A psychometric evaluation of the Barrett-Lennard Relationship Inventory Obs-40 (Version 3) in humanistic counselling for young people

Bhatti, Kiranjeet

Award date:
2020

Awarding institution:
University of Roehampton
A Psychometric Evaluation of the Barrett-Lennard Relationship
Inventory Obs-40 (Version 3) in Humanistic Counselling for Young People

By

Kiranjeet Bhatti BSc, MSc

A thesis submitted in partial fulfilment of the requirements for the degree of

PsychD

Department of Psychology

University of Roehampton

2020
Acknowledgments

It is with immense gratitude that I acknowledge the support of my Director of Studies Professor Mick Cooper. I thank him for his constant presence, teaching, and warmth during this project. I also wish to thank my supervisor Dr Gina Pauli, who joined the project in my hour of need and offered her reassurance and knowledge. I feel incredibly lucky to have had such a superb supervisory team whose invaluable guidance and support encouraged throughout this journey.

I thank the participants and staff of the ETHOS project, in particular Gemma Ryan, and the raters on this project, without you none of this would have been possible.

I would like to acknowledge my fellow classmates and friends, Lena, Steph, Viviana, and, Lilit. Their understanding, kindness, and laughter made my time on the PsychD a joy.

I want to thank my mother and my brother, for their immeasurable strength that has sustained me throughout my education and life. And special thanks, for my husband Thomas, who inspired me, believed in me and helped make this dream come true. Finally, this is for my dad, whom I miss and know would have appreciated the statistics.
Abstract

Many young people experience mental health difficulties, which, if left untreated, can have detrimental effects into adulthood. Humanistic therapies are widely used for this population; however, evidence of their effectiveness is limited. Research on humanistic therapy with young people can be supported by the existence of validated, observer-rater adherence measures. Following a literature review, the most appropriate instrument for such a purpose was identified as the Barrett-Lennard Relationship Inventory (BLRI). This aims to measure Empathic Understanding, Congruence, Level of Regard, and Unconditionality of Regard. However, the psychometric properties of the BLRI have yet to be evaluated in young populations. Using 136 audio recordings of humanistic counselling with adolescents, the presence and level of the therapeutic core conditions were rated using the observer-rated BLRI Obs-40 (Version 3), and the psychometric properties of the scale were examined. Results of confirmatory factor analyses supported the original four-subscale dimensionality of the scale. Additionally, there was support for a model in which the first-order subscale factors loaded significantly on a single second-order factor of general ‘person-centeredness’. The internal consistency of the sub-scales and total score was high, and five-month test-retest reliability was satisfactory for total BLRI score. However, inter-rater reliability was found to be unsatisfactory across three independent raters using a sample of 50 recordings. The BLRI total score and subscales correlated strongly and significantly with the Person Centred and Experiential Psychotherapy Scale adapted for young people (PCEPS-YP) with the exception of the unconditionality subscale. Results suggest the BLRI Obs-40 (Version 3) can be used with young populations in training and research environments; however, rigorous training for raters is necessary to strengthen scale reliability and use as an observer-rated measure.
Table of Content

Acknowledgments...........................................................................................................3
Abstract............................................................................................................................4
Abbreviations..................................................................................................................8

Chapter 1: Introduction....................................................................................................9
1.1 Background..............................................................................................................9
1.1.1 Mental Health Experience in Young People.........................................................9
1.1.2 Measurement Scales in Person-Centred Therapy..................................................10
1.2 Key Definitions......................................................................................................12
1.3 Rationale..............................................................................................................14
1.4 Contribution to Counselling Psychology...............................................................14
1.4.1 Contributions to Research..................................................................................14
1.4.2 Contributions to Practice....................................................................................15
1.5 Study Aim............................................................................................................15
1.6 Overview of Structure and Chapters......................................................................16

Chapter 2: Literature Review.........................................................................................17
2.1 Mental Health in Young People............................................................................17
2.1.1 Mental Health Trends in Young People...............................................................17
2.1.2 Consequences of Poor Mental Health in Young People.....................................19
2.2 Mental Health Treatment for Young People.........................................................19
2.2.1 Available Support..............................................................................................19
2.2.2 Effectiveness of Counselling with Young People..............................................20
2.3 Humanistic Therapy for Young People................................................................23
2.3.1 Humanistic Therapy..........................................................................................23
2.3.2 Effectiveness of Humanistic Therapy in Young People.......................................24
2.3.3 Role of Core Conditions in Young People.........................................................25
2.3.4 Rationale for Humanistic Therapy For Young People.......................................26
2.4 Importance of Measurement.................................................................................28
2.5 Current Measures of Humanistic Therapy in Young People................................30
2.5.1 The Basic Empathy Scale..................................................................................34
2.5.2 Parent Questionnaire.........................................................................................36
2.5.3 Barrett-Lennard Relationship Inventory...........................................................38
2.5.4 Truax Scales......................................................................................................39
2.5.5 PCEPS-YP.........................................................................................................42
2.6 Rationale for BLRI................................................................................................43
2.7 Current Psychometric Evidence and Evaluation of BLRI......................................45
2.7.1 Development and Revision of the BLRI............................................................45
2.7.2 Subsequent Psychometric Evaluation of the BLRI..............................................48
2.8 Aims and Objectives.............................................................................................52
2.9 Research Questions...............................................................................................52
2.10 Chapter Summary...............................................................................................53

Chapter 3: Methods.....................................................................................................54
3.1 Epistemology..........................................................................................................54
3.1.1 Overview............................................................................................................54
3.1.2 Positivism..........................................................................................................54
3.1.3 Psychometric Assumptions...............................................................................55
3.1.4 Therapeutic Relationship
3.1.5 Pragmatism
3.1.6 Summary

3.2 Design
3.3 ETHOS Trial and Data Access
3.3.1 ETHOS Study Aims and Overview
3.3.2 Data Access
3.3.3 ETHOS Participants
3.3.4 ETHOS Intervention
3.3.5 Recordings
3.3.6 ETHOS Participants
3.3.7 ETHOS Intervention
3.3.8 Recordings

3.4 Observer Raters
3.4.1 Rater Recruitment
3.4.2 Rater Demographics
3.4.3 Rater Training
3.4.4 Observer Raters
3.4.5 Rater Recruitment
3.4.6 Rater Demographics
3.4.7 Rater Training

3.5 Measures
3.5.1 BLRI Obs-40 (Version 3)
3.5.2 PCEPS-YP
3.5.3 COSMIN Checklist

3.6 Procedure
3.6.1 Selection and Transfer of Recordings
3.6.2 Rating Procedure and Data Entry

3.7 Ethical Considerations and Approval
3.7.1 Measures and Data Use
3.7.2 Confidentiality and Anonymity
3.7.3 Data Storage
3.7.4 Health and Safety Considerations

3.8 Data Analysis
3.8.1 Power Calculation
3.8.2 Internal Reliability
3.8.3 Inter-rater Reliability
3.8.4 Test-retest Reliability
3.8.5 Convergent Validity
3.8.6 Structural Validity
3.8.7 Post Analysis Evaluation
3.8.8 Analysis Software

3.9 Pilot Ratings

Chapter 4: Results
4.1 Pilot Results
4.2 Descriptive Statistics
4.3 Internal Reliability
4.3.1 Inter Scale Correlations
4.4 Structural Validity
4.5 Convergent Validity
4.6 Test-Retest Reliability
4.7 Inter-Rater Reliability
4.7.1 Post-Hoc Analysis
4.8 COSMIN Checklist
Chapter 5: Discussion

5.1 Chapter Introduction

5.2 Summary of Findings

5.2.1 Internal Reliability

5.2.2 Test-Retest Reliability

5.2.3 Inter Rater Reliability

5.2.4 Structural Validity

5.2.5 Convergent Validity

5.3 Evaluation of the BLRI Obs-40 (Version 3)

5.3.1 Scale Design, Content, and Scoring

5.3.2 Scale Revisions

5.3.3 Observer Format

5.4 Limitations

5.4.1 Methodological Limitations

5.4.2 Analysis Limitations

5.5 Further Research

5.6 Contributions to Counselling Psychology

5.7 Reflexivity

Chapter 6: Summary

References

Appendices

Appendix A: Search Strategy for Systematic Review

Appendix B: COSMIN Checklist

Appendix C: Basic Empathy Scale

Appendix D: Parental Questionnaire

Appendix E: BLRI Obs-40 and Scoring Key

Appendix F: Truax Scales

Appendix G: PCEPS-YP

Appendix H: PCEPS

Appendix I: ETHOS Eligibility Criteria - Schools

Appendix J: ETHOS Eligibility Criteria - Young People

Appendix K: ETHOS Eligibility Criteria - Counsellors

Appendix L: ETHOS Adherence Auditing Procedures

Appendix M: Practice Ratings

Appendix N: ETHOS Data Sharing Application Form

Appendix O: CREST Data Storage and Protection Procedures

Appendix P: Parcelling Allocation

Appendix Q: Pilot Data Descriptive Statistics

Appendix R: Sample 1 Descriptive Statistics

Appendix S: Sample 2 Descriptive Statistics

Appendix T: Sample 3 Descriptive Statistics

Appendix U: Sample 4 Descriptive Statistics

Appendix V: Absolute Agreement Per Item

Appendix W: Absolute Agreement Per Recording

Appendix X: Triad Participant Record Form
**Abbreviations**

- **BACP**: British Association for Counselling and Psychotherapy
- **BES**: Basic Empathy Scale
- **BLRI**: Barrett Lennard Relations Inventory
- **BLRI Obs-40**: 40 item Observer Rated Barrett Lennard Relations Inventory (Version 3)
- **CAMHS**: Child and Adolescent Mental Health Services
- **CBT**: Cognitive Behavioural Therapy
- **CFA**: Confirmatory Factor Analysis
- **COSMIN**: Consensus-based Standards for the Selections of Health Measurement Instruments
- **DSM-5**: Diagnostic and Statistical Manual of Mental Disorders
- **ETHOS**: Effectiveness and Cost-Effectiveness of Humanistic Counselling in Schools
- **ICC**: Intraclass Correlation Coefficient
- **K Alpha**: Krippendorff’s Alpha
- **NICE**: National Institute for Health and Care Excellence
- **PCEPS**: Person-Centred Experiential Psychotherapy Scale
- **PCEPS-YP**: Person-Centred Psychotherapy Scale – Young Person Counselling Version
- **PRISMA**: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
- **PROMS**: Patient Reported Outcome Measures
- **RCT**: Randomised Control Trial
- **SBHC**: School-Based Humanistic Counselling
- **SD**: Standard Deviation
- **UK**: United Kingdom
Chapter 1: Introduction

1.1 Background

1.1.1 Mental Health Experience in Young People

The importance of good mental health in young people is becoming more evident (Department of Health, 2015), and it is estimated 10-20% of 16-24-year-olds experience poor mental health (Ford, Hamilton, Meltzer, & Goodman, 2008; McManus, Meltzer, Brugha, Bebbington, & Jenkins, 2009). Additionally, many mental health difficulties experienced in adulthood originate in younger years (Jones, 2013).

Child and Adolescent Mental Health Services (CAMHS) are primarily responsible for the treatment of mental health difficulties in young people. However, the quality and effectiveness of these services are questionable (Birchwood & Singh, 2013), and there appears to be a ‘treatment gap’ when it comes to accessing mental health support. According to the British Child and Adolescent Mental Health Survey, 55% of 12-15-year-olds with mental health difficulties had no contact with mental health support services (Knapp et al., 2016).

Despite this treatment gap, evidence for the effectiveness of therapy for children and adolescences has been found. A systematic review commissioned by the British Association of Counselling and Psychotherapy (BACP), found counselling to be a useful intervention for this population (McLaughlin, Holliday, Clarke, & Ilie, 2013). However, many of the included studies focused on the effects of Cognitive Behavioural Therapy. Therefore evidence for the efficacy of alternative therapies is lacking, in particular, evidence to support the use of humanistic and person-centred therapy is minimal.
Humanistic therapy developed from the work of Carl Rogers (1957) and involves the offering of the six necessary and sufficient conditions of Person-Centred Therapy. These include: 1) empathy, 2) unconditional positive regard, 3) congruence 4) an incongruent client, 5) psychological contact, and 6) the potential for a client to receive these conditions. Work is nondirective and makes use of these conditions to facilitate change and promote growth. Humanistic therapy can be effective for young people in reducing psychological distress, as measured by clinical outcome measures (Pybis et al., 2014). Additionally, qualitative evidence has found humanistic therapy to help improve relationships with peers and increase self-confidence (Lynass, Pykhtina, & Cooper, 2012). Nevertheless, the evidence for humanistic therapy with young people is in its infancy and further research, and randomised control trials (RCTs) are necessary (Hanley & Noble, 2017).

One way to encourage the undertaking of RCTs is to introduce validated psychometric instruments that measure the therapeutic core conditions of person-centred therapy. Having a validated instrument is necessary to give confidence to research findings, which potentially strengthen the evidence base and justification for the use of person-centred and humanistic therapy in younger populations. (For further discussion see Section 2.4.)

1.1.2 Measurement Scales in Person-Centred Therapy

For each of the relationship conditions of empathy, congruence, and unconditional positive regard, Rogers (1957) provided an operationalized definition (Table 1). In doing so, this permitted person-centred theory to be tested and encouraged the development of measures for the core conditions. Over the years, several instruments have been developed to measure the relationship conditions; however, these have been developed and validated in adult populations only. A review of these is found in Freire and Grafanaki (2010). As well as
the BLRI and Truax Scales, which are discussed in section 2.5, there are global measures of the relationship conditions in addition to core-condition specific measures. One of the earliest of these scales is found in an unpublished doctoral dissertation by Halkides (1958), which aimed to measure the relationship conditions. Two empathy specific scales followed, one measuring empathic communication (Cochrane, 1974), the other measuring empathic understanding (Lister, 1970). Other measures of specific core conditions include the Revised Multidimensional Response Empathy Scale (Elliott et al., 1982), an observer-rated instrument used to measure therapist empathy; the Measure of Expressed Empathy (Watson & Prosser, 2002); and the Therapist Empathy Scale (Decker, Nich, Carroll, & Martino, 2014) which assesses affective, cognitive, and attitudinal aspects of therapist empathy. Additionally, there is the Nondirective Client-Centred Rating System (Wilczynski, Brodley, & Brody, 2008), an observer-rated scale that aims to distinguish between therapist nondirective and directive intentions. More general measures of the relationship conditions include the Therapist Presence Inventory (Geller, Greenberg, & Watson, 2010), a self-report instrument that measures in-session process and experience of therapist presence, the Relational Depth Inventory (Wiggins, Elliott, & Cooper, 2012), a client-rated measure, and the Therapeutic Relationship Scale (Sanders & Freire, 2008, as cited in Freire & Grafanaki, 2010), a client and therapist rated scale inspired by the BLRI. More recently, are the PCEPS and PCEPS-YP, which will be discussed further in section 2.5.
Table 1

*Operationalized definitions of the Relationship Conditions as defined by Rogers (1957).*

<table>
<thead>
<tr>
<th>Condition</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Empathy</td>
<td>‘To sense the client’s private world as if it were your own, but without ever losing the “as if” quality.’</td>
</tr>
<tr>
<td>Congruence</td>
<td>‘Within the relationship he is freely and deeply himself, with his actual experience accurately represented by his awareness of himself.’</td>
</tr>
<tr>
<td>Unconditional</td>
<td>‘The extent that the therapist finds himself experiencing a warm acceptance of each aspect of the client’s experience as being a part of that client.’</td>
</tr>
</tbody>
</table>

1.2 Key Definitions

Table 2 presents the definitions of the key terms used within this work. Distinctions are made between various interventions and forms of the Barrett-Lennard Relationship Inventory.

Table 2

*Key Definitions*

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Psychotherapy and Counselling</td>
<td>An umbrella term to define the field of psychological theory, therapy and intervention.</td>
</tr>
<tr>
<td>Psychotherapy</td>
<td>A form of intervention, typically long-term lasting months or years. Intervention is described as in-depth and likely to explore the past.</td>
</tr>
<tr>
<td><strong>Counselling</strong></td>
<td>A form of intervention, typically described as short-term, possibly behavioural and focused on present difficulties.</td>
</tr>
<tr>
<td>-----------------</td>
<td>--------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>Person-Centred Therapy</strong></td>
<td>A form of psychotherapy and counselling intervention developed by Carl Rogers, characterised by the six necessary and sufficient conditions of therapeutic change (Rogers, 1957).</td>
</tr>
<tr>
<td><strong>Humanistic Counselling</strong></td>
<td>A form of psychotherapy and counselling intervention related to the humanistic approach, which holds a person-centred foundation. Central the intervention is the belief in the individuals capacity for growth and potential (Rogers, 1959).</td>
</tr>
<tr>
<td><strong>Core Conditions</strong></td>
<td>Characteristics of the therapeutic relationship hypothesized by Rogers (1957) to facilitate change and promote growth. Typically there are six conditions, but for the purpose of this study, there is a focus on the therapist-offered conditions of empathy, congruence, and unconditional positive regard. These three conditions have been referred and likened to as ‘therapist interpersonal skills’, which in children and adolescent samples have been linked to positive outcomes in therapy (Karver, Handelsman, Fields, &amp; Bickman, 2006).</td>
</tr>
<tr>
<td><strong>BLRI</strong></td>
<td>The Barrett–Lennard Relationship Inventory, used when referring to the inventory in general, i.e. not specific to any particular version of the scale.</td>
</tr>
<tr>
<td><strong>BLRI Obs-40</strong></td>
<td>40 item Observer Rated Barrett Lennard Relations Inventory, Version 3. Used when referring to the 40 item observer rated scale specifically used for the purpose of this research.</td>
</tr>
</tbody>
</table>
1.3 Rationale

Research focused on humanistic therapies in young populations is needed to meet the growing mental health needs of young people. In general, research is lacking for both humanistic therapy and in the young population. The lack of evidence in these areas could be mistaken for lack of effectiveness of humanistic therapy for young people (House, Kalisch, & Maidman, 2017), and therefore undervalue the contributions humanistic therapy can offer children and adolescents. Early research has begun to demonstrate the effectiveness of humanistic counselling for young people (Cooper, 2009), and increase its accessibility; however, further research is required in order to contribute to its growing evidence base. Furthermore, research in these areas can support practitioners in their understanding of how humanistic therapy operates in this population. Additionally, research on the effects of the core conditions is needed to clarify the influence of person-centred and humanistic therapies on younger populations. For this to be investigated valid and reliable instruments that measure these conditions are needed, with such instruments, improvements to humanistic interventions can be developed and subsequently evaluated in RCTs. Currently, no evaluated measurement of the person-centred core conditions for use with young people exists.

1.4 Contributions to Counselling Psychology

1.4.1 Contributions to Research

A refined measurement will enable further research to investigate what specifically facilitates change in child and adolescent therapy, and could make an attempt to find the ‘common factors’ of therapy in this population. If these factors can be established,
counsellors may be able to adapt their practice to integrate these conditions, which could potentially maximise client engagement with therapy and lead to positive outcomes. Future research may also begin to explore the association between the level of therapeutic relationship conditions available and therapy outcomes. More generally, a validated measure will encourage RCTs and start to build upon and strengthen the evidence base and rationale for providing person-centred and humanistic therapies for young people.

1.4.2 Contributions to Practice

Regarding the contribution to counselling psychologists, a validated instrument could be utilised in training and supervision environments. It is possible the scale can be administered as an adherence measure, to ensure the therapeutic relationship conditions are observed when practising. Supervisors could use the scale with trainees to gain an indication of the extent to which the core conditions have been met and understood. Clients of therapy may see an indirect benefit from the research. Using the scale to measure adherence to training, may also mean counsellors can provide the core conditions more readily.

1.5 Study Aim

The aim of the study is to explore the psychometric properties of the Barrett-Lennard Relationship Inventory Obs 40 (Version 3) in humanistic counselling for young people.

The objectives are to examine:

1. Internal reliability of the BLRI Obs-40
2. Inter-rater reliability of the BLRI Obs-40
3. Test-retest reliability of the BLRI Obs-40
4. Convergent validity of the BLRI Obs-40, using the PCEPS-YP
5. Structural validity of the BLRI Obs-40
6. The psychometric methodological quality of the evaluation of the BLRI Obs-40, using the consensus-based standards for the selections of health measurement instruments (COSMIN) checklist (Mokkink et al., 2010).

1.6 Overview of Structure and Chapters

Chapter 2 includes a review of the literature concentrating on the mental health experiences of young people. Additionally, findings from a systematic review of the measures of the person-centred core conditions in young people are presented, and the current psychometric work of the BLRI is considered. Chapter 3 consists of the epistemology and methods used for the study, as guided by the COSMIN Study Design Checklist. Chapter 4 presents the results according to the research questions outlined in Chapter 2. Chapter 5 reviews the findings in the context of previous work and provides an in-depth evaluation of the BLRI Obs-40 for use with young people receiving humanistic counselling. Finally, Chapter 6 concludes the thesis and summarise the key findings.
Chapter 2: Literature Review

This chapter offers a context and rationale for the current research project aims and objectives. The literature is organised around three themes: mental health in young people, humanistic therapy for young people, and the current measurement instruments of humanistic therapy available for use with this population.

2.1 Mental Health in Young People

2.1.1 Mental Health Trends in Young People

The level of mental health difficulties in children and adolescents is rising (World Health Organisation, 2006.) A UK 2017 survey, from the Office of National Statistics, reported that one in eight children aged between 5-19 years old experienced a potentially diagnosable mental health problem (Sadler et al., 2018). Globally, it is estimated 6.7% of 5-17-year-olds have a diagnosed mental disorder, however as prevalence data in low and middle-income countries is limited, these countries are poorly represented, and therefore estimates may not be entirely accurate (Erskine et al., 2017). Regarding high-income countries, where data is more available, mental health problems are thought to affect 10-20% of children and adolescents (Kieling et al., 2011).

The range of mental difficulties is broad, and individuals can experience symptoms from across the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) (Hagell & Maughan, 2017). Furthermore, in a prospective longitudinal study using a sample of 1037 adults, 75% were found to have met the diagnostic criteria for a mental health disorder before the age of 18 (Kim-Cohen et al., 2003), further emphasising the importance of good mental
health in childhood. In the UK, the range of mental health difficulties experienced by young people is grouped into four categories; the most prevalent are emotional disorders, such as anxiety and depression, which are experienced by 1 in 12, 5-19-year-olds. Further categories include behavioural disorders, hyperactive disorders, and less common disorders, such as autism and eating disorders (Sadler et al., 2018). However, many young people may be experiencing difficulties despite not receiving a formal diagnosis. For example, though not officially diagnosed, 1 in 7, 15-year-olds report low life satisfaction (Public Health England, 2016) and therefore are likely to experience symptoms similar to depression and anxiety. Comorbidities are also prevalent in young people; a study of 10,000 American adolescents found 40% of individuals with a diagnosed mental disorder also met criteria for a second disorder (Merikangas et al., 2010; Merikangas, Nakamura, & Kessler, 2009).

Given the increased demand for mental health treatment for children and adolescents (Sarginson et al., 2017), there is an assumption that young people are experiencing higher levels of mental health distress (Murphy & Fonagy, 2012). Within the UK, this assumption is reflected in the data. Between 1999 and 2017, there has been a 1.5% increase in the prevalence of mental disorders in 5-15-year-olds (Sadler et al., 2018). This is particularly true regarding rates of emotional disorders. A comparison of two cross-sectional studies demonstrated an increase in emotional problems as measured by the strengths and difficulties questionnaire in girls between 2009 and 2014 (Fink et al., 2015). A systematic review of 19 studies also found an increase in internalising problems in adolescent girls between the 20th and 21st century (Bor, Dean, Najman, & Hayatbakhsh, 2014). Adolescent girls appear to be at an increased risk for developing mental health difficulties, with approximately 1 in 4 17-19-year-olds having a mental disorder, and almost half of those participating in deliberate self-harm (Sadler et al., 2018).
2.1.2 Consequences of Poor Mental Health in Young People

A myriad of consequences accompanies mental health difficulties experienced in childhood and adolescence. An online survey of 1000 young people found up to 40% missed school due to experiencing a mental health difficulty; in addition, poor mental health prevented engagement with everyday life such as socialising with friends and family (Time To Change, 2012). If left untreated, difficulties in mental health can develop into longer-term problems, as childhood difficulties continue into adulthood (Hofstra & Verhulst, 2000).

Evidence suggests that approximately half of adult mental health difficulties begin during the adolescent years (Jones, 2013). Childhood mental illness can go on to affect many areas of life, from education, occupation, and relationships, resulting in overall poor mental and physical well being (Prince et al., 2007). Additionally, from a financial standpoint, 16-25-year-olds with mental health problems are less likely to be in employment, and twice as likely to receive welfare benefits (Knapp et al., 2016). On a macro level, untreated childhood mental health problems can cause additional costs to health, social, and educational services (Snell et al., 2013; Whiteford et al., 2013). Therefore, there is a demand and an increasing awareness of the need for good mental health in children and adolescents (Department of Health, 2015).

2.2 Mental Health Treatment for Young People

2.2.1 Available Support

Within the UK, there is a wide range of services and professionals who support young people’s mental health. These can include non-specialised universal services, such as school
nurses or primary care workers, to more specialised mental health workers in the community, such as within GP surgeries and schools, or third sector organisations. Regarding counselling support, Child and Adolescent Mental Health Services are primarily responsible for the treatment of mental health difficulties in young people, with specialised inpatients units available for the most severe difficulties. NICE guidelines manage interventions for the treatment of mental health difficulties in young people. Talking therapies are recommended as the first line of treatment, and medication when all other support has failed. Despite current support, there is an increasing consensus that current mental health services for young people are lacking (Birchwood & Singh, 2013; Lamb, Hall, Kelvin, & Van Beinum, 2008). Furthermore, there appears to be a ‘treatment gap’ when it comes to accessing mental health support. A national survey estimated that over half 12-15-year-olds experiencing mental health difficulties had no contact with supportive services (Knapp et al., 2016), indicating that many young people are left without care. However, in addition to CAMHS services, school-based counselling is now routinely offered to support young peoples mental health (Cooper, 2013). School-based counselling can be very accessible as young people can refer themselves and waiting times are minimal in comparison to those seen in CAMHS services (Hanley, Jenkins, Barlow, Humphrey, & Wigelsworth, 2012). Furthermore, evidence suggests school-based counselling is effective in improving young people’s emotional and mental wellbeing (Cooper, 2013), adding to the growing research that counselling and mental health treatment for young people is useful and necessary (McLaughlin et al., 2013).

2.2.2 Effectiveness of Counselling and Psychotherapy with Young People

Over the past five decades, there has been an increase in the number of clinical trials of psychotherapy interventions with young people (Weisz, Ugueto, Cheron, & Herren, 2013).
Trials have varied across treatment methods and target mental health difficulties. Four meta-analyses conducted using broad inclusion criteria allow for an average estimation of effect sizes across a range of mental disorders and treatment modalities. The first, conducted by Casey and Berman (1985), comprised of 75 outcome studies of psychotherapy with children under the age of 12. The average effect size for treatment-control comparison was 0.71, similar to previous findings in adult populations (Cooper, 2008). A similar effect size (0.79) was found in a meta-analysis of 108 outcome studies using adolescent populations (Weisz, Weiss, Alicke, & Klotz, 1987). In Kazdin, Bass, Ayers, and Rodgers (1990) meta-analysis of 223 outcome studies with young people aged 4-18, an effect size of 0.88 was found. Additionally, the authors commented that at the time of reviewing the evidence there was a bias towards research on therapy outcomes rather than moderators of therapy, therefore although therapy appeared effective in younger populations, the causes and contributions to positive outcomes were unknown, highlighting a gap in the research. A further meta-analysis of 150 outcome studies with young people aged between 2-18 found an average effect size of 0.71 (Weisz, Donenberg, Han, & Weiss, 1995). However, effects were more significant for targeted treatment; specifically, therapy for anxiety was more effective for anxiety disorders than for other mental health difficulties, suggesting treatment may be useful for specific problems only, and not in general. In addition, at follow up 5-6 months after treatment ending, effects remain stable. Overall, findings suggest that psychotherapy is effective for supporting young people with their experiences of mental health difficulties, and evidence from meta-analyses suggest moderate to large improvements in mental wellbeing for young people.

Findings are complimented by randomised-control trials and meta-analyses of model and treatment specific trials in young populations. A meta-analysis of 55 psychotherapy trials for anxiety found an overall effect size of 0.65; however, most of the included studies were
based on CBT and therefore served to strengthen the evidence base for CBT for anxiety (Reynolds, Wilson, Austin, & Hooper, 2012). In addition, small to medium effects of psychotherapy for depression have been found (Weisz & Gray, 2008).

Practice-based evidence also supports the findings of clinical trials. Ford, Hutchings, Bywater, Goodman, and Goodman (2009) found modest but significant improvements in mental health outcomes following mental health intervention in practice. In addition, using Patient Reported Outcome Measures (PROMs) from approximately 16,000 encounters between practitioners and young people, Wolpert et al. (2012) found an overall reduction in mental health problems after engaging with CAMH services. Furthermore, a systematic review of 138 mental health interventions, commissioned by the BACP, found counselling to be useful for children and young people across a range of difficulties (McLaughlin et al., 2013). Additionally, the review found factors such as client motivation, child-therapist alliance, and race, and gender matching to be positively associated with therapeutic outcomes. However, 36% of included interventions were CBT-based, suggesting a need for further research into the effects of alternative therapies, such as person-centred and humanistic therapy.

In addition to quantitative evidence, qualitative research with young people has explored the aspects of therapy that are found to be effective when working with children and adolescents. A qualitative meta-analysis of 9 studies found having the opportunity to talk and be listened to was the most helpful factor in their therapy; in addition, the quality of the relationship, the confidential space, and problem-solving strategies were also reported as helpful (Griffiths, 2013). Furthermore, the effects of therapy appear to extend beyond a young person’s mental health; young people report positive impacts on their learning, overall behaviour, and school attendance (Lynass et al., 2012). Overall, both quantitative and
qualitative research supports the effectiveness and importance of counselling and psychotherapy with young people.

2.3 Humanistic Therapy for Young People

2.3.1 Humanistic Therapy

Person-Centred and Humanistic Therapy is based on the work of Carl Rogers (1957). According to Rogers, there are six necessary and sufficient conditions required for therapeutic change to occur within the therapeutic encounter. These are:

1. Two people in psychological contact
2. The client in a state of incongruence
3. The therapist as congruent in the relationship
4. The therapist experiencing unconditional positive regard towards the client
5. The therapist experiencing empathic understanding for the client’s internal frame of reference
6. The client receiving the communication of the therapists’ empathic understanding and unconditional positive regard

If these conditions exist between the client and therapist, Rogers hypothesised the therapeutic relationship would become constructive. Person-centred work is non-directive and makes use of these conditions to facilitate change and promote growth (Rogers, 1963). Humanistic therapies developed from Rogers’s ideas and are grounded in a person-centred foundation (Kirkbride, 2017). Within adult populations, person-centred and humanistic
therapies are effective for a range of common mental health difficulties and found to significantly reduce psychological distress (Elliott, Greenberg, Watson, Timulak, & Freire, 2013). However, the effectiveness of person-centred and humanistic therapy for young people is less established.

### 2.3.2 Effectiveness of Person-Centred and Humanistic Therapy in Young People

Research into person-centred therapy in young people is limited; therefore, studies that include person-centred elements are discussed, to provide a picture of the current status of research in this area. A meta-analysis of 93 studies on child-focused therapy, found an average effect size of 0.80. The majority of these studies focused on play therapy, which can bear a similarity to person-centred therapy. 73 of the 93 studies were described as non-directive, and the effect size for this sample was 0.92 (Bratton, Ray, Rhine, & Jones, 2005). This suggests that non-directive therapy based on person-centred principles can be useful for children.

