Check for updates

Development and validation of a 12-item version of the Barrett-Lennard Relationship Inventory (B-L RI:mini) using item response theory

Shun Chen[1] · Faith Liao[2,3] · David Murphy[1] · Stephen Joseph[1]

Abstract

The Barrett-Lennard Relationship Inventory (B-L RI; Barrett-Lennard, (2015). The Relationship Inventory A Complete Resource and Guide. http://nbn-resolving.de/urn:nbn:de:101:1-201602103917) is the most widely used measure of the facilitative conditions described by Carl Rogers as important for constructive personality development in his person-centered theory. At with 64 items it is however time consuming for participants to complete, and even in its shorter form, the B-L RI is 40-items long. In order to improve its utility, a mini form of the B-L RI was developed. In study 1, we used item response theory to select 12 out of the full 64 items to form the B-L RI:mini based on their discrimination, difficulty, information, and measurement invariance across the English ($n = 298$), Chinese ($n = 658$), and Spanish ($n = 330$) language versions of the inventory. In study 2 ($N = 362$), we validated the reliability and validity of the new 12-item measure. It was found that the B-L RI:mini showed excellent total internal consistency, temporal stability, and construct validity. According to the results of the study, the B-L RI:mini maintains optimal psychometric properties with a small number of items. This scale is recommended for use in further studies.

Keywords Barrett-Lennard Relationship Inventory · Short form · Item response theory · Person-centered theory · Bifactor model

Introduction

Genuineness (congruence), empathic understanding, and unconditional positive regard are well-known as Carl Rogers' facilitative relationship conditions/core conditions. Rogers (1957) theorized that constructive personality development happens when a minimal degree of the facilitative conditions are perceived through psychological contact with another person. The facilitative conditions are widely accepted as common factors that make psychotherapy effective (Bozarth & Motomasa, 2017; McAleavey & Castonguay, 2015). Their positive effects in enhancing personal development and human flourishing are equally applicable in any relationship involving psychological contact (Rogers, 1959, 1961). The Barrett-Lennard Relationship Inventory (B-L RI; Barrett-Lennard, 1962) was developed as a measurement tool to specifically evaluate the extent to which people experience the facilitative conditions of congruence, unconditional positive regard, and empathic understanding. The B-L RI score consists of four subscales: level of regard (R), empathic understanding (E), congruence (C), and unconditionality of regard (U). The subscales can be summed to produce an overall total score that has been referred to as 'facilitativeness' (Cramer, 2003; Davis et al., 2015). Since the 1960s, the B-L RI has wide application in various fields ranging from counseling psychology (Davis et al., 2015; Dolev & Zilcha-Mano, 2019; Murphy & Cramer, 2014), medicine (Chu & Tseng, 2013; Moghaddasian et al., 2013), forensic psychology (Hearn et al., 2020), sport psychology (Oh et al., 2012), education (Bockmier-Sommers et al., 2017; Swan et al., 2020), to business (Janssen, 2012).

✉ Shun Chen
shun.chen1@nottingham.ac.uk

1   University of Nottingham, School of Education, Wollaton Road, Nottingham NG81BB, UK

2   Department of Education, Taipei Medical University Hospital, Taipei, Taiwan

3   Graduate Institute of Humanities in Medicine, Taipei Medical University, Taipei, Taiwan

The B-L RI has been modified many times since its first publication; its length reduced from 92 to 85 items (Barrett-Lennard, 1962), then to 72 items, and finally to 64 and 40 items (Barrett-Lennard, 2015). Although the 64 and 40 item current versions of the B-L RI were greatly shortened compared to its first version, researchers still considered the B-L RI lengthy and suggested a further reduction based on the results of exploratory factor analysis (Cramer, 1986; Gurman, 1977; Wiebe & Barnett Pearce, 1973). Several abbreviated versions of the B-L RI were developed based on the specific purposes of the studies (e.g., Schacht et al., 1988; Schumm, et al., 1980a, b), which were not subject to any validity testing. There has long been a need of a shorter measure of the facilitative conditions for more practical application. An even shorter version than the 40-items of the B-L RI is necessary, especially when research participants have only limited patience or attention span, there is a fixed time period for testing, or there are financial limits for conducting a study (Donnellan et al., 2006). Participants' experience and motivation for completing questionnaires may be improved by providing a shorter measurement scale. Thus, an even shorter form of the B-L RI is warranted to alleviate the burden of completing the questionnaire and enhance its practicality. This study aims to fill the gap by developing and validating a very short scale based on the 64-item B-L RI to measure the facilitative conditions.

Psychometric Properties of the B-L RI

Barrett-Lennard developed the B-L RI to measure the facilitative conditions to test Rogers' theory of constructive personality change (1957) in the clinical setting (Barrett-Lennard, 1962). The current versions of the inventory (the 64-item/40-item B-L RI; Barrett-Lennard, 2015) consist of four subscales: level of regard (*R*), empathic understanding (*E*), congruence (*C*), and unconditionality of regard (*U*). Each subscale has the same number of items (16/10 items). Parallel forms of the B-L RI were developed for respondents who receive (OS:other to self), provide (MO:me/myself to other), or observe (Obs) the facilitative conditions in a relationship (Barrett-Lennard, 2015). Research has consistently shown that the scores of the B-L RI are statistically significantly correlated with a range of psychological and behavioral outcomes, such as positive client outcomes in psychotherapeutic relationships (Bell et al., 2016), greater authenticity as a client outcome in counseling (Bayliss-Conway et al., 2020), women athletes' body appreciation and eating style in coach-athlete relationships (Oh et al., 2012), students' learning experience in student–teacher relationships (Swan et al., 2020), and prisoners' post-traumatic growth in staff-prisoner relationships (Hearn et al., 2020).

The B-L RI has been translated into more than 20 different languages, including Arabic, Chinese, Dutch, French, German, Greek, Hebrew, Italian, Japanese, and Spanish (Barrett-Lennard, 2015; Liao et al., 2018).

The B-L RI has consistently shown high internal consistency and temporal stability reliabilities. For the B-L RI: OS-64, the means of internal consistency coefficients across a number of studies have been found to be 0.93 for R, 0.82 for E, 0.88 for C, 0.74 for U, and 0.85 for the total scale (Chu & Tseng, 2013; Davis et al., 2015; Dolev & Zilcha-Mano, 2019; Dufey & Wilson, 2017; Elkin et al., 2014; Fulton, 2016; Gurman, 1977; Hara et al., 2017; McClintock et al., 2017; Suzuki et al., 2019). For the B-L RI: OS-40, the means of alpha coefficients were reported to be 0.85 for R, 0.89 for E, and 0.82 for C, 0.76 for U, and 0.92 for the total scale (Barrett-Lennard, 2002; Greason & Welfare, 2013). Mean test–retest reliability coefficients for each subscale in the B-L RI: OS-64 were R, 0.83; E, 0.83; C, 0.88; U, 0.80 and for the total scale, 0.90 (Gurman, 1977).

The factor structure of the B-L RI has long been controversial. A detailed discussion can be found in the Electronic Supplementary Material. Both the 64-item and the 40-item B-L RI are considered too lengthy to measure a single construct.

Cross-language equivalency is an important psychological property for a measurement tool with multiple language versions. The presence of language equivalency determines whether comparisons between scores of different language versions of the B-L RI are statistically meaningful (Khojasteh & Lo, 2015). A multigroup CFA demonstrated that the partial scalar invariance was supported across the English, Chinese, and Spanish language versions of the B-L RI: OS-64, and the noninvariant items were recommended to be removed during development of a shorter version of the inventory to ensure the language equivalence of the new scale (Chen et al., 2021).

Current Research

The first goal of this paper was to develop (Study 1) a shorter version of the B-L RI (B-L RI:mini) with a unidimensional construct of facilitativeness based on data collected from English, Chinese, and Spanish language versions of the B-L RI: OS-64 using item response theory (IRT) approaches. The English, Chinese, and Spanish language versions were selected because they were the most widely spoken languages worldwide. Additionally, English, Chinese, and Spanish are spoken by three completely different cultures, items that perform well and consistent across these three language versions of the B-L RI are more likely to be shared understanding in human relationships. To our best knowledge, the psychometric properties of the B-L RI have never

been examined in the context of IRT. As a modern approach to item and test analysis, IRT is recommended for questionnaire development, evaluation, and refinement, which can overcome some of the limitations of classical test theory (CTT; Paek & Cole, 2020). First, IRT models outcomes at the item level, instead of the test level as in CTT. IRT was considered as a more informative and thorough approach to evaluate the items and the person's latent trait (θ). Second, IRT takes into account the specific item characteristics and how a person responds to it when estimating the person's latent trait. Item parameter estimation is dependent on the specific sample that responds to the item, and the estimation of the person's latent trait depends on the specific set of items that were answered within a CTT framework. Unlike IRT, the estimations of person and item parameters are independent to each other. B-L RI is an instrument that has been translated into multiple languages, and its item/test psychometric properties may vary across different language versions. Considering that the B-L RI:mini will also certainly have multiple language versions in the future, only items that show consistently good measurement properties and are invariant across different language versions of the B-L RI: OS-64 should be retained in the new scale. The item and scale properties of the English, Chinese, and Spanish language versions of the B-L RI: OS-64 were analyzed, since they are the most spoken languages in the world.

