to the number of words spoken by the therapist. They concluded that raters on the Accurate Empathy scale depend on superficial objective therapist behaviours, rather than on information about 'an abstract variable such as empathy' (p. 514). Furthermore, Barrow (1977) suggested that ratings on the Truax scales are formulated with the 'aid of ground rules that are not explicitly defined by the scales and that might differ from one rating team to another' (p. 659).

Finally, many studies did not find a significant correlation between client-ratings of empathy and the ratings on the Accurate Empathy scale (e.g., Burstein & Carkhuff, 1968; Caracena & Vicory, 1969; Hansen, Moore, & Carkhuff, 1968), and these findings represent a further challenge to the construct validity of the Accurate Empathy scale.

Reliability

Truax and Carkuff (1967) estimated the inter-rater and inter-item reliability coefficients from 28 studies involving a variety of therapist and client populations. The range of inter-rater reliability values reported for the Accurate Empathy, Nonpossessive Warmth, and Genuineness scales were .43-.79, .48-.84, and .40-.62, respectively, and the median reliability values were .95, .77, and .72, respectively. Thus they concluded that the scales showed a moderate to high degree of reliability.

However, these reliability findings were also object of much dispute. First, Rappaport and Chinsky (1972) pointed out that these reliability scores were 'spuriously inflated' (p. 403) because these studies used a small number of therapists. They considered that if the same therapist was rated more than once by the same rater, then the independence of the ratings was compromised. In addition, they argued that 'the number in the computation of reliability coefficients should be determined by the number of therapists, not the number of patient-therapist interaction samples' (Rappaport & Chinsky, 1972. p. 381).

In response to this criticism, Beutler, Johnson, Neville, and Workman (1973) estimated the consistency of therapists' ratings across sessions and patients and found that accurate empathy is not a stable quality of the therapist, but rather a reflection of the therapist-client dyad. They concluded that these findings supported the use of reliability coefficients based on the number of client-therapist pairs (as done by Truax) rather than on the number of therapists (as suggested by Chinsky & Rappaport, 1970).

Interdependence of the scales

There has been considerable divergence in the report of the inter-correlations between the Accurate Empathy, Nonpossessive Warmth, and Genuineness scales. Initially, Truax et al. (1966) found that Nonpossessive Warmth correlated negatively with Empathy and Genuineness. Subsequently, Truax and Carkhuff (1967) and Rogers, Gendlin, Kiesler, and Truax (1967) found moderate levels of positive correlation between the scales. On the other hand, Garfield and Bergin (1977) found that Genuineness correlated negatively with Accurate Empathy and Non-possessive Warmth. Another study by Barrow (1977) added more divergence to these findings: he obtained very high positive correlations among the three scales (.85 to .93). These large discrepancies could be another indication of the poor reliability and construct validity of the Accurate Empathy, Nonpossessive Warmth and Genuineness scales.

Further comments

The *Scales for Therapist Accurate Empathy, Nonpossessive Warmth, and Genuineness* have been used in most of the research that grew out of the Rogers' conditions theory. The Accurate Empathy scale is the best-known of these scales, and it has been one of the most widely used observer measures of empathy in psychotherapy research (Bohart et al., 2002; Feldstein & Gladstein, 1980). However, in spite of its widespread use, it seems that the validity of these scales have not been established.

3. Other measures

During the first two decades after Rogers' publication of the relationship conditions theory, a few other measures were developed, but they were rarely used. Halkides (1958) developed the first observer-rated measure of the relationship conditions, a few years before the Truax and Carkhuff scales. Halkides' scale was developed for her doctoral dissertation at the University of Chicago, but it remained unpublished. Barrett-Lennard (1998) commented that 'although seldom acknowledged, this study opened the way to a plethora of subsequent work employing judges' ratings' on the relationship conditions (p. 264).

Two other observer-rated measures of empathy derived from Roger's theory were developed in the 70s: Cochrane's (1974) measure of *Empathic communication*

and Lister's (1970) *Scale for the measurement of empathic understanding*. Cochrane's measure was only tested through a therapeutic analogue procedure and Lister's study remained unpublished. Truax also developed a client-rated version of his own scales, the *Truax Relationship Inventory* (TRI) (Truax & Carkhuff, 1967) but this measure has seldom been used in psychotherapy research.

II. FURTHER DEVELOPMENTS

Inspired by the seminal work of these pioneers, other researchers from the 80s onwards embraced the challenge of creating new instruments in order to promote the development of further process-outcome studies on Rogers' relationship conditions. During the last decades, two new observer-rated measures of empathy, the *Multidimensional Response Empathy Scale* (Elliott, Filipovich, Harrigan, Gaynor, Reimschuessel, & Zapadka, 1982) and the *Measure of Expressed Empathy* (Watson & Prosser, 2002), and a rating system for studying nondirective client-centered interviews (Wilczynski, Brodley & Brody, 2008) have been developed.

4. Revised Multidimensional Response Empathy Scale

Elliott, Filipovich, Harrigan, Gaynor, Reimschuessel, and Zapadka (1982) developed a multicomponent rating scale for empathy that focuses on particular counsellor verbal responses. The authors considered that empathic responding is composed of multiple components, and that other measures of expressed empathy "fail to measure components of empathy that do predict client perceptions [of empathy]" and that these measures have "inadequate specification of the empathy construct in terms of specific counsellor behaviors" (p. 380). Therefore, Elliott et al. constructed an observer-rated instrument, the *Response Empathy Rating Scale*, that aimed to measure expressed empathy associated with particular counsellor responses and that divided empathy into a number of components. The instrument was originally constructed based on the components of the Lister Empathy Scale (i.e., *internal frame of reference, perceptual inference, accurate perceptual inference, immediacy, emphasis on personal perceptions, use of fresh words, appropriate voice, and pointing to exploration*). Elliott et al (1982) further defined these components in the development of the *Response Empathy Rating Scale*. The revised version of the instrument (Elliott, Reimschuessel, Filipovich, Zapadka, Harrigan, & Gaynor, 1981) consists of eight components: *Client Feelings, Perceptual Inference & Clarification, Centrality of Topic, Expressiveness, Collaboration, Verbal Allowing vs. Crowding, Exploration, and Impact on Exploration*.