In addition, a systematic review of 30 studies of school-based counselling, found therapy associated with improvements in mental health (Cooper, 2009). Though included studies were not exclusively person-centred, the majority were based on the core principles of person-centred practice. Moreover, a review of three RCTs of school-based humanistic counselling, which is based on the work of Rogers (1957), found, that after 12-weeks of counselling young people showed reduced psychological distress and increased self-esteem (McArthur, Cooper, & Berdondini, 2013). More recently, findings have been supported by additional pilot RCTs (Cooper, 2013; Pearce et al., 2017; Pybis et al., 2014) and the effectiveness and cost-effectiveness of school-based humanistic counselling are currently undergoing further investigation in the ETHOS trial (Stafford et al., 2018).
2.3.3 Role of Core Conditions in Young People

There are very few studies that have investigated the role of the therapeutic core condition specifically in therapy with young people, with the exception of Truax, Altmann, Wright, and Mitchell (1973). This study investigated the effects of accurate empathy, genuineness, and non-possessive warmth in child therapy. Using observer-rated measures, the authors found children who showed more considerable improvement had received higher levels of the therapeutic conditions. In addition, children who received lower levels of the therapeutic conditions were found to deteriorate, suggesting their therapy was harmful. This study supports the use and importance of therapeutic conditions in child therapy; however, as the sample size was small (n=16) it was not possible to differentiate between the effects of the three specific conditions. Siegel (1972) produced a similar finding. In this study, high and low levels of therapist communicated conditions were compared during play therapy for children with learning disabilities. At the end of therapy, those who received high levels of empathy, congruence, and unconditional positive regard made more insightful and positive statements about themselves, demonstrating a positive modelling effect. A more recent study supports these findings. Brouzos, Vassilopoulos, and Baourda (2015) investigated the effects of group-leader perceived conditions on a sample of children participating in a group intervention for social anxiety. It was found that children who perceived higher group-leader level of regard, congruence, and empathy, reported a significant increase in self-reported likability, in comparison to those who perceived low levels of therapeutic conditions. These findings have been supported by a qualitative meta-analysis, in which young people reported helpful and unhelpful aspects of their counselling experience. Providing a non-judgemental space, and the extent to which their therapist was understanding was reported as the most
helpful factors of their counselling (Griffiths & Griffiths, 2013); supporting the positive impact of the core conditions of empathy and unconditional positive regard.

The majority of studies have measured the core conditions together as one entity. However, Green (1996) measured empathy in isolation. It was found that engagement in therapy was associated with higher levels of perceived therapist empathy. However, the parents of the children in therapy, not the children themselves, reported on the level of therapist empathy.

At the time of writing, it appears evidence for the use of Rogers’s therapeutic condition in therapies for children and adolescents is limited.

2.3.4 Rationale for Humanistic Therapy for Young People

As discussed the evidence for person-centred and humanistic therapy for young people is limited. Not all psychological treatments have been researched and evaluated to the same degree. Within the literature, clinical trials for CBT dominate; however, in practice, psychodynamic, systemic and humanistic therapies are routinely offered. Research on the use and potential effectiveness of these alternative forms of therapy is required, particularly when CBT proves to be ineffective for some young people (Southam-Gerow, Kendall, & Weersing, 2001).

Within CAMHS, CBT is most commonly delivered, being recommended by the NICE guidelines for a range of mental health difficulties (Stallard, Udwin, Goddard, & Hibbert, 2007). CBT is typically protocol-based and aimed towards a specific symptom or disorder (e.g. depression or social anxiety). However, many young people report or experience feelings that are not diagnosable or conform to traditional diagnostic criteria, such as exam
stress or relationship difficulties. Therefore, CBT may not always be appropriate for these experiences, and less manualised interventions could be useful for young people.

In addition, the core conditions of person-centred therapy have been likened to the common factor of ‘therapist interpersonal skills’. There appears to be considerable overlap with the core conditions of empathy, congruence and unconditional positive regard, with interpersonal therapist skills (Midgley, Hayes, & Cooper, 2017). Within adult populations, there is evidence to suggest these skills contribute positively to therapeutic outcomes and the therapy process (Lambert & Norcross, 2019). In particular, a meta-analysis of 82 empathy outcome studies in adult populations found a mean effect size of 0.58 (Elliott, Bohart, & Watson, 2019). In addition, small to medium effects were found from a meta-analysis of 64 positive regard outcome studies (Farber, Suzuki, & Lynch, 2019) and a meta-analysis of 22 outcome studies of the core condition of congruence found an effect size of 0.46 (Kolden, Wang, Austin, & Klein, 2018). Therefore, it is possible these skills may also contribute to positive outcomes in child and adolescent therapy also.

Current evidence of these skills with young people is lacking; however, a meta-analysis of 16 studies found therapist interpersonal skills to be strongly related to positive treatment outcomes (Karver, Handelsman, Fields, & Bickman, 2006). However, the included studies were not specific to person-centred or humanistic therapy. Further evidence for both person-centred therapy and therapist interpersonal skill is lacking in young populations. Clinical trials are needed to build the evidence-based and to do this; a validated measurement scale is required.
2.4 Importance of Measurement

The evidence base for psychological therapy is growing, and there is mounting recognition of the benefits to many therapeutic treatments (Kazdin, 2007). As a consequence, mental health treatment is increasingly researched and evaluated in terms of its effectiveness, using outcome studies and randomised-control trials (Kazdin, 2008). Additionally, research has focused on understanding the therapeutic process, with the aim to uncover change mechanisms within therapy (Kazdin & Nock, 2003). The investigation of aspects of the client, therapist, or therapeutic process, which may elicit therapeutic change, has become central to the evidence-base of psychotherapy and counselling. This focus demonstrates a movement from research on specific models of therapy towards the common factors present across all therapies (Wampold, 2015).

Common factors are features of therapy that are not model specific and refer to the elements of therapies that bring about or influence therapeutic change (Wampold, 2015). Within adult populations, common factors are strongly believed to contribute to therapeutic change, with client motivation or the therapy relationship being most effective (Cooper, 2008). However, in child and adolescent populations, empirical evidence on common factors is limited due to the small volume of research in young populations (Hayes, 2017). In order to improve outcomes in therapy, there needs to be an understanding and further study of the components of therapy that work (Castonguay, 2013). Knowledge of the factors that lead to positive outcomes in therapy has the potential to lead to more successful outcomes when incorporated into the therapy (Kazdin, 2007). An awareness of the common factors of therapy can inform and adapt mental health services and treatments to include factors that promote and maximise the potential for change for clients.
The assessment and evaluation of the therapeutic process and outcomes are essential for improving and adapting psychological treatment (Kazdin & Nock, 2003). Quality instruments can encourage research in these areas. The use of a robust instrument in research can strengthen the evidence base and give confidence to findings (Margison et al., 2000). Sound instruments that aim to measure the therapeutic process are necessary for further research into common factors for therapy for young people. In particular, the common factor of therapist interpersonal skills, commonly including empathy, congruence, and regard, which is sparsely measured in child and adolescent populations (Hayes, 2017). Given that the therapeutic relationship is strongly associated with positive outcomes in adult therapy, measuring the level and presence of the relationship conditions in young populations could be useful for further research and practice. Robust instruments can encourage the conducting of clinical trials, with the aim to support the practice of humanistic therapy with young people. RCTs give scientific credence to therapeutic interventions in a field that is dominated by the medical model and can begin to protect and enhance person-centred and humanistic practice within young populations (Castonguay, 2013).

Regarding measurement in young populations, there needs to be a consideration of the context in which instruments are developed and validated, as measurement in young populations encounters complexity due to the varied brain development across childhood (Music, 2017). Developmental capacity varies greatly depending on the young persons' age as brain development changes dramatically over this period. For example, verbal comprehension at six years is vastly different at 16. Therefore, it is essential that when using measures with young people that they are valid, meaning the scale has been developed and tested with young populations (Hanley & Noble, 2017). Typically scales for young people are developed and tested using a select age range to account for brain development; therefore, this may limit the use or suitability of certain client rated measures within young populations.
As a consequence, scales may not be designed or validated for the age range under investigation. As a result, observer or therapist rated measures of therapy with young people that are designed to be administered by adults, may be more practical and suitable, as their use is not dependent on the age of the young person. Furthermore, the use of observer measurement considers Rogers’s perspective on the use of psychometric tests in client-centred counselling (Rogers, 1946). Rogers argues that using psychometric tests for the purpose of research are justified, given that the scores are not shared with the counsellor or client and do not invade the therapeutic space (Rogers, 1946). However, the use of psychometric tests can be seen as incompatible with the phenomenological perspective of person-centred work. Within person-centred practice, there is an honouring of the clients’ subjective experience (Rogers, 1959). Therefore it could be argued that an observer rated measure detracts from the clients’ own experience and instead prioritise the experience of the observer. However, by holding a pragmatic perspective in which the usefulness of information is prioritised, using an observer measure can benefit the client without influencing the therapy relationship (Rogers, 1946); this is discussed further in Section 3.1.

2.5 Current Measures of the Person-Centred Core Conditions in Young People

A systematic search strategy (Appendix A) was used to identify all relevant studies of person-centred instruments used in a population of young people. The search followed methods recommended by the Cochrane Handbook (Higgins & Green, 2017) and accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Moher, Liberati, Tetzlaff, Altman, & Prisma Group, 2009).
Inclusion criteria were as follows, a) population under the age of 18 years, b) studies reporting on measures of three of Rogers’s (1957) core conditions: empathy, unconditional positive regard, and congruence. Instruments could measure condition either in isolation or as a combination. Additionally, c) studies related to psychological therapy d) original studies and e) publications in English or translation available. The study was designed to identify tools used to measure the core conditions of person-centred therapy in a population of children and young people. Therefore the following exclusion criteria were applied a) population older than 18 years, b) qualitative measurements and c) no English version available.

Search terms included a combination of ‘EMPATHY’ ‘UNCONDITIONAL POSITIVE REGARD’ ‘CONGRUENCE’ ‘CHILD*’ ‘ADOLESC*’ ‘CORE CONDITION*’ ‘INTERVENTION’ and ‘THERAP*.’ Different search terms were used to maximise the potential publications found, without losing its specificity. For this reason, the terms ‘THERAP*’ and ‘INTERVENTION*’ were included. This was to differentiate between person-centred care, as seen in nursing, for example, and person-centred therapy. A combined ‘Mesh’ and ‘free text’ search through PubMed (http://www.ncbi.nlm.nih.gov/) was conducted on the 28th December 2016, yielding 246 items. A further search of PsychINFO was conducted using OvidSP (www.ovidsp.uk.ovid.com) on the 6th January 2017, yielding 2061 items. The Web of Science database (http://apps.webofknowledge.com/) was searched on the 26th January 2017, yielding an additional 981 publications. In addition, a free text search in Google Scholar was conducted, to identify studies that met the inclusion criteria, but were missed by the database search; this yielded 12 further items. Titles and abstracts of all publications were reviewed to refine articles relevant to the measurement of core conditions in a population of children and young people. This yielded 227 items, which were read in full and discussed with supervisors regarding their suitability for the inclusion of the
review. This led to the exclusion of 222 publications. Therefore upon search completion, five studies using four different instruments were to be included in the qualitative synthesis. Figure 1 provides an overview of the search and selection procedure.

The four identified instruments include The Basic Empathy Scale, developed by Jolliffe and Farrington (2006). The Parent Questionnaire, found in Green (1996), the Barrett-Lennard Relationship Inventory (Barrett-Lennard, 1962), and the Truax Scales, developed by Truax and Carkhuff (1967). In addition, since the completion of the systematic search, a further updated search has been conducted, and it was found that the Person-Centred and Experiential Psychotherapy Scale (Freire, Elliott, & Westwell, 2014) has been adapted for use with young people, and therefore is included in this review.
Figure 1. Summary of search and selection procedure.
Psychometric Evaluation of Identified Scales

Each of the scales identified was evaluated for their psychometric quality using the criteria of the consensus-based standards for the selection of health measurement instruments (COSMIN) (Mokkink et al., 2010). These standards were developed in an international Delphi study in 2005, and have subsequently been verified. Though initially proposed for the evaluation of patient-reported health outcome measures, the use of the COSMIN checklist has also been recommended for the evaluation of other measurement instruments (Mokkink, Terwee, Knol et al., 2010). A summary of COSMIN criteria is outlined in Appendix B. Quality is assessed and measured on an ordinal scale of excellent, good, fair, and poor (Terwee et al., 2012). A principle of ‘worse score counts’ is observed when describing the overall quality of the development publication for each scale.

2.5.1 The Basic Empathy Scale

Jolliffe and Farrington (2006) developed the Basic Empathy Scale (BES). The BES (Appendix C) is based on Cohen and Strayer’s (1996) conception of empathy: “The understanding and sharing in another’s emotional state or context” (Cohen & Strayer, 1996, p.523). This definition includes the experience of empathy, as described as an affective trait, and a cognitive ability. Using this definition, the authors developed a 40 item scale that aimed to measure cognitive empathy and affective empathy. Following an initial exploratory factor analysis, using 363 participants, the scale was reduced to 20 items. The scale includes items such as, ‘I get caught up in other people’s feelings easily’ and ‘I find it hard to know when my friends are frightened.’ Respondents are asked to indicate the extent to which these statements describe them, using a 5-point Likert-type scale. The authors used two waves of
testing: the first, on 363 adolescents from three schools in Hertfordshire, England. The second wave completed precisely one year after consisted of 357 adolescents from the same schools. The initial exploratory factor analysis, which shortened the scale to 20 items, was based on the evaluation of 363 participants; however, the overall validation of the scale was based on all 720 respondents.

Structural validity of the 20-item scale was tested using a two-factor confirmatory factor analysis. Measures of goodness of fit indices support the separation of cognitive and affect items (goodness-of-fit index [GFI] = 0.89, adjusted goodness-of-fit [AGFI] = 0.86, and root mean square residual [RMS] =0.06). Authors expected positive correlations with sympathy and perspective taking (Interpersonal Reactivity Index), verbal fluency (Thurstone Verbal Fluency Task), and personality (Big Five Inventory), and negative correlations with alexithymia (The Toronto Alexithymia Scale) parental supervision, and low socioeconomic status. These predictions were correct for all, excluding verbal fluency, impulsivity, and parental supervision.

In addition to the original validation presented above, the BES has shown good cross-cultural validity in youth populations across Europe (Albiero, Matricardi, Speltri, & Toso, 2009; D’Ambrosio, Olivier, Didon, & Besche, 2009; Salas-Wright, Olate, & Vaughn, 2013; Topcu, & Erdur-Baker, 2012) and Asia (Ang & Goh, 2010; Geng, Xia, & Qin, 2012). The construct validity of the scale is further supported, as researchers have found positive correlations between the translated BES with the Interpersonal Reflexivity Index (D’Ambrosio et al., 2009) and the Balanced Emotional Empathy Scale (Mehrabian, 1996).

**Evaluation:** The overall COSMIN methodological quality score for the development of the BES, based on Jolliffe and Farrington’s (2006) publication, is poor. No information was provided regarding the assessment of content validity, and details of reliability of internal consistency were absent. However, assessments of the scale’s structural validity were
of high methodological quality. The BES is one of the few measures that has been developed and validated using a child or adolescent population. However, the definition and theory of empathy used for the scale do not match that of Rogers (1957). There is an element of coherence regarding affective empathy; however, this is not as comparable as necessary in order to be considered as measuring the core condition of empathy. Therefore, the Basic Empathy Scale cannot claim to measure any of the core conditions of person-centred therapy with certainty.

2.5.2 Parent Questionnaire

The identification and the development of The Parent Questionnaire (Appendix D) are found in Green (1996). The questionnaire was designed to reflect the complexity of the therapeutic alliance (with regards to empathy). In particular, it was based on the work of Hougaard (1994) and Bordin (1979), in which the therapeutic alliance is divided into components regarding personal relationships and collaborative relationships within the therapy bond. The Parental Questionnaire focuses on the collaborative nature of the therapeutic alliance and is an extension of a follow-up questionnaire that had been used at the Tavistock Clinic. No reference in the text was provided for the follow-up measure, as the author obtained this via his personal correspondence. Therefore, details regarding the scale’s development are lacking. Eleven items are included in the questionnaire, three of which are used to understand empathy and the therapeutic alliance. These include the questions: ‘Do you feel that your therapist(s) accurately understood the nature of the problems in the child or family?’ ‘Did the kind of help offered to make sense to you as a treatment for the problems identified?’ and ‘Did you feel you understood the reasons for the kind of treatment offered?’ Respondents were asked to answer these questions using a 4-point Likert type scale, ranging
from ‘Not at all’ to ‘Completely.’ The scale was designed for both adolescents and adults. However, due to a poor response rate, the scale was tested in adults only. More specifically, it was tested using 25 parents of children who were receiving psychological treatment at Booth Hall Hospital, Manchester, England.

In addition to the 25 participants, therapists used the scale to rate their understanding of the therapeutic alliance. Comparisons between these rating using weighted Kappa suggest at best a good agreement between the different raters (scores between 0.3-0.5) (Cohen, 1968). Furthermore, as no details are given regarding the therapist sample size, it is not possible to fully comment on the inter-rater reliability. Green (1996) indicates some evidence of formal content validation; however, very few details are given, and only three of the 11 items are understood to be relevant to the construct measured. Further exploration of the validation of the Parental Questionnaire has not been performed; therefore, it is difficult to comment further on the psychometric quality of this instrument.

**Evaluation:** The overall COSMIN methodological quality score for the Parental Questionnaire, based on Green (1996) is poor, as there is little to no evidence supporting a robust evaluation of the instrument. Furthermore, although the scale claims to measure empathy, the definition and theory used to develop the scale focuses on the therapeutic alliance and makes little mention of empathy. Therefore we can question the extent to which this scale measures the core condition of empathy. Lastly, it is doubtful that this scale has been validated for the use of child and adolescent populations, even though this was one of the design aims of the scale.
2.5.3 Barrett-Lennard Relationship Inventory

Details regarding the development and validation of the BLRI will be discussed further in section 2.6; however, the use of the BLRI and validation in child and adolescent populations is discussed here.

The BLRI (Appendix E) was developed by Barrett-Lennard (1959; 1962). The instrument was designed to measure the therapist-to-client relationship conditions and is based on Rogers’s (1957) core conditions for therapeutic change. According to the author, the definitions of empathy and congruence reflected those provided by Rogers (1957) however, unconditional positive regard was divided into two separate concepts: ‘level of regard’ and ‘unconditionality of regard’ as initially proposed by Standal (1954). Based on these therapeutic variables, a 64-item scale was developed, comprising of four subscales: ‘Empathic Understanding’, ‘Level of Regard’, ‘Unconditionality of Regard’, and ‘Congruence’. Example items include, ‘I feel that I am genuinely myself with (therapist)’ or ‘(Therapist) is friendly and warm towards (client).’ Respondents use a 6-point Likert Scale to indicate the extent to which they feel the statements are true, ranging from -3 (No, I feel strongly that is not true) to +3 (Yes, I feel strongly that is true). A formal content validation procedure was used to develop the content of the scale. Five judges were approached to classify 85 items for its relevance and importance towards the variable, as a result of the validation procedure, a subscale of 16 items represented each of the four variables. The internal consistency of the original development study as assessed using the split-half method suggested satisfactory internal consistency of the measure, (Spearman-Brown correlation \( r \geq 0.82 \) for each sub-scale). Additionally, test-retest reliability indicated good stability of the measure over four weeks (total BRLI score \( r = 0.95 \)).
Much of the additional validation of the BLRI comes from a review of the evidence, from Gurman (1977), high internal reliability ($r=0.91$ for total BLRI score) has been found across 14 studies. In addition, inter-correlations between the subscales found high correlations between empathy, level of regard and congruence, suggesting good internal consistency. Furthermore, a factor analysis based on three studies suggests good construct validity (Gurman, 1977). With regards to validation in child and adolescent populations, there has been little work. However, a study found via the systematic search (Brouzos et al., 2015) found evidence for satisfactory re-test reliability over six weeks; however, it is important to note that the completion of the BLRI was not conducted in the same settings.

**Evaluation:** The overall COSMIN methodological quality score for the BLRI, based on Barrett-Lennard (1962) is poor, as there is little to no evidence testing the external validity of the measure. However, subsequent evidence for the scale’s validation is promising. In addition, it is important to note that the above is a commentary of the validation of the 64 item BLRI. Subsequent versions of the scale have been developed, and include: a shortened 40 item scale, an observer-rated scale, scales adapted for group use, family therapy, and within educational settings (Barrett-Lennard, 2015). These additional adaptations of the scale have not yet been robustly evaluated in the literature. The BLRI was explicitly developed to test the core conditions of therapeutic change; therefore, it is a highly appropriate instrument for the measurement of person-centred therapies. Nevertheless, the BLRI has yet to be evaluated in populations of children and adolescents.

### 2.5.4 Truax Scales

The Truax Scales include The Scales for Therapist Accurate Empathy, Non-possessive Warmth, and Genuineness (Appendix F). The development and validation of these
scales are described in Truax and Carkhuff (1967). The scales were developed with Rogers’s relationship conditions in mind; however, adjustments were made. For example, empathy was changed to ‘Accurate Empathy,’ unconditional positive regard to ‘Non-possessive Warmth,’ and congruence to ‘Genuineness’ (Truax & Carkhuff, 1967). For each of these relationship conditions, a separate scale was developed. These are observer-rated scales, intended for the measurement of the relationship conditions as demonstrated by the therapist. The first, the ‘Accurate Empathy Scale’ is measured using a nine-point anchored rating scale, in which the observer must rate the therapist on the level of accurate empathy demonstrated. The scale ranges from Stage 1, in which the therapist has ‘no awareness of the clients feeling’, to Stage 9, where ‘the therapist responds to the client’s full range of feelings in their exact intensity’ (Truax & Carkhuff, 1967, p.56). The second scale, measuring ‘Non-Possessive Warmth’ uses a five-point anchored rating scale, from Stage 1 ‘the therapist is actively offering advice or giving clear negative regard’ to Stage 5, where ‘the therapist communicates warmth without restriction’ (Truax & Carkhuff, 1967, p.66). Finally, the scale of therapist ‘Genuineness’ uses a five-point anchored rating scale. Stage 1 indicates a ‘discrepancy between what the therapist says and what he experiences’ and Stage 5 indicates ‘a therapist that is freely and deeply himself in the relationship’ (Truax & Carkhuff, 1967, p.72).

No information was given regarding the original development and validation of the Truax Scales; however, an estimation of the inter-rater reliability from 28 studies report values of r=0.43-0.79 for The Accurate Empathy Scale, r=0.48-0.84 for the Non-possessive Warmth Scale, and r=0.42-0.62 for the Genuineness Scale (Truax & Carkhuff, 1967). These estimations were made using a variety of client and therapist populations, and therefore the authors suggest high reliability for the scales. However, the quality of these 28 studies has been debated, as many studies had inadequate sample sizes, and therefore the same therapists
were evaluated more than once by the same raters. Also, many of the studies measured the conditions in a group therapy setting, which is not what the scale was intended to measure.

Subsequent publications have commented on the psychometric quality of the Truax Scales. Shapiro (1968) tested the construct validity of the scales by correlating the Truax Scales with a 7-point observer scale of empathy, warmth and genuineness. A significant correlation was found with the accurate empathy scale; however, stronger correlations were found with the warmth and genuineness scales. Rappaport and Chinsky (1972) subsequently criticized these finding, by suggesting the accurate empathy scale measures a general therapist quality, as opposed to empathy categorically.

There has been no formal validation of the Truax Scales in a population of children or adolescents; however, studies identified via the systematic search do give some indication of the reliability of the scales when used in child therapy. Siegel (1972) found high rater reliability with the Accurate Empathy Scale (r=0.81), the Unconditional Positive Regards Scale (r=0.70), and the Therapist Congruence Scale (r=0.84). In addition, Truax et al. (1973) found high inter-rater reliability with the Accurate Empathy Scale (r=0.72). However, this trend was not observed with the Genuineness Scale or the non-possessive Warmth Scale, possibly as these scales do depend on a counsellors overt verbal response and therefore may not be a clearly recognised when rating.

**Evaluation:** The overall COSMIN methodological quality score for the Truax Scales, based on Truax and Carkhuff (1967) is poor, as there is little to no evidence regarding the development and evaluation of the Scales. It can be argued that the Truax Scales are a close representation of the core therapeutic conditions. However, evidence of the scale’s validation is lacking and near non-existent, this is especially true when considering its use with children and young people.
2.5.5 PCEPS-YP

The Person-Centred Experiential Psychotherapy Scale – Young Person Counselling Version (Appendix G) is an adaptation of the Person-Centred Experiential Psychotherapy Scale (PCEPS) developed by Freire et al. (2014). The 15 item PCEPS (Appendix H) was developed in accordance with the humanistic therapy competences (Roth, Hill, & Pilling, 2009), and acts to support randomised control trials of person-centred therapy, in addition to acting as an adherence measure of humanistic therapies. The internal reliability of the original PCEPS scale indicated high internal consistency, across all 15 items (Cronbach’s alpha 0.98) (Freire et al., 2014). In addition, the average inter-rater reliability, as tested with six raters, was reported at Cronbach’s alpha 0.87 when all 15 items were averaged together (Freire et al., 2014). However, convergent validity assessed via correlations with the Working Alliance Inventory, the Revised Session Reaction scales and the client and therapist Therapeutic Relationship Scale were found to be inconclusive due to multiple confounding variables (Freire, Elliott, & Westwell, 2012).

The PCEPS-YP is adapted for use with young people and was developed using the BACP’s framework of competences for working with children and young people (British Association for Counselling and Psychotherapy, 2017). The PCESP-YP consists of nine items and is currently undergoing psychometric validation in a population of young people (Ryan et al., 2019). The scale uses an expanded response format; for each item, a 6-point descriptive response is provided. For example, for item 2 ‘Tracking’, responses range from ‘No tracking: Therapist responses divert client from their thoughts/feelings; therapist fails to adapt to client feedback’, to ‘Excellent tracking: Therapist is sensitively and actively follows the client’s track, quickly responding and revising perceptions based on client feedback.’ In
addition, each item is accompanied by a description and further questions designed to explain the content of the item.

2.6 Rationale for BLRI

After reviewing the current instruments of person-centred and humanistic therapy in young people, the BLRI was deemed the most appropriate scale to undergo further validation for the focus of the study.

The BLRI was created and developed specifically to test Rogers’s theory of personality change (Barrett-Lennard, 1959). The scale aims to measure the quality and presence of the therapeutic relationship conditions, including, empathy, congruence and unconditional positive regard (Barrett-Lennard, 1962). The concepts of Empathic Understanding and Congruence essentially mirrored the definitions provided by Rogers (1957). In addition, the subscales of Unconditionality and Level of Regard were derived from Rogers’s concept of unconditional positive regard. Therefore, the latent variables of the scale were defined and developed using Rogers’s operationalized definitions of the core conditions and aimed to match as close as possible to Rogers’s original concepts (Barrett-Lennard, 1959). Furthermore, the item pool for the scale was created using experts in the field of person-centred theory. During development, there was a constant interaction between person-centred theory and the operationalised expressions, leading to a progressive refinement of the meaning and concepts underlying each variable (Barrett-Lennard, 1962). Since the BLRI’s initial conception and initial development, the scale has been refined and continually utilised in many settings and populations (Barrett-Lennard, 2015). The scale has demonstrated robust validity and reliability from a range of studies, further encouraging its use in person-centred, humanistic, and relationship contexts (Barrett-Lennard, 2015).
It could also be argued that the Truax Scales are an appropriate measure of the core conditions and ought to be further validated. The Truax scales were also developed with an aim to measure Rogers’s client-centred theory (Truax & Carkhuff, 1967), and indeed the three scales each measure one of the core conditions: empathy, unconditionality, and unconditional positive regard. However, there is very little information available regarding the initial design of the scales, and information regarding the scale’s formation, validation, and development is lacking. Furthermore, the information that is available demonstrates poor inter-rater reliability and poor convergent validity (Shapiro, 1968). Additionally, sample sizes are small, and much of its use has been in group therapy environments as opposed to individual therapy (Truax & Carkhuff, 1967).

The BES was found to be inappropriate for use in this study as is not a measure of relationship empathy but rather cognitive and affective empathy (Jolliffe & Farrington, 2006). Therefore, the BES cannot be said to measure Rogers’s understanding of the core condition of empathy. Additionally, as a measure of empathy only, the scale is limiting as it fails to measure the conditions of unconditional positive regard and congruence.

The PQ is also inappropriate for use in this study as only one item in the scale aims to measure the therapeutic relationship: item 6 ‘Do you feel that your therapist(s) accurately understood the nature of the problems in the child or family?’ Furthermore, the scale is limited, as it does not measure empathy from the perspective of the young person in therapy but of the parent’s perspective (Green, 1996). Therefore, answers to this item may not be informed by the quality of the relationship, but rather counsellors report of the therapy to the parent.

At the time of reviewing the PCEPS had not yet been adapted to for use with young people and therefore was not considered for further validation in this study. However, at the time of writing the PCEPS-YP is currently undergoing further validation (Ryan et al., 2019).
Therefore, after reviewing the available scales, the BLRI was chosen for this project. More specifically, the 40 item observer rated version of the BLRI was selected for further validation.

2.7 Current Psychometric Evidence and Evaluation of the BLRI

2.7.1 Development and Revision of the BLRI

Aim of the BLRI

The Barrett-Lennard Relationship Inventory was developed and designed to measure the therapist-to-client relationship conditions and is based on Rogers’s (1957) core conditions for therapeutic change: empathy, congruence, and unconditional positive regard. The motivation for the scale worked to connect the ‘cause and effect’ in the therapy process. Based on Rogers’s core conditions, the study (Barrett-Lennard, 1959) focused on the proposition that therapeutic change occurs according to the degree to which a client experiences the therapist-to-client relationship conditions. Therefore, the primary aim of the original 1959 study was to determine whether each of the measured qualities of the relationship significantly predicted therapeutic improvement and whether more experienced therapists’ demonstrated greater levels of the relationship qualities than inexperienced therapists. Specifically, the scale was initially designed to measure the client’s experience of the relationship qualities being communicated by the therapist, as perceived by the client and self-perceived by the therapist.
The author proposed five therapeutic variables that influenced therapeutic change. These included positive regard, non-judgemental accepting, empathic understanding, congruence (or genuineness), and willingness to be known. Definitions of empathy and congruence reflected those provided by Rogers (1957); however, unconditional positive regard was divided into ‘level of regard’ and ‘unconditionality of regard’ as proposed by Standal (1954). The measure of ‘willingness to be known’ was included in the earliest version of the scale; however, this was subsequently removed from later editions due to its lack of predictive power in relation to therapeutic outcomes (Barrett-Lennard, 1962).

Therefore, based on the remaining therapeutic variables, a 64-item scale was developed, comprising of four subscales: ‘Empathic Understanding’, ‘Level of Regard’, ‘Unconditionality of Regard’, and ‘Congruence’. Items representing each variable were dispersed evenly across the scale to obtain maximum independence of answers, and positively and negatively worded items were incorporated at random. Example items include, ‘I feel they (therapist) respect me as a person’ and ‘They (therapist) are indifferent to me (client).’

Content Validation

A formal content validation procedure was used to develop the content of the scale. Five judges were approached to classify 85 items for its relevance and importance towards the variable. Judges agreed on all but four items, which were subsequently removed, and an additional three items were removed due to duplication of content. As a result of the validation procedure, a subscale of 16 items represented each variable (Barrett-Lennard, 2015).
**Scoring Method**

To determine the clarity and strength of the conditions, a 6-point Likert Scale is used to indicate the extent to which respondents feel the statements are true, ranging from -3 (No, I feel strongly that is not true) to +3 (Yes, I feel strongly that is true). There is no neutral response option provided.

**Sampling**

The original sample used for the development and testing of the instrument consisted of 42 clients and 21 therapists, from the Counselling Centre of the University of Chicago. The client sample consisted of student and community members (age range 19–45 years, with a mean of 28 years old). 60% were male, and approximately half were college graduates. Length of therapy ranged from 7 – 96 sessions, with an average therapy length of 33 sessions. Data were collected from both therapists and clients after the first five sessions, the 15th, 25th and final session of therapy (Barrett-Lennard, 2015).

**Initial Reliability Analysis**

The initial reliability analysis of the scale makes use of data collected from the fifth session of therapy, as the sample size was the largest at this point. The internal consistency of the sub-scales was assessed using the split-half method and found to be highly correlated in both client and therapist samples (Spearman-Brown correlation $r \geq 0.82$ for each sub-scale, $r=0.82-96$), suggesting satisfactory internal consistency of the measure. Furthermore, inter-scale correlations between the subscales demonstrated strong associations between empathy and congruence. In contrast, unconditionality and regard were weakly correlated, which the author took to support the division of unconditional positive regard to two distinct variables.
Additionally, the author subsequently found support for test-retest reliability (total BRLI score $r=0.95$) over a four-week period.

**Revisions of the BLRI**

The scale was revised for use in different settings. It was believed that the core factors driving the therapeutic process were qualities of relationships in general. Therefore any helping or developmental relationships are contexts in which these qualities may play a vital role. Therefore, adapted forms of the BLRI were developed for use with groups, observers, and family and school settings. Furthermore, revisions included refining the concepts of congruence and unconditionality, and balancing the number of positive and negatively worded items in each subscale. In addition, an item analysis was conducted to assess the homogeneity across items in each of the subscales and investigate which items contributed more or less to the subscales, following this; the original 72-item scale was reduced to 64 items, and the subscale of ‘willingness to be known’ was removed (Barrett-Lennard, 2015).