The second goal of this paper was to validate (Study 2) the English language version of the B-L RI:mini in a new set of samples. The characteristics of the items in the inventory were thoroughly examined using IRT. In the scale-level analysis, internal consistency, test–retest reliability, construct validity, convergent validity, and criterion-related validity were evaluated for the newly developed scale.

Study 1

Method

Participants and Procedures

Two-hundred and ninety-eight native speakers of English, 658 native speakers of Chinese, and 330 native speakers of Spanish participated in this study. The English-speaking sample and the Spanish-speaking sample were all taken from Chen et al. (2021), which were recruited using social media websites, and Jisc Online Surveys (www.jisc.ac.uk) were used for data collection from June to July 2020. The Chinese-speaking sample was all taken from Liao et al. (2018). In their study, a stratified random sampling technique (Eckman & West, 2016) was used to draw samples in six age strata (18–25, 26–35, 36–45, 46–55, 56–65, > 65) and data

were collected via an online survey. Prior to completing the B-L RI: OS-64, participants were asked to respond to each of the items with reference to a present relationship with a friend.

The English-speaking sample included 261 (87.6%) females, 36 (12.1%) males and one transgender person. The participants were aged 18 to 79 years old, and their mean age was 38 years old (SD = 12.9). Of the participants, 47.0% were aged from 18 to 35 years old, 43.3% of the participants were aged from 36 to 55 years old, and 9.7% were 56 years of age and over. Participants' friendships had lasted on average for 15.4 years (SD = 12.7). The shortest length of friendship was 0.5 years; the longest, 62.9 years. Regarding occupation, 7.0% were researchers, 10.4% were teachers, 14.8% were students, 26.8% were professionals, and 41.0% did not indicate their occupation.

The Spanish-Speaking sample included 284 (86.1%) females and 46 (13.9%) males. Of the participants, 61.1% were between the age of 18 and 35 years, 32.2% of the participants ranged from age 36 to 55, and 6.7% of the participants were over 65 years old. In terms of occupation, 37.0% were researchers, 11.5% were professionals, 10.6% were teachers, 4.3% were students, and 36.6% did not indicate their occupation. Participants' friendships had lasted on average for 14.6 years (SD = 9.2). The shortest length of friendship was 2.2 years; the longest, 30.5 years.

The Chinese-speaking sample was also predominantly female, consisted of 495 (75.2%) females, 162 (24.6%) males, and one other. With regard to age, 32.1% of the participants were aged from 18 to 25, 33.6% of the participants aged from 26 to 35, 16.7% of the participants were aged from 36 to 45, 12.3% of the participants ranged from age 46 to 55, 5% of the participants ranged from age 56 to 65, and 0.3% of the participants were over 65 years old. In terms of occupation, 28.0% of participants were students, 21.4% were professional occupations, 11.6% were sales and customer service workers, 10.2% were administrative and secretarial occupations, 8.7% were elementary occupations, 5.8% were skill trades, and 4.4% were unemployed. Of the participants, 1.7% were in-relationship with their friends for less than six months, 6.2% had friendships lasting from 6 to 12 months, 12.5% had friendships lasting from 1 to 3 years, 16% had friendships lasting from 3 to 5 years, and 63.7% had friendships lasting more than five years.

Measures

The B-L RI: OS-64 (Barrett-Lennard, 2015) was developed to assess the experience of the facilitative conditions in a relationship. The facilitative conditions were deemed important for constructive personality change in Rogers' theory (1957). The B-L RI: OS-64 is composed of four 16-item subscales: 1) level of regard (R); 2) empathic understanding

(E); 3) congruence (C); and 4) unconditionality of regard (U). Examples of items from the subscales include: R, "___ respects me as a person"; E, "___ wants to understand how I see things"; C, "___ is comfortable and at ease in our relationship"; and U, "___'s interest in me depends on the things I say or do" (negatively worded item). The participant is asked to think about their relationship with a particular person and to answer each of the items with that person in mind.

Participants answer each item on a six-point Likert-type scale (-3 = NO, I strongly feel that it is not true; -2 = No, I feel it is not true; –1 = (No) I feel that it is probably untrue, or more untrue than true; + 1 = (Yes) I feel that it is probably true, or more true than untrue; + 2 = Yes, I feel it is true; + 3 = YES, I strongly feel that it is true). Each subscale includes an equal number of positively and negatively worded items. After responses to negatively worded items are reverse-coded, sum scores for each subscale are calculated with higher scores representing higher levels of perceived regard, empathy, congruence, and unconditionality in a relationship.

In this study, three language versions of the B-L RI: OS-64 were used, namely English, Chinese, and Spanish. The original English version of the B-L RI: OS-64 was back-translated into the Chinese and Spanish versions by bilingual translators (Barrett-Lennard, 2015; Celis, 1999; Liao et al., 2018). In the previous study (Chen et al., 2021), the alpha coefficients in the range of 0.70 to 0.94 were reported for scores from the subscales in the English, Chinese, and Spanish language versions of the B-L RI: OS-64, and ranging from 0.95 to 0.96 for scores from the total scales. The unidimensional model was confirmed through CFA and bifactor model evaluation for the three language versions of the B-L RI: OS-64. Eight (U3, U11, C24, E26, U27, U31, U43, and E46) out of sixty-four items were reported to be noninvariant across the three language versions of the B-L RI: OS-64.

Data Analysis

Assumptions of IRT First, the basic assumptions of IRT include unidimensionality and local independence. As mentioned above, unidimensionality for the B-L RI: OS-64 (the English, Chinese, and Spanish language versions) was previously supported in CFA and bifactor model evaluation (Chen et al., 2021). The datasets used in this study were taken from Chen et al. (2021). The one-factor CFA showed satisfactory fit across the three language versions: the comparative fit index (CFI) ranged from 0.970 to 0.985, the Tucker-Lewis index (TLI) ranged from 0.969 to 0.985, the root mean square error of approximation (RMSEA) ranged from 0.054 to 0.072, and the standardized root mean square residual (SRMR) ranged from 0.077 to 0.081. The bifactor model evaluation rejected the existence of the specific factors (R, E, C, and U) beyond the general factor

(facilitativeness) across the three language versions of the inventory: omega hierarchical (ω_h) was ranged from 0.946 to 0.978, but omega hierarchical subscale (ω_{hs}) was ranged from 0.000 to 0.304, explained common variance (ECV) was reported ranging between 0.834 and 0.866, percent of uncontaminated correlations (PUC) were 0.762 and 0.635, and high construct replicability (H) ranging from 0.982 to 0.986. Local independence means that when latent trait is held constant across respondents, the observed responses are statistically independent (Samejima, 2015). The inter-item residual correlations (Yen's Q3; Yen, 1984) were evaluated to test local independence. Christensen et al. (2017) recommended the criterion for high likelihood of local dependence was greater than |.30|. At least one item in a pair of locally dependent items was suggested to be trimmed in order to strengthen the unidimensionality of the scale.

IRT Second, we used the graded response model (GRM; Samejima, 2015) to assess the scale and item-level functioning since the item responses are ordered, polytomous, and categorical. The sample size needed to estimate the GRM is 250 (Reeve & Fayers, 2004); thus the requirement was met in this study. In the GRM, one common item slope (a/ item discrimination) and a set of k-1 location (b/item difficulty) parameters are estimated. k is the number of response categories; hence five b parameters for each item can be produced from the six response options in the B-L RI. Item discrimination is the degree to which an item differentiates respondents with similar levels of the same latent trait (Embretson & Reise, 2013). According to Baker's (Baker, 2001) discrimination classification: very high discrimination, $a > 1.7$; high discrimination, $1.35 < a < 1.69$; moderate discrimination, $0.65 < a < 1.34$; low discrimination, $0.35 < a < 0.64$; very low discrimination, $0.01 < a < 0.34$; no discrimination, $a = 0$. Only items with high or very high discrimination across the three language versions of the inventory were considered for retention. Item difficulty is the amount of the latent trait that is necessary for the respondent to have a 50% chance to endorse a given category (Embretson & Reise, 2013). The means of b parameters (b1-b5) for each item were calculated. Items with different levels of difficulty were retained to best differentiate respondents with different levels of latent trait. Then, the location and slope parameters were used to compute item information curves, which describe how much information an item relative to the total information of the latent construct. Items that have high discrimination and have a difficulty parameter close to the respondents' latent trait will provide relatively high information, whereas items that have low discrimination and have a difficulty parameter far away from the respondents' latent trait will provide relatively low information (Zickar & Broadfoot, 2009). The item information curves across items can be aggregated as the test information curve. In IRT, the

information curves were used to depict measurement precision. In order to maximize the precision of measurement at different levels of the latent trait, items with high information levels at different parts along the continuum were retained (Reeve & Fayers, 2004).