Examples of items are listed below:

- *Clients Feelings*. To what extent does the therapist address the client's feelings?
- *Perceptual Inference & Clarification*. Does the therapist make inferences to tell the client something the client hasn't said yet to add to the client's frame of reference or to bring out implications?
- *Exploration*. Does the therapist actively encourage client's exploration by the content of what therapist says?

All items are rated on 5-point behaviourally anchored rating scales. The scoring unit is a single therapist's response.

Validity

The validity of the first version (1982) of the instrument was tested using client's ratings on how understood they felt during the session. Client's ratings were obtained with the use of 'Interpersonal Process Recall' technique. Immediately after the session, 28 clients viewed the videotape of their session and rated how understood they felt at particular counsellor responses, using a 6-point scale ranging from 1 ('not at all understood') to 6 ('extremely understood'). The correlation between client's perceptions of being understood and the ratings on *Response Empathy Rating Scale* was small ($r = .10$ to $.27$). However, when client's ratings were aggregated into episodes or sessions the correlation increased to a medium-sized effects on the total scale ($r = .42$).

Reliability

Psychometric data is available only for the first version of the instrument (Elliott et al., 1982). Five raters participated in the study. Interrater reliabilities (Cronbach's alpha) for all but two components were very good, ranging from .75 to .91. Two components failed to reach the .70 criterion. In the revised version of the instrument one of these components was dropped (*Voice*) and the other (*Manner*) was split into two separate components (*Collaboration* and *Exploration*). Inter-item reliability for the total scale was .82, which indicates that the instrument has a high degree of internal consistency.

Factor analysis

Using a principal component method with Varimax rotation, two factors were extracted, accounting for 62% of the total variance. The first of these factors was described as 'Depth-Expressiveness' and the second factor as 'Empathic Exploration'.

Further comments

The instrument was extensively revised after the publication of these reliability and validity results. However, no further tests were reported on the revised version of the

instrument. Although it would be expected that the reliability and validity results would improve in relation to the first version, the fact that these results were not reported is a limitation of the instrument.

5. Measure of Expressed Empathy (MEE)

The *Measure of Expressed Empathy* is an observer-rated measure of therapist communicated empathy developed by Watson and Prosser (2002). The construction of the instrument was based on behavioural correlates of empathy identified in previous research: therapists' verbal and non-verbal behaviours, speech characteristics, and response modes. The MEE uses a nine-point Likert scale ranging from 0 ('Never') to 8 ('All the time'). Items are rated on five-minute video-taped therapy segments. The first version of the instrument (Watson, 1999) had 22 items, but a shorter version of the instrument with 9 items was developed subsequently (Watson & Prosser, 2002). Examples of items from the revised MEE are listed below:

- Do the therapist's responses convey an understanding of the client's cognitive framework and meanings?
- Does the therapist look concerned?
- Is the therapist responsive to the client?
- Is the therapist's voice expressive?

Validity

The validity of the scale was tested by correlating the MEE ratings with clients' ratings on the Barrett-Lennard Relationship Inventory (BLRI). However, the original (22-item) scale failed to correlate significantly with client's BLRI empathy ratings. Therefore, analysis of the correlations of individual components led to a refinement of the instrument. Scores on the MEE were recalculated using only the ratings of the nine items that had correlated with the BLRI at the .36 level or higher (i.e., *Voice conveys concern*, *Captures intensity of feelings*, *Emotional words*, *Attuned*, *Cognitive understanding*, *Warmth*, *Looks concerned*, *Expressive voice*, and *Responsive*). The revised MEE composed of these 9 items showed an overall significant correlation with BLRI empathy ratings of .66 (Watson & Prosser, 2002).

Reliability

The psychometric properties of the 22-item version of the MEE have been tested with the use of archival data from the York Psychotherapy of Depression Research Project (Greenberg & Watson, 1998). Forty video-tape 20-minute segments from the middle of the 7th and 16th sessions of twenty clients (i.e., 2 sessions for each client) were randomly selected. Each 20-minute segment was further divided into 5-minute

segments that were then rated on all 22 items of the scale. In total, eight sets of ratings were collected for each client (Watson & Prosser, 2002).

The interrater reliability for most of the items ranged from .80-.90, although four of the 22 items did not reach the recommended .7 level (*Rate of speech, Interruptions, Clarity, and Emotional words*). The inter-item reliability (Cronbach's alpha) for the overall scale (components summed across raters) was .88, which indicates a high degree of internal consistency (Watson & Prosser, 2002).

Further comments

The significant correlation obtained between the revised 9-item MEE and the BLRI client's empathy ratings is a very promising result. However, further replication and testing of this revised version of the MEE with a larger sample is advisable for effectively assessing the psychometric properties of this instrument (Watson & Prosser, 2002). According to Watson and Prosser (2002), the MEE can be also a useful tool in the training of empathy. Learning to use the scale can help trainees to improve their empathy and appreciate different aspects of counsellor behaviour that have an impact on expressed levels of empathy.