**2.7.2 Subsequent Psychometric Evaluation of the BLRI**

**Reliability**

Additional evidence for the validation of the BLRI comes from a review of studies using adult populations, from Gurman (1977). Regarding the internal consistency of the scales, 14 studies across therapy, friendship and teaching settings found high internal reliability. An average internal reliability coefficient of 0.91 for total BLRI score, 0.84 for Empathy, 0.91 for Level of Regard, 0.74 for Unconditionality of Regard and 0.88 for Congruence was found across the 14 studies (Gurman, 1977). Additionally, inter-scale correlations have been used to assess internal reliability. Based on findings from seven
studies, the average inter-correlations between the subscales of the BLRI have been measured. Findings demonstrate high positive correlations between the subscales of empathy, congruence, and level of regard. However, low correlations with the subscale of unconditionality were found, suggesting a high level of independence for this condition (Gurman, 1977).

Reports of test-retest reliability across ten studies demonstrate high reliability over time, with mean correlations of 0.90 for total BLRI score, 0.83 for Empathy, 0.83 for Level of Regard, 0.80 for Unconditionality of Regard, and 0.85 for Congruence. The majority of the included studies re-administered the scale at a short interval (typically within a month), however, three studies found a high degree of stability over a period of up to one year (Hollenbeck, 1961; Kiesler, Mathieu, & Klein, 1967; Kiesler, Klein, & Mathieu, 1967).

Validity

Since its development, many studies have assessed the BLRI’s construct validity across many settings. Using Principal Component Analysis, Mills and Zytowski (1967) extracted three factors from the BLRI. They concluded that there appeared to be a single dominant factor that the four subscales contributed to; with the subscale of Unconditionality of Regard found to have the lowest loading of all subscales. This finding was further supported by McClanahan (1974). Using Principal Component Analysis on data from 83 clients from a university counselling centre, a one-factor solution was found, suggesting unidimensionality of the relationship inventory, again the Unconditionality of Regard scale had the lowest loading. Support for a one-factor model has also been found in Lin (1973) and Gross, Curtin, and Moore (1970), suggesting non-independence among the subscales of the BLRI. Evidence to support the independence of the Unconditionality of Regard scale has also been found, Lewis and Krauss (1971) found a two-factor solution of the BLRI. One factor
had positive loadings for Empathy, Level of Regard, and Congruence. The second factor showed positive loading for Unconditionality of Regard only. Walder and Little (1969) using item inter-correlations and Principal Component Analysis, found a three-factor solution for the BLRI, including ‘Non-Evaluative Acceptance’ ‘Psychological Insight’ and ‘Likability’.

Potentially the most in-depth factor analysis study of the BLRI comes from two studies by Lietaer (1974a; 1974b). An extended BLRI of 123 items was administered to 800 college students who were asked to rate their relationship with either their mother or father. Using a rotated Principal Component Analysis, five factors emerged: Empathy, Positive Regard, Unconditionality, Transparency, and Directivity. In the second study, the 123 item BLRI was completed by 100 client-therapist dyads, half of which were receiving client-centred therapy, the other half receiving psychoanalytic therapy. Six factors emerged: Empathy, Regard, Unconditionality, Incongruence, Transparency, and Directivity. Unconditionality was found to be independent from Empathy, Regard, and Incongruence in the psychoanalytic population but not in the client-centred population. Using the scale to rate the relationship with group leaders, Bebout’s 1971 study (as cited in Gurman, 1977) found a five factors solution similar to Lietaer (1974a) including: Empathic Understanding, Positive Regard, Conditionality, Open-Genuine, and Non-Imposing.

Cramer (1986a) assessed the original 69 item relationship inventory in a sample of 169 participants who rated their relationship with their closest friend. Using Principal Component Analysis, Cramer extracted 19 factors, the first four of which supported the four subscales of the BLRI. Cramer (1986b) found similar results when assessing the 64-item scale and took this as support for the construct validity of the BLRI.

Most recently, the structural validity of the BLRI has been assessed via Confirmatory Factor Analysis. Liao, Murphy, and Barrett-Lennard (2018) used both Principal Component Analysis and Confirmatory Factor Analysis to assess the Chinese-Mandarin translated 64
item and 40 item BLRI in a sample of 658 students who measured their relationship with a specific friend. Results from both scales supported the original four subscale dimensionality of the BLRI.

**Summary**

Since its development, the BLRI has undergone considerable assessment regarding its psychometric quality (Gurman, 1977). However, it can be argued that much of the psychometric evidence of the BLRI is out dated. In recent years there has been very little study into the reliability and validity of the scale with the exception of Liao et al. (2018). However, this study used a translated version of the BLRI, and did not focus on a therapeutic relationship but rather a friendship. Therefore, an updated evaluation of the scale’s reliability and validity is necessary in order to preserve its continued use within therapeutic contexts.

Furthermore, much of the evidence from Gurman (1977) reports the reliability and validity as an average value, across multiple studies. Many findings have been averaged across studies that have not used the same populations or study contexts (e.g. parental relationships and therapeutic relationships). Therefore, there is not a clear picture of the scale’s psychometric merits for specific populations or presentations.

Moreover, in the existing literature, there is very little distinction between the different versions of the BLRI forms in terms of their psychometric quality. Specifically, there is almost no validation of the observer rated form with any population or relationship setting, and even less information available regarding its validation with young people.

Finally, the methods of statistical analysis used to evaluate the psychometric quality of the BLRI do not conform to the current recommended standards (Mokkink et al., 2010). Many past evaluations of the structural validity have used exploratory factor analysis, when, in fact, confirmatory factor analysis would be better suited. In addition, tests of reliability
have previously been conducted using correlation methods, whereas Cronbach’s alpha and intraclass correlation coefficients are now preferred (Perinetti, 2018).

2.8 Aims and Objectives

This study aims to evaluate the psychometric properties of the Barrett-Lennard Relationship Inventory Obs-40 (Version 3) in a population of young people receiving humanistic counselling.

The objectives are to examine:

1. Internal reliability of the BLRI Obs-40
2. Inter-rater reliability of the BLRI Obs-40
3. Test-retest reliability of the BLRI Obs-40
4. Convergent validity of the BLRI Obs-40, using the PCEPS-YP
5. Structural validity of the BLRI Obs-40
6. The psychometric methodological quality of the evaluation of the BLRI Obs-40, using the COSMIN checklist (Mokkink et al., 2010).

2.9 Research Questions

1. What is the reliability of the BLRI Obs-40 in humanistic counselling for young people? More specifically:
   a. What is the internal reliability as measured by Cronbach’s alpha for each of the subscales and the overall scale of the BLRI Obs-40?
b. What is the inter-rater reliability as measured by Cronbach’s alpha, Intraclass correlation coefficient, and Krippendorff’s Alpha, using three independent raters?
c. What is the test-retest reliability as assessed using Pearson’s Product Moment Correlation, Cronbach’s alpha, and Intraclass correlation coefficient?

2. What is the validity of the BLRI Obs-40 in humanistic counselling for young people?

More specifically:

a. What is the convergent validity against the PCEPS-YP as assessed using Pearson’s Product Moment Correlation?
b. What is the structural validity of the BLRI Obs-40 as assessed using Confirmatory Factor Analysis?

3. What is the overall psychometric methodological quality of the BLRI Obs-40 as evaluated by the COSMIN criteria?

2.10 Chapter Summary

Children and adolescents can experience a range of mental health difficulties, which can have detrimental effects lasting into adulthood. Psychological intervention can be effective in reducing psychological distress in this population, including person-centred and humanistic therapies. However, currently, there is very little research available to support these claims. The presence of a validated instrument that measures the core conditions of person-centred therapy can encourage further research and can begin to build the evidence base for the effectiveness of person-centred and humanistic therapy for young people. The BLRI is a suitable measure for this purpose; however, despite the existing literature on the evaluation of the BLRI, there is no validated BLRI scale for use with young people.
Chapter 3: Methods

3.1 Epistemology

3.1.1 Overview

Section 3.1 will discuss the epistemological foundations of this project, which rely on different sources of information. The methods and subsequent analysis are grounded in positivism and empiricism. However, the use of observational measurement instruments invites a discussion on the appropriateness of positivism for this project as subjective evaluations and interpretations are introduced through the items of BLRI Obs-40. Furthermore, the subjective experience that is prized within a therapeutic relationship is considered. Subsequently, the challenges of holding a positivistic approach when measuring person-centred therapy are examined with reference to Rogers’s origins in logical positivism. Finally, as a consequence of this struggle, pragmatism is introduced in an attempt to reconcile the disagreement between the positivistic methods adopted in the project and the constructionist setting in the therapy room.

3.1.2 Positivism

The methods and analysis of this project are based on the assumptions of empiricism, with foundations in positivist epistemology (Cacioppo, Semin, & Berntson, 2004). Central to empiricism is the belief that the world can be known via observation and objective truth can be uncovered via the scientific method (Ponterotto, 2005). Within a positivist epistemology, knowledge is meaningful only when it is obtained via scientific observation (Kutney, 2006).
Furthermore, this knowledge is believed to have originated from sensory-based experiences that can be experienced by all people from a singular reality (Smythe, 1992).

Furthermore, positivism is related to a realist ontology, where truths and reality are believed to be identifiable and measurable objectively (Ponterotto, 2005), key to positivism is the belief that anything that exists can be measured via the scientific method (Thorndike, 1918). Many of these epistemological and ontological assumptions can be clearly connected to the research aims and objectives of this project. There is an assumption that the therapeutic relationship can be measured via sensory observation. In addition, by using a large sample, there is an underlying intention of generalising the findings to the general adolescent population, implying the research can make universal laws and conclusions (Kitchener, 2004). Furthermore, the use of statistical analysis to solve practical problems, and determine reliability and validity, subscribes to a positivistic epistemology (Kim, 2003).

However, despite the cohesion between the methods and a positivist epistemology, the underlying philosophical assumptions of psychometrics conflict with a purely positivist approach (Robinson, 2014).

### 3.1.3 Psychometric Assumptions

Central to positivism is the quantifications of observation; however, much of psychometrics relies on self-report and introspection, in which participants are required to pass judgements on their own behaviours, thoughts and feelings (Robinson, 2014). Therefore, the extent to which self-report measures are objective and meaning free is questionable (Westerman, 2006). According to Comte (1855), self-observation as a psychological method is logically impossible, as the observer and the observed can never be separated, and therefore objectivity cannot be guaranteed. The lack of objectivity present in self-report
measure casts doubt on the appropriateness of a positivist epistemology that values objectivity of observation.

As the project makes use of observer-reported measures, it could be argued that it is still possible and appropriate to hold a positivist stance. However, the interpretation of self and observer report measures poses an additional problem with the relationship between psychometrics and positivism. Many psychometric measures require interpretation from both the researcher and the participant (Cook & Beckman, 2006). Furthermore, items on measures can refer to terms and constructs that may be ambiguous or have subjective meanings (Parker, 1999). For example, item 29 on the BLRI Obs-40, ‘S/he is friendly and warm towards B (client)’ can be interpreted in many ways depending on the raters’ understanding of what it means to be friendly and warm. The interpretation of a construct can become dependent on the raters’ personal understanding. In particular, there may be cultural differences or norms that contribute to this understanding, and therefore determine what score and judgement would be given.

Additionally, though it is possible for items on a scale to be useful, they cannot be exhaustive (Westerman, 2006), hence it is not possible to include items that cover every meaning of a construct in order to reduce interpretation differences, and in fact, scales with too many items are open to criticism as they can begin to include redundant items. Therefore, though, understandably, some interpretation will be required on these measures, the variability of interpretation cannot be dismissed.

Furthermore, a consequence of using observer-rated measures implies twice the amount of interpretation required. This double hermeneutic means the observer rater will interpret both the observed therapist and the items on the scale. This chain of interpretation is demonstrated by item 12 on the BLRI Obs-40, ‘My sense is that s/he is genuine and honest with B (client)’. In this example, the rater is first interpreting the item content, i.e. what it
means to be genuine and honest, and next interpreting the behaviours and actions of the therapist that may or may not demonstrate this construct. A chain of interpretation makes the different layers of evaluation dependent upon each other, meaning one evaluation in the chain (i.e. the raters’ understanding of the construct) may strongly influence the other (the extent the rater perceived the therapist demonstrating the construct).

Furthermore, as the recordings are secondary data, the opportunity to engage in open-ended discussion with the observed therapist is lost, meaning there is little opportunity to discover the varied interpretations of the constructs measured. At most ratings could be justified through an evaluation of test-retest reliability; however, there is the danger of this confirming (potentially biased) interpretations already made. It is possible that examining inter-rater reliability could evaluate the interpretations made; however, as these raters will also be susceptible to biases when making interpretations (Hoyt, 2000).

The difficulty in holding a purely positivist stance for this research is further amplified by positivism’s ontological foundation in naïve realism (Ponterotto, 2005). The positivistic belief that there is one true reality that can be identified and measured in a value and context free environment is in direct opposition to the nature of the therapeutic relationship (Freire, 2009). Furthermore, the belief that positivistic approaches can be free of bias is also flawed, as the extent to which a researcher can be detached and objective is limited within psychometrics (Robinson, 2014). In this context there are multiple subjective interpretations occurring across different observers and therapist-client dyads over which the researcher will have some influence over. Therefore, a positivist stance does not appear to compliment the project entirely, and an alternative perspective must be sought.
3.1.4 Therapeutic Relationship

The methods and analysis of the project are grounded in empiricism. However, attention also needs to be paid to the environment in which measurements and observations are made. In particular, the BLRI Obs-40 will be used to measure the characteristics and qualities of the therapeutic relationship. Within the therapeutic relationship, there is, according to Rogers a strong emphasis on the clients’ subjective experience (Rogers, 1961); however, a positivistic approach assumes research is value, context, and bias free, and reality exists independent to social contexts (Plack, 2005). Therefore an incompatibility exists between the objective scale measurement and the subjective and phenomenological nature of person-centred therapy.

Furthermore, within the therapeutic relationship and environment, there is an element of relativism. Relativism is an ontological position in which there exist multiple constructed realities rather than one singular reality (Ponterotto, 2005). The belief of numerous and subjective realities are both at the centre of therapeutic practice and constructionism (Corrie & Callahan, 2000). Constructionism assumes multiple valid realities (Schwandt, 1994), which are formed in the mind of an individual (Hansen, 2004). According to constructionism, meaning is not entirely obvious but is developed and made clear through reflection (Schwandt, 2000). Constructionism, much like therapeutic contact, seeks to understand the ‘lived experiences’ of an individual, typically within their social and historical foundations (Ponterotto, 2005). Therefore given the epistemological differences in the foundations of measurement and that of person-centred therapy, a purely positivist approach which fails to account for the clients’ experience is limiting and naïve. Furthermore, it is possible that given the phenomenological foundations of person-centred theory (Freire, 2009) one can begin to
question if there is even a quantitative structure to the therapeutic relationship that can subsequently be measured (Freire & Grafanaki, 2010).

This reservation is reflected in the foundations of Person-centred theory and Rogers’s early research. Person-centred theory originated within an environment of logical positivism, which was dominant with objectivism and empiricism (Nolan, 2008). Rogers wished to test the validity of his clinical ideas and further investigate the therapeutic process empirically, with an aim to understand what promoted positive change within therapy (Elliott & Farber, 2010). In this positivistic context, Rogers provided operationalized definitions of the core conditions of the therapeutic relationship (Freire & Grafanaki, 2010). These definitions encouraged the theory to be tested with the traditional scientific methods of quantitative research dominant at the time (Freire, 2009), and as a result, stimulated the development of instruments that could be used to measure the core conditions. Though later in his life, Rogers shifted towards alternatives to positivism that would embrace the individual’s subjectivity (Elliott & Farber, 2010), a strong emphasis on research remained throughout his career.

3.1.5 Pragmatism

There appears to be a divide in epistemological foundations among the different domains of the project. The methods and analysis are founded in empiricism and rely on the objectivity of measurement; however, the topic and focus of the therapeutic relationship within person-centred therapy are grounded in subjective constructionism. Hence, there is a need for a framework that permits cohesion and appreciation of the differing epistemological aspects within the project. Therefore in order to achieve this solidity, it is necessary to hold a pragmatic stance.
Pragmatism was first conceptualised by philosophers James and Dewey (Rorty, 1982). Central to this approach is the functionality of knowledge, whereby the “what works is what is true” (James, 1909). Therefore, knowledge is understood in terms of its consequences, and its ability to influence effective action (McLeod, 2010). Furthermore, knowledge is appraised with regard to its usefulness beyond the scope of the project (Kvale, 1992), meaning, research should not only meet the study aims but also impact the broader society, such as generating solutions to address relevant socio-political concerns. With regards to this project, the practical contributions of the project fit well with a pragmatic stance. For example, establishing a validated measure of the core conditions has practical use in both therapy and research environments. The scale could be used for training and supervision to ensure the therapeutic relationship conditions are observed when practising. Additionally, a validated measure will encourage future outcome and process research. In this sense, a pragmatic approach can justify the application of this research to a broader context (Yardley & Bishop, 2008).

Using a pragmatic perspective aims to resolve the conflict between positivists methods and a person-centred focus. From a pragmatic position, the investigation aims to achieve a richer understanding of experience; meaning truth is not separate from human experience (Maxcy, 2003). Therefore, holding a pragmatic perspective in this project may mean the experience of the relationship is respected and valued more so than from a purely positivist perspective. This “truth” can be achieved by both quantitative and qualitative methods, depending on what is required of the research, e.g. the aims and objectives (Tashakkori & Teddlie, 1998). Therefore pragmatism is not concerned with which method is chosen, but rather the consequences of that method, which can improve and enrich our understanding of reality (Rorty, 1982). By holding a pragmatic stance, in which knowledge and the consequences of this knowledge are prioritised, (Morgan, 2007), a pragmatic
approach can begin to justify the methods used to measure the experience of the therapeutic relationship. Additionally, unlike a traditional epistemology, in which one’s ontology may dictate which methods are appropriate, in pragmatism, the methodology will be dependent on how best to achieve a particular research goal and answer a specific question (Yardley & Bishop, 2008).

3.1.6 Summary

A pragmatic approach towards the project may be useful to develop a measured and nuanced approach to the research. Holding a pragmatic stance does not remove the problems associated with subjective interpretation, but will provide flexibility and awareness of these issues without fully subscribing to the assumptions of a positivistic epistemology. Furthermore, the use of a pragmatic approach can take into account both the objective nature of the methods and the subjective environment conceptualised in the therapeutic relationship (Guyon, Kop, Juhel, & Falissard, 2018). Additionally, according to Rogers (1946), the use of psychometric measures within person-centred therapy can be useful, as long as the tests are utilised for research purposes rather than counselling purposes, and both the client and therapist are not made aware of the results. In this way, potential damage to the therapeutic space is avoided, and the relationship is preserved (Rogers, 1946). In this sense, a pragmatic approach can respect the clients’ experience and give appreciation to empirical observation beyond quantitative labels.
3.2 Design

Studies determining the reliability and validity of an instrument have typically been described as cross-sectional studies. However, validation studies can be study designs in their own right, as there are specific design and methodology associated with the validation procedure (Arafat, 2016). Therefore, this enquiry is labelled as a validation study for the use of the BLRI Obs-40 (Version 3) in a sample of young people.

3.3 ETHOS Trial and Data Access

This project made use of audio-recorded extracts from client-therapist dyads of humanistic counselling with adolescents from the ETHOS trial. Details of the trial relevant to the study and access to recordings follow. For further information, see Stafford et al. (2018).

3.3.1 ETHOS Study Aims and Overview

The ETHOS trial is a randomised control trial comparing school-based humanistic counselling (SBHC) with pastoral care as usual. The trial aims to evaluate the effectiveness and cost-effectiveness of school-based humanistic counselling in reducing psychological distress on young people, in comparison to pastoral care as usual. In addition, the trial aims to evaluate the effectiveness of school-based humanistic counselling on a range of additional outcomes, to identify the mechanisms of change in school-based counselling. The trial is based in 14 secondary schools in England and involves up to 306 young people aged 13-16 years old, who are experiencing moderate to severe levels of emotional symptoms of psychological difficulties. The school-based humanistic intervention is delivered face-to-face.
in up to ten weekly individual counselling sessions. Sessions are between 45 to 60 minutes long and audio recorded. Those in the control group will make use of the schools current pre-existing pastoral care services. Throughout the trial participants, counsellors, pastoral care teams, schools, supervisors, and the research team will be required to complete a variety of assessments. These include measures relating to outcomes, demographics, educational engagement, adherence and attendance, all to be used in the subsequent data analysis.

3.3.2 Data Access

Prior to data collection, an ETHOS Data Sharing Application Form was submitted to the ETHOS Chief Investigator. Following approval for data sharing, 136 audio-recorded extracts from client-therapist dyads of humanistic counselling with young people from the ETHOS trial were made available for use in this study.

3.3.3 ETHOS Participants

Schools

Secondary schools in England were invited to participate in the ETHOS trial. If interested, information regarding the trial was made available, and schools were assessed for eligibility to participate. Eligibility criteria are found in Appendix I and 14 schools were included in the trial.

Young People

From the schools included, the pastoral care team identified potentially eligible young people and discussed the project to ascertain their interest and eligibility for the trial.
Inclusion and exclusion criteria are found in Appendix J. A total of 306 young people met the trial criteria and were randomly assigned to either SBHC intervention or pastoral care as usual.

**Counsellors**

Potential counsellors were recruited via advertisements in the recruitment section of ‘Therapy Today’. Counsellors were assessed for their eligibility (Appendix K) and appropriately qualified. In addition, counsellors were required to undergo additional training in SBHC competences and practice, and have an understanding of the ETHOS protocol and research procedures. 19 counsellors met these requirements and were recruited for the trial. Table 3 presents the demographics of the counsellors and young people from the ETHOS trial including in the project.

**Table 3**

*Young People and Counsellor Demographics*

<table>
<thead>
<tr>
<th></th>
<th>Counsellor (N=19)</th>
<th>Young People (N=167)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>3 (16%)</td>
<td>37 (22%)</td>
</tr>
<tr>
<td>Female</td>
<td>16 (84%)</td>
<td>127 (76%)</td>
</tr>
<tr>
<td>Other</td>
<td>0 (0%)</td>
<td>3 (2%)</td>
</tr>
<tr>
<td><strong>Age (Mean/SD)</strong></td>
<td>45.0 (9.0)</td>
<td>13.7 (0.8)</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White British</td>
<td>14 (74%)</td>
<td>90 (54%)</td>
</tr>
<tr>
<td>Asian/Asian British</td>
<td>-</td>
<td>16 (10%)</td>
</tr>
<tr>
<td>African / Caribbean /</td>
<td>5 (26%)</td>
<td>27 (16%)</td>
</tr>
<tr>
<td>Mixed</td>
<td>-</td>
<td>29 (17%)</td>
</tr>
<tr>
<td>Other</td>
<td>-</td>
<td>4 (2%)</td>
</tr>
<tr>
<td>Missing</td>
<td>-</td>
<td>1 (&lt;1%)</td>
</tr>
<tr>
<td><strong>Years of Practice (Mean/SD)</strong></td>
<td>7.2 (6.6)</td>
<td>N/A</td>
</tr>
</tbody>
</table>
3.3.4 ETHOS Intervention

School-based humanistic counselling was offered to participants of the ETHOS trial. The model is based on the competencies of humanistic counselling with young people and believes young people can overcome their difficulties if given the opportunity to talk them through with an empathic and supportive therapist (Hill, Roth, & Cooper, 2014). Participants were offered up to 10 weekly, individual, face-to-face sessions, lasting between 45-60 minutes long. In addition, participants continued to have access to their schools’ original pastoral care support.

3.3.5 Recordings

To ensure counsellors adhered to the theoretical model of school-based humanistic counselling upon delivery, therapy sessions were recorded during the ETHOS trial. Audio recordings were created and selected in accordance with the ETHOS Adherence Auditing Procedure (Appendix L). Recordings were 20 minutes long, excluding the first and last five minutes of a session. Recordings were chosen at random from the first and second half of a counsellors’ work with their client, excluding the first and final sessions. These procedures resulted in 136 audio-recordings from 19 counsellors and 167 young people (Stafford et al., 2018). Each young person had two recordings, one taken from the first half of their therapy sessions, and one from the second half of therapy. To ensure counsellors practiced in adherence to the SBHC model, audio recordings were rated by a team of ETHOS auditors, using the young persons Person-Centred Experiential Psychotherapy Scale (PCEPS-YP).
3.4 Observer Raters

3.4.1 Rater Recruitment

To examine the inter-rater reliability of the BLRI Obs-40, two raters were recruited in addition to the lead researcher (Kiranjeet Bhatti - Rater 1). Raters were recruited via the existing team of ETHOS auditors who had rated the audio-recordings with the PCEPS-YP and via the University of Roehampton Temp Job Website. Due to the low response, the pool of potential raters was limited and therefore the two raters with the most experience of person-centred work were selected for the role. Raters were paid £10/hour and spent up to 25 hours rating 50 audio-recordings using the BLRI Obs-40. Table 4 reports the demographic information for the three observer raters.

3.4.2 Rater Demographics

Table 4

Observer Rater Demographics

<table>
<thead>
<tr>
<th></th>
<th>Rater 1</th>
<th>Rater 2</th>
<th>Rater 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>Female</td>
<td>Female</td>
<td>Female</td>
</tr>
<tr>
<td>Age</td>
<td>28</td>
<td>46</td>
<td>25</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>Indian</td>
<td>White English</td>
<td>Indian</td>
</tr>
<tr>
<td>Level of Training</td>
<td>PsychD</td>
<td>Diploma</td>
<td>MSc</td>
</tr>
<tr>
<td>Length of Training</td>
<td>7</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>Highest Degree</td>
<td>PsychD</td>
<td>Diploma</td>
<td>MSc</td>
</tr>
<tr>
<td>Years of Practice</td>
<td>3</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>Orientation</td>
<td>Integrative</td>
<td>Person-Centred</td>
<td>Movement Psychotherapy</td>
</tr>
</tbody>
</table>
3.4.3 Rater Training

Prior to rating, the additional raters attended a two-hour workshop led by Rater 1 to discuss the project and rating procedure. Time was spent considering the project aims and objects, and a background to the project was provided. The purpose of the workshop was to review the BLRI Obs-40 and gain familiarity with the scale. Time was spent discussing the subscales of the BLRI, understanding the interpretation of items, being aware of the combination of negatively and positively worded items, and gaining familiarity with the scoring procedure. During this workshop, raters completed practice ratings on two extracts of humanistic therapy, the raw subscale and total BLRI scores from these ratings are found in Appendix M. During the workshop, Rater 1 reviewed the practice ratings and a discussion followed regarding the rating procedure and scoring. In addition, raters were given the opportunity to express any queries regarding the process or ambiguous items in the scale. Particular attention was paid to negatively worded items, and the items of the Unconditionality of Regard and Congruence subscales, as scores were most divergent here.

3.5 Measures

3.5.1 BLRI Obs-40 (Version 3)

The BLRI Obs-40, developed by Barrett-Lennard (1959; 1962), is an observer-rated scale consisting of 40 items. This is a shorted version, adapted from the original 64 item scale. The instrument aims to measure therapist-to-client relationship conditions and is based on the core conditions of person-centred therapy, proposed by Rogers (1957). The instrument has four subscales: ‘Empathic Understanding’, ‘Level of Regard’, ‘Unconditionality of
Regard’, and ‘Congruence.’ Each of the four subscales consists of ten items. 20 of the items in the scale were negatively worded and placed in a random order throughout the scale. Observers rate the extent therapists demonstrate these conditions using a 6-point Likert-type scale. Much of the evidence regarding the scale’s reliability and validity has come from the evaluation of the 64-item scale from the original development study (Barrett-Lennard, 1962), and a review of the literature from Gurman (1977). There is very little validation in young populations; however, evidence for satisfactory re-test reliability over six weeks has been found (Brouzos et al., 2015). Further details are found in section 2.6.

3.5.2 Person-Centred Experiential Psychotherapy Scale – Young Person Counselling Version

The Person-Centred Experiential Psychotherapy Scale – Young Person Counselling Version (PCEPS-YP) is an adapted measure from the original adult version of the PCEPS. The PCEPS-YP aims to measure competency and adherence to person-centred or experiential practice with young people (Ryan et al., 2019).


The PCEPS-YP is an observer-rated measure, using a 6-point anchored rating scale. For each item on the scale, a description of the therapist quality is provided. In addition,
example questions that the rater can hypothetically ask of the therapist are given, to help determine the raters’ response. For example, ‘Item 5 – Genuineness: To what degree is the therapist able to skilfully express their congruent experience of the young person in a facilitative manner?’ Each item is rated using a 6-point scale, from 1 ‘no evidence (e.g. of genuineness)’ to 6 ‘excellent evidence (e.g. of genuineness)’. The rated therapist is then provided with an overall average score ranging from 1 (non-adherent) to 6 (excellent adherence). Scores between 1-3 are considered to be unacceptable levels of competence, and scores between 4-6 are considered to be acceptable adherence and competence (Ryan et al., 2019).

The internal reliability and inter-rater reliability of the original PCEPS scale was found to be satisfactory (Freire et al., 2014). In addition, an exploratory factor analysis demonstrated good structural validity of the scale (Freire et al., 2014). The psychometric properties of the adapted PCEPS-YP are currently undergoing investigation (Ryan et al., 2019). Preliminary findings of internal consistency appear to be satisfactory with a Cronbach’s alpha of 0.94 across the nine items. Furthermore, correlations between the nine items of the PCEPS-YP were found to be significant (p < 0.05), ranging from r = 0.48 – 0.84. The average inter-rater reliability using eight raters had a Cronbach’s alpha of 0.61 and correlations between the raters’ average score were high and significant (r = 0.774; p < 0.001). Furthermore, an exploratory factor analysis found one strong factor explaining 67.9% of the variance (Ryan et al., 2019).

3.5.3 COSMIN Checklist

The COSMIN checklist (Mokkink et al., 2010) is a tool used to assess the methodological quality of outcome measures (Appendix B). The checklist consists of 10
items related to various aspects of reliability and validity; criteria are provided for each of these aspects and judgements are made on their quality and testing (Mokkink et al., 2012). This use of the COSMIN checklist is permitted to all, provided the authors are adequately referenced. The checklist was initially proposed for the evaluation of patient-reported health outcomes measures; however, its use has been recommended for the evaluation of other measurement instruments (Mokkink, Terwee, Knol et al., 2010). Therefore this checklist will be used towards the end of the project to determine the overall quality of the methodology used to conduct the psychometric evaluation, as stated in objective six.

3.6 Procedure

3.6.1 Selection and Transfer of Recordings

As specified in the ETHOS Audition Adherence Procedure (Appendix G), audio recordings were randomly selected from a counsellor’s work with their client. In total, 136 audio recordings were used for this project. These recordings were transferred from the ETHOS trial to the lead researcher (Rater 1) and additional raters (Rater 2 and Rater 3). This was done using a cloud-based server with a two-stage authentication program, accessible only to the raters. All recordings were stored on an encrypted external hard drive and deleted once data collection was completed.

3.6.2 Rating Procedure and Data Entry

Audio was rated using either the BLRI Obs-40 or PCEPS-YP. Rater 1 rated all 136 recordings using the BLRI Obs-40, and a selection of 50 recordings using the PCEPS-YP.
Five months after completing the ratings, the selections of 50 recordings were re-rated using the BLRI Obs-40 again by Rater 1 in order to collect data for the test-retest analysis. Raters 2 and 3 each also rated the selection of 50 recordings using the BLRI Obs-40. All raters listened to each audio recording once and subsequently rated the audio for the presence and level of therapeutic core conditions by completing the BLRI Obs-40. Ratings took place in a quiet and private location, where there was no risk of interruption and no one other than the rater would hear the audio. Each audio recording was labelled with a unique code, which raters would indicate on the completed BLRI Obs-40 form. Following this completed BLRI Obs-40 forms were transferred electronically to and collected by Rater 1. Rater 1 then inputted all scores into an SPSS database to allow for subsequent analysis.

3.7 Ethical Considerations and Approval

This research was submitted for ethics consideration under the reference PSYC 18/292 in the Department of Psychology and was approved under the procedures of the University of Roehampton’s Ethics Committee on 18.04.18.