Differential item functioning (DIF) Third, DIF was conducted to detect item equivalence across the English, Chinese, and Spanish language versions of the B-L RI: OS-64 by using likelihood-ratio tests (Lopez Rivas et al., 2009). For accurate DIF detection, all others as anchors approach (Thissen et al., 1993) was used: a baseline model with all the parameter (item difficulty and discrimination) constrained to be equal across groups was specified first; then the parameters of each item, in turn, were freed and constraints on the parameters of the other items remained; and changes in model fit compared to the baseline model was examined. Those items with significant chi-square values ($p < 0.05$) are considered to exhibit differential functioning——the opposite of measurement invariance. Both uniform and non-uniform DIF can be detected using the method. Uniform DIF indicates that a consistent systematic difference in the response to the item between the groups across levels of the latent trait spectrum. The inequivalence of the magnitude of focal item difficulty across groups indicates the presence of uniform DIF. Whereas non-uniform DIF indicates varying differences across levels of the trait and the inequivalence of the magnitude of focal item discrimination across groups indicates the presence of non-uniform DIF (Tay et al., 2015). Items exhibiting non-uniform DIF across the three language versions of the B-L RI: OS-64 were suggested to be removed from the scale (O'Neill & McPeek, 1993). Items that display measurement invariance or only uniform DIF across the three language versions of the B-L RI: OS-64 were considered to be retained.

All the analyses mentioned above were conducted in R statistics (R Core Team, 2020) using the RStudio interface (version 1.3.1093; RStudio Team, 2020), using the multidimensional item response theory (MIRT) package (Chalmers, 2012).

Results

Local Independence

Out of the 4096 item pairs, only forty-six (1.12%) in the English version of the B-L RI: OS-64, twenty-eight (0.68%) in the Chinese version, and twenty-two (0.54%) in the Spanish version had Q3 values greater than |.30|. For example, the Q3 value, between item U27 ("___likes or accepts certain things about me, and there are other things s/he does not like in me") and item U43 ("___approves of me in some ways or sometimes, and plainly disapproves of me in other ways/other times"), was 0.43. The violation of local independence may be resulting from the similar content, and the amount of locally dependent item pairs were small, the local independence assumptions might not hold strictly but closely enough for using IRT advantageously (Kolen & Brennan, 2004). Besides, locally dependent items were removed to ensure the unidimensionality of the B-L RI:mini.

Item Discrimination and Difficulty

Table S1 presents the discrimination and difficulty parameters from the GRM for the English, Chinese, and Spanish language versions of the B-L RI: OS-64. The values of the discrimination parameters for the English version fell within the range 0 to 3.44; for the Chinese version, the range was between -1.17 and 3.06; for the Spanish version, the range was from -0.68 to 4.37. Item R25 ("___ cares for me"; Barrett-Lennard, 2015, p. 102) demonstrated the highest discrimination values in both the English and Spanish language versions of the inventory. The most discriminative item in the Chinese version of the inventory was Item R37 ("___ is friendly and warm with me"; Barrett-Lennard, 2015, p. 102). Twenty-three items had *a* parameters less than 1.35 at least in one of the three language versions of the inventory, which were considered for removal from the B-L RI:mini.

The means of difficulty parameters ranged from -2.45 to 75.69 in the English version of the B-L RI: OS-64; ranged from -17.14 to 6.23 in the Chinese version; ranged from -3.23 to 10.86 in the Spanish version of the inventory. Further inspection of the b values for each item showed that the item difficulties were well spread out across the latent continuum. Items with extreme values (e.g., Item E46 in the Chinese version of the inventory had b_{1-5} values ranged from -105.46 to 60.36) were not considered to be included in the new scale. Item C4 ("___ is comfortable and at ease in our relationship"; Barrett-Lennard, 2015, p. 102) was the 'easiest' item across the three language versions of the inventory. Item E14 ("___ looks at what I do from their own point of view"; Barrett-Lennard, 2015, p. 101) was the most 'difficult' item in both the English and Spanish versions of the inventory. In the Chinese version, item U35 ("If I show I am angry with ___ they become hurt or angry with me, too"; Barrett-Lennard, 2015, p. 102) was the most 'difficult' item.

Item and Test Information Curves

Figure S1 displays item information curves (IIC) for all the 64 items across the English, Chinese and Spanish versions of the B-L RI: OS-64. The 64 items provided the most information at the middle to lower levels of facilitativeness across the three language versions of the inventory. Most of the

items had the highest information in the range between $\theta = -4$ and $\theta = -2$. Items that showed peak information values less than 1 in any one of the language versions of the inventory were removed (e.g., Item U3, "___'s interest in me depends on the things I say or do." (Barrett-Lennard, 2015, p. 101), yielded almost no information across the three language versions of the inventory). The test information curves (TIC) for the three language versions of the inventory were shown in Figure S2, which peaked in the range of -3 SD and 0/1 SD from the mean. The Spanish version of the inventory appeared to have the highest test information of all the three versions at $\theta = -2$.

Differential Item Functioning

We conducted tests of DIF across the English, Chinese and Spanish language versions of the B-L RI: OS-64. As shown in Table S2, both the item discrimination and difficulty parameters of E2, R5, and E10 are equivalent across the three language versions of the inventory. Forty-nine items showed non-uniform DIF as indicated by the significant discrimination parameters, and fifty-seven items showed uniform DIF as indicated by the significant difficulty parameters. This is consistent with the previous study (Chen et al., 2021) that examined the measurement invariance of the three language versions of the inventory by using multigroup CFA. The noninvariant items found in their study showed both uniform and non-uniform DIF.

Item Selection

As a result, 12 items, R5, E10, E18, E30, E34, C36, R41, C44, U51, U55, R57, and R61 were retained to form the B-L RI:mini (see Appendix 1). These 12 items were highly discriminative and sufficiently informative across the English, Chinese, and Spanish language versions of the B-L RI: OS-64. Besides, they did not exhibit non-uniform DIF across the languages.

Study 2

The aim of study 2 was to analyze the dimensionality, reliability, and construct validity of the B-L RI:mini. It was expected that the scale shows acceptable reliability and validity. Specifically, the value of the Cronbach's alpha coefficient should be greater than 0.70 with regard to internal consistency reliability (Onwuegbuzie & Daniel, 2002), the intraclass correlation coefficient should be greater than 0.75 in terms of test–retest reliability (Koo & Li, 2016). Regarding validity, the factor structure of the B-L RI:mini should be unidimensional, the scale should be moderately related to other measure of the same construct (convergent validity)

and meaningful outcome (criterion-related validity), and the scale should not related be to measure that is conceptually unrelated to it (discriminant validity).

Method

Participants and Procedures

Participant demographics are shown in Table 1. The average duration of participants' relationship was 108.14 months ($SD = 102.28$).

Participants were recruited using social media (Benfield & Szlemko, 2006). Two longitudinal surveys were administered via Jisc Online Surveys (www.jisc.ac.uk). The first survey included all the measurements mentioned below,

Table 1 Demographic characteristics

Demographic Variables	Count (N = 362)	Percentage
Gender		
Male	316	87.3
Female	30	8.3
Trans	4	1.1
Genderqueer	6	1.7
Prefer not to say	6	1.7
Ethnicity		
Caucasian	274	75.7
Asian	46	12.7
Latino/Hispanic	6	1.7
Mixed	12	3.3
Other	24	6.6
Age (years)		
18–25	81	22.4
26–35	143	39.5
36–45	73	20.2
46–55	47	13.0
56–65	16	4.4
> 65	2	.6
Types of relationship		
Spouse	122	33.7
Partner	195	53.9
Friend	39	10.8
Other	6	1.6
Duration of relationship (years)		
< .5	9	2.5
.5–1	18	5.0
1–3	102	28.2
3–5	43	11.9
5–10	87	24.0
10–20	64	17.7
20–50	39	10.8

whereas the second survey only included the B-L RI:mini. Participants were asked to provide their email addresses in order to receive the invitation to complete the B-L RI:mini again within seven days of initial administration for assessment of test–retest reliability. We included measures of social support and experiences in close relationship in order to establish convergent validity and criterion-related validity, respectively. Empathic understanding, unconditional positive regard, and genuineness can be seen as a form of emotional support. And people who continuously and consistently perceive the facilitative conditions in their close relationships tend to feel more comfortable and secure, less anxious and avoidant in relationships, and to be more authentic (Rogers, 1961). We would expect the B-L RI:mini to be moderately associated with higher ratings of social support and closeness of relationship (Rogers, 1957, 1961). Whereas the BL-RI: mini is developed on the basis of Rogers' (1957) theory, other measures are derived from different theoretical perspectives, such as attachment theory. We also included a test for social desirability in order to establish that the B-L RI:mini discriminant validity.