6. Nondirective Client-Centered Rating System

The *Nondirective Client-Centered Rating System* (Wilczynski, Brodley, & Brody, 2008) is an observer-rated scale that takes into account the therapists' nondirective intention or attitude. This instrument was initially developed by Brodley and Brody (1990) for study of the psychotherapy sessions conducted by Carl Rogers that were available through audio and video recordings, film, and transcripts. The primary aim of this scale is to distinguish the therapist's nondirective intentions or attitudes from the directive ones.

The responses are rated into five major mutually exclusive categories, distinguishing five different apparent therapists' intentions: *Empathic Understanding Response, Therapist Comment, Interpretation/explanation, Therapist Agreement, and Leading Question*. Responses are rated as *Empathic Understanding Response* when the 'therapist's apparent intention is to check his or her understanding of the experience, feelings, or point of view immediately expressed by the client' (Wilczynski et al., 2008, p. 39); responses are rated as *Therapist Comment* when the therapist's apparent intention is 'to offer his or her observation or opinion, or to express the therapist's own feelings about the client or a general point' (p. 44); responses are rated as *Interpretation/explanation* when the therapist's apparent intention is 'to explain the client to the client' (p. 45); a *Therapist Agreement* is identified when the therapist's apparent intention is to verbally agree with the client; and a *Leading Question* is identifiable by the presence of the therapist's apparent intention, 'in the form of a question, to direct the client's feelings, responses, thoughts, or considerations' (p.

47). The mean percentage of agreement between raters in the latest version of this instrument has been reported 90% (Wilczynski, Brodley, & Brody, 2008); unfortunately, the authors failed to report the reliability values using standard statistics (i.e., Cohen's kappa).

III. DEVELOPMENTS FOR THE FUTURE

In the last few years, new instruments have started to emerge, bringing new perspectives in the investigation of the Rogerian relationship conditions. Geller, Greenberg and Watson (in press) have developed the *Therapeutic Presence Inventory* (TPI), which aims to measure an overarching relationship condition that goes beyond the three core conditions of empathy, unconditional positive regard, and congruence. Two new measures are currently being developed by the research group of the University of Strathclyde in Scotland: The *Therapeutic Relationship Scale* (Sanders & Freire, 2008) that encompasses Roger's therapeutic conditions and incorporates the dimensions of client's deference and therapist's directivity; and the *Person-Centred and Experiential Psychotherapy Scale* (Freire, Elliott, & Westwell, 2010), a competence/adherence measure for person-centered and experiential psychotherapies.

7. Therapeutic Presence Inventory (TPI)

Geller et al (in press) developed two self-report measures of the in-session process and experience of therapeutic presence – the *Therapeutic Presence Inventory*, with a therapist version (TPI) and a client version (TPI-C). These instruments used a model of therapeutic presence developed by Geller and Greenberg (2002) from a qualitative study of interviews with expert therapists. According to this model, the therapist's presence provides not only a 'foundation' for the relationship conditions of empathy, congruence and unconditional positive regard, but it also 'encompasses' them. According to the authors, the 'therapeutic presence prepares the ground for a therapist to be empathic, genuine, and unconditionally accepting' (Geller et al., in press). Moreover, in this model, therapeutic presence is seen as the 'larger condition by which empathy, congruence, and unconditional regard can be expressed', reflecting a global quality that encompasses all relationship conditions and 'yet goes beyond them'.

Description of the measures

The TPI consists of 21 items presented on a 7-point likert-scale ranging from 'completely' to 'not at all'. Ten items reflect the 'Process' aspects of therapeutic presence (*Receptivity, Inwardly Attending, and Extending and Contact*), and 11 items represent the 'Experience' of therapeutic presence (*Immersion, Expansion, Grounding, and Being*

With and For the Client). Examples of items of the TPI are listed below:

- I was aware of my own internal flow of experiencing
- The interaction between my client and I felt flowing and rhythmic
- I was able to put aside my own demands and worries to be with my client
- My responses were guided by the feelings, words, images, or intuitions that emerged in me from my experience of being with my client

The measure of the client perceived therapeutic presence (TPI-C) contains three items with the same 7-point likert scale used for the therapist TPI:

1. My therapist was fully there in the moment with me
2. My therapist's responses were really in tune with what I was experiencing in the moment
3. My therapist seemed distracted

Intercorrelation between TPI and TPI-C

Geller et al. (in press) found that the relationship between therapists' and clients' ratings on the presence measures was not strong. Although there was a statistically significant correlation between the TPI and the TPI-C, this finding was a result of the large sample size ($n = 358$), as the correlation was small and not clinically significant ($r = .20$).

Reliability and Validity

The item-total reliability (Cronbach's alpha) computed for the TPI and TPI-C were .94 and .82 respectively, which indicate that these measures have good internal reliability.

Construct validity was established in the process of construction, selection and refinement of the TPI items. The 21 items of the final version of the TPI were selected from an initial pool of 32 items based on ratings of nine expert therapists. Only items that discriminated between sessions perceived by experts as having high presence and low presence, and were rated by the experts as clearly reflecting the experience of presence, were selected.

Convergent validity was further assessed by the relationship between therapists' ratings on the TPI and the therapist's ratings on the Barrett-Lennard Relationship Inventory (BLRI). Correlations between TPI scores and the BLRI subscales of Empathy, Congruence, Level of regard, and Unconditionality were .59, .41, .34, and .20 respectively (all statistically significant).

In terms of concurrent validity of the TPI-C, Geller et al (in press) found that

clients “reported a positive change following a therapy session where they felt their therapist was present with them, regardless of theoretical orientation of the therapy”. Also, when clients rated their therapist as present with them, the therapeutic alliance was also rated as positive.