3.7.1 Measures and Data Use

Permission to use ETHOS data and relevant measures was sought prior to data collection. Permission was gained from the original ETHOS study (reference: PSYC16/227) to allow for subsequent data analysis, and involved submitting an ETHOS Data Sharing Application Form (Appendix N), which was reviewed and subsequently approved by the chief investigator for ETHOS (Prof Mick Cooper), prior to the start of the project. Permission for the free use of the BLRI Obs-40 for the purpose of this project was attained from the
author. The use of the COSMIN checklist is permitted to all, provided the authors are adequately referenced. Therefore, no further permission for the use of the checklist was necessary. Finally, permission for the use of the PCEPS-YP for the purpose of this project was attained via the ETHOS trial.

3.7.2 Confidentiality and Anonymity

All audio recordings were anonymised prior to raters receiving them. Therefore there was no participant identifying data on the recordings, except for the participant’s voice. Therefore audio recordings were treated as partially anonymised data and stored in accordance with the CREST Data Storage and Protection Procedures (Appendix O).

3.7.3 Data Storage

Audio Recordings

Audio recordings included the participant’s voice were treated as partially anonymised data. Audio recordings were kept electronically using encrypted files and stored on password-protected devices. Audio recordings were electronically transferred between ETHOS and Rater 1, or between Rater 1 and either Rater 2 or 3. This was done through a cloud-based server with a two-stage authentication program accessible only to the raters. All recordings were further stored on an encrypted external hard drive, which was held securely in a locked box at the raters’ home. Once the ratings were complete, the recordings were double deleted from both the hard drive and the recycling bin, ensuring there was no trace of the files left on any devices. In addition, raters held no further information regarding the participants of the recordings or any further information on additional ETHOS data. Audio
recordings were kept by Rater 1 only for the duration of the project and once completed were returned to the chief investigator of ETHOS for storage.

*Completed BLRI Obs-40 (Version 3) Forms*

Completed BLRI Obs-40 forms were treated as anonymised data. Non-electronic ratings were kept in a locked and secure location at all times in the raters’ home. Electronic ratings from Rater 2 and 3 were anonymised and transferred to Rater 1 and stored on a password-protected device. Data from the completed scales were entered into SPSS and password-protected. All physical copies of the scales were securely kept for the duration of the project and subsequently destroyed upon completion.

*Completed PCEPS-YP Forms*

Completed PCEPS-YP forms were treated as anonymised data. All completed PCEPS-YP forms were stored electronically and held securely on an encrypted external hard drive. Data from the completed PCEPS-YP scales were entered into SPSS and password-protected.

*Statistical Analysis*

Processed data, including the statistical analysis, was numerical and entered into an SPSS spreadsheet with no identifying characteristics; therefore, data were treated as anonymous. The data was stored in an encrypted folder and on a password protected device. When and if data needed to be transferred, this was done using a USB memory stick, which had been encrypted. Data was then saved on the new device in an encrypted folder with a password. Once the data had been transferred, it was removed from the USB stick in such a
way that it could not be recovered. Data will be stored securely for at least ten years after the completion of the project, and all copies destroyed.

3.7.4 Health and Safety Considerations

The health and safety concerns for this project were evaluated in a risk assessment prior to data collection. Recommendations to promote a safe working environment were assessed and approved as part of the ethics application for this project. One potential risk included harm to raters’ emotional wellbeing. As raters were listening to therapy sessions, there was the possibility of raters hearing upsetting and emotionally distressing content. Consideration was paid to the potential costs of this on the raters’ mental wellbeing, and supportive services were offered if and when necessary. Supportive services included discussing emotional responses with the lead rater (Kiranjeet Bhatti) who was a counselling psychologist in training, and signposting to relevant agencies or the university counselling service. However, these services were not required or utilised by any of the raters. In addition, to minimise manual handling of the paperwork needed for the project, scales were digitalised and sent electronically so minimise the load and weight carried. Furthermore, all raters were recommended to take frequent breaks when using computer screens to ensure a comfortable working environment.

3.8 Data Analysis

The COSMIN checklist (Mokkink et al., 2010) informed the design and proposed analysis of the evaluation of the BLRI Obs-40. In addition, the overall methodological quality of the validation of BLRI Obs-40 was examined post analysis using this checklist. This
evaluation followed the recommended steps outlined in the COSMIN manual (Mokkink et al., 2012) and served as a self-assessment on the quality of the psychometric evaluation procedure.

3.8.1 Power Calculation

The sample size of 136 recordings was estimated with a prior power analysis (Faul, Erdfelder, Buchner, & Lang, 2009) on the basis of: \( \alpha=0.05 \), \( \beta=0.8 \) (Cohen, 2013), two-tailed, and an assumed inter-scale correlation rate of 0.8. This was based on the results of the internal reliability of the BLRI (OS-64) found in Barrett-Lennard (1962), where the internal consistency of the subscales of the BLRI was \( r \geq 0.82 \). Using a confidence interval calculator (Evans, 2016), the estimated 95% confidence interval was 0.73-0.85 for a sample size of 136. Earlier simulations (Guadagnoli & Velicer, 1988; Velicer & Fava, 1998) also informed the sample size.

3.8.2 Internal Reliability

The internal reliability of the BLRI Obs-40 was assessed using Cronbach’s Alpha. This is the most common measure of internal consistency (Tavakol & Dennick, 2011), and was used to gauge the reliability of the 40 items of the BLRI, the four subscales, and the internal consistency of each of the ten items in the subscales of the BLRI. A Cronbach’s alpha between 0.70 and 0.95 is considered adequate to excellent (Hair, Black, Babin, & Anderson, 2010; Terwee et al., 2007). Item total statistics were also assessed to make judgements on the specific items that make up the BLRI Obs-40. Examining the ‘scale if item deleted’ statistics allowed an assessment of the items value and retention in the scale. If
Cronbach’s alpha increased with the removal of an item, a judgement was made to delete the item. If no improvement to Cronbach’s alpha were found, then the item would remain in the scale. In addition, to further assess the internal reliability, inter-scale correlations between the four subscales of the BLRI were assessed using Pearson’s Product Moment Correlation.

3.8.3 Inter-Rater Reliability

Using 50 scores each from three independent raters, the inter-rater reliability of the BLRI Obs-40 was analysed. Cronbach’s Alpha was used to measure the consistency of the three raters’ scores across item, subscale, and total BLRI score. In addition, the intraclass correlation coefficient (ICC) was used to assess the extent to which raters agreed with one another in their ratings of the BLRI. A two-way mixed-effect model based on single ratings and absolute agreement assessed the inter-rater reliability. Mean estimations along with 95% confidence intervals (CI) are reported for each ICC for each item and subscale of the BLRI. ICC estimations of <0.40 are poor, 0.40-0.60 fair, 0.60-0.75 good, and above 0.75 excellent (Perinetti, 2018). Furthermore, Krippendorff’s Alpha (K Alpha) was calculated to assess the level of rater disagreement; a score of <0.8 is found to be unsatisfactory (Hayes & Krippendorff, 2007).

3.8.4 Test-retest Reliability

Fifty recordings were rated using the BLRI Obs-40 at Time 1 (T1: October 2018) and subsequently re-rated at Time 2 (T2: February 2019). The test-retest reliability of the BLRI Obs-40 and the four subscales were examined using Pearson’s Product Moment Correlation. In addition, Cronbach’s Alpha was estimated BLRI Obs-40 subscale and total, to measure the
consistency of the ratings over time. To assess repeatability, intraclass correlation coefficient (ICC) was used for the analysis. A two-way mixed-effect model based on single ratings assessed the test-retest reliability.

### 3.8.5 Convergent Validity

Fifty randomly selected audio recordings were rated using the BLRI Obs-40 and the PCEPS-YP; these ratings were used to assess convergent validity. Prior to analysis, a 9X4 correlation matrix was estimated (Table 5), based on the nine items of the PCEPS-YP and the four subscales of the BLRI, qualitative estimates of the expected direction of correlations were made, and subsequently tested using Pearson’s Product Moment Correlation.

#### Table 5

*Estimated Direction of Correlations Between BLRI Obs-40 and PCEPS-YP*

<table>
<thead>
<tr>
<th></th>
<th>Regard</th>
<th>Empathy</th>
<th>Unconditionality</th>
<th>Congruence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Client Frame of Reference</td>
<td>Positive</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tracking</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Empathic Resonance</td>
<td>Positive</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accepting Presence</td>
<td>Positive</td>
<td>Positive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Genuineness</td>
<td></td>
<td></td>
<td>Positive</td>
<td></td>
</tr>
<tr>
<td>Emotion Focused</td>
<td></td>
<td></td>
<td>Positive</td>
<td></td>
</tr>
<tr>
<td>Emotion Symbolisation</td>
<td></td>
<td></td>
<td>Positive</td>
<td></td>
</tr>
<tr>
<td>Facilitation Of Client Self-Development</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Developmental Responsiveness</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3.8.6 Structural Validity

The structural validity of the BLRI Obs-40 was assessed using a Confirmatory Factor Analysis. A single-factor model of person-centeredness with the 40 items of the BLRI Obs-40 as latent factors was first assessed. Based on previous assessments of the construct validity of the scale (Barrett-Lennard, 1962; Gurman, 1977), a four-factor correlated model with Empathic Understanding, Congruence, Unconditionality of Regard, and Level of Regard as latent factors was also assessed. Additionally, based on previous reviews of inter-scale correlations and exploratory factor analysis findings (Gurman, 1977; Kiesler, 1973), a three-factor correlated model with Empathic Understanding, Congruence, and Level of Regard as latent factors were assessed. A higher-order model was also assessed, based on previous exploratory factor analyses of the BLRI (Gross et al., 1970; Lin, 1973; McClanahan, 1974). This was a four-factor correlated model of Empathic Understanding, Congruence, Level and Unconditionality of Regard, loading on to one higher-order factor of person-centeredness.

Finally, a higher-order model omitting Unconditionality of Regard was also assessed, based on findings from Mills and Zytowski (1967).

Maximum Likelihood was used to estimate model parameters and goodness-of-fit for all CFA Models. The goodness-of-fit of each of the models was assessed using goodness-of-fit index, comparative-fit index, normed-fit index, root mean square residual, standardised root mean square residual, chi-squared, and chi-Squared/degrees of freedom. Using a criteria of Chi-Square p>0.05 (Wheaton, Muthen, Alwin, & Summers, 1977), RMSEA <0.08 (Browne & Cudeck, 1993), SRMR <0.08 (Schreiber, Stage, King, Nora, & Barlow, 2006) GFI > 0.90 (Joreskog & Sorbom, 1984), CFI > 0.90 (Bentler, 1990), NFI > 0.90 (Bollen, 1989), and Chisq/df <3.0 (Marsh & Hocevar, 1985).
Parcelling

For four models, the observed variables were parcelled in order to create fewer variables. Items were parcelled for each subscale; each of the ten items was split into three parcels, and the sums of the parcels were used in the analysis (Matsunaga, 2008). This resulted in the 40 items of the BLRI being transformed into 12 parcels (or three parcels per subscale).

Rationale for Parcelling

The quantitative nature of the analysis and confirmatory factor analysis implies an empirical perspective, in which the model data and the collected data should mirror as much as possible and avoid any potential bias from the researcher (Little, Cunningham, Shahar, & Widaman, 2002). However, by using a pragmatist approach, it is possible to argue that the flexibility of parceling data can at times be appropriate and provides useful observations (Little, Rhemtulla, Gibson, & Schoemann, 2013).

Parcelling can have many benefits, for instance, by combining items, it is more likely parcels represent an underlying construct in comparison to a single item, which can therefore, improve model fit (Rushton, Brainerd, & Pressley, 1983). In addition, given that measurement error increases with the increase in the number of items, parceling can reduce the absolute amount of measurement error and improve model fit (Little et al., 2002). Furthermore, parcels are more likely to normalise the distribution (Bandalos & Finney, 2001) and can be helpful when sample sizes are small, as the number of parameters estimated in the model is reduced (Little et al., 2013).
Parcelling Method

To ensure variables were equally representative of the latent variables, items were parcelled by subscale. Such as, the ten items of the Level of Regard subscale were parcelled between each other to form three parcels that were representative of the Level of Regard variable. Furthermore, following recommendations in the literature (Matsunaga, 2008) each factor consisted of three parcels only.

In addition, items were parcelled in accordance to their factor loadings, following methods of factorial algorithm parcel building (Rogers & Schmitt, 2004). This method was chosen in an attempt to distribute item-specific components evenly across the parcels. The factor loadings of the ten items of a subscale were ordered from highest to lowest and allocated to a parcel dependent on the factor loading size. For example, the highest loading factor was allocated to parcel one, the second highest to parcel two, the third highest to parcel three, this then continued with the fourth highest factor allocated to parcel one and the pattern continued. Appendix P illustrates the factor loading allocation for each subscale.

3.8.7 Post Analysis Evaluation

Using the COSMIN checklist (Mokkink et al., 2010), the methodological quality of the validation of BLRI Obs-40 was examined post-analysis. This evaluation followed the recommended steps outlined in the COSMIN manual (Mokkink et al., 2012) and served as a self-assessment on the quality of the psychometric evaluation procedure.

3.8.8 Analysis Software

Data were analysed using IBM SPSS Statistics for Mac, Version 24.0 and AMOS, Version 25 (Arbuckle, 2006). SPSS Statistics was used for calculating Pearson’s Product
Moment Correlations, Cronbach’s alpha, intraclass correlation coefficients and Krippendorff’s alpha. AMOS was used to conduct the confirmatory factor analyses used to assess structural validity.

3.9 Pilot Ratings

Prior to data collection, a random selection of 15 recordings was rated using the BLRI Obs-40 by the lead researcher and the project supervisor. The aim of this was to trial the rating procedure, become familiar with the scale, and begin an initial assessment of the inter-rater and internal reliability of the scale. The pilot ratings were also used to plan the statistical analysis for the project.
Chapter 4: Results

4.1 Pilot Results

Means and Standard Deviations of the pilot ratings are found in Appendix Q. The internal consistency of the BLRI Obs-40 was preliminarily estimated using Cronbach’s alpha on a sample of 15 ratings. The internal reliability was found to be high for both raters ($\alpha=0.96$, $\alpha=0.94$). To investigate inter-rater reliability in this sample, Pearson’s Product Moment Correlations were assessed between the two raters’ scores. Table 6 reports the findings. Medium to weak positive correlations were found across all subscales and total BLRI score. All were insignificant, with the exception of the Congruence subscale; however, this may be a result of the small sample size.

Based on the preliminary results of the pilot study, and reflecting on the rating procedure, adaptations were made to the format and content of the training workshop for additional raters. The inter-rater reliability results from the pilot study indicated discrepancies in the interpretations of items, in particular, the items of the Unconditionality of Regard subscale and the negatively worded items. Therefore, adaptations to the rater training were made during the planning stages for the workshop, before rater recruitment. There was a greater focus on gaining an understanding of the different subscales of the BLRI Obs-40 and interpreting specific items of the scale. In addition, based on the experience from the pilot study, greater attention was paid to negatively worded items and how these were scored.
Table 6

Summary of Pilot Correlations between Raters

<table>
<thead>
<tr>
<th></th>
<th>Regard (R₁)</th>
<th>Empathy (R₁)</th>
<th>Unconditionality (R₁)</th>
<th>Congruence (R₁)</th>
<th>BLRI Total (R₁)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regard (R₂)</td>
<td>0.30</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Empathy (R₂)</td>
<td></td>
<td>0.45</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unconditionality (R₂)</td>
<td></td>
<td></td>
<td>0.42</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Congruence (R₂)</td>
<td></td>
<td></td>
<td></td>
<td>0.64*</td>
<td></td>
</tr>
<tr>
<td>BLRI Total (R₂)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.43</td>
</tr>
</tbody>
</table>

*Significant At P <0.01 (2-Tailed)

4.2 Descriptive Statistics

The rating procedure (as outlined in the methods section) resulted in a total of 336 observations each involving 40 ratings taken from 136 recordings by three raters using either the BLRI Obs-40 or the PCEPS-YP. There were no missing data present and observations were divided into four samples dependent on which research questions were evaluated. Table 7 illustrates how each sample was rated and subsequently assessed. Specifically, to assess structural validity and internal reliability, sample one used 136 ratings by the lead research (Rater 1). Sample two used 50 recordings rated using the PCEPS-YP and BLRI Obs-40 to assess convergent validity. Sample three used 50 recordings rated at Time 1 (October 2018) and Time 2 (March 2019) to assess test-retest reliability, and finally, sample four used 50 recordings rated by three independent raters to examine inter-rater reliability. Means and standard deviations for each of the four samples are found in the Appendix (R-U), descriptive statistics of the BLRI Obs-40 subscales and total are found in Table 8. Note, possible scores for each of the subscales of the BLRI range between -30 to +30. Possible scores for the total BLRI, range between -120 to +120.
Table 7

Division of Ratings by Sample and Assessment

<table>
<thead>
<tr>
<th>Sample</th>
<th>Ratings (N)</th>
<th>Rating Measure</th>
<th>Raters</th>
<th>RQs Assessed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample 1</td>
<td>136</td>
<td>BLRI Obs-40</td>
<td>Rater 1</td>
<td>1a. Internal Reliability 2b. Structural Validity</td>
</tr>
<tr>
<td>Sample 2</td>
<td>100*</td>
<td>50 with BLRI Obs-40 50 with PCEPS-YP</td>
<td>Rater 1</td>
<td>2a. Convergent Validity</td>
</tr>
<tr>
<td>Sample 3</td>
<td>100*</td>
<td>BLRI Obs-40 at Time 1 and Time 2</td>
<td>Rater 1</td>
<td>1c. Test –Retest Reliability</td>
</tr>
<tr>
<td>Sample 4</td>
<td>150*</td>
<td>BLRI Obs-40</td>
<td>Rater 1 Rater 2 Rater 3</td>
<td>1b. Inter-Rater Reliability</td>
</tr>
</tbody>
</table>

*50 of these ratings were taken from Sample 1, resulting in 336 unique ratings.
Table 8

Means, Standard Deviations, Range, Median, and Mode of Ratings from 136 Audio Recordings

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>Range</th>
<th>Median</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regard</td>
<td>17.65</td>
<td>8.20</td>
<td>25</td>
<td>21</td>
<td>21</td>
</tr>
<tr>
<td>Empathy</td>
<td>20.44</td>
<td>4.46</td>
<td>50</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>Unconditionality</td>
<td>22.09</td>
<td>2.56</td>
<td>18</td>
<td>22</td>
<td>22</td>
</tr>
<tr>
<td>Congruence</td>
<td>12.63</td>
<td>6.91</td>
<td>36</td>
<td>14</td>
<td>11</td>
</tr>
<tr>
<td>BLRI Total</td>
<td>72.80</td>
<td>18.33</td>
<td>105</td>
<td>78.5</td>
<td>82</td>
</tr>
</tbody>
</table>

4.3 Internal Reliability

Sample 1 was used to examine the internal consistency of the BLRI Obs-40. The 40 items of the BLRI were found to have a high level of internal consistency with Cronbach’s alpha = 0.93. Most items appeared to be worthy of retention, resulting in a decrease in the alpha if deleted. Any exceptions to this would result in minimal improvement if deleted (alpha increase by 0.01-0.03).

Separate reliabilities were calculated for the four subscales of the BLRI. The ten items that make up the Level of Regard subscale had a Cronbach’s alpha of 0.84, indicating good internal consistency between the items. The Cronbach’s alpha increased to 0.85 if item 25 ‘S/he disapproves of B - as s/he sees him/her’ was deleted.

The Empathy subscale had ten items and a Cronbach’s alpha of 0.94, indicated high internal reliability. Cronbach’s alpha increased to 0.95 when item 14, ‘A’s personal reaction to something about B gets in the way of his/hers understanding’ is deleted.

The 10-item Unconditionality of Regard subscale had a Cronbach’s alpha of 0.59, suggesting poor internal reliability, and there was no improvement to Cronbach’s alpha if any items were deleted.
The Congruence subscale, which also consisted of 10 items, had a Cronbach’s alpha of 0.83 suggested good internal reliability of the subscale. This increased to 0.84 when item 16 ‘S/he keeps quiet about his/her own inner impressions and feelings with B’ was deleted.

The internal consistency for the four subscales of the BLRI as measured by Cronbach’s alpha was 0.77. Cronbach’s alpha increased to 0.84 if the subscale of Unconditionality of Regard were removed.

4.3.1 Inter-Scale Correlations

All scales were significantly correlated with each other (p<0.05). The subscales of Level of Regard, Empathy, and Congruence demonstrated strong positive correlations with each other, suggesting good internal reliability. However, the subscale of Unconditionality of Regard weakly correlated with the remaining subscales.

Table 9

*Summary of Cronbach’s Alpha and Inter-Scale Correlations of the BLRI*

<table>
<thead>
<tr>
<th></th>
<th>Cronbach’s Alpha</th>
<th>Regard</th>
<th>Empathy</th>
<th>Unconditionality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regard</td>
<td>0.84</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Empathy</td>
<td>0.94</td>
<td>0.74**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unconditionality</td>
<td>0.59</td>
<td>0.31**</td>
<td>0.17*</td>
<td></td>
</tr>
<tr>
<td>Congruence</td>
<td>0.83</td>
<td>0.73**</td>
<td>0.65**</td>
<td>0.23**</td>
</tr>
</tbody>
</table>

**Significant at p <0.01 (2-tailed); *Significant at p <0.05 (2-tailed)**
4.4 Structural Validity

Confirmatory Factor Analysis was conducted using sample 1 (descriptive statistics in Appendix R). Confirmatory Factor Analysis was conducted to assess the structural validity of the BLRI Obs-40 and determine whether the factor structure of the measure reflected the four-sub scales and relationship dimensions of person-centred theory. Based on previous literature, and findings from the inter-scale correlations, six models, were compared. Fit Indices for each model are found in Table 10.

Model 1 (Single-factor, 40 observed variables)

Model 1 was a single factor model, examining the degree to which the items reflected a single dimension on person-centeredness. Model 1 indicated a poor fit of the data with all fit indices not meeting the appropriate cut-offs for all except Chisq/df 2.09.

Model 2 (Four-factor correlated, 40 observed variables)

Model 2 was a four-factor correlated model originally proposed by Barrett-Lennard (1962) and later validated by Cramer (1986b) and Liao et al. (2018). This model fixed the factor loadings, so subscale items loaded uniquely on the latent factor (relationship factor) that the item was hypothesised to reflect. Factors were allowed to inter-correlate. Model 2 indicated a poor fit of the data with all fit indices not meeting the appropriate cut-offs for all except RMSEA 0.08 and Chisq/df 1.78.

Model 3 (Four-factor correlated, 12 parcels)

Model 3 is a variation of Model 2; this is a four-factor correlated model with 12 parcels rather than 40 observed variables (Figure 2). The ten items of each of the subscales in
the BLRI were reduced to three parcels per subscale. According to the correlation matrix of the four latent factors Empathy, Level of Regard, and Congruence correlated highly with each other ($r=0.78-0.88$, $p <0.001$), however, Unconditionality of Regard weakly correlated with Level of Regard ($r=0.38$, $p=0.01$), Empathy ($r=0.19$, $p =0.10$), and Congruence ($r=0.28$, $p=0.04$). Model 3 demonstrated a very good fit for the data. All fit indices indicated a satisfactory fit of the model for the data (Table 10), with a satisfactory Root Mean Square Error of Approximation (RMSEA) = 0.05 (<0.08 for acceptance), Standardised Root Mean Square Residual (SRMR) = 0.04 (<0.08 for acceptance), Goodness of Fit Index (GFI) = 0.93 (>0.90 for acceptance), Comparative Fit Index (CFI) =0.99 (>0.90 for acceptance), and Normed Fit Index (NFI) = 0.95 (>0.90 for acceptance). Almost all parcels had heavy loadings ($r>0.65$); however, two of the three Unconditionality parcels revealed lower loadings, unconditionality parcel 1 ($r=0.46$) and unconditionality parcel 2 ($r=0.54$).
Figure 2. CFA Model 3 of The Barrett-Lennard Relationship Inventory Obs-40. Latent Factors: Level of Regard, Empathy, Unconditionality of Regard, and Congruence.

Model 4 (Three-factor correlated, 9 parcels)

Model 4 is a three-factor correlated model, omitting the Unconditionality of Regard scale as a latent factor. This is based on the finding from Gurman (1977), Kiesler (1973) and inter-scale correlation findings. Three latent factors of Level of Regard, Empathy, and Congruence were allowed to correlate, and item parcels acted as observed variables, rather than the 30 scale items. The correlation matrix of the three latent factors found Empathy, Level of Regard, and Congruence to significantly correlate strongly with each other (r=0.78-0.88, p <0.001), and all parcels had high factor loadings. Fit indices of RMSEA, SRMR, GFI,
CFI, NFI and Chisq/df suggest a good fit of the model for the data. However, the chi-square statistic was significant $\chi^2(24)=45.48 \ p=0.01$ suggesting the parameters of the model differed significantly from those in the data. Therefore the model is not an exact fit to the data.

**Model 5 (Four-factor correlated loading on to one higher-order factor, 12 parcels)**

Based on findings from Gross et al. (1970), Lin (1973), and McClanahan (1974), the higher-order model examined the degree to which inter-correlation between factors could be explained by a single second-order factor representing person-centeredness (Figure 3). Model 5 examined a higher-order model, using parcels rather than individual items. Model 5 demonstrated a very good fit for the data. All fit indices indicated a satisfactory fit of the model for the data (Table 10), with a satisfactory Root Mean Square Error of Approximation (RMSEA) = 0.05 (<0.08 for acceptance), Standardised Root Mean Square Residual (SRMR) = 0.04 (<0.08 for acceptance), Goodness of Fit Index (GFI) = 0.93 (>0.90 for acceptance), Comparative Fit Index (CFI) =0.98 (>0.90 for acceptance) and Normed Fit Index (NFI) = 0.94 (>0.90 for acceptance). Level of Regard ($r=0.98$), Congruence ($r=0.90$), and Empathy ($r=0.85$) loaded heavily on Person-Centeredness. However, Unconditionality of Regard had a lower loading ($r=0.35$). Furthermore, two of the three Unconditionality parcels loaded low on the Unconditionality sub factor: unconditionality parcel 1 ($r=0.47$) and unconditionality parcel 3 ($r=0.56$).
Figure 3. CFA Model 5 of The Barrett-Lennard Relationship Inventory Obs-40. Four-factor correlated loading on to one higher-order factor of Person-Centeredness.

Model 6 (Three-factor correlated loading on to one higher-order factor, 9 parcels)

Based on Mills and Zytowski (1967), a higher-order model which examined the degree to inter-correlations between the factors of Level of Regard, Empathy, and Congruence could be explained by a single second-order factor of person-centeredness was conducted. Fit indices of RMSEA, SRMR, GFI, CFI, NFI, and Chisq/df suggest a good fit of
the model for the data. However, again, the chi-square statistic was significant $\chi^2(24)=45.48$ p=0.01 suggesting an inexact fit of the model to the data. Additionally, all factors loaded heavily on Person-Centeredness (Level of Regard, r=0.97, Congruence, r=0.90, and Empathy, r=0.86) furthermore, all parcels also loaded heavily on the factors of Level of Regard, Congruence, and Empathy.

**Comparison of Models**

Based on the findings, Model 3 and Model 5 appear to fit the data best, with all fit indices indicating a good fit of the models to the data. In order to determine which of the models provided a better fit of the data, the change in the chi-squared statistic in relation to the degrees of freedom were compared for model 3 and model 5 (Model 3 vs. Model 5: $\Delta \chi^2=4.56$, $\Delta df=2$, p=0.11). The change between the four factors correlated and the four factors – one second-order factor model was not significant, suggesting neither model is a better fit than the other. In addition, the single-factor model was compared to model 3 and model 5. A significant difference was found between the models, suggesting either of the four-factor models were superior to the single-factor model. Comparisons are reported in Table 11.
### Table 10

**Fit Indices For Each Model**

<table>
<thead>
<tr>
<th>Model</th>
<th>$\chi^2$</th>
<th>df</th>
<th>P</th>
<th>RMSEA</th>
<th>SRMR</th>
<th>GFI</th>
<th>CFI</th>
<th>NFI</th>
<th>Chisq/df</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1, Single factor, 40 items</td>
<td>1545.04</td>
<td>740</td>
<td>0.001</td>
<td>0.09</td>
<td>0.09</td>
<td>0.62</td>
<td>0.69</td>
<td>0.55</td>
<td>2.09</td>
</tr>
<tr>
<td>Model 2, Four factor correlated 40 items</td>
<td>1308.75</td>
<td>734</td>
<td>0.001</td>
<td>0.08</td>
<td>0.84</td>
<td>0.69</td>
<td>0.78</td>
<td>0.62</td>
<td>1.78</td>
</tr>
<tr>
<td>Model 3, Four factor correlated 12 parcels</td>
<td>60.85</td>
<td>48</td>
<td>0.10</td>
<td>0.05</td>
<td>0.04</td>
<td>0.93</td>
<td>0.99</td>
<td>0.95</td>
<td>1.27</td>
</tr>
<tr>
<td>Model 4, Three factor correlated 9 parcels</td>
<td>45.48</td>
<td>24</td>
<td>0.01</td>
<td>0.08</td>
<td>0.04</td>
<td>0.93</td>
<td>0.98</td>
<td>0.96</td>
<td>1.90</td>
</tr>
<tr>
<td>Model 5, Four first order factors, one second-order factor, 12 parcels</td>
<td>65.41</td>
<td>50</td>
<td>0.07</td>
<td>0.05</td>
<td>0.04</td>
<td>0.93</td>
<td>0.98</td>
<td>0.94</td>
<td>1.31</td>
</tr>
<tr>
<td>Model 6, Three first order factors, one second-order factor, 9 parcels</td>
<td>45.48</td>
<td>24</td>
<td>0.01</td>
<td>0.08</td>
<td>0.04</td>
<td>0.93</td>
<td>0.98</td>
<td>0.96</td>
<td>1.90</td>
</tr>
</tbody>
</table>

### Table 11

**Comparison of Models**

<table>
<thead>
<tr>
<th>Model Comparison</th>
<th>$\Delta \chi^2$</th>
<th>$\Delta df$</th>
<th>P of $\Delta \chi^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1 versus Model 3</td>
<td>1484.19</td>
<td>692</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Model 1 versus Model 5</td>
<td>1479.63</td>
<td>690</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Model 3 versus Model 5</td>
<td>4.56</td>
<td>2</td>
<td>0.11</td>
</tr>
</tbody>
</table>
4.5 Convergent Validity

Sample 2 was used to examine convergent validity using ratings of the BLRI Obs-40 and the PCEPS-YP. Item and total scale means and standard deviations for both the BLRI and PCEPS-YP are shown in Appendix S. Correlations between the BLRI Obs-40 subscales, total score, and PCEPS-YP items are reported in Table 12. Pearson’s Product Moment Correlation between BLRI total score and PCEPS-YP total score was 0.83 (p<0.01). Each item of the PCEPS-YP also correlated significantly with the BLRI total score (r=0.47-0.69, p<0.01). The BLRI subscales of Level of Regard, Empathy, and Congruence significantly correlated with PCEPS-YP total score (Level of Regard, r=0.73, p<0.01; Empathy r=0.85, p<0.01; and Congruence r=0.58, p<0.01). Correlations between the unconditionality subscale and all PCEPS-YP items were low or insignificant.
Table 12

*Summary of Correlations Between PCEPS-YP and BLRI Subscales*

<table>
<thead>
<tr>
<th></th>
<th>Regard</th>
<th>Empathy</th>
<th>Unconditionality</th>
<th>Congruence</th>
<th>BLRI Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frame of Reference</td>
<td>0.46*</td>
<td>0.63*</td>
<td>0.08</td>
<td>0.19</td>
<td>0.47*</td>
</tr>
<tr>
<td>Tracking</td>
<td>0.61*</td>
<td>0.47*</td>
<td>0.42*</td>
<td>0.42*</td>
<td>0.59*</td>
</tr>
<tr>
<td>Empathic Resonance</td>
<td>0.54*</td>
<td>0.80*</td>
<td>0.12</td>
<td>0.41*</td>
<td>0.67*</td>
</tr>
<tr>
<td>Accepting Presence</td>
<td>0.61*</td>
<td>0.46*</td>
<td>0.51*</td>
<td>0.58*</td>
<td>0.67*</td>
</tr>
<tr>
<td>Genuineness</td>
<td>0.61*</td>
<td>0.47*</td>
<td>0.41*</td>
<td>0.64*</td>
<td>0.69*</td>
</tr>
<tr>
<td>Emotion Focused</td>
<td>0.46*</td>
<td>0.81*</td>
<td>0.01</td>
<td>0.35</td>
<td>0.61*</td>
</tr>
<tr>
<td>Emotion Symbolisation</td>
<td>0.44*</td>
<td>0.72*</td>
<td>-0.03</td>
<td>0.26</td>
<td>0.53*</td>
</tr>
<tr>
<td>Client Self-Development</td>
<td>0.59*</td>
<td>0.65*</td>
<td>-0.08</td>
<td>0.55*</td>
<td>0.66*</td>
</tr>
<tr>
<td>Developmental Response</td>
<td>0.51*</td>
<td>0.41*</td>
<td>-0.01</td>
<td>0.46*</td>
<td>0.51*</td>
</tr>
<tr>
<td>PCEPS-YP Total</td>
<td>0.73*</td>
<td>0.85*</td>
<td>0.20</td>
<td>0.58*</td>
<td>0.83*</td>
</tr>
</tbody>
</table>

*Significant at p<0.01 (2-Tailed)
4.6 Test Re-Test Reliability

Test-retest reliability was assessed using sample 3, which consisted of 50 BLRI Obs-40 ratings at T1 (October 2018) and T2 (March 2019) by the lead researcher (Rater 1). Means and standard deviations of BLRI items and subscales at T1 and T2 are found in Appendix T.