Measures

B-L RI: mini The new 12-item B-L RI measure was developed from the 64-item B-L RI (Barrett-Lennard, 2015) in Study 1. Cronbach's alpha for B-L RI:mini was 0.91 in Study 2.

Multidimensional Scale of Perceived Social Support (MSPSS) We used the MSPSS scale (Zimet et al., 1988) to measure perceived social support from family (e.g., My family really tries to help me), friends (e.g., My friends really try to help me) and significant others (e.g., There is a special person who is around when I am in need). The scale consists of 12 items, all measured on a Likert scale ranging from 1 (very strongly disagree) to 7 (very strongly agree). The mean rating across all items is computed. Higher scores indicate greater perceived social support. A Cronbach alpha of 0.88 was reported by Zimet et al. (1988).

Socially Desirable Response Set Five-Item Survey (SDRS-5) The SDRS-5 (Hays et al., 1989) measures the extent to which participants respond in a socially desirable manner (e.g., No matter who I'm talking to, I'm always a good listener). This scale contains five items scored using a Likert scale from 1 (definitely true) to 5 (definitely false). Only some extreme responses are scored 1, and all other responses are scored 0. The alpha coefficients in the previous study ranged from 0.66 to 0.68 (Hays et al., 1989).

The Experiences in Close Relationship Scale – Short Form (ECR-S) The ECR-S (Wei et al., 2007) is a 12-item scale derived from the original 36-item ECR (Brennan, Clark, & Shaver, 1998). The ECR-S was used to assess a general pattern of adult attachment by measuring the level of attachment anxiety (e.g., I need a lot of reassurance that I am loved by my partner) and attachment avoidance (e.g., I want to get close to my partner, but I keep pulling back). Each subscale contains six items. The response is scored using a Likert scale from 1 (Strongly Disagree) to 7 (Strongly Agree). According to the previous study, the alpha coefficients for the anxiety subscale was 0.77 and 0.78 for the avoidance subscale. Also, 1-month test–retest reliability = 0.80 and 0.83 for anxiety and avoidance subscales, respectively (Wei et al., 2007).

Data Analysis

Reliability The internal consistency reliability and test–retest reliability of the B-L RI:mini were examined using Cronbach's alpha and intra-class correlation coefficient, respectively. SPSS version 26.0 was used to conduct correlation and reliability analyses.

Factor Structure CFA is commonly used to evaluate the internal construct validity and dimensionality of assessments (Harrington, 2009). A common rule of thumb is that the minimum sample to variable ratio of 10:1 is necessary for performing factor analysis, while the ideal ratio might be 15:1 or 20:1 (Clark & Watson, 1995). The ratio between the number of participants and the items turned out to be as high as 30:1 ($N = 362$). CFA was conducted to evaluate the adequacy of three potential models: (1) a unidimensional model with the 12 items loading on a single latent variable, facilitativeness; (2) a correlated four-factor model that includes R, E, C, and U; (3) a bifactor model with a general factor (facilitativeness), along with four specific factors (R, E, C, and U). Multiple indices of fit were used to evaluate and compare these models: The Root Mean Square Error Approximation (RMSEA), the Confirmatory Fit Index (CFI), and finally by the Standardized Root Mean Square Residual (SRMR). The acceptable fit was evaluated based on the following standards (Hooper et al., 2008; Kline, 2015): RMSEA < 0.08, CFI ≥ 0.90, and SRMR ≤ 0.08.

Unidimensionality Bifactor model analysis was used to examine the unidimensionality of the B-L RI:mini (Neff et al., 2017), which allows each item to load on a general factor (facilitativeness) and a group factor (R, E, C, and U). The following statistical indices were calculated using the Bifactor Indices Calculator (Dueber, 2017). The omega index (ω) indicates the proportion of total score variance that is attributable to all sources of common variance included in the model (Reise et al., 2013). By the same logic, the omega subscale ($ω_s$) indicates the amount of each subscale score's

total variance that is attributable to the blend of general and group factor variance (Watkins, 2017). Omega hierarchical (ω_h) indicates the ratio of variance in the total scores that is attributable to the single general factor (McDonald, 2013). Omega hierarchical subscale (ω_{hs}) is a reliability estimate that gives the proportion of a subscale score variance that is attributable to the specific factor after accounting for the general factor (Reise et al., 2013). ω_h values greater than 0.80 suggest that the most of the explained variance was attributed to the general factor, rather than a specific factor. High ω_{hs} values and low ω_h values suggest that the scale is multidimensional, instead of unidimensional. The construct replicability (H; Rodriguez et al., 2016a) informs the degree to which a latent factor is well defined by a set of items, and a cut-off value of greater than 0.70 was recommended. Factor determinacy (FD; Rodriguez et al., 2016a) represents the correlation between factor scores and the factors, which indicates the validity of factor scores for independent use. The values of FD greater than 0.90 demonstrate the factor score estimates are trustworthy (Gorsuch, 2013). Explained common variance (ECV) is an indicator of unidimensionality, which is calculated by dividing the variance attributable to the general factor by the variance attributable to both the general and the subgroup factors. Percent of uncontaminated correlations (PUC) represents the percentage of item correlations contaminated by variance that is attributed to the general factor and specific factor, which is computed by dividing the number of correlations between items from different group factors by the total number of correlations (Rodriguez et al., 2016a). Rodriguez and colleagues (2016b) recommended the criteria for the essential unidimensionality: both ECVs and PUCs are greater than 0.70.

Convergent and Criterion-Related Validity As a part of construct validity, convergent and criterion-related validity of the B-L RI:mini were also tested in this study. MSPSS (Zimet et al., 1988) was used to test the convergent validity of B-L RI:mini and ECR-S (Wei et al., 2007) was used for the test of criterion-related validity. Theoretically, perceived facilitativeness should be positively related to perceived social support and negatively related to attachment-related anxiety and avoidance. Thus, the convergent validity and criterion-related validity were examined by assessing associations between B-L RI:mini, MSPSS and ECR-S. Social desirability bias is a common threat to the validity of self-report data (King & Bruner, 2000). Therefore, social desirability was evaluated to check and control for its impact on participants' responses.

Results

Internal Consistency Reliability

The internal consistency reliability of the B-L RI:mini was examined using a Cronbach's alpha index. The alpha coefficient for total B-L RI:mini score was 0.91. The item-total correlation coefficients ranged from 0.25 to 0.84 (See Table 2). The score of item U55N displayed a low but significant correlation ($r = 0.25$, $p < 0.001$) with the total score. Another unconditionality item, U51's score, presented the second least correlation ($r = 0.64$, $p < 0.001$) with the total score.

Table 2 B-L RI:mini scale items, mean (M), standard deviation (SD), and item-total correlation (r)

(n = 362)	M	SD	r
1. "_____ feels a true liking for me." (R5)	2.40	1.07	.73*
2. "_____ nearly always knows exactly what I mean." (E10)	1.36	1.44	.79*
3. "Whether the ideas and feelings I express are "good" or "bad" seems to make no difference to _____'s feeling toward me." (U51)	1.63	1.58	.64*
4. "_____ expresses their true impressions and feelings with me." (C36)	1.68	1.48	.77*
5. "I feel that _____ really values me." (R41)	1.98	1.36	.83*
6. "_____ usually senses or realizes what I am feeling." (E18)	1.26	1.62	.80*
7. "Sometimes I am more worthwhile in _____'s eyes than I am at other times." (U55N)	-.03	2.09	.25*
8. "_____ realizes what I mean even when I have difficulty in saying it." (E30)	.81	1.69	.80*
9. "_____ is willing to express whatever is actually in their mind with me, including personal feelings about themself or me." (C44)	1.22	1.70	.75*
10. "_____ is truly interested in me." (R57)	2.12	1.32	.84*
11. "_____ usually understands the whole of what I mean." (E34)	1.31	1.58	.82*
12. "_____ feels affection for me." (R61)	2.31	1.19	.72*

Total alpha = .91; N = negatively worded item; In parenthesis: the 64-item B-L RI item number (Barrett-Lennard, 2015, pp. 101–103); * significant at the .01 level

Test–Retest Reliability

The test–retest interval was one week and resulted in 216 verifiably matchable responses. The B-L RI:mini showed excellent test–retest reliability ($r = 0.87$).