Factor Analysis

Items on the TPI and TPI-C were submitted to a principal-axis analysis. On the TPI, the 21 items fell under one main factor that accounted for 50.01% of the variance. On the TPI-C, the three items resulted in one factor that accounted for 67.59% of the variance. These findings indicated that both the TPI and the TPI-C are unidimensional measures, i.e., they reflect one single factor, as predicted - the therapeutic presence - which support the construct validity of the measures.

Further comments

The *Therapeutic Presence Inventory* represents a very important contribution to the development of a new frontier of research on person-centred and experiential therapies. As Rogers pointed out in an interview later in his life (Baldwin, 2000): ‘perhaps it is something around the edges of those conditions that is really the most important element of therapy – when my self is very clearly, obviously present’ (p. 30). Therefore, an instrument that aims to tap into these ‘edges’ of the relationship conditions certainly brings a promising contribution to the development of research on person-centered and experiential psychotherapies.

8. Relational Depth Inventory (RDI-C)

The *Relational Depth Inventory* (RDI-C) is an instrument developed by Wiggins, Elliott, and Cooper (2010) that aims to assess the experience of *relational depth* in therapy, as defined by Mearns and Cooper (2005). The concept of *relational depth* was initially proposed by Mearns (1996, 1997) as the coming together of all six of Rogers’ relationship conditions at it best, representing a distinctive hallmark of person-centred therapy. Later on, Mearns and Cooper (2005) defined *relational depth* 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’ (p. xii).

Description of the measure

The Relational Depth Inventory (RDI-C) consists of 24 items presented on a 5-point likert-scale ranging from ‘not at all’ to ‘completely’. Clients are initially asked to describe a ‘particularly helpful moment or event’ that they had during a therapy session. Then, they are asked to rate ‘how accurately’ each of the 24 items fits with their experience of this event. Examples of items are listed below:

- I felt I was going beyond my ordinary limits
- I felt a warm personal bond between myself and my therapist as fellow human beings
- I felt more alive
- I felt my therapist and I were equal

Reliability and Validity

The inter-item reliability (Cronbach's alpha) for the RDI-C was .93, which indicates a high degree of internal consistency (Wiggins, 2010). The construct validity of the instrument was sought through the process of construction, selection and refinement of the items. The initial pool of items was developed out of client and therapist's descriptions of relational depth experiences in therapy collected by Cooper (2005) and Knox (2008). The pilot version of the RDI contained 64 items and it was responded online by 189 therapists and 152 clients. The events described by the participants were subsequently rated by three judges to the extent that they represented a 'relational depth' event according to Mearns and Cooper (2005) definition cited above. The correlations between the RDI items and these ratings of 'relational depth presence' ranged from -.37 to .47. The items whose correlations were \geq .30 were selected for the second version of the instrument (Wiggins, Elliott, and Cooper, in press).

Discriminant validity was assessed by the relationship between clients' ratings on the RDI-C and the Working Alliance Inventory (WAI-SR). Wiggins (2010) reported a significant but moderate correlation between RDI-C and WAI-SR of .34, which seems to indicate that 'working alliance' and 'relational depth' are different although related constructs. Moreover, RDI-C scores were found to predict therapeutic outcome, a finding that supports the concurrent validity of the instrument (Wiggins, 2010).

Factor Analysis

An exploratory factor analyses using principal component method with Varimax rotation found two factors that accounted for 47% of the total variance. The first of these factors was described as 'Therapist genuine/available' and the second factor was described as 'Transcendence'. These factors were named according to their most highly loaded items: the first factor was named after the item '*I felt my therapist was being genuine with me*' and the second factor was named after the items '*The atmosphere was kind of awesome*' and '*I felt a kind of magic happened*'.

Further comments

The RDI-C is an instrument still in development, as part of the Ph.D studies of Susan Wiggins. A current limitation of the measure that might be addressed in the course of its future developments is that it aims to evaluate the 'depth' of one single

moment in therapy – a moment defined by the client as helpful – rather than the overall level of ‘depth’ attained across the relationship. Also, although the process of construction and validation of the RDI-C was very complex and elaborated, it is also not entirely clear what construct the RDI-C is really measuring. The factor analysis would seem to suggest that the RDI-C is not measuring a unidimensional construct, as the definition of ‘relational depth’ would suggest. In fact, the RDI-C items encompass a width of characteristics that goes from mundane aspects as ‘I felt my therapist respected me’ to out-of-ordinary experiences as ‘I felt as if time had stopped’; some of which are not clearly related to the concept of ‘relational depth’ as defined by Mearns and Cooper. This observation is supported by the moderate correlations found between the RDI-C items and the ratings of ‘relational depth presence’.

9. Therapeutic Relationship Scale (TRS)

Sanders and Freire (2008) are developing a new instrument, the *Therapeutic Relationship Scale* (TRS), which aims to capture the client’s and therapist’s experience of the quality of their relationship, focusing particularly upon the core conditions of empathy, unconditional positive regard and congruence, the experience of client’s deference, and therapist’s directivity. The authors recognize the importance and value of other measures of the Rogerian relationship conditions, particularly the Barrett-Lennard Relationship Inventory, however they consider that the BLRI presents some limitations that they aim to overcome with their new measure. They consider that the BLRI is too long, and many of its items are too complex with some items asking more than one question. Also the type of bipolar rating scale used in the BLRI is no longer recommended in modern theories of measure development (Yorke, 2001). Moreover, the TRS aims to include other dimensions of the therapeutic relationship that goes beyond Rogers’ core conditions (e.g., ‘Dynamics of power’, and ‘Trust/feeling safe’).