Pearson’s Product Moment Correlations between T1 and T2 were examined across the four subscales and Total BLRI score (Table 13). Level of Regard, Congruence, Unconditionality of Regard, and Total BLRI score were significantly correlated between T1 and T2 (r= 0.62 for Level of Regard, 0.60 for Congruence, 0.43 for Unconditionality of Regard, and 0.55 for Total BLRI score). The subscale of Empathy showed a weak non-significant correlation across T1 and T2 (r=0.24).

Cronbach’s alpha was estimated per subscale and total to assess consistency over time (Table 14). Total BLRI, Level of Regard, and Congruence subscales were found to have adequate reliability over five months (Cronbach’s alpha = 0.69 for total BLRI score, 0.75 for Level of Regard, and 0.75 for Congruence). The subscale of Unconditionality of Regard, and Empathy, had Cronbach’s alphas of 0.59 and 0.39, suggesting poor reliability over time.

The ICC for Total BLRI score was between fair and good at 0.69 (0.56-0.83), and good to excellent for the Level of Regard scale 0.74 (0.55-085) and Congruence scale 0.75 (0.55-0.85). The subscale of Unconditionality of Regard had an ICC of 0.53 (0.14-0.74) suggesting fair reliability. However, the Empathy subscale showed poor reliability of 0.38 (-0.08-0.65).
### Table 13

*Summary of Correlations Between Ratings at T1 and T2*

<table>
<thead>
<tr>
<th></th>
<th>Regard (T2)</th>
<th>Empathy (T2)</th>
<th>Unconditionality (T2)</th>
<th>Congruence (T2)</th>
<th>BLRI Total (T2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regard (T1)</td>
<td>0.61*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Empathy (T1)</td>
<td>0.24</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unconditionality (T1)</td>
<td>0.43*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Congruence (T1)</td>
<td></td>
<td>0.60*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BLRI Total (T1)</td>
<td></td>
<td></td>
<td></td>
<td>0.55*</td>
<td></td>
</tr>
</tbody>
</table>

*Significant At P <0.01 (2-Tailed)*

### Table 14

*Summary of Cronbach’s Alpha and Intraclass Correlations Between T1 and T2*

<table>
<thead>
<tr>
<th>Scale</th>
<th>Cronbach’s α</th>
<th>ICC 2-Way Mixed (CI)</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regard</td>
<td>0.75</td>
<td>0.74 (0.55-0.85)</td>
<td>20.19 (3.66)</td>
</tr>
<tr>
<td>Empathy</td>
<td>0.39</td>
<td>0.38 (-0.08-0.65)</td>
<td>18.06 (5.70)</td>
</tr>
<tr>
<td>Unconditionality</td>
<td>0.59</td>
<td>0.53 (0.14-0.74)</td>
<td>21.27 (1.93)</td>
</tr>
<tr>
<td>Congruence</td>
<td>0.75</td>
<td>0.75 (0.55-0.85)</td>
<td>14.00 (6.57)</td>
</tr>
<tr>
<td>Total BLRI</td>
<td>0.69</td>
<td>0.69 (0.46-0.83)</td>
<td>73.51 (13.67)</td>
</tr>
</tbody>
</table>
4.7 Inter-Rater Reliability

Ratings from sample 4 were used to assess inter-rater reliability. Descriptive statistics are shown in Appendix U. Further descriptive statistics are found in Appendix V and W, and report the percentage of absolute agreement of ratings across all three raters per items and per recording. Across the 50 recordings, raters scored items of the BLRI Obs-40 identically an average of 29.3% of the time. In particular, the three raters scored item 1 of the BLRI Obs-40 identically in 31 of the 50 recordings. Additionally, per recording, the three raters scored items the same an average 11.7% of the time. In particular, for recording 26, all three raters scored the same on 23 of the 40 items of the BLRI Obs-40.

Cronbach’s alpha was estimated for individual items and ranged from -1.16 to 0.59 (shown in Table 15). For subscales, the inter-rater reliability for Level of Regard was 0.43, Empathy 0.55, Unconditionality of Regard -0.29, and Congruence 0.32. For the total score, Cronbach’s alpha was 0.42.

Additionally, mean estimations along with 95% confidence intervals (CI) are reported for each ICC for each item and subscale of the BLRI. The ICC for Total BLRI score was poor at 0.28 (-0.41-0.53) and between poor and fair for the Level of Regard scale, 0.39 (0.02-0.36). ICC for Empathy and Congruence scale was fair with 0.55 (0.28-0.73) for Empathy and 0.31(-0.10-0.58) for Congruence. The ICC for Unconditionality of Regard was poor at -0.01 (-0.02-0.02) suggesting poor inter-rater reliability. This is reflected in the ICC estimates for the items of the Unconditionality of Regard scale.

K alpha was calculated to assess inter-rater reliability; K Alpha estimates are also shown in Table 15 Across subscales and total BLRI score K Alpha was found to be unsatisfactory, i.e. <0.8 (Hayes & Krippendorff, 2007).
### Table 15

**Summary of Cronbach’s Alpha, Krippendorff’s Alpha, Intraclass Correlations, Means, and Standard Deviations for Inter-Rater Reliability Using Three Raters**

<table>
<thead>
<tr>
<th>Item</th>
<th>Cronbach’s α</th>
<th>K α</th>
<th>ICC 2-Way Mixed (CI)</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.03</td>
<td>-0.01</td>
<td>-0.03 (-0.65-0.38)</td>
<td>2.05 (0.37)</td>
</tr>
<tr>
<td>2</td>
<td>0.58</td>
<td>0.33</td>
<td>0.60 (0.35-0.76)</td>
<td>1.77 (0.77)</td>
</tr>
<tr>
<td>3</td>
<td>-0.01</td>
<td>-0.27</td>
<td>-0.02 (-0.15-0.19)</td>
<td>1.43 (0.73)</td>
</tr>
<tr>
<td>4</td>
<td>-0.48</td>
<td>-0.29</td>
<td>-0.18 (-0.48-0.14)</td>
<td>1.38 (0.69)</td>
</tr>
<tr>
<td>5</td>
<td>0.39</td>
<td>0.16</td>
<td>0.38 (0.03-0.62)</td>
<td>1.63 (0.72)</td>
</tr>
<tr>
<td>6</td>
<td>0.55</td>
<td>0.29</td>
<td>0.55 (0.28-0.73)</td>
<td>1.50 (1.11)</td>
</tr>
<tr>
<td>7</td>
<td>0.09</td>
<td>-0.40</td>
<td>0.01 (-0.04-0.09)</td>
<td>1.00 (0.62)</td>
</tr>
<tr>
<td>8</td>
<td>0.35</td>
<td>0.13</td>
<td>0.34 (-0.04-0.60)</td>
<td>1.44 (1.04)</td>
</tr>
<tr>
<td>9</td>
<td>-0.50</td>
<td>-0.14</td>
<td>-0.47 (-1.29-0.10)</td>
<td>2.08 (0.57)</td>
</tr>
<tr>
<td>10</td>
<td>0.41</td>
<td>0.19</td>
<td>0.41 (0.06-0.65)</td>
<td>1.71 (0.81)</td>
</tr>
<tr>
<td>11</td>
<td>0.09</td>
<td>-0.07</td>
<td>0.10 (-0.29-0.37)</td>
<td>1.95 (0.39)</td>
</tr>
<tr>
<td>12</td>
<td>-0.09</td>
<td>-0.04</td>
<td>-0.10 (-0.70-0.34)</td>
<td>2.00 (0.54)</td>
</tr>
<tr>
<td>13</td>
<td>0.35</td>
<td>0.15</td>
<td>0.35 (-0.04-0.61)</td>
<td>1.87 (0.54)</td>
</tr>
<tr>
<td>14</td>
<td>-0.03</td>
<td>-0.04</td>
<td>-0.03 (-0.56-0.36)</td>
<td>1.78 (0.65)</td>
</tr>
<tr>
<td>15</td>
<td>0.001</td>
<td>-0.45</td>
<td>0.0001 (-0.02-0.04)</td>
<td>0.74 (0.56)</td>
</tr>
<tr>
<td>16</td>
<td>0.35</td>
<td>-0.01</td>
<td>0.26 (-0.06-0.51)</td>
<td>0.24 (1.49)</td>
</tr>
<tr>
<td>17</td>
<td>0.21</td>
<td>0.06</td>
<td>0.20 (-0.25-0.51)</td>
<td>2.15 (0.56)</td>
</tr>
<tr>
<td>18</td>
<td>0.24</td>
<td>0.10</td>
<td>0.24 (-0.23-0.54)</td>
<td>1.63 (0.82)</td>
</tr>
<tr>
<td>19</td>
<td>0.15</td>
<td>0.03</td>
<td>0.14 (-0.32-0.47)</td>
<td>2.06 (0.63)</td>
</tr>
<tr>
<td>20</td>
<td>0.38</td>
<td>0.11</td>
<td>0.34 (-0.01-0.59)</td>
<td>1.00 (1.43)</td>
</tr>
<tr>
<td>21</td>
<td>0.25</td>
<td>-0.24</td>
<td>0.10 (-0.09-0.30)</td>
<td>1.34 (0.71)</td>
</tr>
<tr>
<td>22</td>
<td>0.34</td>
<td>0.15</td>
<td>0.34 (-0.06-0.60)</td>
<td>1.87 (0.95)</td>
</tr>
<tr>
<td>23</td>
<td>-1.16</td>
<td>-0.24</td>
<td>-1.10 (-3.00-0.11)</td>
<td>2.01 (0.41)</td>
</tr>
<tr>
<td>24</td>
<td>0.35</td>
<td>0.13</td>
<td>0.34 (-0.04-0.60)</td>
<td>1.73 (0.87)</td>
</tr>
<tr>
<td>25</td>
<td>-0.01</td>
<td>-0.17</td>
<td>-0.01 (-0.26-0.26)</td>
<td>2.20 (0.41)</td>
</tr>
<tr>
<td>26</td>
<td>0.42</td>
<td>0.18</td>
<td>0.41 (0.07-0.65)</td>
<td>1.82 (0.55)</td>
</tr>
<tr>
<td>27</td>
<td>0.003</td>
<td>-0.46</td>
<td>0.0001 (-0.02-0.03)</td>
<td>0.77 (0.50)</td>
</tr>
<tr>
<td>28</td>
<td>0.02</td>
<td>-0.03</td>
<td>0.01 (-0.48-0.38)</td>
<td>1.63 (0.82)</td>
</tr>
<tr>
<td>29</td>
<td>0.51</td>
<td>0.26</td>
<td>0.52 (0.22-0.71)</td>
<td>1.93 (0.63)</td>
</tr>
<tr>
<td>30</td>
<td>0.39</td>
<td>0.17</td>
<td>0.39 (0.02-0.63)</td>
<td>1.71 (0.96)</td>
</tr>
<tr>
<td>31</td>
<td>0.02</td>
<td>-0.30</td>
<td>0.01 (-0.12-0.17)</td>
<td>1.87 (0.50)</td>
</tr>
<tr>
<td>32</td>
<td>0.11</td>
<td>0.01</td>
<td>0.11 (-0.37-0.45)</td>
<td>1.73 (0.62)</td>
</tr>
<tr>
<td>33</td>
<td>0.28</td>
<td>0.06</td>
<td>0.25 (-0.13-0.53)</td>
<td>2.15 (0.66)</td>
</tr>
<tr>
<td>34</td>
<td>0.59</td>
<td>0.31</td>
<td>0.58 (0.34-0.75)</td>
<td>1.75 (0.94)</td>
</tr>
<tr>
<td>35</td>
<td>0.20</td>
<td>0.02</td>
<td>0.18 (-0.22-0.48)</td>
<td>1.96 (0.66)</td>
</tr>
<tr>
<td>36</td>
<td>-0.16</td>
<td>-0.08</td>
<td>-0.14 (-0.70-0.29)</td>
<td>1.86 (0.59)</td>
</tr>
<tr>
<td>37</td>
<td>0.40</td>
<td>0.02</td>
<td>0.30 (-0.03-0.55)</td>
<td>1.11 (1.09)</td>
</tr>
<tr>
<td>38</td>
<td>-0.27</td>
<td>-0.12</td>
<td>-0.22 (-0.79-0.22)</td>
<td>1.91 (0.68)</td>
</tr>
<tr>
<td>39</td>
<td>-0.19</td>
<td>-0.37</td>
<td>-0.04 (-0.13-0.10)</td>
<td>1.19 (0.68)</td>
</tr>
<tr>
<td>40</td>
<td>0.27</td>
<td>0.10</td>
<td>0.27 (-0.15-0.56)</td>
<td>0.47 (0.91)</td>
</tr>
</tbody>
</table>

- **Regard**: Cronbach’s α = 0.43, K α = 0.14, ICC = 0.39 (0.02-0.36), Mean = 18.53 (4.28)
- **Empathy**: Cronbach’s α = 0.55, K α = 0.29, ICC = 0.55 (0.28-0.73), Mean = 17.45 (5.96)
- **Unconditionality**: Cronbach’s α = -0.29, K α = -0.47, ICC = -0.01 (-0.02-0.02), Mean = 14.99 (1.74)
- **Congruence**: Cronbach’s α = 0.32, K α = 0.12, ICC = 0.31 (-0.10-0.58), Mean = 14.70 (5.66)

**Total BLRI Score**: Cronbach’s α = 0.42, K α = -0.03, ICC = 0.28 (-0.41-0.53), Mean = 65.68 (14.25)
4.7.1 Post-Hoc Analysis

Results from the inter-rater reliability analysis, and in particular the negative Cronbach’s alpha reported, encouraged further investigation. The internal reliability of the scores from each individual rater was calculated using Cronbach’s alpha. In addition, correlations between the pairs of raters were examined using Pearson’s Product Moment Correlation.

Appendix U reports findings of internal consistency for each rater. The internal consistency of the total BRLI score, Level of Regard, Empathy, and Congruence subscales for the three raters had a Cronbach’s alpha of 0.65-0.91, suggesting good internal consistency. However, the internal reliability for the Unconditionality of Regard subscale was poor for all three raters.

Table 16 reports correlations between the pairs of raters. Correlations between Rater 1 and Rater 2 demonstrated significant positive correlations on the subscales of Level of Regard, Empathy, and Congruence, and total BLRI score. However, Rater 3 did not significantly correlate with Rater 1 or Rater 2 on any of the subscales or total BLRI score.

### Table 16

<table>
<thead>
<tr>
<th></th>
<th>Regard</th>
<th>Empathy</th>
<th>Unconditionality</th>
<th>Congruence</th>
<th>Total BLRI</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1 vs. R2</td>
<td>0.42* 0.02</td>
<td>0.64* &lt;0.01</td>
<td>-0.01 0.92</td>
<td>0.44* &lt;0.01</td>
<td>0.51* &lt;0.01</td>
</tr>
<tr>
<td>R1 vs. R3</td>
<td>-0.02 0.88</td>
<td>0.03 0.81</td>
<td>-0.13 0.36</td>
<td>-0.03 0.82</td>
<td>-0.05 0.75</td>
</tr>
<tr>
<td>R2 vs. R3</td>
<td>0.26 0.70</td>
<td>0.19 0.18</td>
<td>-0.09 0.55</td>
<td>0.01 0.98</td>
<td>0.20 0.16</td>
</tr>
</tbody>
</table>

Findings suggested Rater 3 to be unreliable; therefore, an analysis of the inter-rater reliability between Rater 1 and Rater 2 was performed. Findings are reported in Table 17.
Cronbach’s alpha was estimated for subscales, the inter-rater reliability for Level of Regard was 0.59, Empathy 0.78, Unconditionality of Regard -0.03, and Congruence 0.57. For the total score, Cronbach’s alpha was 0.66.

Additionally, mean estimations along with 95% confidence intervals (CI) were reported for each ICC for the subscales of the BLRI and the total BLRI score. The ICC for Total BLRI score was good at 0.67 (0.41-0.83) and between fair and good for the Level of Regard scale, 0.59 (0.28-0.76). ICC for Empathy was excellent with 0.78 (0.62-0.88) and fair for the Congruence scale with 0.54 (0.20-0.74). However, the ICC for Unconditionality of Regard was poor at -0.01 (-0.29-0.26).

K alpha was also calculated to assess inter-rater reliability between Rater 1 and Rater 2; however, this was found to be unsatisfactory across all subscales and total BLRI score (K Alpha <0.8).

Table 17

*Summary of Post-Hoc Analysis including Cronbach’s Alpha, Intraclass Correlations, and Krippendorff’s Alpha for Inter-Rater Reliability Using Two Raters*

<table>
<thead>
<tr>
<th></th>
<th>Cronbach’s α</th>
<th>K α</th>
<th>ICC 2-Way Mixed (CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regard</td>
<td>0.59</td>
<td>0.41</td>
<td>0.59 (0.28-0.76)</td>
</tr>
<tr>
<td>Empathy</td>
<td>0.78</td>
<td>0.64</td>
<td>0.78 (0.62-0.88)</td>
</tr>
<tr>
<td>Unconditionality</td>
<td>-0.03</td>
<td>-0.34</td>
<td>-0.01 (-0.29-0.26)</td>
</tr>
<tr>
<td>Congruence</td>
<td>0.57</td>
<td>0.34</td>
<td>0.54 (0.20-0.74)</td>
</tr>
<tr>
<td>Total BLRI Score</td>
<td>0.66</td>
<td>0.50</td>
<td>0.67 (0.41-0.81)</td>
</tr>
</tbody>
</table>
4.8 COSMIN Checklist

Post analysis, the COSMIN checklist was used to assess the methodological quality of the validation procedures used to evaluate the BLRI Obs-40. Judgments were made on the assessment of internal consistency, inter-rater and test rest reliability, hypothesis testing, and structural validity. Evaluations are summarised in Table 18. The overall COSMIN methodological quality score for the BLRI Obs-40, as assessed in this study, is good. The assessments of the scale’s internal consistency were of high methodological quality, as this was assessed using Cronbach’s alpha and calculated for each subscale, as required in the COSMIN criteria. In addition, the scale’s inter-rater and test-retest reliability meet the COSMIN criteria for very good methodological quality, as intraclass correlations were used in favour of Pearson’s correlation. The assessment of the scale’s structural validity using Confirmatory Factor Analysis is good, as is the hypothesis testing assessing convergent validity. However, additional psychometric properties such as divergent validity, content validity, and cross-cultural validity have not been assessed in this study, therefore the psychometric evaluation of the BLRI-Obs 40 is limited to the specific aspects of reliability and validity assessed in this study.
<table>
<thead>
<tr>
<th><strong>Table 18</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Psychometric Characteristics and Quality Assessment of BLRI Obs-40 (Version 3)</strong></td>
</tr>
<tr>
<td><strong>Barrett-Lennard Relationship Inventory Obs-40 (Version 3)</strong></td>
</tr>
<tr>
<td><strong>Number of Items</strong></td>
</tr>
<tr>
<td><strong>Theoretical Basis</strong></td>
</tr>
</tbody>
</table>
| **Construct** | (1) Empathic Understanding  
(2) Level of Regard  
(3) Unconditionality of Regard  
(4) Congruence |
| **Study Population and Country** | 167 clients and 19 counsellor, UK |
| **Internal Consistency** | Cronbach’s Alpha and Inter-Scale Correlations |
| **COSMIN Study Quality** | VERY GOOD: Cronbach’s alpha calculated for each subscale |
| **Reliability** | Inter-Rater Reliability and Test-Retest Reliability assessed via ICC |
| **COSMIN Study Quality** | VERY GOOD: ICC calculated and appropriate time scale used |
| **Structural Validity** | Confirmatory factor analysis supports the four-factor structure |
| **COSMIN Study Quality** | GOOD: CFA calculated and confirms original structure |
| **Hypothesis Testing (Convergent Validity)** | Comparison with PCEPS-YP |
| **COSMIN Study Quality** | GOOD: Hypotheses formulated a priori with expected direction but not magnitude |
| **Overall COSMIN Evaluation** | GOOD |
Chapter 5: Discussion

5.1 Chapter Introduction

The present study examined the psychometric properties of the BLRI Obs-40 Version 3, with a population of young people. The discussion from the psychometric assessment of the scale will follow, with a focus on the contributions and value to future research.

5.2 Summary of Findings

This project aimed to explore the psychometric properties of the BLRI Obs-40 in humanistic counselling for adolescents. Specifically, the objectives were to examine and evaluate the reliability and validity of the scale, including internal reliability, test-retest reliability, inter-rater reliability, structural validity and convergent validity. Despite the extensive assessment and use of the BLRI, no other study to date has assessed the psychometric quality of the observer-rated BLRI in a population of young people.

Furthermore, there has been little investigation into the inter-rater reliability and convergent validity of the scale. Additionally, the majority of evidence for the construct validity of the BLRI has been established via Exploratory Factor Analysis, with the exception of one Confirmatory Factor Analysis study supporting the theoretical model underpinning the scale (Liao et al., 2018). However, this is the first study to use Confirmatory Factor Analysis to confirm the theoretical model of the therapeutic relationship in a sample of young people.

Overall, the findings of this study suggest the BLRI Obs-40 is a reliable measure of Level of Regard, Empathy, and Congruence in therapeutic relationships with young people. The reliability of the Unconditionality of Regard subscale is less established, suggesting its
independence with the remaining subscales in the BLRI. In addition, reliability across raters and over time is less supported. The BLRI has also been confirmed theoretically via Confirmatory Factor Analysis and assessment of convergent validity with the PCEPS-YP.

5.2.1 Internal Reliability

Internal reliability of the BLRI Obs-40 was established using Cronbach’s alpha and inter-scale correlations. Findings suggest the scale demonstrates good reliability for the measurement of the therapeutic relationship condition via observation and can be applied to young people.

The scale, overall, had high internal consistency with a Cronbach’s alpha of 0.93. This exceeds the reliability estimates of the original BLRI validation (Barrett-Lennard, 1962) and indicates a high internal consistency across all items. This finding also replicates the reliability estimates of 14 studies from a review conducted by Gurman (1977). The internal consistencies of the subscales of the BLRI were also assessed. The subscales of Level of Regard, Empathy, and Congruence had high internal consistency ($\alpha=0.84–0.94$), suggesting the items that make up each of the subscales measure a similar latent variable. The exception to this was the Unconditionality of Regard subscale, which had a Cronbach’s alpha of 0.59, suggesting poor internal consistency among the items of the subscale.

Comparing to previous studies of the internal consistency of the subscales of the BLRI, there appears to be a trend in which the Unconditionality of Regard scale demonstrates poorer internal reliability than the remaining subscales (Gurman, 1977; Liao et al., 2018). Results from this study found no improvements could be made to the internal consistency of the Unconditionality of Regard subscale if any items were deleted; however, findings from the internal consistency for the four subscales of the BLRI show an improved Cronbach’s
alpha from 0.77 to 0.84 if the subscale of Unconditionality of Regard were removed. This suggests a level of independence with regards to the Unconditionality of Regard subscale, and that perhaps the subscale is measuring a different construct than the remaining Level of Regard, Empathy, and Congruence subscales.

Internal consistency findings are further supported when considering the strength of correlations between the subscales. The subscales of Level of Regard, Empathy, and Congruence demonstrated strong positive correlations with each other; however, the subscale of Unconditionality of Regard weakly correlated with the remaining subscales. This finding replicates the trend seen in Gurman’s (1977) review of the BLRI. Based on seven studies, average inter-scale correlations were assessed, and the subscale of Unconditionality of Regard was found to correlate with the subscales weakly. These findings imply the independence of the Unconditionality of Regard scale and could be taken as further support for the division of the core condition of unconditional positive regard into two subscales (Barrett-Lennard, 1959).

5.2.2 Test-Retest Reliability

Test-retest reliability of a subset of recordings revealed moderate stability in Level of Regard, Unconditionality of Regard, Congruence, and total BLRI scores over a five-month period, as assessed by Cronbach’s alpha and ICC. However, poor stability over time was found for the Empathy subscale. To compare data to previous research Pearson’s Product Moment Correlation coefficient was also assessed between T1 and T2. A weak non-significant correlation for Empathy at T1 and T2 was found. All other subscales and total BLRI score demonstrated fair significant correlations over the 5-month period. Taken together, findings suggest the Empathy subscale is not stable over time. However, there
appears to be fair to good stability over a 5-month period for total BLRI score, and the subscales of Congruence and Level of Regard and Unconditionality of Regard.

Past test-retest data from 10 studies included in Gurman’s (1977) review, found mean correlations of 0.80 to 0.85 for subscales, suggesting good stability over time. However, these findings are based on correlation coefficients, which, unlike ICC, do not take into account systematic error (Streiner & Kottner, 2014).

Nevertheless, findings from the current study support the stability of the Congruence, Level of Regard, and Unconditionality of Regard subscales for use with young people, but not the Empathy subscale. It is possible that the time frame between the first and second set of ratings was not ideal for all subscales. There was a need for negotiation between reducing memory bias and maintaining near-identical ratings conditions and environment. However, a shorter time frame may have been more appropriate, given the majority of past test-retest studies using the BLRI have a time frame of approximately four weeks. Three test-retest studies included in Gurman’s (1977) review were conducted over a more extended period of time (up to one year) and did show good stability. However, these studies used non-therapeutic samples, alternative BLRI formats and used correlations to analyses results.

Furthermore, the poor stability of the Empathy subscale could be a result of the version of the scale used. It is possible that when using an observer rated scale empathy is harder to detect as the rater is positioned externally to the therapy relationship. Empathy as a core condition is described from the perspective of the therapist as a felt sense (Rogers, 1957). Therefore, it is unlikely that observers are able to fully comprehend the level of empathy present in the relationship, especially when this may not always be verbalised.

It is possible the lack of stability in the Empathy subscale is due to rater attributes rather than the scale (Yu, 2005). Over time, the rater had become more familiar with the BLRI Obs-40 and rating procedure. In addition, as the rater was a therapist in training, their
clinical skills and practice would have improved over time; meaning interpretation of the items or the understanding of how the core conditions were assessed by the scale may have shifted and most likely influenced their ability to rate therapy segments consistently. It is possible the lack of stability is a consequence of this increased rating proficiency, and the Empathy subscale is more sensitive to changes in an observers ability to rate sessions over time than the Level of Regard, Congruence, and Unconditionality of Regard subscales.

5.2.3 Inter-Rater Reliability

Inter-rater reliability was assessed using a subset of recordings (sample 4). Regarding absolute rater agreement, across 50 recordings, all three raters scored identically on an average 29.3% of items, in addition across the 40 items, all three raters scored the same on average of 11.7% of the recordings. Given the 6-point rating scale, this could suggest a moderate level of inter-rater agreement. However, when looking at the average scores for the subscales of the BLRI Obs-40, as well as the total BLRI score, there is a clear discrepancy between Raters 1 and 2, with Rater 3. Across all subscales, Rater 3 scored on average lower than both Rater 1 and Rater 2, suggesting less consistency across the three raters.

Regarding the findings from the inter-rater analysis, Cronbach’s alpha for total BLRI score, Level of Regard, and Empathy subscales demonstrated consistent ratings across three raters. However, Cronbach’s alpha for the Unconditionality of Regard, and Congruence subscales were poor. This trend is similar to the findings of internal reliability conducted with Sample 1 (Section 5.2.1). Furthermore results from K alpha suggest poor agreement across raters for all items and subscales of the BLRI. This is further supported by ICC findings, which imply poor agreement and correlations between raters scoring of the BLRI. Therefore, taken together, these results suggest poor inter-rater reliability of the BLRI Obs-40 across all
items and subscales. Additionally, many individual items had negative reliability estimates, indicating that items were interpreted differently by each of the raters.

Further post-hoc analyses revealed Rater 3 to be unreliable, correlations between the pairs of raters found Rater 1 and Rater 2 to significantly correlate with each other, whereas Rater 3 did not significantly correlate with either rater. Additional analyses between Rater 1 and Rater 2, found an improvement in the inter-rater reliability as measured by Cronbach’s alpha and ICC; however, K alpha was still unsatisfactory.

The low level of agreement across raters may reflect the differences in the raters’ clinical training and client facing experience. Out of the three raters, Rater 1 and Rater 2 had approximately seven years of training each, whereas Rater 3 had five years of training. It is possible differences in rater agreement were influenced by the duration of training. Furthermore, Rater 1 and Rater 2 trained and practised therapy verbally, whereas Rater 3 practised using a combination of verbal and non-verbal therapy, meaning Rater 1 and Rater 2 were likely to be more familiar with the format of therapy present in the audio recordings. This could explain why Rater 1 and Rater 2 correlated in their scores when Rater 3 did not.

In addition, low agreement across raters may be a consequence of the raters understanding of the person-centred and humanistic approach as a result of their training orientation. The raters included an integrative counselling psychologist trainee (Rater 1), a person-centred counsellor (Rater 2), and a movement psychotherapist (Rater 3), each possessing varying levels of familiarity with the person-centred approach. It is possible Rater 3 may have interpreted items on the BLRI Obs-40 from a different perspective than Rater 1 and Rater 2. As Rater 3 was a movement psychotherapist, their understanding of the core conditions may have differed and included examples or criteria of the conditions that are illustrated non-verbally. For example, when rating the level and presence of empathy, Rater 3 may have been working with a definition of empathy that included embodied empathy or
mirroring. As such, these expressions of empathy would not have been present in the audio recordings. This may also account for why the average subscale and total BLRI score across recordings for Rater 3 were lower than the scores for Rater 1 and Rater 2. Rater 3 may have adopted a more stringent criterion when scoring and would have required stronger behavioural evidence when rating.

Differences across the raters were also found in their familiarity with the recordings, rating procedure and BLRI Obs-40. Rater 1 and Rater 2 were familiar with the recordings and the project procedure prior to data collection as they both worked as auditors on the ETHOS trial. The ETHOS trial used the same therapy extracts and a very similar rating procedure to this project. Therefore, the agreement between Rater 1 and Rater 2 could be attributed to familiarity with the recordings, and it may be that raters are more reliable in their scoring with greater experience. Furthermore, Rater 1 had the most practice using the BLRI Obs-40, having used the scale to collect the data for the rest of the project. It is possible the varying levels of familiarity with the scale and recordings, along with the other between rater differences discussed, acted as confounding variables and may account for the low level of rater agreement.

Results of inter-rater reliability from the pilot study ratings, however, indicate that rater differences such as level of person-centred understanding or familiarity with the scale and recordings are not the sole explanation for poor rater agreement. Poor agreement may also result from biases in the interpretation of scale items. Similar levels of rater agreement were found in the pilot study even when one of the raters was an expert in the person-centred field, and had also gained experience with the recordings and rating procedure. Therefore, even though rater experience and training may impact their understanding of the therapeutic relationship factors, which impact their ratings, there also appears to be a discrepancy in the interpretation of items that may originate in the scale. Specific scale and item interpretations
are discussed in section 5.3; however, differences were found in rater interpretation. Looking further at the differences between the raw scores across the three raters, Rater 3 often scored oppositely to Raters 1 and 2 on the negatively worded items. It is possible, Rater 3 misinterpreted these negatively worded items, this may have been due to the fact that Rater 3 was a non-native English speaker and therefore the scale is less adapted or less sensitive to those with English as a second language, and items are more likely to be misinterpreted.

The inter-rater reliability findings suggest raters with similar training backgrounds and level of understanding of the person-centred model may interpret and score the BLRI Obs-40 more consistently with each other, this indicates further training on the scale could therefore improve reliability. However, item interpretation, which originates in the scale and externally to the raters, can also impact the level of inter-rater agreement.