Confirmatory Factor Analysis

The factorial validity and unidimensionality of the B-L RI:mini were investigated using CFA techniques, and confirmatory bifactor modeling. We could not confirm a good fit for the model with the significance of χ^2 (< 0.05) for all solutions. However, the previous study has demonstrated that this statistic is very sensitive to sample size (Kline, 2015). Thus, other fit indices were analyzed. The results indicate that all the models fitted the data sufficiently well (Unidimentional model: $\chi^2(54) = 388.127$, $p < 0.001$, CFI $= 0.952$, RMSEA $= 0.051$, SRMR $= 0.061$; Four-factor model: $\chi^2(48) = 130.883$, $p < 0.001$, CFI $= 0.988$, RMSEA $= 0.069$, SRMR $= 0.030$; Bifactor model: $\chi^2(37) = 58.213$, $p = 0.015$, CFI $= 0.997$; RMSEA $= 0.040$; SRMR $= 0.016$). However, the bifactor model had the best overall fit indices.

Bifactor Model Evaluation

The overall omega index (ω) was 0.95, indicating that 95% of total score variance could be attributed to both the general factor (facilitativeness) and specific factors (*R*, *E*, *C*, and *U*). Thus, the proportion of error score in the total score was only 5%. The omega hierarchical (ω_h) index was 0.87 greater than 0.80 (Reise et al., 2013), indicating that the B-L RI:mini's total score predominantly reflects the general factor (See Table 3).

In contrast, omega hierarchical subscale (ω_{hs}) scores were low for all four factors (0.19—0.20); none of them met the minimum standard of 0.50 suggested by Reise (2012). This result indicates that most of the reliable variance of each subscale score was due to the general factor rather than the specific factors. Even though the omega subscale for *R*, *E*, and *C* were high, there were only small proportions of variance in subscale scores that were attributed to the group

factors alone. Within the bifactor model, H of the general factor was greater than 0.70, and no specific factor met the criteria for adequate construct replicability, which suggested that only the general factor was considered well defined by its items. Additionally, only the general factor showed FD value greater than 0.90, which suggested that only the total scale score should be used. On the scale level, both ECV and PUC were greater than 0.70, supporting the unidimensional nature of the B-L RI:mini.

Convergent and Criterion-Related Validity

Both convergent and criterion-related validity were supported by the significantly moderate correlations between B-L RI:mini and relative measures. Social desirability was controlled in convergent and criterion-related validity analyses. Convergent validity was demonstrated by a positive correlation between the B-L RI:mini and the MSPSS ($r = 0.34$, $p < 0.001$). For criterion-related validity, the B-L RI:mini was found to be negatively related to both the anxiety subscale ($r = -0.25$, $p < 0.001$) and the avoidance subscale ($r = -0.36$, $p < 0.001$) of the ECR-S.

Discriminant Validity

There were only low and non statistically significant correlations found between the B-L RI:mini and the SDRS-5, indicating that no social desirability bias is present. Pearson's correlation between the total B-L RI:mini scale and the SDRS-5 was—0.07 ($p = 0.193$).

Discussion

The B-L RI is a well-known instrument to measure facilitative conditions for constructive personality development, which has been used in various fields and been translated into a variety of languages (Barrett-Lennard, 2015; Liao et al., 2018). The B-L RI provides information which can help us to improve the quality of relationships with others and our interpersonal and communication skill to facilitate

Table 3 Bifactor Evaluation Indices for Bifactor Model with Four Specific Factors

Factors	ω/ω_s	ω_h/ω_{hs}	H	FD	ECV	PUC
F	.945	.867	.934	.944	.765	.788
R	.921	.204	.468	.740	-	-
E	.914	.188	.425	.719	-	-
C	.816	.194	.282	.650	-	-
U	.365	.177	.190	.461	-	-

R = level of regard, E = empathy, U = unconditionality, C = congruence, F = facilitativeness (general factor), ω = omega, ω_s = omega subscale, ω_h = omega hierarchical, ω_{hs} = omega hierarchical subscale, FD = factor determinacy, ECV = Explained common variance, PUC = Percent of uncontaminated correlations

other's personal growth. In psychotherapy research, the B-L RI has generally been used to measure the facilitativeness of the therapeutic relationship and the supervisory relationship (e.g., Carey & Williams, 1986; Lawson, 1982; Wade & Bernstein, 1991). The B-L RI enables counselors, counseling students, and educators to examine if counselors: 1) have perceived sufficient facilitative conditions for their personal/professional development in counselor education/supervision/group settings; 2) have provided sufficient facilitative conditions for an effective therapeutic relationship. Noteworthily, the application of the B-L RI is common across different types of relationships, both clinical and nonclinical, which enables us to evaluate the facilitative conditions that individuals perceived in their relationships with several significant others by a single instrument.

However, the length of the B-L RI may be seen as excessive by researchers and practitioners who want to use the B-L RI in combination with a large battery of instruments, particularly when the B-L RI needs to be administered on multiple occasions. The short measurement instrument has the advantage of reducing the difficulty for participants to remain focused on completing the questionnaire, so the research compliance rate and participants' motivation in responding to the questionnaire can be improved.

To meet the practical need for a short form of the B-L RI, this study aimed to develop a mini form of the B-L RI in a scientifically meaningful manner. This research involved two studies: 12 items were selected from the 64-item B-L RI to be included in the shortest form of the B-L RI using IRT in Study 1. Following the numbering of the 64-item B-L RI, the abbreviated B-L RI included items R5, E10, E18, E30, E34, C36, R41, C44, U51, U55, R57, and R61. To our best knowledge, this is the first study that employed IRT-based techniques to investigate the psychometric properties of the English, Chinese, and Spanish language versions of the B-L RI: OS-64. IRT provides more detailed information on the item level comparing to CTT, which is more suitable for scale development. Besides, DIF analysis revealed that most of the items function differently across the three language versions of the inventory. In Study 2, the validation process was implemented for the B-L RI: mini. With the reduced structure, the reliability analyses showed good results for the inventory, with a Cronbach's alpha of 0.91. All of the items showed high item-total correlations except for the *U55*, which may be caused by the ambiguous wording and the reverse scoring. Da Rocha Bastos et al., (1979) argued that a high degree of unconditionality of regard could be represented either as unconditional acceptance or as inexorable rejection. The semantic ambiguity is more likely to occur in the real-life setting because the therapist is expected to be related to the client in a positive way. Besides, *U*55 was the only reverse worded item in the B-L RI:mini, which may increase the difficulty for participants to understand the

statement. This finding was consistent with the context of B-L RI literature, indicating that significant and relatively high correlations between *R*, *E,* and *C* except for *U* (Da Rocha Bastos et al., 1979). The test–retest reliability of the inventory after one week was 0.87. In summary, both the internal consistency and test–retest reliabilities of the B-L RI:mini were proved to be satisfactory.

The B-L RI:mini was demonstrated to be unidimensional using CFA and bifactor model evaluation. Empathic understanding, congruence, and unconditional positive regard are meaningful in theory, but not in psychometric testing. The prior research has only used EFA to examine the factor structure of the B-L RI. Several possible models have been explored, but they have never been confirmed and compared. With the help of CFA, the bifactor model evaluation showed that most of the variances in the subscales' scores were attributed to the general factor (facilitativeness), and the specific factors are invalid and unreliable. Even though previous research found that the subscales scores were internally consistent and temporally stable, which can be derived from the reliability of the general factor. The facilitative conditions are conceptually distinguishable, but also synchronous in close relationships. Both total scale and subscale scores revealed only the overall facilitativeness of relationships, the use of subscale score is meaningless. Consistent with the only previous report for bifactor model evaluation of the B-L RI: OS-64 (Chen et al., 2021), this new form of B-L RI can be considered as an efficient and useful tool to evaluate the levels of perceived facilitativeness instead of measuring the relationship conditions separately.

Assessments of convergent validity and criterion-related validity further supported the construct validity of the B-L RI:mini. Facilitativeness is a special form of social support that serves to promote personality change (Rogers, 1957). Perceiving facilitativeness constantly in a relationship would reduce the experience of anxiety and avoidance (Rogers, 1961). As expected, we found a positive association of relationship quality with perceived social support as well as the negative associations of relationship quality with attachment avoidance and anxiety. Also, low and non statistically significant correlations between scores on the B-L RI:mini and the measure of social desirability suggested no evidence of social desirability bias in our study.