The pilot version of the TRS contains 27 items, which aims to cover the following dimensions: *Empathy, Acceptance, Warmth, Collaboration/Partnership, Trust/Feeling Safe, Genuineness, Dynamics of power, and Self-disclosure*. Examples of items of the TRS are listed below:

- I am happy with the way that my therapist and I are working together (*Collaboration/Partnership*)
- I felt that I could say everything that was in my mind (*Trust/Feeling safe*)
- I felt it was ok for me to correct or disagree with my therapist (*Dynamics of power*)
- My therapist revealed something personal about themselves to me (*Self-disclosure*)

There is also a parallel therapist's form of the TRS.

Some of the items on this instrument were derived from items of the BLRI, and some other items were derived from Rennie's (1994) study on client's experience of deference in psychotherapy. The instrument uses a 4-point rating-scale ranging from 'not at all' (0) to 'a great deal' (3). The reliability and validity of the TRS is currently being tested with clients of the Strathclyde Therapy Research Centre.

10. Person-Centered and Experiential Psychotherapy Scale (PCEPS)

The assessment of 'treatment integrity' is an essential component of psychotherapy trials (Waltz, Addis, Coerner, & Jacobson, 1993). The test of 'treatment' integrity includes both an assessment of therapist adherence to the therapy manual and an assessment that the therapy is being performed competently. Waltz et al. (1993) recommended that integrity checks be undertaken through analysis of audio or video recordings of the therapy sessions by independent researchers/practitioners. In view of the absence of an appropriate adherence/competence measure of person-centred and experiential therapies and given the fundamental importance of this kind of measure in the development of efficacy trials of person-centred/experiential therapies, Freire, Elliott, and Westwell (2010) are developing a new instrument to fill that gap: the *Person-Centred and Experiential Psychotherapy Scale* (PCEPS). This instrument consists of two subscales corresponding to (a) Person-centred Process and (b) Experiential process.

The *Person-centred Process* scale consists of 10 items: *Client Frame of Reference*, *Client Track*, *Core Meaning*, *Client Flow*, *Warmth*, *Clarity of Language*, *Content Directiveness*, *Accepting Presence*, *Judgment*, and *Ungenuineness*. Examples of items of this scale are:

- Do the therapist's responses convey an understanding of the client's experiences as the client themselves understands or perceives it? (*Client Frame of Reference*)
- Do the therapist's responses reflect the core, or essence, of what the client is communicating or experiencing in the moment? (*Core Meaning*)
- Do the therapist's responses convey judgments of the client's experiences behaviour? (*Judgment* – reversed item)

The *Experiential Process* scale consists of 9 items: *Collaboration*, *Experiential Specificity*, *Emotion Focus*, *Articulation of Emotions*, *Core Client Experiences*, *Emotion Regulation*, *Resolution*, *Client Self-development*, and *Therapeutic Indicators*. Examples of items of this scale are:

- Does the therapist actively try to facilitate client-therapist collaboration and mutual involvement in the goals and tasks of therapy? (*Collaboration*).

- Does the therapist actively try to help the client achieve and maintain an optimal level of emotional arousal for exploring their feelings? (*Emotion Regulation*)
- Do the therapist's responses aim to help the client identify, persist in, and resolve key goals or tasks, within and across sessions? (*Resolution*)

The reliability and validity of this new measure is currently being tested using one-hundred and eighty audio-recorded 10-minute segments of therapy sessions systematically selected from the archive of taped therapy sessions of the Strathclyde Therapy Research Centre.

IV. MEASURING THE RELATIONSHIP CONDITIONS: A PARADOX?

For more than five decades since the publication of Rogers' hypothesis of the 'necessary and sufficient conditions' for therapeutic personality change, many researchers have attempted to tackle the challenge of testing its validity. The development of instruments to 'measure' the therapist-provided conditions of empathy, unconditional positive regard and congruence, and the client's perception of these conditions became the first step or the essential tool for any viable test of Rogers' hypothesis. However, how close have we come to really measuring these 'core conditions'? Or perhaps that is the wrong question, perhaps one should ask: are these conditions in fact measurable?

In a traditional, positivist approach to psychology there is an omnipresent credo that: "Whatever exists at all exists in some amount. To know it thoroughly involves knowing its quantity" (Thorndike's, 1918, p. 167). However, is that true that all psychological attributes can be measured? In quantitative science, like physics, attributes such as temperature, length, velocity, etc., are taken to be measurable because they have a distinctive kind of internal structure: a quantitative structure. According to Mitchel (1997, 2000), there is no logical necessity that any attribute should have a quantitative structure³, in fact, conceptualizing an attribute as quantitative is a scientific hypothesis to be tested in itself.

3. A 'quantitative structure' is characterized by the following: 'A range of instances of an attribute Q constitutes continuous quantity if and only if for any a and b in Q, (1) One and only one of the following is true: (i) $a = b$, (ii) There exists c in Q such that $a = b + c$; (iii) There exists c in Q such that $b = a + c$; (2) $a + b = b + a$; (3) $a + (b + c) = (a + b) + c$, (4) If $a > b$, there exists c in Q, such that $a > c > b$; (5) every non-empty subset of Q that has an upper bound has a least upper bound. It is because of this quantitative structure, that magnitudes of a quantity are measurable, that is, they stand in relations to one another that can be expressed as real numbers' (Mitchel, 1997: 356).

The paradigm that underpins measure development in psychology is characterised by the assumption that measurement is simply 'the assignment of numerals to objects or events according to rule' (Mitchel, 2000, p. 650). However, Mitchel (1997, 2000) points out that this definition of measurement is quite unlike the traditional concept used in the physical sciences. Therefore, he concludes that quantitative psychology is a 'pathology of science', and that psychometrics has a 'methodological thought disorder', since it endorses an anomalous definition of measurement (Mitchel, 2000, p. 639). Perhaps Mitchel's criticisms of the concept of measurement in Psychology could shed some light on the reasons why the endeavor to measure Rogers' relationship conditions seems to always fall short of a complete success. Perhaps trying to measure the relationship conditions is somewhat like a Sisyphian challenge, an absurd and futile task that will never be fully accomplished?