5.2.4 Structural Validity

Data were consistent with the original theoretical model of the therapeutic relationship proposed by Barrett-Lennard (1962). Strong support was found for a four-factor correlated model (model 3) consisting of the four subscales of the BLRI. All items loaded significantly upon their hypothesised factor, with high factor loadings for most items. However, findings from the correlation matrix of the factors do show a weak correlation between the Unconditionality of Regard factor and Level of Regard, Empathic Understanding, and Congruence factors. This supports previous findings (Gurman, 1977; Lin, 1973; McClanahan, 1974) that imply the independence of the Unconditionality of Regard factor. Therefore, a three-factor correlated model (model 4), omitting the Unconditionality of Regard factor, was tested to examine the fit to the data. This model provided a good fit for
the data with high factor loadings; however, chi-squared was found to be significant, suggesting the model was a good but not an exact fit.

Based on previous Exploratory Factor Analyses of the BLRI (Mills & Zytowski, 1967), and findings from the inter-scale correlations found in the present study, the possibility of a higher-order factor or a single-factor model was suggested. The data might have reflected a single dimension of person-centeredness, in which the subscales or individual items of Level of Regard, Empathic Understanding, Congruence, and Unconditionality all contributed to. The single-factor model fitted the data poorly. However, the higher-order factor (model 5) produced a good fit for the data. The good fit of the higher-order factor supports the work of previous explorative factor analyses, in which the four subscales of the BLRI contribute to an overall latent factor of the person-centred relationship (Mills & Zytowski, 1967). These findings may also support claims from relational depth theory, in which the core conditions of Empathy, Congruence, and Unconditional Positive Regard are thought to integrate and blend together to represent a single variable of relational depth (Knox, Murphy, Wiggins, & Cooper, 2012). However, upon closer inspection of the higher-order model, only three factors (Level of Regard, Congruence, and Empathy) loaded significantly on the higher-order factor of person-centeredness. Therefore, even though there was no significant difference found between the fit of model 3 and model 5, model 3 may provide a better overall fit for the data.

Nevertheless, both the four-factor correlated model and the four-factor one second-order factor can be taken as support for the original theoretical model of the therapeutic relationship as outlined by Barrett-Lennard (1962). Therefore, results suggest the scale can be used to measure the therapeutic relationship conditions in young people receiving school-based humanistic counselling.
5.2.5 Convergent Validity

The convergent validity of the BLRI Obs-40 was supported by a strong positive correlation between the total BLRI score and total PCEPS-YP score. Positive significant correlations ($r > 0.50$) were found for all hypothesised correlations between the items of the PCEPS-YP and the BLRI subscales, further supporting the construct validity of the subscales of the BLRI Obs-40. The Level of Regard and Unconditionality of Regard subscales correlated highly with item 4 ‘Accepting Presence’ of the PCEPS-YP. Suggesting these items measure a similar or related construct. Additionally, the Congruence subscale and item 5 ‘Genuineness’ of the PCEPS-YP strongly correlated, and many of the Congruence subscale items closely matched the descriptions and example questions found in Genuineness item 5, such as item 12 in the BLRI Obs-40, ‘My sense is that s/he is genuine and honest with B’ and ‘Therapist sounds consistently natural or genuine’ in the PCEPS-YP. Positive correlations were found between the Empathy subscale and many of the PCEPS-YP items, in particular, strong correlations were found with items 6 and 7 of the PCEPS-YP, which focused specifically on the emotional, and feeling based responses.

Additionally, high positive correlations were found between items of the PCEPS-YP and BLRI Obs-40 that had not been previously hypothesised. The Level of Regard subscale correlated highly with item 2 ‘Tracking’ and item 5 ‘Genuineness’ of the PCEPS-YP. It is possible that these PCEPS-YP items encapsulate a sense of being with the clients and adapting and accepting to whatever they may bring to the room. This is related to item 13 of the BLRI Obs-40, ‘The way that A responds conveys a personal appreciation and valuing of B.’ Additionally the Empathy subscale of the BLRI Obs-40 also correlated with item 8 of the PCEPS-YP ‘Facilitation of Client Self-Development’ which measures a therapists expression of a clients inner world. It is possible the Empathy subscale of the BLRI Obs-40, and item 8 of the PCEPS-YP are measuring the expression of therapist emotion within the relationship.
The Level of Regard, Empathy, and Congruence subscales of the BLRI correlated significantly with the PCPES-YP total score; suggesting good convergent validity across the scales. This is supported by findings of convergent validity from the Measure of Expressed Empathy Scale (Watson & Prosser, 2002). This is an observer rated measure of therapist communicated empathy, which was correlated with the Empathy subscale of the client rated BLRI, which found a significant positive correlation \((r=0.66)\). Conversely, the Unconditionality of Regard subscale did not significantly correlate with the PCEPS-YP total score or many PCEPS-YP individual items; suggesting a difference in the underlying constructs of the PCEPS-YP and Unconditionality of Regard subscale. This is supported by previous findings of convergent validity with the Therapist Presence Inventory (TPI) (Geller et al., 2010). The TPI is a therapist self report measure of in-session process and experience of therapeutic presence. Correlations between TPI and therapist BLRI subscale scores found significant positive correlations for all subscales (Level of Regard \(r=0.34\), Empathy \(r=0.59\), Unconditionality of Regard \(r=0.20\), and Congruence \(r=0.41\)); however, the correlation between TPI and Unconditionality of Regard was weakest among all subscales.

One possible explanation for the lack of compatibility between the PCEPS-YP and the Unconditionality of Regard scale could be due to the independence of the Unconditionality subscale. The Unconditionality of Regard subscale poorly correlates with the remaining BLRI subscales indicating an independence from the rest of the BLRI Obs-40, this is replicated with its correlations with the PCEPS-YP. Therefore, due to this independence, the Unconditionality subscale may be an inappropriate assessment of convergent validity.

Furthermore, the PCEPS-YP operationalises classic person-centred theory and experiential humanistic approaches; as such, the instrument consists of two subscales. There is a divide of items within the scale across person-centred process and experiential process;
therefore, not all items of the PCEPS-YP will focus on the therapist relationship conditions as the BLRI Obs-40, many will focus on the facilitation and exploration of emotion instead (Freire et al., 2014). These differences may account for the imperfect convergent validity of the BLRI Obs-40 and PCEPS-YP. In addition, the BLRI aims to measure the therapeutic relationship conditions, whereas the PCEPS-YP acts as an adherence and competency measure of person-centred practice. Therefore, one scale focuses on the relationship between the therapist and client, whereas the other focuses on the therapist’s skill. Looking at the correlations in Table 12, it appears that the majority of PCEPS items that aim to measure relationship are more strongly correlated with the BLRI Obs-40 subscales. Given that the BLRI was developed with to measure classical person-centred theory, it is expected that the person-centred PCESP-YP items will have stronger and significant correlations with the BLRI subscales than the experiential process items. This holds true for the Unconditionality of Regard and Congruence subscales; however, the Level of Regard and Empathy subscales do appear to significantly and strongly correlate the experiential process items of the PCEPS-YP also. Therefore, it is possible that the Level of Regard and Empathy subscales could potentially measure experiential process as well as the classic person-centred relationship conditions.

Additionally, there is a practical difference between the two instruments. Given that the PCEPS-YP acts as a therapist adherence measure, it is possible the scale is better suited as an observer measure than the BLRI Obs-40, which measures the relationship conditions. When scoring the PCEPS-YP, there may be a reliance on more explicit and external demonstrations of therapist skill that is more suited to observer measurement. Whereas, the focus on the relationship in BLRI Obs-40 may mean there are greater assumptions made about the relationship and the internal world of both therapist and client that cannot be clearly observed, and one must rely on behavioural observable indicators instead.
5.3 Evaluation of the BLRI Obs-40 (Version 3)

The majority of this paper has focused on testing the psychometric properties of the BLRI Obs-40 in a population of young people. What follows are the conclusions of this testing, with an aim to provide an overall assessment of the scale’s suitability and potential use with young people.

5.3.1 Scale Design, Content, and Scoring

Certain characteristic of response scales can affect the psychometric quality of a measurement instrument (DeCastellarnau, 2018). One such characteristic is the scales’ evaluative dimension; in the BLRI, this refers to the ‘agree-disagree’ Likert-type response. Agree-disagree responses have been found to increase the presence of acquiescence bias (Billiet & McClendon, 2000). The balance of both negatively worded items and positively worded items can counter the risk of acquiescence bias. However, respondents can become confused by the inclusion of negatively worded items, or by their order in the scale and therefore misinterpret these items. For example, when looking at the inter-rater reliability results, Rater 3 frequently misinterpreted the negatively worded items. In addition, some items in the BLRI can be interpreted as both positive and negative, such as item 40, and as a result of this ambiguity, the quality of responses to this item is threatened. Furthermore, within the BLRI there is the potential for greater ambiguity as the Likert response refers to ‘true-not true’ evaluations. The varying degrees of ‘true-not true’ could cause difficulty when rating as there may be greater cognitive effort required to assess the extent to which an item can be true. Raters are required to differentiate between the different levels of truth with certainty, which could cause doubt, as typically truth is understood as a complete construct.
without fluctuating levels. Therefore, the use of a ‘true-not true’ evaluation may be better suited to a dichotomous rating response, and potentially reduce the scale’s response quality.

Looking at the mode of scores per item, it appears some items produce responses with more variability than other items, which in turn will impact subscale scores. For example, items that make up the Empathy subscale had a greater range of responses than those that made up the Unconditionality of Regard subscale. However, a closer examination of the descriptive statistics of this study reveals a lack of variability of responses in a majority of items of the BLRI, and a positive skew in scoring across the subscales, suggesting that not all response options on the Likert scale are used when rating the recordings. This could suggest high acquiescence bias or the possibility that not all score responses are required. For example, scoring on the majority of the empathy items across all recordings fell within the positive range (+1 to +3); meaning the scale may cause a bias towards the favourable scoring of empathy items. However, given the context of the recording – i.e. a humanistic therapy setting, it is expected and almost assumed that empathy will be present in the majority of cases; therefore, it is unsurprising that not all of the response options will be used. Perhaps in different relationship contexts, such as friendship groups or parent-child relationships, there is greater variability in responses, as empathy is not always assumed. This begins to question the suitability and applicability of the observer rated BLRI in therapeutic contexts, whereby certain characteristics of the therapy relationship are expected.

Furthermore, the use of agree-disagree response scales has been found to require a great cognitive effort from respondents (Kunz, 2015). This, in combination with the length of the scale, can increase the risk of respondent fatigue, which has been found to impact rater response quality (Wetzel & Greiff, 2018). Given that the BLRI includes 40 items, it is possible raters became tired and this could have invited careless responding, endangering response quality. In addition, having numerous response options, as in the BLRI, can increase
the level of interpretive effort required, which can be detrimental to the consistency of ratings (Kieruj & Moors, 2013). With regards to the BLRI, the interpretation required for both the scale items and Likert responses could account for the inconsistency across raters, as different raters may have arrived at different interpretations. Ambiguity in interpretation can increase measurement error, which reduces overall reliability and validity (Zhang, Tse, & Savalei, 2019). Ambiguity may be further amplified given the absence of a fixed reference point and a neutral alternative option within the Likert scale of the BLRI. Responses range from +3 to -3; excluding ‘0’ as a neutral alternative, furthermore, there are no fixed references points such as ‘completely, exactly, or never’. The inclusion of a fixed reference point can prevent variations in responses as reference points can minimise doubts about the position of the additional responses for raters. Also, reference points can encourage comparability across different respondents, and evidence finds the inclusion of fixed reference points have a positive impact on improving a measurements psychometric quality (Revilla, 2015). Regarding the BLRI, in which no clear reference points are included, there is an element of doubt introduced concerning the interpretation of the Likert labels. This could partly explain the poor inter-rater reliability, as raters did not have a clear reference point to judge the degree of the presence of the therapeutic core conditions. Additionally, the BLRI offers no neutral alternative response, which may have forced the raters to select an option that does not reflect their true opinion.

5.3.2 Scale Revisions

The BLRI Obs-40 has merit for use with young people, and its psychometric quality in terms of internal reliability, test-rest reliability, convergent validity, and structural validity is promising. However, the scale’s inter-rater reliability is poor, and the Unconditionality of
Regard subscale problematic. Therefore, as an observer rated measure of the person-centred relationship conditions, the scale may require further revision and adaptation for use with young people.

One adaption concerns the response scale of the BLRI Obs-40. Evidence suggests Likert response scales, as seen in the BLRI, can be vulnerable to response style behaviours causing non-random response error (Moors, Kieruj, & Vermunt, 2014). Whereas, item specific scales, such as that seen in the PCEPS-YP, reduces this bias and can provide higher measurement quality (DeCastellarnau, 2018). Furthermore, given the response labels in the BLRI are minimal and non-conceptual, there is a higher risk of acquiescence bias (Billiet & McClendon, 2000). Evidence shows that fully labelled scales have a positive impact on psychometric reliability (DeCastellarnau, 2018). In addition, DeCastellarnau (2018) has suggested an expanded response format is superior to the traditional Likert format. Expanded format responses can reduce ambiguity and confusion with interpretation and negatively worded items (Zhang et al., 2019). Revising the response format may be particularly useful for the unconditionality of Regard subscale, as poor interpretation of these negatively worded items in conjunction with the ambiguous Likert responses may have required a greater degree of interpretation and cognitive effort. Therefore, revising the response format of the BLRI Obs-40 from a likert-type response to an expanded format response with conceptual labels may reduce the level of item interpretation required when scoring, which as a consequence may improve scale reliability.

In addition, the length of the scale can impact response quality (Wetzel & Greiff, 2018); therefore shortening the BLRI Obs-40 may reduce rater fatigue and strengthen the scale. However, adapting the scale length would need to take into account the length of the subscales. Removing items and shortening the subscales to less than ten items, risks reducing internal reliability (DeVellis, 2016). Furthermore, there is a benefit to including the four
dimensions, though not formally evaluated in this project, the items that make up each of the four subscales do suggest good face validity (Barrett-Lennard, 2015).

Throughout the assessment of the BLRI Obs-40, the Unconditionality of Regard scale has shown poorer reliability and validity. The split of the unconditional positive regard core condition into the Level of Regard and Unconditionality of Regard subscales had previously been justified by findings of inter-scale correlations, which demonstrate an independence of unconditionality from the other dimensions of the scale. However, psychometrically, its continued inclusion in the BLRI Obs-40 is questionable given that the internal consistency of the scale would improve if the subscale of Unconditionality of Regard were removed. In addition, rater interpretations of the items of the Unconditionality of Regard subscale were more diverse and dissimilar than any other subscale items. Indicating the biggest source of ambiguity in the scale may originate from the Unconditionality of Regard items. There is the potential for the Unconditionality of Regard and Level of Regard subscales to be combined as seen in the Triad Participant Ratings Forms of the BLRI (Barrett-Lennard, 2015). These are a set of three scales, measuring Empathy, Congruence or Unconditional Positive Regard in a counselling training triad situation. Each of these scales is divided into three subscales, which are then rated by the speaker, listener and observer. In these scales, Unconditional Positive Regard is not divided into two subscales, and there is the possibility to triangulate ratings across the speaker, listener and observer (Appendix X). However, currently, there has been no psychometric evaluation of these scales, and therefore the potential psychometric benefits of measuring unconditional positive regard, as one construct, is unknown.
5.3.3 Observer Format

There needs to be a consideration for the use of an observer-rated scale for the measurement of the level and presence of the therapeutic relationship conditions.

Firstly, as the observer is positioned externally to the therapeutic relationship, there is a limit to the extent one can adequately rate the level and presence of the therapeutic core conditions.

Furthermore, not all items of the BLRI Obs-40 require the same level of observer judgment when scoring. For example, the items from the Level of Regard subscale indicate a rating judgment based on the verbal responses of the therapist, for example, item 37 ‘S/he responds very warmly to B, even with a kind of affection.’ Whereas ratings for the Unconditionality of Regard subscale require the observer to reflect and make a judgment on the internal thoughts or feelings of the therapist, such as in item 7 ‘Whether B is feeling happy or unhappy with him/herself doesn’t (or wouldn’t) affect A’s own feeling toward him/her.’ When the therapist does not verbalise their internal feelings, it can be very difficult for the observer to adequately rate the level of Unconditionality of Regard, and ratings for these items may be based on assumptions that cannot be justified by the therapists’ verbal responses. In addition, if the therapist has negative feelings about the relationship or is making a judgement on the client, it is unlikely that these will be communicated. There may be an element of social desirability bias or an attempt to preserve the relationship with the client that prevents the therapist from communicating any negative feelings. Therefore, the scale would be unable to rate moments of non-verbal incongruence.

Additionally, the Empathy subscale items rely on the observer to recognise the therapist’s emotional responses to the client. For example, item 2 ‘She/he senses or realizes how B is feeling.’ Again, when these are verbalised by the therapist, rating empathy can be
straightforward. Items from the Congruence subscale also depend on the extent to which the therapist communicates to the client, their perspective on the state of the therapeutic relationship. For example, item 16 ‘S/he keeps quiet about his/her own inner impressions and feelings with B.’ When this is discussed, items are more easily rated and the information is made available. However, when therapists do not discuss the relationship, raters can find it difficult to judge the level of congruence present and may have to assume its level.

There appears to be implicit criteria necessary for rating the different items and subscales of the BLRI Obs-40, and these criteria can be simple or more complex to meet. Furthermore, rating for many items is dependent on the level of verbal communication from the therapist. It is possible many of the difficulties observers face when rating could be remedied by using video and auditory recordings. In this project, recordings were audio based only. Therefore, there is no information regarding non-verbal communicated relationship conditions. Evidence supports non-verbal communication such as eye contact, body gesture to have a positive impact on the therapeutic alliance (Bobie & Jones, 2012), counselling relationship (Toriello & Strohmer, 2004) and even perceived therapist skill and competency (Jones, 2004). Non-verbal cues can offer the observer greater insights into the presence and level of the therapeutic core conditions, and items which do not solely rely on a therapist verbal response may be easier to rate, such as item 40 ‘There are things going on but unspoken in their communication that make the relationship somewhat unreal.’

5.4 Limitations

The study has methodological limitations; therefore, findings and interpretations obtained from the data are dependent on the study context. Additionally, there are limitations
with the analysis, which may be remedied with subsequent validation of the BLRI Obs-40, or with revisions to the scale.

5.4.1 Methodological Limitations

Limitations in the rating procedure are apparent; the rating procedure outlined in the methodology required raters to listen to the recording once before rating. The justification for this was to ensure a uniform rating procedure with all recordings listened to once only. However, given that recordings are 20 minutes long and ratings are based on a short extract of work, it is possible that with repeated listening raters may have become more sensitive to presence and level of the therapeutic relationship conditions.

In addition, the five-month time frame between ratings for the test-retest analysis may have influenced the results. There was a need to negotiate between reducing any effects of memory bias and preserving a near-identical rating environment. However, based on previous findings of the test-retest reliability of the BLRI it is possible five months was too long a period given that the average length of time between raters in previous work was four weeks (Gurman, 1977). It is possible test-rest reliability may have improved with a shorter interval between testing.

Furthermore, there are limitations with the rater training procedures. Training for raters was limited to a two-hour workshop, in which the project aims and BLRI were introduced. Raters were trained to use the BLRI Obs-40 and attempt practice ratings of sessions. However, given the limited timeframe, it is unlikely raters felt proficient and comfortable using the scale, and competency with the instrument likely developed while listening to rating segments for the project (on the job, so to speak). The minimal training provided may mean the inter-rater reliability was lower than if raters had more in-depth
training. Furthermore, there were limited opportunities for raters to provide feedback on the rating process or the use of the scale. Throughout the project, raters were offered to provide qualitative feedback on the scale. However, this was informal and not mandatory. Therefore, discussions regarding the scale or recordings were minimal and could account for the discrepancy in the interpretation of scale items across raters. Given that the observer rated form of the BLRI was used, a more detailed and extensive rater training would have been appropriate. For example, extending the two-hour workshop, providing additional literature of the BLRI, and scheduling formal feedback or supervisory appointments to discuss the rating process and scale. This is similar to the observer training used for the PCEPS, which involves a 12 hour workshop and fortnightly supervisions to discuss ratings and provide feedback (Freire et al., 2014). An observer training protocol similar to the guidance provided for the use of the PCEPS would have been beneficial to the project, given that supplementary observation training has been shown to significantly improve the accuracy and competence of observations (Thornton & Zorich, 1980).

A potential limitation of the study can also be found in the variation of raters. There are differences in the background, training and clinical experience of raters. This could potentially explain the disagreement between raters regarding the findings of the inter-rater reliability of the scale. Ideally, raters should have shared a similar training background and level of clinical experience to determine whether instability across ratings were as a result of the scale reliability or differences among raters. It is possible, testing with raters who share similar training and clinical experience can improve the inter-rater reliability. When testing reliability with the exclusion of Rater 3 (arguably the most dissimilar of the raters), inter-rater improved. Additionally, differences in the level of familiarity with the scale across the raters may have also contributed to the poor inter-rater reliability. The lead researcher (Rater 1) had used the BLRI previously to rate all 136 segments, whereas the remaining raters rated only
50 recordings. Furthermore, differences in the level of familiarity with the recordings may be problematic. Two of the three raters had previously heard a proportion of the recordings due to their involvement with the ETHOS trial before joining this study. Therefore there may have been an expectation or over-familiarity with some of the recordings used for the inter-rater reliability analysis.

Finally, there is a limitation to the generalizability of the findings. Due to the sample size used, findings are restricted to young people aged between 13 and 16 years old. Furthermore, findings cannot be generalised to clinical samples or clients presenting with severe psychological distress and high risk.

5.4.2 Analysis Limitations

The analysis followed the guidelines and recommendations as outlined by the COSMIN checklist; however, limitations in the analysis are apparent. The sample size for the test-retest analysis, inter-rater reliability analysis and convergent validity analysis were limited to 50 recordings, which reduced statistical power, in particular for the inter-rater correlations. A post hoc power analysis based on the inter-rater correlations results of 0.51, 0.05, and 0.20, with $\alpha=0.05$, and $n=50$, calculated power $\beta$ at 0.99, 0.35, and 0.92 respectively. This limited sample size was due to the practical and financial limitations of the PsychD, meaning that there was a limit of the number of raters that could be recruited, and therefore a limit on the number of recordings that could be rated. Furthermore, regarding the convergent validity of the BLRI Obs-40 with the PCEPS-YP, a priori hypotheses could have specified the magnitude in addition to the expected direction of correlations between the PCEPS-YP items and the BLRI subscales.
Finally, given the potential for subjectivity when using measurement scales (Robinson, 2014), it could have been useful to triangulate the observer ratings with client and therapist ratings via the BLRI 40 item client and therapist scales (Barrett-Lennard, 2015), to further investigate the convergent validity of the BLRI Obs-40.

5.5 Further Research

In the context of the project’s findings, future research should attempt to further validate the BLRI Obs-40 in a larger sample with video and audio recordings. This would be particularly useful in the investigation of test-retest reliability, inter-rater reliability, and convergent validity. Subsequent validation should use raters with similar levels of clinical experience and training backgrounds. In addition, a stringent rating and training protocol should be in place to ensure uniformity across different raters. The development of a training protocol to ensure observed assessments are reliable is paramount for the continued use of the observer rated BLRI form. Furthermore, further validation should triangulate ratings across the observer, client, and therapist rated BLRI Obs-40, in order to further assess the validity of the scale in this population.

In addition to further validating the BLRI Obs-40, subsequent research could begin to adapt the scale. In particular, adaptation may be required for items of the Unconditionality of Regard scale and could be adapted using the observer sections of the BLRI triad forms (Barrett-Lennard, 2015). In addition, conceptual labels and an expanded response format could be added to provide more explicit descriptions of items and minimise misinterpretation.

Finally, future research could use the BLRI Obs-40 in outcome studies with young populations, to further investigate whether the level and presence of the therapeutic relationship conditions affect the psychological outcomes of therapy. Outcomes studies with
the BLRI Obs-40 could then be used to establish the common factors of psychological therapy for young people.

5.6 Contributions to Counselling Psychology

The validated BLRI Obs-40 in an adolescent population has implications for both practice and research. In regards to practice, the BLRI Obs-40 can be used in supervision when listening to an extract of a supervisee practice. Together, the supervisor and supervisee can use the scale to gain an understanding of the extent the core conditions are present in practice. The scale could also be used in triad work for practitioners training to work with young people. However, in light of the findings, if raters have not received in-depth training, the BLRI Obs-40, ratings may be exploratory and not entirely accurate. However, ratings can be used to provide feedback and the scale as a tool to develop practitioners’ use of the core conditions.

The instrument can also contribute to future research. With observer training, the scale can be used in outcome studies and RCTs, which can build the evidence base for person-centred and humanistic therapy with young people. This can potentially contribute to the NICE guidelines; leading to the widening accessibly of talking therapies for children and adolescents.

5.7 Reflexivity

Engaging with this project and the PsychD as a whole has been a transformative endeavour. During my time on the project, I have developed a deeper appreciation and greater understanding of the person-centred theory and the therapeutic relationship. I was
taken by surprise when I found my experience with the project began to influence and shape my clinical practice. Over the past years, my clinical practice has developed and altered across three models of therapy, mainly, person-centred, psychodynamic and cognitive-behavioural approaches. Despite the changing of models, I have found the use of the core conditions in my practice to be a constant, which I attribute to this project. Time spent reading person-centred literature and listening to humanistic counselling sessions have, I feel, given me a greater understanding and respect for the core conditions when I deliver therapy and embark upon a therapeutic relationship. I feel my time on this project may have fused with my practice and shaped my way of working.

I found I encountered an epistemological shift when working on this project. I begin with a positivistic perspective with regards to my research topic and methods; however, I found this to gradually waiver the further I progressed in my training. As a counselling psychologist I subscribe to the Scientist-Practitioner Model (Blair, 2010), in which, ideally, research and practice are equally integrated and expressed (Vespia, Sauer, & Lyddon, 2006). However, I found engaging in a quantitative project caused a divide in this dual role as a therapist and a researcher. I felt torn between the empirical approach of my project and my identity as a counsellor. Although a positivist stance met my research needs, it did not satisfy my requirements as a practitioner. Therefore in an attempt to find a solution that would integrate the scientist and practitioner roles I was holding, I was led to pragmatism.

While participating in this project, I kept a reflexive journal, and similar to how my researcher role came into conflict with my therapist role, I noticed my role as a therapist has made itself present in my research. When listening to the extracts of therapy, I was aware of experiencing the core conditions towards the clients. I wonder then when rating if I was rating the therapists’ level of the core conditions or my own and if my therapeutic engagement with the recordings was somehow influencing the accuracy of the ratings. This
effect may have revealed itself in the results of the test-rest analysis, as empathy was found to be unstable over the five months. It is possible my empathy skills developed throughout the project, as my practitioner training continued in this time, and therefore I could have become more sensitive to either the empathy in the relationship in the extracts or my empathy with the clients I was listening to.

Overall, I feel my involvement with this project has given much to reflect on. I am grateful for the opportunity I have had to commit to a project I feel will contribute to improving psychological support for young people, something I passionately believe is necessary for every young person. I have enjoyed the collaboration with my supervisory team and fellow raters and am grateful for the increased confidence I have found in my research abilities and myself.
Chapter 6: Summary

The project aimed to evaluate the psychometric properties of the 40 item observer rated Barrett-Lennard Relationship Inventory in a population of young people receiving school-based humanistic counselling. After rating 136 audio-recorded sessions, the internal reliability, test-retest reliability, and inter-rater reliability were examined. In addition, structural validity, and convergent validity were assessed.

The BLRI Obs-40 was found to have satisfactory internal reliability and five-month test-retest reliability regarding total BLRI score and Empathy, Level of Regard, and Congruence subscales. However, finding from the subscale of Unconditionality of Regard were poor. Correlations between ratings from the BLRI Obs-40 and the PCEPS-YP were promising, indicating good convergent validity of the scale. In addition, the Confirmatory Factor Analysis findings support Barrett-Lennard’s original hypothesis of the therapeutic relationship conditions and support the structural validity of the scale. However, finding from both the pilot study inter-rater correlations and inter-rater reliability analysis indicate poor inter-rater reliability of BLRI Obs-40. This could be attributed to many factors, including variations in raters’ clinical experience, lack of formal observer training, and subjectivity of item interpretation.

Furthermore, the qualities of scale format may contribute to the poor quality of responses. For example, the use of an agree-disagree response can increase the cognitive effort required of raters, in addition to the scale’s length. Moreover, the lack of descriptive labels for both the scale items and the scale response can introduce ambiguity into the scale. Therefore, it may be necessary for the observer rated BLRI to be adapted for subsequent use.

The underlying motivation for this study was to validate a measurement scale in a young population, which would then begin to encourage further research and build the
evidence base for providing therapy to children and adolescents. The findings and work in this study can contribute to the further validation of the BLRI Obs-40 and the PCEPS-YP, which can then be used in outcome studies and RCTs of person-centred therapy. In addition, this work can contribute to the practice of humanistic therapy with young people, as practitioners can use the scale to monitor and self-assess their level and presence of the core conditions in their therapeutic relationships with clients.


British Association for Counselling and Psychotherapy. (2017). *Competences for work with British association for counselling and psychotherapy, competences and curricula: Children and young people*.


Comte, A. (1855). *The positive philosophy of auguste comte*.


Griffiths, G. (2013). Helpful and unhelpful factors in school-based counselling: Clients’ perspectives. Lutterworth: British Association for Counselling and Psychotherapy,


Hill, A., Roth, A., & Cooper, M. (2014). The competences required to deliver effective humanistic counselling for young people. Lutterworth: BACP,


Lister, J. (1970). A scale for the measurement of empathic understanding. Unpublished Manuscript, University of Florida,


McLeod, J. (2010). *Case study research in counselling and psychotherapy* Sage Publications.


from the national comorbidity survey Replication–Adolescent supplement (NCS-A).


psychological distress in young people: Outcomes and methodological reflections.

Counselling and Psychotherapy Research, 1-10.


International Journal of Market Research, 57(2), 225-238.


doi:https://doi.org/10.1016/j.jclinepi.2006.03.012


Time To Change. (2012). *Time to change children and young people’s programme development.*


Appendix A: Search Strategy for Systematic Review of Person-Centred Instruments in a Population of Young People

1. Research Question or Aim:

The aim of this systematic review is to investigate the tools used to measure the core conditions of person-centred therapy in a population of children and young people.