Scoring Methods

The use of the B-L RI:mini is convenient and can save administration time, especially for longitudinal study and monitoring of the maturity of personality development where the participant may be required to complete the form on a number of occasions. We recommend using a total score for the B-L RI:mini. Based on the bifactor model

evaluation, it was concluded that it was not appropriate to use the subscales independently.

On the one hand, B-L RI can be used to measure one's perceived facilitativeness in various types of relationships. On the other hand, the wide application of the inventory makes the establishment of norms and standards difficult. After reversing negatively worded item U55 (-3/-2/-1 = 3/2/1), a total score varying from -36 to 36 can be obtained by summing all the item scores. For psychotherapy relationships, Barrett-Lennard (2015) suggested to "utilize a three-fold approach to assembling comparison data and working standards." (p.42). The first two components of this approach can be directly applied in the B-L RI:mini. First, means and variance data can be organized from available studies that reporting such data from the B-L RI:mini. Second, a local data pool should be built up by gathering data systematically from participants in the same local setting. Then, the mean and variance of the local data could be selected and organized. The standard scoring method was suggested as a complementary or alternative to the two components above. Comparison standards for the 64-item B-L RI were established by this method. Applying the same method to establish scoring interpretation for the B-L RI:mini: a total score of 30 and above are "as high as one could plausibly expect in any relationship context, in terms of honest, discriminating perception." (p.42); a total score of 24 implies that the facilitative conditions were substantially perceived in the referent relationship; a total score of 18 is probably the minimal level that should be achieved in fruitful helping relationships; any score below 12 "would be expected to represent a less than adequate level in therapy relationships." (p.42).

Strengths and limitations

The interpretation of the findings should take the strengths and limitations of this study into account. Alternative confirmatory analytic models of the B-L RI have been little explored; we compared the original correlated four-factor model with unidimensional and bifactor models in this study. The bifactor modeling approach was used to further our understanding of the scoring of the B-L RI. The findings demonstrated that the B-L RI:mini produces the same unidimensional structure as the B-L RI: OS-64. Further strengths are its large sample size, as it provided an ideal participant sample ratio to conduct factor analysis. However, like all research, this study had limitations. The homogeneous nature of the sample might limit the findings of this study. A majority of the participants were female Caucasian. Another limitation might be that only one negatively worded item was contained in the scale; the use of alternating item wording in questionnaires has been recommended in order to reduce acquiescent bias and extreme response bias (Rorer,

1965). The necessity of negatively worded items is still under discussion. Sonderen et al. (2013) found the negatively worded items not only did not prevent such bias but also caused confusion and inattention. Despite controversial opinions regarding the inclusion of reverse-worded items, the B-L RI:mini showed adequate reliability and validity.

Conclusions

In summary, the results from these studies indicated that the 12-item B-L RI is a valid and reliable instrument of facilitativeness in the non-clinical setting. The B-L RI:mini was proven to retain the good psychometric properties of the 64-item B-L RI and to require less time to complete. Our findings indicated that the B-L RI:mini should only be used to obtain a total score for facilitativeness and should not be separated into its subscales. It is recommended that future studies assess the reliability and validity of the B-L RI:mini in the clinical setting.

Appendix 1

The Barrett-Lennard Relationship Inventory:Mini (B-L RI:mini).

"Below are listed a variety of ways that one person may feel or behave in relation to another person. Please consider each numbered statement with reference to your present relationship with _____ (name), mentally adding their name in the space provided. If the other person's name is John, for example, then read statement number 1 as "John feels a true liking for me."

Mark each statement in the answer column on the right, according to how strongly you feel that it is true, or not true, in this relationship. EXAMPLE: Please be sure to mark every one. Write in a plus number (+3, +2, or +1), or a minus number (–1, –2, or –3), to stand for the following answers:

–3: NO, I strongly feel that it is not true.

–2: No, I feel it is not true.

–1: (No) I feel that it is probably untrue, or more untrue than true.

+1: (Yes) I feel that it is probably true, or more true than untrue.

+2: Yes, I feel it is true.

+3: YES, I strongly feel that it is true.

1. _____ feels a true liking for me

2. _____ nearly always knows exactly what I mean

1. _____ feels a true liking for me

3. Whether the ideas and feelings I express are "good" or "bad" seems to make no difference to _____'s feeling toward me

4. _____ expresses their true impressions and feelings with me

5. I feel that _____ really values me

6. _____ usually senses or realizes what I am feeling

7. Sometimes I am more worthwhile in _____'s eyes than I am at other times

8. _____ realizes what I mean even when I have difficulty in saying it

9. _____ is willing to express whatever is actually in their mind with me, including personal feelings about themself or me

10. _____ is truly interested in me

11. _____ usually understands the whole of what I mean

12. _____ feels affection for me

" (Barrett-Lennard, 2015, pp. 101–103).

Scoring Instructions.

A total score can be obtained by summing up item scores after reversing of the item 7.

Declarations

Informed Consent Informed consent was obtained from all individual participants included in the study.

Ethics Approval Ethical approval was obtained at the School of Education Research Ethics Committee of the University of [blind note] (2018/52). The study was performed in compliance with GDPR and the Code of Research Conduct and Research Ethics at the University of [blind note].

Conflicts of Interest The authors have no conflict of interest.

References

Baker, F. B. (2001). *The basics of item response theory* (2nd ed.). ERIC clearinghouse on assessment and evaluation.

Barrett-Lennard, G. T. (1962). Dimensions of therapist response as causal factors in therapeutic change. *Psychological Monographs: General and Applied, 76*(43), 1–36. https://doi.org/10.1037/h0093918

Barrett-Lennard, G. T. (2002). Perceptual variables of the helping relationship: A measuring system and its fruits. In G. Wyatt & P. Sanders (Eds.), *Rogers' Therapeutic Conditions: Evolution, Theory and Practice. Vol. 4, Vol. 4, (pp. 25–50).* PCCS.

Barrett-Lennard, G. T. (2015). The relationship inventory A complete resource and guide. Retrieved Dec 11, 2018, from http://nbn-resolving.de/urn:nbn:de:101:1-201602103917

Bayliss-Conway, C., Price, S., Murphy, D., & Joseph, S. (2020). Client-centred therapeutic relationship conditions and authenticity: a prospective study. *British Journal of Guidance & Counselling*. Advance online publication. https://doi.org/10.1080/03069885.2020.1755952

Bell, H., Hagedorn, W. B., & Robinson, E. H. M. (2016). An exploration of supervisory and therapeutic relationships and client outcomes. *Counselor Education and Supervision, 55*(3), 182–197. https://doi.org/10.1002/ceas.12044

Benfield, J. A., & Szlemko, W. J. (2006). Internet-based data collection: promises and realities. *Journal of Research Practice*, *2*(2), 15. Retrieved Mar 8, 2019, from http://jrp.icaap.org/index.php/jrp/article/view/30/51

Bockmier-Sommers, D., Chen, C. C. B., & Martsch, M. (2017). Student perception of teacher empathy, high regard and genuineness and the impact on student engagement. *E-Mentor, 70*(3), 66–72. https://doi.org/10.15219/em70.1310

Bozarth, J. D., & Motomasa, N. (2017). *Searching for the core: The interface of client-centred principles with other therapies. In The handbook of person-centred therapy and mental health. Theory, Research and Practice (pp. 391–408)*. Ross-on-Wye: PCCS Books.

Brennan, K. A., Clark, C. L., & Shaver, P. R. (1998). Self-report measurement of adult attachment: An integrative overview. In J. A. Simpson & W. S. Rholes (Eds.), *Attachment theory and close relationships* (pp. 46–76). The Guilford Press.

Carey, J. C., & Williams, K. S. (1986). Cognitive style in counselor education: A comparison of practicum supervisors and counselors in training. *Counselor Education and Supervision, 26*(2), 128–136. https://doi.org/10.1002/j.1556-6978.1986.tb00707.x

Celis, A. (1999). Proceso de adaptación del inventario de la relación de ayuda de GT Barrett-Lennard. *Revista De Psicología, 8*(1), 163. https://doi.org/10.5354/0719-0581.2011.17296

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, *48*(6), 1–29. Retrieved Jan 21, 2020, from: https://www.jstatsoft.org/rt/captureCite/v048i06/0

Chen, S., Liao, F., Murphy, D., & Joseph, S. (2021). Measurement Invariance of the English, Chinese, and Spanish Versions of the Barrett-Lennard Relationship Inventory. Measurement and Evaluation in Counseling and Development. Advance online publication. https://doi.org/10.1080/07481756.2021.1955212

Christensen, K. B., Makransky, G., & Horton, M. (2017). Critical values for Yen's Q_3: Identification of local dependence in the Rasch model using residual correlations. *Applied Psychological Measurement, 41*(3), 178–194. https://doi.org/10.1177/0146621616677520

Cramer, D. (1986). An item factor analysis of the revised Barrett-Lennard Relationship Inventory. *British Journal of Guidance &*

Counselling, 14(3), 314–325. https://doi.org/10.1080/0306988860 8253521.