There is an inherent paradox (Freire, 2009) in Rogers' theory, in that 'it was cast in the objectivism of logical positivism while also encompassing foundational phenomenological constructs and an almost 'mystical' view of the therapeutic relationship' (p. 228). Nowadays, we can be critical of the naïve objectivism of Rogers' formulation of the relationship conditions and his search for 'operational definitions' of these conditions. We recognize today that the overall "if-then" framework of his theory is a 'deterministic, mechanistic, and reductionistic account of the complexity, richness and unpredictability of the therapeutic process' (p. 228).

Perhaps the time has come for a new formulation of Rogers' theory on which the qualities of the therapeutic relationship won't be defined as 'attributes' provided by the therapist regardless or independently of the client's active participation and engagement in the therapeutic process (Butler & Strupp, 1986; Grafanaki, 2001, 2002; Stiles & Shapiro, 1989). In this new framework, the qualities of the therapeutic relationship would certainly be best investigated by qualitative methodologies, on which the individual differences and contextual variations would be seen as the 'very effects' to be investigated, and not excluded as 'error variance' as it is the case with quantitative methodologies (Yardley, 2008).

We appreciate the monumental work that has been undertaken by a number of researchers in the development of measures of the relationship conditions that would allow for the testing of Rogers' hypothesis. However, we recognise that this might have been a Sisyphian task, as the therapeutic relationship might not have a quantitative structure necessarily that would permit a rigorous mathematical measurement. Therefore, we recommend that a new framework for the investigation of the therapeutic relationship be developed, in which the client wouldn't be considered a passive recipient of the therapist-provided conditions, and in which the uniqueness and singularity of each therapeutic encounter would be considered as the core and essence of any therapeutic process.

V. CONCLUSION

To end this review with the critical tone above would be extremely unfair with the importance and magnitude of the work developed by Rogers and his successors on the investigation of the therapeutic relationship. Despite the inherent epistemological paradoxes, Rogers' hypothesis of the relationship conditions was a fundamental contribution that revolutionised the field of psychotherapy research. Rogers' theory gave rise to a new paradigm, which emphasized the importance and significance of the therapist's attitudes rather than the application of techniques (Rogers, 1951). In fact, all instruments reviewed in this chapter share this underlying assumption that 'it is the relationship that heals'. These instruments are essential and invaluable tools in the process of gathering evidence of the paramount importance of the Rogerian relationship conditions for effective psychotherapy.

REFERENCES

- Asay, T., & Lambert, M. (2001). Therapist relational variable. In D. Cain & J. Seeman (Eds.) *Humanistic psychotherapies: Handbook of research and practice* (pp. 531–558). Washington DC: APA.
- Baldwin, M. (2000). Interview with Carl Rogers on the use of the self in therapy. In M. Baldwin (Ed.), *The use of self in therapy: 2nd Edition* (pp. 29–38). New York: The Haworth Press.
- Barrett-Lennard, G. T. (1959). *Dimensions of perceived therapist response related to therapeutic change*. Doctoral dissertation, University of Chicago.
- Barrett-Lennard, G. T. (1962). Dimensions of therapist response as causal factors in therapeutic change. *Psychological Monographs*, 76 (43, Whole No. 562)
- Barrett-Lennard, G. T. (1978). The Relationship Inventory: Later development and adaptations. *JSAS Catalog of Selected Documents in Psychology*, 8, 68.
- Barrett-Lennard, G. T. (1984). The world of family relationships: A person-centered systems view. In R. F. Levant & J. M. Shlien (Eds.), *Client-centered therapy and the person-centered approach: New directions in theory, research and practice* (pp. 222–242). New York: Praeger.
- Barrett-Lennard, G. T. (1986). The Relationship Inventory now: Issues and advances in theory, method and use. In L. S. Greenberg & W. M. Pinsof (Eds.), *The psychotherapeutic process: A research handbook* (pp. 439-476). New York: Guilford.
- Barrett-Lennard, G. T. (1998). *Carl Rogers' helping system: Journey and substance*. London: Sage.
- Barrett-Lennard, G. T. (2002). Perceptual variables of the helping relationship: A measuring system and its fruits. In G. Wyatt, G. & P. Sanders P. (Eds.), *Rogers' therapeutic conditions: Evolution, theory and practice. Vol. 4, Contact and Perception* (pp. 65–90). Ross-on-Wye, UK: PCCS Books.
- Barrett-Lennard, G. T. (2003). *Steps on a mindful journey: Person-centred expressions*. Ross-on-Wye, UK: PCCS Books.
- Barrow, J. C. (1977). Interdependence of scales for the facilitative conditions: Three types of correlational data. *Journal of Consulting and Clinical Psychology*, 45, 654–659.
- Butler, L. E., Johnson, D. T., Neville, C. W., & Workman, S. N. (1973). Some sources of