2. Inclusion and Exclusion Criteria

Population: Participants under 18 years
Intervention: Person-centred or humanistic therapy / Therapy using the core conditions (or equivalent)
Outcomes: Quantitative measures only
Additionally: English language
Original papers
Human studies

3. Search Methods

Electronic Databases:
• PubMed
• PsychINFO
• OvidSP
• Web of Science database

Additionally: Free text and citation search using Google Scholar

4. Terminology and Search Terms

MESH Search:
• Child
• Adolescent
• Person centred therapy
  AND
• Major Topic Search
• Person centred therapy

Free-Text Search, using a combination of terms:
• Child*
• Adolec*
• ‘Person Cent*’
• ‘Core Conditions’
• Empathy
• Congruence
• ‘Unconditional Positive Regard’
• Intervention*
• Therap*
• ‘Person Centr*’
5. **Review Method**
   - I. Initial **screening** via titles and abstract
   - II. Full text **review**
   - III. **Data extraction** into Excel worksheet – note reasons for exclusion or name and presence of core condition measure
   - IV. **Quality assessment** using COSMIN Criteria Checklist to be included in narrative synthesis
### Appendix B: Consensus-Based Standards for the Selection of Health Measurement Instruments (COSMIN) Criteria, as adapted from Mokkink et al., (2010)

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal Consistency</td>
<td>The extent of interrelatedness among the items (or subscales)</td>
</tr>
<tr>
<td>Reliability</td>
<td>Consistency of measurement over time</td>
</tr>
<tr>
<td>Measurement Error</td>
<td>The systematic and random error of the respondents score that is not attributed to true changes in the construct</td>
</tr>
<tr>
<td>Content Validity</td>
<td>The extent of the instrument accurately reflects the construct that is measured</td>
</tr>
<tr>
<td>Structural Validity</td>
<td>The ability of the scores to reflect the dimensionality of the instrument</td>
</tr>
<tr>
<td>Hypotheses Testing</td>
<td>The extent to which the instrument relates to other measures in an expected way</td>
</tr>
<tr>
<td>Cross-cultural Validity</td>
<td>The degree to which scores on culturally adapted scales adequately reflect those from the original scale</td>
</tr>
<tr>
<td>Criterion Validity</td>
<td>The extent to which the scores reflect a ‘gold standard’</td>
</tr>
<tr>
<td>Responsiveness</td>
<td>The ability to detect change over time</td>
</tr>
<tr>
<td>Interpretability</td>
<td>The extent to which a meaningful interpretation can be assigned to scores on the instrument</td>
</tr>
</tbody>
</table>
Appendix C: Basic Empathy Scale (Jolliffe & Farrington, 2006)

<table>
<thead>
<tr>
<th>Basic Empathy Scale (20 items)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rate each statement on a 5-point scale with 1=strongly agree and 5=strongly disagree.</td>
</tr>
</tbody>
</table>

1. My friends’ emotions don’t affect me much ________________________________________
2. After being with a friend who is sad about something, I usually feel sad _____________
3. I can understand my friend’s happiness when they do well at something ______________
4. I get frightened when I watch characters in a good scary movie _____________________
5. I get caught up in other people’s feelings easily _________________________________
6. I find it hard to know when my friends are frightened ___________________________
7. I don’t become sad when I see other people crying ______________________________
8. Other people’s feeling don’t bother me at all _____________________________________
9. When someone is feeling ‘down’ I can usually understand how they feel _____________
10. I can usually work out when my friends are scared _____________________________
11. I often become sad when watching sad things on TV or in films __________________
12. I can often understand how people are feeling even before they tell me____________
13. Seeing a person who has been angered has no effect on my feelings________________
14. I can usually work out when people are cheerful _______________________________
15. I tend to feel scared when I am with friends who are afraid_______________________
16. I can usually realize quickly when a friend is angry ____________________________
17. I often get swept up in my friends’ feelings____________________________________
18. My friend’s unhappiness doesn’t make me feel anything__________________________
19. I am not usually aware of my friends’ feelings__________________________________
20. I have trouble figuring out when my friends are happy__________________________
Appendix D: Parental Questionnaire (Green, 1996)

FIGURE 1. The questionnaire

1. What do you think was/were the main problem(s) that led to the referral to this clinic?

2. Since treatment here, have this (these) problem(s) changed?
   Much worse □  worse □  the same □  better □  much better □

3. Do you have any comments about how the clinic was managed? (e.g. appointments, decor, comfort, friendliness, efficiency, etc)

4. What kind of treatment did your child and/or other family members receive? (you may need to tick more than one box)
   Individual treatment for the referred child □
   Individual treatment for any other family member □
   Meetings with the whole family □
   Meetings with parents only □
   Medicine for the child □
   Advice with housing, benefits etc □
   Help in other ways (please specify) □

5. Was this the kind of help you expected before you came?
   Yes □  No □

6. Do you feel that your therapist(s) accurately understood the nature of the problems in the child or family?
   Completely □  Quite well □  To some extent □  Not at all □

7. Did the kind of help offered make sense to you as a treatment for the problems identified?
   Not at all □  To some degree □  Very much □  Completely □

8. Did you feel you understood the reasons for the kind of treatment offered?
   Not at all □  To some degree □  Very much □  Completely □

9. Quite apart from whether or not you thought the treatment was appropriate, do you think it was helpful?
   Completely □  Quite well □  To some extent □  Not at all □

10. Have you sought help elsewhere for the same problem since leaving the clinic?
    Yes □  No □  If so, where? ____________________________

11. Does anyone in the family have any other comments they would like to make on the clinic and/or the service received? Do you have any suggestions for improvement?
Appendix E: BLRI Obs-40 (Version 3) and Scoring Key (Barrett-Lennard, 1962)

Barrett-Lennard Relationship Inventory: Form Obs-40 (Version 3)
Developed by Godfrey T. Barrett-Lennard, PhD

Listed below are different ways that one person may feel and behave in relation to another. Please consider each numbered statement with reference to your present sense and perception of person A’s (__________) attitude, feeling and way of relating to person B (__________). First, note in the spaces here who “A” and “B” are.

Try to suspend any reaction of approval or disapproval you may have in giving your own best sense and estimate of how A is responding to B, as you focus on each item-aspect. The ‘right’ answer is whatever is truest to your personal sense and view from what you have observed and have to go on.

Mark each statement in the left margin, according to how strongly you feel that it is true, or not true, in this relationship. Please be sure to mark every one. Write in a minus number (−3, −2, or −1) when your answer is on the “no” side, and a plus number (+1, +2, or +3) when your answer is a “yes.” Here is the exact meaning of each answer number:

+3: YES, I strongly feel that it is true
+2: Yes, I feel it is true
+1: (Yes) I feel that it is probably true, or more true than untrue
−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. A is personally respecting of [write in name/identity of B, as above]
2. She/He senses or realizes how B is feeling.
3. Her/his interest in B depends on B’s communication and style of behavior.
4. She assumes a role or front with B.
5. She evidently feels a responsive personal warmth and liking for B.
6. She reacts to B’s words but does not see the way B feels inside.
7. Whether B is feeling happy or unhappy with him/herself doesn’t (or wouldn’t) affect A’s own feeling toward him/her.
8. She doesn’t avoid or shy away from anything that’s important in the relationship with B.
9. She is indifferent to B as a person.
10. She nearly always sees exactly what B means.
11. Depending on B’s behavior, s/he has a better (or worse) opinion of B sometimes or at some moments than s/he has at other times/moments.
12. My sense is that s/he is genuine and honest with B.
13. The way that A responds conveys a personal appreciation and valuing of B.
14. A’s personal reaction to something about B gets in the way of her/his understanding.
15. A’s attitude and responsiveness (or lack of response) stays essentially the same no matter what feelings and self-qualities B expresses.
16. S/he keeps quiet about his/her own inner impressions and feelings with B.
17. S/he evidently finds B pretty dull and uninteresting.
18. S/he somehow grasps what B means even when B has difficulty in saying it.
19. S/he judges B's reactions and style, with definite preference for certain characteristics.
20. A is willing to say what is on her/his mind with B, including personal feelings and sense of how they are getting along.
21. S/he isn't going through motions; s/he cares about B.
22. S/he doesn't listen and pick up on what B actually thinks and feels.
23. S/he seems to like and respect B in some ways, but not to like other things about her/him.
24. S/he is personally straightforward and open in their relationship.
25. S/he disapproves of B--as s/he sees her/him.
26. S/he usually understands the whole of B's expression and meaning.
27. No matter whether B's feelings/reactions are 'good' or 'bad', healthy or unhealthy, her/his attention and response to B remains the same.
28. At some moments s/he is not comfortable, but they go on, outwardly ignoring it.
29. S/he is friendly and warm toward B.
30. S/he really doesn't understand B's experienced world and feelings.
31. Her/his reaction to B seems to range from acceptance to disapproval.
32. I sense that s/he is quite in touch with her/his own feelings and reactions with B; s/he is not covering up inside.
33. S/he just tolerates B, because that's part of her/his job or role.
34. S/he appreciates just how the things B is going through feel to her/him.
35. At some moments she warns to B and then at other times she seems cool or indifferent to what B expresses or feels.
36. Her/his outward response is different from my sense of the way she/he is feeling underneath.
37. S/he responds very warmly to B, even with a kind of affection.
38. Her/his response with B is so fixed and automatic that B does not get through to him/her.
39. Nothing B expresses alters A's basic attitude and felt response to her/him.
40. There are things going on but unspoken in their communication that make the relationship somewhat unreal.

Have you answered all the items, even ones that were not straightforward or where you had to go out on a limb a bit? Please double check and make sure there are no gaps. Add qualifying comments if you wish. (Godfrey T. Barrett-Lennard. Adapted 2007 from RI forms OS-40 & MO-40 (and a prior 64-item observer form)).
Barrett-Lennard Relationship Inventory: Scoring Sheet – 40-item OS, MO, and Obs–40 (adult) forms
Developed by Godfrey T. Barrett-Lennard, PhD

Name/code ........................................ Date answered ..................................................
Class of relationship .............................. Form: OS MO Obs40 (circle one) ...........................
Position of respondent ......................... Note if other RI data on same relationship ...........

Enter the answer for each item in its space, below, after reversing the sign (from + to −, or − to +) for the item numbers shown in bold. (The bold items are worded “negatively.”)

<table>
<thead>
<tr>
<th>Level of Regard (R)</th>
<th>Empathy (E)</th>
<th>Unconditionality (U)</th>
<th>Congruence (C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item No.</td>
<td>Answer</td>
<td>Item No.</td>
<td>Answer</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>Reverse sign</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>Reverse sign</td>
<td>7</td>
</tr>
<tr>
<td>9</td>
<td>Reverse sign</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>13</td>
<td>14</td>
<td>Reverse sign</td>
<td>15</td>
</tr>
<tr>
<td>17</td>
<td>Reverse sign</td>
<td>18</td>
<td>19</td>
</tr>
<tr>
<td>21</td>
<td>22</td>
<td>Reverse sign</td>
<td>23</td>
</tr>
<tr>
<td>25</td>
<td>Reverse sign</td>
<td>26</td>
<td>27</td>
</tr>
<tr>
<td>29</td>
<td>Reverse sign</td>
<td>30</td>
<td>Reverse sign</td>
</tr>
<tr>
<td>33</td>
<td>Reverse sign</td>
<td>34</td>
<td>35</td>
</tr>
<tr>
<td>37</td>
<td>38</td>
<td>Reverse sign</td>
<td>39</td>
</tr>
</tbody>
</table>

Add the 10 values in each answer column to obtain the score for that scale, which might be a positive or negative number. The possible range of scale scores is −30 (or −3 × 10) to +30 (+3 × 10). If avoidance of negative values is necessary, add a constant of +30 to each obtained scale score, to yield ‘converted scores’ with a possible range of 0 to 60.
### Appendix F: Truax Scales for Accurate Empathy, Non-Possessive Warmth and Genuineness (Truax & Carkhuff, 1967)

<table>
<thead>
<tr>
<th>Level 1</th>
<th>Genuineness</th>
<th>Accurate Empathy</th>
<th>Non-Possessive Warmth</th>
</tr>
</thead>
<tbody>
<tr>
<td>A considerable discrepancy exists between the counsellor’s overt responses and his/her actual feelings. Alternatively, congruent responses are negative and retaliatory or judgemental. The counsellor is guarded and defensive and attempts to conceal feelings.</td>
<td>The counsellor’s verbal and behavioural responses are irrelevant, subtract significantly from the feelings expressed and the content of the client’s material, and do not attend appropriately to the client’s expression. Responses include advice giving, arguing, and changing the subject, criticising or expressing opinions.</td>
<td>Verbal and non-verbal responses communicate overt disrespect, or negative regard, declaring verbally or non-verbally the other person’s feelings and experiences unworthy of consideration. The counsellor becomes the focus of evaluation, actively disapproves of behaviour.</td>
<td></td>
</tr>
</tbody>
</table>

| Level 2 | Incongruence exists between the counsellor’s behaviour and feelings, and self-disclosure identifies with rather than empathises. The counsellor withholds appropriate responses. Rather than being genuine and human, the counsellor responds from an artificial ‘professional’ role, altogether lacking in spontaneity. | The counsellor responds to at least part of the surface feelings of the other person, but his/her responses noticeably do not respond to the intensity of feelings expressed. The counsellor may respond from her/his frame of reference rather than to what the client expresses. | The counsellor communicates little respect for the feelings, potentials or experiences of the other person. He/she may ignore what the other says, respond in a casual, passive, clichéd or mechanical manner, and withhold from involvement. |
| Level 3 | The counsellor shows no incongruence between behaviour, statements and feelings. She/he sounds present as s/he listens, attends, reflects, or clarifies, i.e., the counsellor engages with the client but does not necessarily share own feelings or experience. | The counsellor offers open, relaxed attention and stays alongside the client in his/her frame of reference rather than becoming empathic to other people in the client’s story. The counsellor mirrors and names the obvious, surface current feelings and communicates understanding at the level of feeling the client expresses. | Both through general attitude, words and demeanour, the counsellor communicates a positive, genuine concern and respect for the other person’s feelings, ability to act constructively and express him/herself. The counsellor suspends his/her own judgement of the other and communicates an openness or willingness to enter into a relationship. |
| Level 4 | It is clear the counsellor is present and authentic and is willing to engage in the here and how relationship with the client. This may involve sharing appropriate personal feelings or sensations in response to the client. These may be positive or negative or ambivalent, but must be relevant to the client’s needs and shared in such a way that facilitates exploration. | The counsellor’s responses accurately yet tentatively identify implicit, underlying feelings somewhat beyond the expressions of the client and complement feelings with content that enable the client to go deeper in his/her own exploration. He/she may reflect the client’s non-verbal cues. This level of empathy arises out of a felt experience, an organismic sensing of the client’s world. | The counsellor demonstrates trust in the client’s actualising tendency i.e., stays engaged with an alongside the client’s process even when difficult. He/she is without agenda, maintaining an attitude of creative indifference to the client’s choice of direction, and stays comfortably in the client’s frame of reference. The counsellor’s pacing is sensitive. |
A high degree of self-awareness and counsellor presence is evident, creating good contact with the client. Good rapport and trust are not sacrificed in the interventions despite the possibility of discomfort. The counsellor displays a high degree of realness, vitality and transparency, which are acutely facilitative within the counselling relationship.

The counsellor’s presence and responses significantly add to the feelings and meaning expressed by the client. A quality of naturalness is present. The counsellor moves with all the complex nuances of the client’s changing and fluid process. It is as if there are no seams between them. The counsellor is fully in tune with the client and responses visibly touch or impact the client.

The counsellor is able to integrate honest, fresh expression with caring and sensitivity for the client. Strong personal responses such as boredom and irritation may be owned and expressed within the content of a sound relationship. It will be clear that they are shared in goodwill and have a helpful outcome. Deep respect for the clients as an equal human being is visible in a choice of phrase, tone of voice and body language.
Appendix G: Person-Centred and Experiential Psychotherapy Scale – Young Person Counselling Version (PCEPS-YP)

Client ID______ Session _____________
Rater_________ Segment_____________

Rate the items according to how well each activity occurred during the therapy segment you’ve just listened to. It is important to attend to your overall sense of the therapist’s level of skill. Try to avoid forming a ‘global impression’ of the therapist early on in the segment.

1. CLIENT FRAME OF REFERENCE:

How much do the therapist’s responses convey an understanding of the client’s frame of reference (i.e., their world view) and the ways in which the young person understands themselves within it?

Do the therapist’s responses convey an understanding of the client’s inner world, as immediately expressed by the client? Or, conversely, is the therapist only able to respond from their own frame of reference?

<table>
<thead>
<tr>
<th>1. No understanding:</th>
<th>Therapist’s responses convey no understanding of the client’s frame of reference; or, therapist adds meaning based completely on their own frame of reference.</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. Minimal understanding:</td>
<td>Therapist’s responses convey a poor understanding of the client’s frame of reference; or, therapist adds meaning partially based on their own frame of reference rather than the client’s.</td>
</tr>
<tr>
<td>3. Slight understanding:</td>
<td>Therapist’s responses begin to approach an adequate understanding of the client’s frame of reference but are consistently somewhat ‘off’.</td>
</tr>
<tr>
<td>4. Adequate understanding:</td>
<td>Therapist’s responses convey an adequate understanding of the client’s frame of reference.</td>
</tr>
<tr>
<td>5. Good understanding:</td>
<td>Therapist’s responses convey a good understanding of the client’s frame of reference.</td>
</tr>
<tr>
<td>6. Excellent understanding:</td>
<td>Therapists’ responses convey an accurate understanding of the client’s frame of reference and therapist adds no meaning from their own frame of reference.</td>
</tr>
</tbody>
</table>
2. TRACKING:

To what extent is the therapist following the client’s track?

Are the therapist’s responses closely following the client’s expressed thoughts, feelings and story? While following the client’s track, is the therapist able to check and responsively revise their perceptions of the client’s world view based on client feedback?

Conversely, are the therapist’s responses a diversion from the client’s own train of thoughts and feelings? Is the therapist inflexible in their perspective?

1. **No tracking**: Therapist responses divert client from their thoughts/feelings; therapist fails to adapt to client feedback.

2. **Minimal tracking**: Therapist is only occasionally on client’s track and lacks responsiveness to client feedback.

3. **Slight tracking**: Therapist responses are partially on client’s track and inconsistently adaptive to client feedback.

4. **Adequate tracking**: Therapist is imperfectly but adequately on client’s track and reasonably responsive to client feedback.

5. **Good tracking**: Therapist responses consistently follow the client’s track and are adapted to client feedback.

6. **Excellent tracking**: Therapist is sensitively and actively follows the client’s track, quickly responding and revising perceptions based on client feedback.
3. EMPATHIC RESONANCE:

How well is the therapist able to resonate with, and communicate their understanding of, the young person’s spoken and unspoken feelings and perceptions?

How accurate and consistent is the therapist’s understanding of the client’s inner world? Is the therapist able to tune into, and reflect back, the young person’s unspoken or non-verbal communication such as body language or tone of voice (when this is possible to observe) in addition to the client’s verbally expressed feelings and thoughts?

Conversely, to what extent does the therapist miss or dismiss the client’s feelings, or assume the client shares their feelings?

1. No resonance: Therapist misses or dismisses client feelings and perceptions; makes assumptions based on therapist’s own perceptions and is completely out of tune with the client.

2. Minimal resonance: Therapist is occasionally able to resonate with, and communicate understanding of, some of the client’s feelings and perceptions.

3. Slight resonance: Therapist resonance with client’s feelings and perceptions is inconsistent and based on therapist’s own feelings.

4. Adequate resonance: Therapist is generally, but not always, able to resonate with, and communicate accurate understanding of, client’s feelings and perceptions.

5. Good resonance: Therapist is consistently and accurately attuned to the client and clearly communicates their understanding of the client’s spoken and unspoken feelings.

6. Excellent resonance: Therapist is especially in tune with the client and capable of deeply sensing, and resonating with, the feelings that are both unspoken and spoken.
4. ACCEPTING PRESENCE:

Do the therapist's responses convey a fundamentally accepting attitude toward the young person?

• How well does the therapist’s attitude convey acceptance of the young person’s world view regardless of their behaviour, attitudes and beliefs?

• How well does the therapist’s way of being and tone of voice convey genuine acceptance to the young person?

• To what degree is the therapist able to hold a consistent welcoming and non-judgmental attitude?

1. **Explicit nonacceptance**: Therapist explicitly communicates disapproval or criticism of client’s experience/meaning/feelings.

2. **Implicit nonacceptance**: Therapist implicitly or indirectly communicates disapproval or criticism of client experience/meaning/feelings.

3. **Incongruent/inconsistent nonacceptance**: Therapist acceptance is inconsistent and slightly judgmental.

4. **Adequate acceptance**: Therapist demonstrates at least some degree of acceptance of the client’s experience.

5. **Good acceptance**: Therapist clearly conveys unconditional acceptance, even in face of the client’s challenging behaviours or thoughts.

6. **Excellent acceptance**: Therapist skilfully conveys clear, grounded acceptance of the client’s experience and does not demonstrate any kind of judgment towards client experiences or behaviours, even when these might be criticised by others.
5. GENUINENESS:

How well does the therapist respond in a way that genuinely and naturally conveys their moment to moment experiencing of the client?

- How much is the therapist able to relate to the young person without adopting a professional façade? Does the therapist sound artificial, overly professional, formal, stiff, pedantic or affected vs. genuine, idiosyncratic, natural or real?

- Is the therapist able to relate to the young person in a genuine person-to-person manner? Or, conversely, is the therapist patronising or parental in their responses?

- To what degree is the therapist able to skilfully express their congruent experience of the young person in a facilitative manner?

| 1 No genuineness: Therapist sounds completely fake, artificial or patronising and does not seem aware of their own experiencing of the client. |
| 2 Minimal genuineness: Therapist sounds somewhat wooden, stiff, formal or technical; unable to relate in a person-to-person manner with the client. |
| 3 Slight genuineness: Therapist sounds a little distant or affected and only occasionally aware of their own experiencing of the client; rarely able to connect to the client in a person-to-person manner. |
| 4 Adequate genuineness: Therapist generally sounds natural, unaffected and able to some degree to maintain a person-to-person stance; some congruence with occasional lapses. |
| 5 Good genuineness: Therapist sounds consistently natural or genuine, in touch with their experiencing of the client at a person-to-person level and expresses this in a facilitative manner. |
| 6 Excellent genuineness: Therapist sounds completely genuine, very real or idiosyncratically present, without any façade or pretence; comfortably, sensitively and appropriately conveys their experience of the client in a person-to-person manner. |
6. EMOTION FOCUS

How much does the therapist actively work to help the young person focus on their emotional experiences and meanings, both explicit and implicit?

Does the therapist facilitate the client to:

- focus their attention inwards in order to become more aware of their feelings?
- focus their attention on bodily sensations?
- reflect toward emotionally poignant content?
- intensify, heighten, evoke or deepen their emotions?

Does the therapist help by:

- making empathic conjectures about feelings that have not yet been expressed?
- enquiring about client feelings?

Lower scores reflect ignoring implicit or explicit emotions; staying with non-emotional content; focusing on or reflecting generalised emotional states ('feeling bad') or minimising emotional states (e.g., reflecting 'angry' as 'annoyed'); failing to recognise, or ignoring, the young person's attempt to verbalise a feeling.

<table>
<thead>
<tr>
<th>1. No emotion focus: Therapist consistently ignores emotions or responds instead in a highly intellectual manner while focusing entirely on non-emotional content. When the client expresses emotions, the therapist consistently deflects the client away from them.</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. Minimal emotion focus: Therapist seems to have a concept of emotion focus but doesn't implement adequately, consistently or well; therapist may generally stay with non-emotional content; sometimes deflects client way from their emotion; reflects only general emotional states ('bad') or minimises client emotion.</td>
</tr>
<tr>
<td>3. Slight emotion focus: Therapist often or repeatedly ignores or deflects client away from emotion; therapist only slightly or occasionally helps client to focus on emotion; while they sometimes respond in a way that points to client emotions, at times they fail to do so, or do so in an awkward manner.</td>
</tr>
<tr>
<td>4. Adequate emotion focus: Where appropriate, therapist generally encourages client focus on emotions (by either reflections or other responses), with only minor, temporary lapses or slight awkwardness.</td>
</tr>
<tr>
<td>5. Good emotion focus: Therapist does enough of this and does it skilfully, where appropriate, trying to help the client to evoke, deepen and express particular emotions.</td>
</tr>
<tr>
<td>6 Excellent emotion focus: Therapist does this consistently, skilfully, and even creatively where appropriate, offering the client powerful, evocative reflections or questions, while at the same time enabling the client to feel safe while doing so.</td>
</tr>
</tbody>
</table>
7. EMOTION SYMBOLISATION

How well does the therapist assist the young person to articulate their emotions and other experiences?

How skilful is the therapist in facilitating the young person to:

- find appropriate words to describe their emotions, especially those that seem difficult to access?
- verbalise the concerns, meanings and memories which emerge out of emotional arousal?
- identify and verbalise the wishes, needs, behaviours and goals associated with feelings and emotions?

Is the therapist able to offer imagery and metaphor to help the young person accurately articulate the meaning of their experiences?

1. No emotion symbolisation: Therapist imposes own language on the client, anticipates and assumes client’s meaning and leaves things unexpressed; does not attempt to help client symbolise experience.

2. Minimal emotion symbolisation: Therapist attempts minimally to facilitate the client’s symbolisation of their difficult-to-access feelings and experience but lacks patience and misses the underlying meanings and needs.

3. Slight emotion symbolisation: Therapist is able to facilitate the client’s symbolisation of emotion and other experiences to some degree but is inconsistent and mostly unimaginative in their approach; has slight sensitivity to underlying meanings and needs.

4. Adequate emotion symbolisation: Therapist is generally able to facilitate client symbolisation of emotion and experiences in a patient manner.

5. Good emotion symbolisation: Therapist is skilful and imaginative in facilitating client emotions and experiences and communicates patience when feelings are difficult to symbolise; helps client to identify and express the needs and concerns associated with their emotions.

6. Excellent emotion focus: Therapist is especially sensitive to the client’s pace in symbolising emotions and other experiences; works closely and creatively with the client to fine-tune the expression of even difficult-to-access experience and emotion so that understanding matches symbolisation.
8. FACILITATION OF CLIENT SELF-DEVELOPMENT

How much does the therapist actively work to facilitate new client awareness, growth, perspectives and narratives?

Does the therapist:

• reflect toward, support, or symbolise emerging new client emotions or other experiences?

• facilitate the young person to translate new perspectives into alternative ways of understanding their experiences and actions?

• facilitate the young person to develop new narratives about themselves and their world?

Lower ratings are used when the therapist ignores new awareness, insight or shifts, or focuses on client despair or stickness.

1 No facilitation: Therapist consistently ignores new client awareness or emerging movement towards translating new perspectives into change; generally responds instead to client despair or stickness in recycling old narratives. When the client expresses new, emerging experiences, the therapist consistently deflects the client away from them.

2 Minimal facilitation: Therapist has the concept of client self-development facilitation but doesn’t implement it adequately, consistently or well; therapist generally stays with old or stuck content; or often deflects client away from new experience or changed narrative.

3 Slight facilitation: Therapist has a tendency to deflect client away from emerging new experiences, narratives and behaviours; while they sometimes respond in a way that points to client self-development, at times they fail to do so, or do so in an awkward manner.

4 Adequate facilitation: Where appropriate, therapist generally encourages focus on emerging client experiences or new narratives and actions (by either reflections or other responses), with only minor, temporary lapses or slight awkwardness.

5 Good facilitation: Therapist does enough of this and does it skilfully; where appropriate, trying to help the client to focus on emerging new experiences, narratives or actions, perhaps by offering client choices or implicitly or explicitly communicating trust in the client’s process.

6 Excellent facilitation: Therapist does this consistently, skilfully, and even creatively, where appropriate; for example, offering responses that accurately identify client hope in the midst of despair or implicitly convey trust in the client’s self-development potential.
9. DEVELOPMENTAL RESPONSIVENESS

How skilful is the therapist in adapting to the young person’s developmental capacity?

• Is the therapist able to communicate at a developmentally appropriate level while respecting the young person’s emotional, communicative and self-reflective capacities, as opposed to talking over the young person’s head or patronising them?

• Is the therapist able to employ a range of symbolic communication modes appropriate to the young person’s developmental level, e.g., drawing, play or other creative methods?

• Is the therapist able to understand and work with the client’s developmentally appropriate modes of expressing emotion even when these may be challenging? For example, does the therapist respond with empathy and acceptance when the young person expresses themselves in ways that might be considered unacceptable in another context (within limits of safety for both client and therapist).

Conversely, does the therapist inflexibly insist on adult ways of communicating and acting, or underestimates the client’s capacities and expect something too childish?

<table>
<thead>
<tr>
<th>1. No developmental responsiveness: Therapist does not adapt to the developmental capacity of the client in any way; offers no alternative methods of communication or symbolisation aside from talking; is unable to tolerate expression of client feelings outside of the ‘acceptable’ adult range.</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. Minimal developmental responsiveness: Therapist attempts to respond at the appropriate developmental level but is unable to do so adequately, consistently or well; sounds awkward, stilted, uncomfortable or patronising in adapting their language or way of working with the young person.</td>
</tr>
<tr>
<td>3. Slight developmental responsiveness: Therapist is somewhat able to respond to the young person’s developmental capacity but is slightly ‘off’ e.g. trying too hard or overestimating the young person’s developmental level.</td>
</tr>
<tr>
<td>4. Adequate developmental responsiveness: Therapist is mostly sensitive to, and able to respond appropriately to, the client’s developmental capacity through language and creative methods of symbolising experience; therapist shows some ability to empathise with and accept challenging, but developmentally appropriate, client actions and feelings.</td>
</tr>
<tr>
<td>5. Good developmental responsiveness: Therapist consistently matches client’s developmental capacity through language and creative methods of symbolising experience; they respond non-defensively and openly to challenging but developmentally appropriate client feelings and actions.</td>
</tr>
<tr>
<td>6. Excellent developmental responsiveness: Therapist’s responses are comfortably, consistently, and intuitively matched with the client’s developmental capacities; therapist shows an understanding of the meaning in the client’s developmentally appropriate feelings and actions, even when these are challenging or puzzling.</td>
</tr>
</tbody>
</table>
# Appendix H: Person-Centred and Experiential Psychotherapy Scale (PCEPS)

## PERSON–CENTRED & EXPERIENTIAL PSYCHOTHERAPY SCALE (v. 10.5, 01/03/11)

<table>
<thead>
<tr>
<th>Client ID:</th>
<th>Session:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rater:</td>
<td>Segment:</td>
</tr>
</tbody>
</table>

### Part 1: PERSON-CENTRED PROCESS Subscale

Rate the items according to how well each activity occurred during the therapy segment you’ve just listened to. It is important to attend to your overall sense of the therapist’s immediate experiencing of the client. Try to avoid forming a ‘global impression’ of the therapist early on in the session.

### PC1. CLIENT FRAME OF REFERENCE/TRACK:

How much do the therapist's responses convey an understanding of the client’s experiences as the client themselves understands or perceives it? To what extent is the therapist following the client's track?

- **Do the therapist’s responses convey an understanding of the client’s inner experience or point of view immediately expressed by the client? Or conversely, do therapist’s responses add meaning based on the therapist’s own frame of reference?**
- **Are the therapist’s responses right on client’s track? Conversely, are the therapist’s responses a diversion from the client’s own train of thoughts/feelings?**

1. **No tracking:** Therapist’s responses convey no understanding of the client’s frame of reference; or therapist adds meaning based completely on their own frame of reference.
2. **Minimal tracking:** Therapist’s responses convey a poor understanding of the client’s frame of reference; or therapist adds meaning partially based on their own frame of reference rather than the client’s.
3. **Slightly tracking:** Therapist’s responses come close but don’t quite reach an adequate understanding of the client’s frame of reference; therapist’s responses are slight “off” of the client’s frame of reference.
4. **Adequate tracking:** Therapist’s responses convey an adequate understanding of the client’s frame of reference.
5. **Good tracking:** Therapist’s responses convey a good understanding of the client’s frame of reference.
6. **Excellent tracking:** Therapist’s responses convey an accurate understanding of the client’s frame of reference and therapist adds no meaning from their own frame of reference.

### PC2. CORE MEANING:

How well do the therapist's responses reflect the core, or essence, of what the client is communicating or experiencing in the moment?

Responses are not just a reflection of surface content but show an understanding of the client’s central/core experience or meaning that is being communicated either implicitly or explicitly in the moment; responses do not take away from the core meaning of client’s communication.

1. **No core meaning:** Therapist’s responses address only the cognitive content or stay exclusively in the superficial narrative.
2. **Minimal core meaning:** Therapist’s responses address mainly the cognitive content or the superficial narrative but bring occasional glimpses into the underlying core feeling/experience/meaning.
3. **Slight core meaning:** Therapist’s responses partially but incompletely address the core meaning/feeling/experience that underlies the client’s expressed content.
4. **Adequate core meaning:** Therapist’s responses were close to the core meaning/feeling/experience that underlies the client’s expressed content, but do not quite reach it.
5. **Good core meaning:** Therapist’s responses accurately address the core meaning/feeling/experience that underlies the client’s expressed content.
6. **Excellent core meaning:** Therapist’s responses address with a high degree of accuracy the core meaning/feeling/experience that underlies the client’s expressed content.
PC3. CLIENT FLOW:

In terms of the pacing of the client's process, how well is the therapist responsively attuned to the client's flow moment by moment in the session?

_E.g.,_ therapist does not interrupt client's flow and allows reflective, self-exploratory silences; therapist does not respond too late, too seldom, or too early.

<table>
<thead>
<tr>
<th>No attunement: Therapist is not attuned at all with the client's pace; for example, therapist constantly interrupts clients or consistently leaves long, inappropriate silences, or therapist consistently misses opportunities to respond.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimal attunement: Therapist is not generally attuned with the client's pace, for example often rushing the client or responding somewhat late or early.</td>
</tr>
<tr>
<td>Slight attunement: Therapist is inconsistently attuned with the client's pace, sometimes responding a bit too early, or a bit too late, or not responding enough.</td>
</tr>
<tr>
<td>Adequate attunement: Therapist has adequate attunement with the client's pace, for example consistently allowing the client to finish their thoughts.</td>
</tr>
<tr>
<td>Good attunement: Therapist has a good attunement with the client's pace, for example allowing reflective, self-exploratory silences.</td>
</tr>
<tr>
<td>Excellent attunement: Therapist is has an excellent attunement with the client's pace, sensing the client's need for fast or slow pacing with a high degree of accuracy and grace.</td>
</tr>
</tbody>
</table>

PC4. WARMTH:

How well does the therapist's tone of voice convey appropriate warmth?