Cramer, D. (2003). Facilitativeness, conflict, demand for approval, self-esteem, and satisfaction with romantic relationships. *The Journal of Psychology, 137*(1), 85–98. https://doi.org/10.1080/00223980309600601

Da Rocha Bastos, L., Alves, A. J., Bahiense, V. M., & Michael, W. B. (1979). Empirical test of the validity of a unitary construct underlying a modified form of the Barrett-Lennard Relationship Inventory. *Educational and Psychological Measurement, 39*(2), 499–504. https://doi.org/10.1177/001316447903900234

Davis, T. A., Ancis, J. R., & Ashby, J. S. (2015). Therapist effects, working alliance, and African American women substance users. *Cultural Diversity and Ethnic Minority Psychology, 21*(1), 126–135. https://doi.org/10.1037/a0036944

Dolev, T., & Zilcha-Mano, S. (2019). The role of the therapeutic relationship in the association between interpersonal behaviors and outcome: Comparison of two competing models. *Psychotherapy Research, 29*(5), 553–564. https://doi.org/10.1080/10503307.2017.1422215

Donnellan, M. B., Oswald, F. L., Baird, B. M., & Lucas, R. E. (2006). The mini-IPIP scales: tiny-yet-effective measures of the big five factors of personality. *Psychological Assessment, 18*(2), 192–203. https://doi.org/10.1037/1040-3590.18.2.192

Dueber, D. M. (2017). *Bifactor indices calculator.* University of Kentucky Libraries. https://doi.org/10.13023/edp.tool.01

Dufey, M., & Wilson, J. E. (2017). Short-term effects of the empathic talk: The role of perceived empathy, interpersonal dispositions, and vagal tone on experiencing depth. *The Spanish Journal of Psychology, 20*, E61. https://doi.org/10.1017/sjp.2017.58

Eckman, S., & West, B. T. (2016). Analysis of Data from Stratified and Clustered Surveys. In C. Wolf, D. Joye, T. Smith, & Y. Fu, The SAGE Handbook of Survey Methodology (pp. 477–487). SAGE Publications Ltd. https://doi.org/10.4135/9781473957 893.n31

Elkin, I., Falconnier, L., Smith, Y., Canada, K. E., Henderson, E., Brown, E. R., & McKay, B. M. (2014). Therapist responsiveness and patient engagement in therapy. *Psychotherapy Research, 24*(1), 52–66. https://doi.org/10.1080/10503307.2013.820855

Embretson, S. E., & Reise, S. P. (2013). *Item response theory for psychologists.* Taylor and Francis. Retrieved Jan 26, 2020, from http://www.123library.org/book_details/?id=95840

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*(3), 309–319. https://doi.org/10.1037/1040-3590.7.3.309

Fulton, C. L. (2016). Mindfulness, self-compassion, and counselor characteristics and session variables. *Journal of Mental Health Counseling, 38*(4), 360–374. https://doi.org/10.17744/mehc.38.4.06

Gorsuch, R. L. (2013). Factor analysis. In J. A. Schinka & I. B. Weiner (Eds.), *Research methods in psychology (2nd ed., pp. 143–164).* Wiley.

Greason, P. B., & Welfare, L. E. (2013). The impact of mindfulness and meditation practice on client perceptions of common therapeutic factors. *The Journal of Humanistic Counseling, 52*(2), 235–253. https://doi.org/10.1002/j.2161-1939.2013.00045.x

Gurman, A. S. (1977). The patient's perception of the therapeutic relationship. *Effective Psychotherapy: A Handbook of Research*, (pp. 503–543). Pergamon Press.

Hara, K. M., Aviram, A., Constantino, M. J., Westra, H. A., & Antony, M. M. (2017). Therapist empathy, homework compliance, and outcome in cognitive behavioral therapy for generalized anxiety disorder: Partitioning within- and between-therapist effects. *Cognitive Behaviour Therapy, 46*(5), 375–390. https://doi.org/10.1080/16506073.2016.1253605

Harrington, D. (2009). *Confirmatory factor analysis*. Oxford University Press.

Hays, R. D., Hayashi, T., & Stewart, A. L. (1989). A five-item measure of socially desirable response set. *Educational and Psychological Measurement, 49*(3), 629–636. https://doi.org/10.1177/001316448904900315

Hearn, N., Joseph, S., & Fitzpatrick, S. (2020). Post traumatic growth in prisoners and its association with the quality of staff–prisoner relationships. *Criminal Behaviour and Mental Health*, cbm.2173. https://doi.org/10.1002/cbm.2173

Hill, C. E., Siegelman, L., Gronsky, B. R., Sturniolo, F., & Fretz, B. R. (1981). Nonverbal communication and counseling outcome. *Journal of Counseling Psychology, 28*(3), 203–212. https://doi.org/10.1037/0022-0167.28.3.203

Holland, P. W., Wainer, H., & Educational testing service (Eds.). (1993). *Differential item functioning*. Lawrence Erlbaum Associates.

Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods, 6*(1), 53–60.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Janssen, J. H. (2012). A three-component framework for empathic technologies to augment human interaction. *Journal on Multimodal User Interfaces, 6*(3–4), 143–161. https://doi.org/10.1007/s12193-012-0097-5

Khojasteh, J., & Lo, W.-J. (2015). Investigating the sensitivity of goodness-of-fit indices to detect measurement invariance in a bifactor model. *Structural Equation Modeling: A Multidisciplinary Journal, 22*(4), 531–541. https://doi.org/10.1080/10705511.2014.937791

King, M. F., & Bruner, G. C. (2000). Social desirability bias: A neglected aspect of validity testing. *Psychology and Marketing, 17*(2), 79–103. https://doi.org/10.1002/(SICI)1520-6793(200002)17:2%3c79::AID-MAR2%3e3.0.CO;2-0

Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford publications.

Kolen, M. J., & Brennan, R. L. (2004). Test Equating, Scaling, and Linking. *Springer, New York*. https://doi.org/10.1007/978-1-4757-4310-4

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine, 15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Lance, C. E., & Vandenberg, R. J. (Eds.). (2009). *Statistical and methodological myths and urban legends: Doctrine, verity and fable in the organizational and social sciences*. Routledge.

Lanning, W. L., & Lemons, S. L. (1974). Another Look at the Factor Structure of the Barrett-Lennard Relationship Inventory. *Measurement and Evaluation in Guidance, 6*(4), 228–231. https://doi.org/10.1080/00256307.1974.12022602

Lawson, G. (1982). Relation of counselor traits to evaluation of the counseling relationship by alcoholics. *Journal of Studies on Alcohol, 43*(7), 834–839. https://doi.org/10.15288/jsa.1982.43.834

Liao, F., Murphy, D., & Barrett-Lennard, G. (2018). A confirmatory factor analysis of the Mandarin-Chinese version of the Barrett-Lennard Relationship Inventory. *Asia Pacific Journal of Counselling and Psychotherapy, 9*(1), 20–45. https://doi.org/10.1080/21507686.2017.1419269

Lopez Rivas, G. E., Stark, S., & Chernyshenko, O. S. (2009). The Effects of Referent Item Parameters on Differential Item Functioning Detection Using the Free Baseline Likelihood Ratio Test. *Applied Psychological Measurement, 33*(4), 251–265. https://doi.org/10.1177/0146621608321760

McAleavey, A. A., & Castonguay, L. G. (2015). The Process of Change in Psychotherapy: Common and Unique Factors. In O. C. G. Gelo, A. Pritz, & B. Rieken (Eds.), *Psychotherapy Research (pp. 293–310).* Vienna: Springer. https://doi.org/10.1007/978-3-7091-1382-0_15

McClintock, A. S., Perlman, M. R., McCarrick, S. M., Anderson, T., & Himawan, L. (2017). Enhancing psychotherapy process with common factors feedback: A randomized, clinical trial. *Journal of Counseling Psychology, 64*(3), 247–260. https://doi.org/10.1037/cou0000188

McDonald, R. P. (2013). *Test Theory: A Unified Treatment (1st ed.).* Psychology Press. https://doi.org/10.4324/9781410601087

Mills, D. H., & Zytowski, D. G. (1967). Helping relationship: A structural analysis. *Journal of Counseling Psychology, 14*(3), 193–197. https://doi.org/10.1037/h0024480

Moghaddasian, S., Lak Dizaji, S., & Mahmoudi, M. (2013). Nurses empathy and family needs in the intensive care units. *Journal of Caring Sciences; EISSN 2251–9920.* https://doi.org/10.5681/JCS.2013.024

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

Neff, K. D., Whittaker, T. A., & Karl, A. (2017). Examining the factor structure of the self-compassion scale in four distinct populations: Is the use of a total scale score justified? *Journal of Personality Assessment, 99*(6), 596–607. https://doi.org/10.1080/00223891.2016.1269334

Oh, K. H., Wiseman, M. C., Hendrickson, J., Phillips, J. C., & Hayden, E. W. (2012). Testing the Acceptance Model of Intuitive Eating With College Women Athletes. *Psychology of Women Quarterly, 36*(1), 88–98. https://doi.org/10.1177/0361684311433282

O'Neill, K. A., & McPeek, W. M. (1993). Item and test characteristics that are associated with differential item functioning. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 255–276). Lawrence Erlbaum Associates Inc.