- variance in 'accurate empathy' ratings. *Journal of Consulting and Clinical Psychology, 40*, 167–169.
- Bickhard, M. H. (1992). Myths of science: Misconceptions of science in contemporary psychology. *Theory & Psychology, 2*, 321–337.
- Bohart, A. C., Elliott, R., Greenberg, L. S. & Watson, J. C. (2002). Empathy. In J. C. Norcross (Ed.), *Psychotherapy relationships that work: Therapist contributions and responsiveness to patients* (pp. 89-108). Oxford: Oxford University Press.
- Bown, O. H. (1954). *An investigation of therapeutic relationship in client-centered psychotherapy*. Unpublished doctoral dissertation.
- Brodley, B. T., & Brody, A. (1990, August). *Understanding client-centered therapy through interviews conducted by Carl Rogers*. Paper presented in the panel on Fifty Years of Client-Centered Therapy: Recent Research, at the annual conference of the American Psychological Association, Boston, MA.
- Burstein, J. W., & Carkhuff, R. R. (1968). Objective therapist and client ratings of therapist offered facilitative conditions of moderate to low functioning therapists. *Journal of Clinical Psychology, 24*, 233–236.
- Butler, S., & Strupp, H. (1986). Specific and non-specific factors in psychotherapy: A problematic paradigm for psychotherapy research. *Psychotherapy, 23*, 30–40.
- Caracena, P. F., & Vicory, J. R. (1969). Correlates of phenomenological and judged empathy. *Journal of Counseling Psychology, 16*, 510–515.
- Carkhuff, R. R. (1969). *Helping and human relations: A primer for lay and professional helpers*, Vol. 2, *Practice and research*. New York: Holt, Rinehart, & Winston.
- Cooper, M. (2005). Therapists' experiences of relational depth: A qualitative interview study. *Counselling and Psychotherapy Research, 5*, 87–95
- Chinsky, J. M., & Rappaport, J. (1970). Brief critique of the meaning and reliability of "accurate empathy" ratings. *Psychological Bulletin, 73*, 379–382.
- Cochrane, C. T. (1974). Development of a measure of empathic communication. *Psychotherapy: Theory, Research and Practice, 11*, 41–47.
- Cramer, D. (1986a). An item factor analysis of the revised Barrett-Lennard relationship inventory. *British Journal of Guidance & Counselling, 14*, 3, 314–325.
- Cramer, D. (1986b). An item factor analysis of the original relationship inventory. *Journal of Social and Personal Relationships, 3*(1), 121–127.
- Elliott, R., Reimschuessel, C., Filipovich, H., Zapadka, J., Harrigan, L., & Gaynor, J. (1981). *Revised multidimensional response empathy scale*. Unpublished manuscript, University of Toledo.
- Elliott, R., Filipovich, H., Harrigan, L., Gaynor, J., Reimschuessel, C., & Zapadka, J. K. (1982). Measuring response empathy: The development of a multicomponent rating scale. *Journal of counselling Psychology, 29*, 379–387.
- Engram, B. E., & Vandergoot, D. (1978). Correlation between the Truax and the Carkhuff scales for measurement of empathy. *Journal of Counseling Psychology, 25*, 349–351.
- Feldstein, J. C., & Gladstein, G. A. (1980). A comparison of the construct validities of four measures of empathy. *Measurement and Evaluation in Guidance, 13*, 49–57.
- Freire, E. S. (2009). A quiet revolution... or swimming against the tide? *Person-Centered and Experiential Psychotherapies, 8*, 224–232.
- Freire, E. S., Elliott, R., Westwell, G. (2010, July). *Development and validation of a competence/adherence measure for person-centred and experiential psychotherapies*. Paper presented at the

- 9th World Conference for Person-Centered and Experiential Psychotherapy and Counseling, Rome, Italy.
- Garfield, S. L., & Bergin, A. E. (1971). Therapeutic conditions and outcome. *Journal of Abnormal Psychology, 77*, 108–114.
- Green, C. D. (1992). Of immortal mythological beasts: Operationism in psychology. *Theory & Psychology, 2*, 291–320.
- Greenberg, L., & Watson, J. C. (1998). Experiential therapy of depression: Differential effects of client-centered relationship conditions and process experiential interventions. *Psychotherapy Research, 8*, 210–224.
- Geller, S. M. & Greenberg, L. S. (2002). Therapeutic presence: Therapists' experience of presence in the psychotherapeutic encounter. *Person-Centered & Experiential Psychotherapies, 1*, 71–86.
- Geller, S. M., Greenberg, L. S., Watson, J. C. (in press). Therapeutic Presence: The Development of a Measure. *Psychotherapy Research*.
- Grafanaki, S. (2001). What counselling research has taught us about the concept of congruence: Main discoveries and unresolved issues. In G. Wyatt (Ed.) *Rogers' therapeutic conditions, vol 1, Congruence* (pp.18–35). Ross-on-Wye, UK: PCCS Books.
- Grafanaki, S. (2002). On becoming congruent: How congruence works in person-centred counselling and practical applications for training and practice. In J. Watson, R. Goldman, & M. Warner (Eds.) *Client-centered and experiential psychotherapy in the 21st century* (pp. 278–290). Ross-on-Wye, UK: PCCS books.
- Gurman, A. S. (1977). The patient's perceptions of the therapeutic relationship. In A.S Gurman & A.M. Razin (Eds.). *Effective Psychotherapy: A handbook of research* (pp. 503–543). New York: Pergamon.
- Halkides, G. (1958). *An experimental study of four conditions necessary for therapeutic change*. Unpublished doctoral dissertation. University of Chicago.
- Hansen, J. C., Moore, G. D., & Carkhuff, R. R. (1968). The differential relationship of objective and client perceptions of counseling. *Journal of Clinical Psychology, 24*, 244–246.
- Horvath, A. O. (1994). Empirical validation of Bordin's pantheoretical model of the alliance: The working alliance inventory perspective. In A. O. Horvath & L. S. Greenberg (Eds.), *The working alliance: Theory, research and practice* (pp. 109-128). New York: John Wiley & Sons.
- Kiesler, D. J. (1973). *The process of psychotherapy: Empirical foundations and systems of analysis*. Chicago: Aldine.
- Knox, R. (2008). Clients experiences of relational depth in person-centred counselling. *Counselling and Psychotherapy Research, 8*, 182–188.
- Lietaer, G. (1976). Nederlandstalige revisie van Barrett-Lennard's Relationship Inventory voor individueel-therapeutische relaties. *Psychologica Belgica, 16*, 73–94.
- Lister, J. L. (1970). *A scale for the measurement of empathic understanding*. Unpublished manuscript, University of Florida.
- Lockhart, W. (1984). Rogers' "necessary and sufficient conditions" revisited. *British Journal of Guidance and Counselling, 12*(2), 112–123.
- Mearns, D. (1996). Working at relational depth with clients in person-centred therapy. *Counselling, 7*, 307–311.
- Mearns, D. (1997). *Person-centred counselling training*. London : Sage.
- Mearns, D. & Cooper, M. (2005). *Working at relational depth in counselling and psychotherapy*. London: Sage.