Onwuegbuzie, A. J., & Daniel, L. G. (2002). A Framework for Reporting and Interpreting Internal Consistency Reliability Estimates. *Measurement and Evaluation in Counseling and Development, 35*(2), 89–103. https://doi.org/10.1080/07481756.2002.12069052

Chu, C.-I., & Tseng, C.-C.A. (2013). A survey of how patient-perceived empathy affects the relationship between health literacy and the understanding of information by orthopedic patients? *BMC Public Health, 13*(1), 155. https://doi.org/10.1186/1471-2458-13-155

Paek, I., & Cole, K. (2020). *Using R for item response theory model applications.* Routledge.

Ponterotto, J. G., & Furlong, M. J. (1985). Evaluating counselor effectiveness: A critical review of rating scale instruments. *Journal of Counseling Psychology, 32*(4), 597–616. https://doi.org/10.1037/0022-0167.32.4.597

R Core Team. (2020). *R: A Language and Environment for Statistical Computing* (4.0.3) [Computer software]. R Foundation for Statistical Computing. Retrieved Jul 23, 2020, from https://www.R-project.org

Reeve, B. B., & Fayers, P. (2004). Applying item response theory modelling for evaluating questionnaire item and scale properties. In P. M. Fayers & R. D. Hays (Eds.), *Assessing Quality of Life in Clinical Trials: Methods and practice (2nd ed, pp. 55–73).* Oxford University Press.

Reise, S. P. (2012). The Rediscovery of Bifactor Measurement Models. *Multivariate Behavioral Research, 47*(5), 667–696. https://doi.org/10.1080/00273171.2012.715555

Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and Modeling Psychological Measures in the Presence of Multidimensionality. *Journal of Personality Assessment, 95*(2), 129–140. https://doi.org/10.1080/00223891.2012.725437

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016a). Applying Bifactor Statistical Indices in the Evaluation of Psychological Measures. *Journal of Personality Assessment, 98*(3), 223–237. https://doi.org/10.1080/00223891.2015.1089249

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016b). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods, 21*(2), 137–150. https://doi.org/10.1037/met0000045

Rogers, C. (1957). The necessary and sufficient conditions of therapeutic personality change. *Journal of Consulting Psychology, 21*(2), 95–103. https://doi.org/10.1037/h0045357

Rogers, C. (1959). *A Theory of Therapy Personality and Interpersonal Relationships as Developed in the Client-centered Framework.* McGraw-Hill.

Rogers, C. (1961). *On Becoming a Person: A Therapist's View of Psychotherapy.* Houghton Mifflin Harcourt.

Rorer, L. G. (1965). The great response-style myth. *Psychological Bulletin, 63*(3), 129–156. https://doi.org/10.1037/h0021888

RStudio Team. (2020). *RStudio: Integrated Development for R.* (1.3.1093) [Computer software]. PBC. Retrieved Jul 23, 2020, from http://www.rstudio.com/

Samejima, F. (2015). Graded Response Model. In W. J. van der Linden (Ed.), *Handbook of item response theory (pp. 95–107).* CRC Press, Taylor & Francis Group.

Schacht, A. J., Howe, H. E., & Berman, J. J. (1988). A Short Form of the Barrett-Lennard Relationship Inventory for Supervisory Relationships. *Psychological Reports, 63*(3), 699–706. https://doi.org/10.2466/pr0.1988.63.3.699

Schumm, W. R., Bollman, S. R., & Jurich, A. P. (1980a). Marital Communication or Marital Conventionality? A Brief Report on the Relationship Inventory. *Psychological Reports, 46*(3_suppl), 1171–1174. https://doi.org/10.2466/pr0.1980.46.3c.1171

Schumm, W. R., Jurich, A. P., & Bollman, S. R. (1980b). The Dimensionality of an Abbreviated Relationship Inventory for Couples. *The Journal of Psychology, 105*(2), 225–230. https://doi.org/10.1080/00223980.1980.9915155

Steiger, J. H. (1990). Structural Model Evaluation and Modification: An Interval Estimation Approach. *Multivariate Behavioral Research, 25*(2), 173–180. https://doi.org/10.1207/s15327906mbr2502_4

Suzuki, J. Y., Mandavia, A., & Farber, B. A. (2019). Clients' perceptions of positive regard across four therapeutic orientations. *Journal of Psychotherapy Integration.* https://doi.org/10.1037/int0000186

Swan, K., Chen, C.-C. B., & Bockmier-Sommers, D. K. (2020). Relationships Between Carl Rogers' Person-Centered Education and the Community of Inquiry Framework: A Preliminary Exploration. *Online Learning, 24*(3). https://doi.org/10.24059/olj.v24i3.2279

Tay, L., Meade, A. W., & Cao, M. (2015). An overview and practical guide to IRT measurement equivalence analysis. *Organizational Research Methods, 18*(1), 3–46. https://doi.org/10.1177/1094428114553062

Thissen, D., Steinberg, L., & Wainer, H. (1993). Detection of differential item functioning using the parameters of item response models. In P. W. Holland & H. Wainer (Eds.), *Differential Item Functioning* (pp. 67–113). Lawrence Erlbaum Associates Inc.

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38*(1), 1–10. https://doi.org/10.1007/BF02291170

van Sonderen, E., Sanderman, R., & Coyne, J. C. (2013). Ineffectiveness of reverse wording of questionnaire items: Let's learn from Cows in the Rain. PLoS One, 8(7), e68967. https://doi.org/10.1371/journal.pone.0068967

Wade, P., & Bernstein, B. L. (1991). Culture sensitivity training and counselor's race: Effects on Black female clients' perceptions and

attrition. *Journal of Counseling Psychology, 38*(1), 9–15. https://doi.org/10.1037/0022-0167.38.1.9

Walker, B. S., & Little, D. F. (1969). Factor analysis of the Barrett-Lennard Relationship Inventory. *Journal of Counseling Psychology, 16*(6), 516–521. https://doi.org/10.1037/h0028452

Watkins, M. W. (2017). The reliability of multidimensional neuropsychological measures: From alpha to omega. *The Clinical Neuropsychologist, 31*(6–7), 1113–1126. https://doi.org/10.1080/13854046.2017.1317364

Wei, M., Russell, D. W., Mallinckrodt, B., & Vogel, D. L. (2007). The experiences in close relationship scale (ECR)-short form: Reliability, validity, and factor structure. *Journal of Personality Assessment, 88*(2), 187–204. https://doi.org/10.1080/00223890701268041

Wiebe, B., & Barnett Pearce, W. (1973). An item-analysis and revision of the barrett-lennard relationship inventory. *Journal of Clinical Psychology, 29*(4), 495–497. https://doi.org/10.1002/1097-4679(197310)29:4%3c495::AID-JCLP2270290432%3e3.0.CO;2-7

Yen, W. M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement, 8*(2), 125–145. https://doi.org/10.1177/014662168400800201

Zickar, M. J., & Broadfoot, A. A. (2009). The partial revival of a dead horse? Comparing classical test theory and item response theory. In C. E. Lance & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends: Doctrine, verity and fable in the organizational and social sciences* (pp. 37–59). Routledge/Taylor & Francis Group.

Zimet, G. D., Dahlem, N. W., Zimet, S. G., & Farley, G. K. (1988). The multidimensional scale of perceived social support. *Journal of Personality Assessment, 52*(1), 30–41. https://doi.org/10.1207/s15327752jpa5201_2