- Michell, J. (2000). Normal science, pathological science and psychometrics. *Theory & Psychology, 10* (5), 639–667.
- Michell, J. (1997). Quantitative science and the definition of measurement in psychology. *British Journal of Psychology, 88*, 355–383.
- Patterson, C. H. (1984). Empathy, warmth, and genuineness: A review of reviews. *Psychotherapy, 21*, 431–438.
- Ponterotto, J., & Furlong, M. (1985). Evaluating counsellor effectiveness. A critical review of rating scale instruments. *Journal of Counseling Psychology, 32* (4) 597–616.
- Rappaport, J., & Chinsky, J. M. (1972). Accurate empathy: A confusion of a construct. *Psychological Bulletin, 77*, 400–404.
- Rennie, D. L. (1994). Client's deference in psychotherapy. *Journal of Counseling Psychology, 41*, 427–437.
- Rogers, C. R. (1951). *Client-centered therapy*. Boston: Houghton-Mifflin.
- Rogers, C. R. (1957). The necessary and sufficient conditions of therapeutic personality change. *Journal of Consulting Psychology, 21*, 95–103.
- Rogers, C. R., Gendling, E. T., Kiesler, D. J., & Truax, C. (1967). *The therapeutic relationship and its impact: A study of psychotherapy with schizophrenics*. Madison, WI: University of Wisconsin Press.
- Sanders, T., & Freire, E. S. (2008, July). *Researching the relationship: Developing the therapeutic relationship scale*. Paper presented at the 8th World Conference for Person-Centered and Experiential Psychotherapy and Counseling, Norwich, UK.
- Shapiro, J. G. (1968). Relationship between expert and neophyte ratings of therapeutic conditions. *Journal of Consulting and Clinical Psychology, 32*, 87–88.
- Stiles, W., & Shapiro, D. (1989). Abuse of the drug metaphor in psychotherapy process - outcome research. *Psychology Review, 9*, 521–543.
- Thorndike, E. L. (1918). The nature, purposes, and general methods of measurements of educational products. In G. M. Whipple (Ed.), *Seventeenth yearbook of the National Society for the Study of Education, Vol. 2* (pp. 16–24). Bloomington, IL: Public School Publishing.
- Truax, C. B. (1966). Influence of patient statements on judgments of therapist statements during psychotherapy. *Journal of Clinical Psychology, 22*, 335–337.
- Truax, C. B., Wargo, D. G., Frank, J. D., Imber, S. D., Battle, C. B., Hoehn-Saric, R., Nash, E. H., & Stone, A. R. (1966). Therapist empathy, genuineness, and warmth and patient therapeutic outcome. *Journal of Consulting Psychology, 30*, 395–401.
- Truax, C. B., & Carkhuff, R. R. (1967). *Toward effective counseling and psychotherapy: Training and practice*. Chicago: Aldine.
- Truax, C. B. (1972). The meaning and reliability of accurate empathy rating: A rejoinder. *Psychological Bulletin, 77*, 397–399.
- Waltz, J., Addis, M. E., Coerner, K., & Jacobson, N. S. (1993). Testing the integrity of a psychotherapy assessment of adherence and competence. *Journal of Consulting and Clinical Psychology, 61*, 620–630.
- Watson, J. (1999). *Measure of expressed empathy*. Unpublished manuscript, Toronto: OISE/UT.
- Watson, J. & Prosser, M. (2002). Development of an observer-rated measure of therapist empathy. In J. Watson, R. Goldman, & M. Warner (Eds.) *Client-centered and experiential psychotherapy in the 21st century* (pp. 303–314). Ross-on-Wye, UK: PCCS books.
- Wilczynski, J., Brodley, B. T., & Brody, A. (2008). A rating system for studying nondirective client-centered interviews – Revised. *The Person-Centered Journal, 15*, 34–57.

- Yardley, L. (2008). Demonstrating validity in qualitative psychology. In J. A. Smith (Ed.), *Qualitative Psychology: A practical guide to research methods* (2nd ed.) (pp. 235–251). London: Sage.
- Yorke, M. (2001). Bipolarity... or not? Some conceptual problems relating to bipolar rating scales. *British Educational Research Journal*, *27*, 171–186.
- Wiggins, S. (2010, July). *Relational Depth and Outcome*. Paper presented at the 9th World Conference for Person-Centered and Experiential Psychotherapy and Counseling, Rome, Italy.
- Wiggins, S., Elliott, R., & Cooper, M. (2010). *The prevalence and characteristics of relational depth events in psychotherapy*. Unpublished manuscript.
- Wyatt, G. (2001). Introduction to the Series. In G. Wyatt (Ed.) *Rogers' therapeutic conditions: Evolution, theory and practice. Vol 1: Congruence* (pp. i–vi). Ross-on-Wye: PCCS Books.