_How well does the therapist's tone of voice convey gentleness, caring, or receptiveness?_

<table>
<thead>
<tr>
<th>No warmth: Therapist is cold and aloof in their tone of voice and manner, conveys a sense of being closed or withholding from the client.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimal warmth: Therapist conveys a bit but not nearly enough warmth</td>
</tr>
<tr>
<td>Inappropriate/inconsistent warmth: Therapist conveys too much warmth/over-involvement (for example, offers inappropriate reassurance, praise or sympathy) or therapist conveys insufficient/inconsistent warmth.</td>
</tr>
<tr>
<td>Adequate warmth: Therapist conveys enough warmth and receptiveness.</td>
</tr>
<tr>
<td>Good appropriate warmth: Therapist conveys a good, facilitative level of warmth and receptiveness.</td>
</tr>
<tr>
<td>Excellent appropriate warmth: Therapist conveys facilitative warmth and excellent receptiveness.</td>
</tr>
</tbody>
</table>

PC5. CLARITY OF LANGUAGE:

How well does the therapist use language that communicates simply and clearly to the client?

_E.g.,_ therapist's responses are not too wordy, rambling, unnecessarily long; therapist does not use language that is too academic or too abstract; therapist's responses do not get in the client's way.

<table>
<thead>
<tr>
<th>No clarity: Therapist's responses are long-winded, tangled, and confusing.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimal clarity: Therapist's responses are wordy, rambling or unfocused.</td>
</tr>
<tr>
<td>Slight clarity: Therapist's responses are somewhat clear, but a bit too abstract or long.</td>
</tr>
<tr>
<td>Adequate clarity: Therapist's responses are clear but a bit too long.</td>
</tr>
<tr>
<td>Good clarity: Therapist's responses are clear and concise.</td>
</tr>
<tr>
<td>Excellent clarity: Therapist's responses are very clear and concise, even elegantly capturing subtle client experiences in a few choice words.</td>
</tr>
</tbody>
</table>
**PC6. CONTENT DIRECTIVENESS:**

How much do the therapist’s responses intend to direct the client’s content?

*Do the therapists’ responses introduce explicit new content?* e.g., *do the therapist’s responses convey explanation, interpretation, guidance, teaching, advice, reassurance or confrontation?*

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>&quot;Expert&quot; directiveness: Therapist overtly and consistently assumes the role of expert in directing the content of the session.</td>
</tr>
<tr>
<td>2</td>
<td>Overt directiveness: Therapist’s responses direct client overtly towards a new content.</td>
</tr>
<tr>
<td>3</td>
<td>Slight directiveness: Therapist’s responses direct client clearly but tentatively towards a new content.</td>
</tr>
<tr>
<td>4</td>
<td>Adequate nondirectiveness: Therapist is generally nondirective of content, with only minor, temporary lapses or slight content direction.</td>
</tr>
<tr>
<td>5</td>
<td>Good nondirectiveness: Therapist consistently follows the client’s lead when responding to content.</td>
</tr>
<tr>
<td>6</td>
<td>Excellent nondirectiveness: Therapist clearly and consistently follows the client’s lead when responding to content in a natural, inviting and unforced manner, with a high level of skill.</td>
</tr>
</tbody>
</table>

**PC7. ACCEPTING PRESENCE:**

How well do the therapist’s attitudes convey an unconditional acceptance of whatever the client brings?

*Does the therapist’s responses convey a grounded, centred, and acceptant presence?*

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Explicit nonacceptance: Therapist explicitly communicates disapproval or criticism of client’s experience/meaning/feelings.</td>
</tr>
<tr>
<td>2</td>
<td>Implicit nonacceptance: Therapist implicitly or indirectly communicates disapproval or criticism of client experience/meaning/feelings.</td>
</tr>
<tr>
<td>3</td>
<td>Incongruent/inconsistent nonacceptance: Therapist conveys anxiety, worry or defensiveness instead of acceptance; or therapist is not consistent in the communication of acceptance.</td>
</tr>
<tr>
<td>4</td>
<td>Adequate acceptance: Therapist demonstrates calm and groundedness, with at least some degree of acceptance of the client’s experience.</td>
</tr>
<tr>
<td>5</td>
<td>Good acceptance: Therapist conveys clear, grounded acceptance of the client’s experience; therapist does not demonstrate any kind of judgment towards client’s experience/behaviour.</td>
</tr>
<tr>
<td>6</td>
<td>Excellent acceptance: Therapist skilfully conveys unconditional acceptance while being clearly grounded and centred in themselves, even in face of intense client vulnerability.</td>
</tr>
</tbody>
</table>

**PC8. GENUINENESS:**

How much does the therapist respond in a way that genuinely and naturally conveys their moment to moment experiencing of the client?

*E.g., How much does the therapist sound phony, artificial, or overly professional, formal, stiff, pedantic or affected vs. genuine, idiosyncratic, natural or real?*

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>No genuineness: Therapist sounds very fake or artificial.</td>
</tr>
<tr>
<td>2</td>
<td>Minimal genuineness: Therapist sounds somewhat wooden, stiff or technical.</td>
</tr>
<tr>
<td>3</td>
<td>Slight genuineness: Therapist sounds a bit distant or affected.</td>
</tr>
<tr>
<td>4</td>
<td>Adequate genuineness: Therapist sounds natural and unaffected.</td>
</tr>
<tr>
<td>5</td>
<td>Good genuineness: Therapist sounds very natural or genuine.</td>
</tr>
<tr>
<td>6</td>
<td>Excellent genuineness: Therapist sounds completely genuine, very real or idiosyncratically present, without any façade or pretence.</td>
</tr>
</tbody>
</table>
PC9. PSYCHOLOGICAL HOLDING:
How well does the therapist metaphorically hold the client when they are experiencing painful, scary, or overwhelming experiences, or when they are connecting with their vulnerabilities?

High scores refer to therapist maintaining a solid, emotional and empathetic connection even when the client is in pain or overwhelmed.
Low scores refer to situations in which the therapist avoids responding or acknowledging painful, frightening or overwhelming experiences of the client.

| No holding: | Therapist oblivious to client's need to be psychologically held; avoids responding, acknowledging or addressing client's experience/feelings. |
| Minimal holding: | Therapist seems to be aware of the client's need to be psychologically held but is anxious or insecure when responding to client and diverts or distracts client from their vulnerability. |
| Slight holding: | Therapist conveys a bit of psychological holding, but not enough and with some insecurity. |
| Adequate holding: | Therapist manages to hold sufficiently the client's experience. |
| Good holding: | Therapist calmly and solidly holds the client's experience. |
| Excellent holding: | Therapist securely holds client's experience with trust, groundedness and acceptance, even when the client is experiencing, for example, pain, fear or overwhelmedness |

PC10. DOMINANT OR OVERPOWERING PRESENCE:
To what extent does the therapist project a sense of dominance or authority in the session with the client?

Low scores refer to situations in which the therapist is taking charge of the process of the session; acts in a self-indulgent manner or takes over attention or focus for themselves; interrupting, talking over, silence or controlling the process; or acting in a definite, lecturing, or expert manner.
High scores refer to situations in which the therapist offers the client choice or autonomy in the session; allows the client space to develop their own experience, waits for the client to finish their thoughts, is patient with the client, or encourages client empowerment in the session.

| Overpowering presence: | Therapist overpowers the client by strongly dominating the interaction, controlling what the client talks about or does in the session; clearly making themselves the centre of attention; or being patronizing toward the client. |
| Controlling presence: | Therapist clearly controls the client's process of the session, acting in an expert, or dominant manner. |
| Subtle control: | Therapist subtly, implicitly or indirectly controls what and how the client is in the session. |
| Noncontrolling presence: | Therapist generally respects client autonomy in the session; therapist does not try to control client's process. |
| Respectful presence: | Therapist consistently respects client autonomy in the session. |
| Empowering presence: | Therapist clearly and consistently promotes or validates the client's freedom or choice, allowing client space as they desire. |
Part 2: EXPERIENTIAL PROCESS Subscale

E1. Collaboration: How much does the therapist appropriately and skillfully work to facilitate client-therapist collaboration and mutual involvement in the goals and tasks of therapy?

How much does the therapist communicate a sense of working together or companionship in a shared effort? For example, this may include trying to locate and maintain a therapeutic focus with the client, while adapting this focus as necessary; informing the client about how therapy works or why it is important to explore feelings; also clarification and negotiation of client’s primary therapeutic goals or tasks; use of meta-communication or reference to therapeutic partnership (“we”).

1. No collaboration: therapist is authoritarian or unilateral.
2. Minimal collaboration: therapist seems to have a concept of collaboration but doesn’t implement adequately, consistently or well; therapist is either slightly controlling or generally fails to offer collaboration where appropriate.
3. Slightly collaborative: therapist is not authoritarian; however, they often act in a collaborative manner, but at times fail to do so, or do so in an awkward manner.
4. Adequate collaboration: therapist generally collaborative, with only minor, temporary lapses or slight awkwardness.
5. Good collaboration: therapist does enough of this and does it skillfully.
6. Excellent collaboration: therapist does this consistently, skillfully, and even creatively.

E2. Experiential Specificity: How much does the therapist appropriately and skillfully work to help the client focus on, elaborate or differentiate specific, idiosyncratic or personal experiences or memories, as opposed to abstractions or generalities?

E.g., By reflecting specific client experiences using crisp, precise, differentiated and appropriately empathic reflections; or asking for examples or to specify feelings, meanings, memories or other personal experiences.

1. No specificity: therapist consistently responds in a highly abstract, vague or intellectual manner.
2. Minimal specificity: therapist seems to have a concept of specificity but doesn’t implement adequately, consistently or well; therapist is either somewhat vague or abstract or generally fails to encourage experiential specificity where appropriate.
3. Slight specificity: therapist is often or repeatedly vague or abstract; therapist only slightly or occasionally encourages experiential specificity; sometimes responds in a way that points to experiential specificity, at times they fail to do so, or do so in an awkward manner.
4. Adequate specificity: where appropriate, therapist generally encourages client experiential specificity, with only minor, temporary lapses or slight awkwardness.
5. Good specificity: therapist does enough of this and does it skillfully, where appropriate trying to help the client to elaborate and specify particular experiences.
6. Excellent specificity: therapist does this consistently, skillfully, and even creatively, where appropriate, offering the client crisp, precise reflections or questions.
E3. Emotion Focus: How much does the therapist actively work to help the client focus on and actively articulate their emotional experiences and meanings, both explicit and implicit?

E.g., By helping clients focus their attention inwards; by focusing the client’s attention on bodily sensations; by reflecting toward emotionally poignant content, by inquiring about client feelings, helping client intensify, heighten or deepen their emotions, by helping clients find ways of describing emotions; or by making empathetic conjectures about feelings that have not yet been expressed. Lower scores reflect ignoring implicit or explicit emotions; staying with non-emotional content; focusing on or reflecting generalized emotional states (“feeling bad”) or minimizing emotional states (e.g., reflecting “angry” as “annoyed”).

| No emotion focus | therapist consistently ignores emotions or responds instead in a highly intellectual manner while focusing entirely on non-emotional content. When the client expresses emotions, the therapist consistently deflects the client away from them. |
| Minimal emotion focus | therapist seems to have a concept of emotion focus but doesn’t implement adequately, consistently or well; therapist may generally stay with non-emotional content; sometimes deflects client away from their emotion; reflects only general emotional states (“bad”) or minimizes client emotion. |
| Slight emotion focus | therapist often or repeatedly ignores or deflects client away from emotion; therapist only slightly or occasionally helps client to focus on emotion; while they sometimes respond in a way that points to client emotions, at times they fail to do so, or do so in an awkward manner. |
| Adequate emotion focus | where appropriate, therapist generally encourages client focus on emotions (by either reflections or other responses), with only minor, temporary lapses or slight awkwardness. |
| Good emotion focus | therapist does enough of this and does it skilfully, where appropriate trying to help the client to evoke, deepen and express particular emotions. |
| Excellent emotion focus | therapist does this consistently, skilfully, and even creatively, where appropriate, offering the client powerful, evocative reflections or questions, while at the same time enabling the client to feel safe while doing so. |

E4. Client Self-development: How much does the therapist actively work to facilitate client new awareness, growth, self-determination or empowerment?

Does the therapist reflect toward, support, or symbolize emerging new aspects of client emotions or other experiences? E.g., This may include offering the client choices; by reflecting/emphasizing client agency, focusing on the client’s emerging sense of inner strength or emerging new client changes or insights ways of experiencing self or others; or accurately reflecting the longing for change in client despair without dismissing the client’s pain. Lower ratings are used when the therapist ignores new awareness, insight or shifts, or focuses on client despair or stuckness.

| No client self-development | therapist consistently ignores or deflections of new awareness, agency, or emerging changes, or generally responds instead to client despair or stuckness. When the client expresses new, emerging experiences, the therapist consistently deflects the client away from them. |
| Minimal client self-development | therapist has the concept of self-development focus but doesn’t implement it adequately, consistently or well; therapist generally stays with old or stuck content; or often deflects client away from new experience or agency. |
| Slight client self-development | therapist often or repeatedly deflects client away from emerging new experiences or agency, therapist only slightly facilitates client self-development; while they sometimes respond in a way that points to client self-development, at times they fail to do so, or do so in an awkward manner. |
| Adequate client self-development | where appropriate, therapist generally encourages focus on emerging client experiences or agency (by either reflections or other responses), with only minor, temporary lapses or slight awkwardness. |
| Good client self-development | therapist does enough of this and does it skilfully, where appropriate trying to help the client to focus on emerging new experiences or agency, perhaps by offering client choices or implicitly or explicitly communicating trust in the client’s process. |
| Excellent client self-development | therapist does this consistently, skilfully, and even creatively, where appropriate; for example, offering responses that accurately identify client hope in the midst of despair or implicitly convey trust in the client’s self-development potential. |
E5. Emotion Regulation Sensitivity: How much does the therapist actively work to help the client adjust and maintain their level of emotional arousal for productive self-exploration?

Client agency is central; this is not imposed by the therapist. There are three possible situations:

(a) If the client is overwhelmed by feelings and wants help in moderating them, does the therapist try to help the client to manage these emotions? E.g., by offering a calming and holding presence; by using containing imagery; or by helping the client self-soothe vs. allowing the client to continue to panic or feel overwhelmed or unsafe.

(b) If the client is out of touch with their feelings and wants help in accessing them, does the therapist try to help them appropriately increase emotional contact? E.g., by helping them review current concerns and focus on the most important or poignant; by helping them remember and explore memories of emotional experiences; by using vivid imagery or language to promote feelings vs. enhancing distance from emotions.

(c) If the client is at an optimal level of emotional arousal for exploration, does the therapist try to help them continue working at this level, rather than deepening or flattening their emotions?

<table>
<thead>
<tr>
<th></th>
<th>No facilitation: therapist consistently ignores issues of client emotional regulation, or generally works against client emotional regulation, i.e., allowing client to continue feel overwhelmed or distanced.</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Minimal facilitation: therapist seems to have a concept of facilitating client emotional regulation but doesn’t implement adequately, consistently or well; therapist either generally ignores the client’s desire to contain overwhelmed emotion or to approach distanced emotion; sometimes they misdirect the client out of a productive, optimal level of emotional arousal, into either stuck or overwhelmed emotion or emotional distance or avoidance.</td>
</tr>
<tr>
<td>3</td>
<td>Slight facilitation: therapist often or repeatedly ignores or deflects client away from their desired level of emotional regulation productive for self-exploration; therapist only slightly facilitates productive self-exploration. While they sometimes respond in a way that facilitates client productive emotional regulation, at times they fail to do so, or do so in an awkward manner.</td>
</tr>
<tr>
<td>4</td>
<td>Adequate facilitation: Where appropriate, therapist generally encourages client emotional regulation (e.g., by helping them approach difficult emotions or contain excessive emotional distress as desired by client), with only minor, temporary lapses or slight awkwardness.</td>
</tr>
<tr>
<td>5</td>
<td>Good facilitation: therapist does enough emotional regulation facilitation and does it skilfully and in accordance with client’s desires, where appropriate trying to help the client to maintain a productive level of emotional arousal.</td>
</tr>
<tr>
<td>6</td>
<td>Excellent facilitation: therapist does this consistently, skilfully, and even creatively, where desired, offering the client evocative or focusing responses to help the client approach difficult emotions when they are too distant and to contain overwhelming emotions, all within a safe, holding environment.</td>
</tr>
</tbody>
</table>

© 2011, Elizabeth Freire, Robert Elliott & Graham Westwell. (Permission is granted to reproduce this form for educational, training, or supervision purposes, on the condition that it is not changed or sold).
Appendix I: ETHOS Eligibility Criteria for Participating Schools

Schools

Eligibility criteria included:

- A secondary school with young people between the ages of 13-16 years of age
- Able to commit for the duration of the trial (up to 2 years)
- Have no current or future plans for a counselling service for the duration of the trial
- English as a first language
- Understand that the decision for a young person to be suitable for the ETHOS trial rests with the research team
- Able to continue to provide support to young people not suitable for the trial, e.g. pastoral care/referral
- Able to provide a quiet and private room for counselling sessions
- Agree and consent to a ‘parental opt-in consent’ process for young people involved in the trial
- Able to timetable ETHOS counsellors to see young people on a weekly basis
- Able to timetable meeting between young people and the ETHOS research team for assessment and follow up procedures
- Understand limitations to counselling provided as ETHOS is primarily a research project
- Understand the randomisation of young people to immediate counselling or immediate pastoral care followed by counselling in 6-9 months
- Agree to allow counselling sessions being audio recorded and understand that only appropriate ETHOS staff will have access to these recordings
• Able to provide data on attendance, exclusions, academic performance and projections for all young people in the trial

• Able to fully begin participation in September 2016 or January 2017

• Be located within an appropriate and convenient location to allow access for the ETHOS team
Appendix J: ETHOS Eligibility Criteria for Young People

Young People

Inclusion and exclusion:

- Aged 13-16 at the time of assessment
- Experiencing moderate to severe levels of psychological distress as assessed by the Strengths and Difficulties Questionnaire Emotional Symptoms (SDQ-ES) Scale
- English reading age of 13
- Desire to participate in counselling
- Not currently receiving therapeutic intervention
- Have a school attendance record of at least 85% as assessed by the school

Exclusion criteria included:

- Unable to provide informed consent
- Parent/carer of the young person has not provided their informed consent
- At risk of harm to self or others at the time of assessment
- Plans to leave school within the academic year
- Unwilling to complete all required assessments
- Unwilling to allow sessions to be audio recording
Appendix K: ETHOS Eligibility Criteria for Counsellors

**Counsellors**

Counsellors held the following:

- At a minimum, a professional, diploma-level training in humanistic, person-centred or humanistic-integrative counselling
- A minimum of 450 hours of counselling or psychotherapy practice covered by at least 1.5 hours of supervision per month
- Up to date DBS certificate
- Membership to the BACP
- Abidance to the BACP’s Ethical Framework for the Counselling Professions (2015)
- Full insurance to practice
Appendix L: ETHOS Adherence Auditing Procedures

Counsellor rating

- The function of obtaining data at the counsellor level is to ensure that the scientific integrity of the delivery of SBHC in the ETHOS trial – that is, that it adheres to and is consistent with the theoretical model of SBHC. The aim of the ratings is both to determine the level of adherence achieved, and also to feed back into the system to ensure that the delivery of the intervention is at the required level.

- Counsellor adherence will be rated using the full Person-Centred and Experiential Psychotherapy Scale (Freire et al., 2014) adapted for work with young people (PCEPS-YP) (Appendix M17).

- Ratings will be based on session segments of 20 minutes.

- ‘Calibration tapes’ will first be developed, which will consist of six PCEPS-YP ratings by independent Raters with expertise in humanistic counselling for children and young people as well as experience with the PCEPS/PCEPS-YP. These will then be used as the basis for training of prospective Raters, and for assessing their rating accuracy.

- Prospective Raters will need to demonstrate an inter rater reliability (Cronbach’s alpha) of .7 against the calibration tapes before they are used to formally assess adherence for the ETHOS trial.

- Audio recordings are recorded on the Dropbox spreadsheet ‘ETHOS Session Recordings Logs’. Research Intern will have access via BACP.

- All recordings are saved on device ‘External encrypted hard drive 1’, VeraCRYPT protected (stored in ETHOS filing cabinet in CREST Clinic) (backup stored in Meg/Mick cabinet)
  - Recordings are stored by counsellor and device (CO...), and then by client
  - Research intern will set up encrypted folder on external hard drive, and come to Roehampton to obtain initial set of recordings, then on ad hoc basis

- For each counsellor, we will assess work from a minimum of four clients, spread across the duration of their involvement in the trial. We will aim to be representative across intakes first (Autumn 2016, Spring 2017, Summer 2017, Autumn 2017, Spring 2018), taking into account prospective intakes as well as completed and completing intakes. If a counsellor has been involved in:
  - 1 intake: 4 from one intake
  - 2 intakes: 2 from each intake
  - 3 intakes: 1 for each intake, plus one additional intake randomly selected (random.org, Random Integer Generator, Generate 1 random integer, value 1-3, 1 column)
  - 4 intakes: 1 from each intake
  - 5 intakes: 4 randomly selected from 5 (random.org, Random Integer Set Generator, Generate 1 set with 4 unique random integers, value 1-5)

- Auditing should only be done for clients who have completed their counselling

- Select client(s) within each selected intake using random.org, Random Integer Set Generator
  - If x clients in intake and y clients needing to be selected, Generate 1 set with y unique random integer, value 1-x. Output identifies client(s) in that
order, with clients organised in ascending numerical order of code. E.g., if number is 2, choose 2nd client in ascending order of code

- The 20 minute segments will be randomly selected from one session in the first half of their work with each selected client (excluding the first session), and one session from the second half of their work with that client (excluding the last session).
  - Research Intern should conduct a median split for each client, dividing countable sessions (i.e., attended, DNAs – check against Session Attendance spreadsheet) into ‘first half’ and ‘second half’ (do not take into account missing recordings at this point). Exclude first and last session where possible, and middle session, if odd number of total sessions, where possible
    - 1 session: use 1 session, choose two non-overlapping extracts from session
    - 2 sessions: choose both sessions
    - 3 sessions: choose session 2 and randomly select 1 or 3
    - 4 sessions: 2 and 3
    - 5 sessions: 2 and 4
    - 6 sessions: randomly choose 2 or 3, 4 or 5
    - 7 sessions: randomly choose 2 or 3, 5 or 6
    - 8 sessions: randomly choose 2-4, 5-7
    - 9 sessions: randomly choose 2-4, 6-8
    - 10 sessions: randomly choose 2-5, 6-9
  - If selected recording is missing (either due to missed session or missed recording), repeat random selection process within that half again until recording is present

- The 20 minute segment will also be selected at random, but we will exclude the first 5 minutes and the final 5 minutes of the session, such that, if the session is 50 minutes in length, the starting time will be a randomly generated time point between 5 and 25. Generate 1 set with 1 unique integer, with value 5 to 25. Output gives starting time for recording, ending point is this time + 20.
  - For other session lengths, if length is x, value = 5 to (x-25). Assume session length is length of recording. If x < 30, but x >20, use final 20 mins recording. If x < 20 but > 10, use whole recording. If x < 10, do not use session and randomly select alternative session.

- 20 mins segmented will be created using the Audacity software programme.
- Each segment will be rated by two Raters working independently
- Research Intern will need to liaise with Raters re how many segments they will rate. Raters can be mixed across pairings, using matrix approach
- Research Intern will email recordings as password protected 7-zip files to rater, who should download them immediately and save to VeraCrypt protected storage. Files must be ‘double deleted’ from email programme. Passwords sent through phone/text
- Research intern is responsible for ensuring data protection procedures are followed.
- Raters should complete electronic PCEPS-YP form for client, save, and return with appropriately titled file (e.g., client code-date)
- Ratings should be returned from Raters to Intern within specified period (e.g., 2 weeks).
• Raters should delete audio files once ratings returned.
• Intern should transfer information to Excel Template Rating spreadsheet.
• The Project Manager should be the Intern of any instances in which a counsellor’s segment is scored an average of 4 or less. This will be reviewed with the Chief Investigator and further action may be taken to support the counsellor in achieving adherence to SBHC competences. This is likely to involve, in the first instance, informing the counsellor and supervisor; but may also include establishing provision for further training, general feedback to the counsellor cohort, or reviewing the suitability of the counsellor for the trial.
• At intervals over the duration of the trial, Raters will be provided with calibration tapes to rate, to re-assess reliability of their ratings
• Ratings will be collated during the trial and key points will be fed back to counsellors and supervisors. This will be as general learning points to enhance the adherence of practice.
Appendix M: Scores for practice ratings used in observer training, for Raters 1, 2, and 3

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Rater 1 Practice 1</th>
<th>Rater 1 Practice 2</th>
<th>Rater 2 Practice 1</th>
<th>Rater 2 Practice 2</th>
<th>Rater 3 Practice 1</th>
<th>Rater 3 Practice 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regard</td>
<td>20</td>
<td>4</td>
<td>23</td>
<td>-6</td>
<td>17</td>
<td>-6</td>
</tr>
<tr>
<td>Empathy</td>
<td>24</td>
<td>-18</td>
<td>24</td>
<td>-18</td>
<td>17</td>
<td>-12</td>
</tr>
<tr>
<td>Unconditionality</td>
<td>21</td>
<td>6</td>
<td>16</td>
<td>12</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Congruence</td>
<td>15</td>
<td>14</td>
<td>18</td>
<td>-10</td>
<td>14</td>
<td>7</td>
</tr>
<tr>
<td>BLRI Total</td>
<td>80</td>
<td>6</td>
<td>81</td>
<td>-22</td>
<td>50</td>
<td>-6</td>
</tr>
</tbody>
</table>
Appendix N: ETHOS Data Sharing Application Form

**ETHOS**
**Research Project Proposal Form**

Submitted by (Name & Organisation): .................................................................

Date submitted: .................................................................................................

Title of proposed research: ................................................................................

<table>
<thead>
<tr>
<th>Question</th>
<th>Yes / No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Have you discussed this research proposal with a member of the ETHOS Principal Research Team (PRT)?</td>
<td>Yes / No</td>
</tr>
<tr>
<td>If yes, who?</td>
<td></td>
</tr>
<tr>
<td>Do you plan to utilise existing ETHOS data?</td>
<td>Yes / No</td>
</tr>
<tr>
<td>Do you plan to contact ETHOS participants?</td>
<td>Yes / No</td>
</tr>
<tr>
<td>Will your research project require additional ethical approval?</td>
<td>Yes / No</td>
</tr>
<tr>
<td>Please briefly describe your research project (e.g. research question, design, analysis, timeline):</td>
<td></td>
</tr>
<tr>
<td>Is this research project a student project?</td>
<td>Yes / No</td>
</tr>
<tr>
<td>Will your research project require support from a member of the ETHOS team?</td>
<td>Yes / No</td>
</tr>
<tr>
<td>If yes, please describe briefly (e.g. responsibilities, time commitment)?</td>
<td></td>
</tr>
<tr>
<td>Please provide information of your plans for the safe and secure storage of any data you are given access to (e.g. will the data be stored on a secure network with access only to named individuals, will the data be password-protected etc.)?</td>
<td></td>
</tr>
<tr>
<td>Do you have funds secured to support these activities?</td>
<td>Yes / No / Not yet / N/A</td>
</tr>
<tr>
<td>Do you plan to publish the findings?</td>
<td>Yes / No</td>
</tr>
<tr>
<td>If yes, please give details of: proposed title, target journal, proposed author(s) (including members of the ETHOS PRT):</td>
<td></td>
</tr>
</tbody>
</table>
Appendix O: CREST Data Storage and Protection Procedures

CENTRE FOR RESEARCH IN SOCIAL AND PSYCHOLOGICAL TRANSFORMATION (CREST)
DEPARTMENT OF PSYCHOLOGY

DATA STORAGE AND PROTECTION PROCEDURES

Sources
These procedures are informed by, and consistent with, the following sources:

• Roehampton University Data Protection Policy, University of Roehampton, May 2010 (revised).
• Ethical Guidelines for Researching Counselling and Psychotherapy, British Association of Counselling and Psychotherapy, 2004.
• Encrypting Confidential Data using Windows XP
• Ethical Principles for Conducting Research with Human Participants, British Psychological Society (accessed Sept. 2008).
• Personal communications with Ralph Weedon, Data Protection Officer, University of Strathclyde

Responsibilities

• The Chief Investigator has overall responsibility to ensure that the appropriate data storage and protection guidelines are followed.

Non-anonymised/personal data

• Non-anonymised (or ‘personal’) data refers to any form of documentation or media – electronic or otherwise – in which an individual is identifiable. This includes, but is not limited to:
  • signed consent forms
  • client identity forms (including DOB, GP details, gender etc)
  • video recordings

Note: even if no name or other obvious data is involved that would identify an individual, data such as date of birth, student matriculation number, national insurance number can be ‘triangulated’, perhaps with other data a third party has acquired, in such a way as to effectively identify someone. Anything that can be used in this way is therefore to be considered personal data.

• Collection of non-anonymised data will be kept to a minimum, and will only be obtained where it is ethically necessary (as in the case of signed consent forms), or where it clearly adds to the scientific value of a project (for instance, the video recording of counselling sessions).
• Non-anonymised data will be kept for ten years.
• All non-anonymised data will be clearly labelled with a date at which it should be destroyed.
• Non-anonymised data will be destroyed in a way which ensures that the data cannot be recovered in any way.
• Non-anonymised data will be kept physically and/or electronically separate from related anonymised data so that links can not be made between the two sets of data.
Non-electronic personal data, such as tape recordings and signed consent forms, should be kept in a locked and secure location at all times, and, wherever possible, at the University of Roehampton.

Electronic personal data will be encrypted and should always be kept on a password protected storage device: wherever possible a PC or network drive located at the University of Roehampton.

Personal data should not be kept on – or transferred to – laptops, USB sticks, CDs or other mobile/ portable devices unless absolutely necessary. As soon as such data is transferred to a secure University location, it must be removed from the portable device such that it cannot be recovered in any way.

*Should it be necessary to transfer personal data from person to person, this should be done in a secure manner (i.e., by hand or by recorded delivery), always separate from any anonymised data. Any posted materials should be marked ‘private and confidential’ and sent recorded delivery.*

For the duration of a study, non-anonymised data may, if absolutely necessary, be stored (in the manner identified above) by investigators other than the Chief Investigator (for instance, where a student is analysing video tapes of counselling sessions). However, on completion of the write-up of the research, all non-anonymised data will be returned to the Chief Investigator for storage, and any copies destroyed.

**Anonymised data**

Anonymised data refers to any form of documentation or media – electronic or otherwise – in which an individual is in no way identifiable. This includes, but is not limited to:

- SPSS spreadsheets in which identifying characteristics (such as age) are not recorded
- completed questionnaires: qualitative or quantitative

Anonymised data may be kept for an unlimited period, and may be used for subsequent research projects and data analyses at the discretion of the Chief Investigator (provided that this is made explicit to participants in consent forms).

Non-electronic anonymised data will be kept in a locked and secure location at all times, ideally at the University of Roehampton.

Electronic anonymised data may be stored electronically. This should always be to the highest possible standard of confidentiality: for instance, storage in an encrypted folder. It may also be kept on a password protected storage device, ideally at the University of Roehampton and, wherever possible, will be encrypted. Transfer and storage on portable/mobile devices (such as USB pens) should be kept to a minimum.

Transfer of anonymised data should be conducted to the highest standards of confidentiality, always separate from any non-anonymised data. Any posted materials should be marked ‘private and confidential.’ If anonymised data is transferred via email, it should be transferred by the receiver to an encrypted portion of a hard disk as soon as possible, and both sender and receiver should hard delete the email/ attachments from their email server.

For the duration of a study, anonymised data may be stored (in the manner identified above) by investigators other than the Chief Investigator. However,
on completion of the write-up of the research, all anonymised data will be returned to the Chief Investigator for storage, and any copies destroyed.

**Partially anonymised data (also known as Pseudo-anonymised data)**

- This section refers to any form of documentation or media – electronic or otherwise – in which it is highly unlikely that research participants can be identified, but in which the possibility of triangulation exists. This may include, but is not limited to:
  - audio recordings
  
  Note, if such media includes clearly identifying content (for instance, an interviewee reveals their name or that of their husband on an audio recording), then it will be treated as non-anonymised data until those identifying characteristics are removed.

- Wherever possible, partially anonymised (and non-anonymised) data should be scrutinised and all identifying details should be deleted/erased (for instance, identifying features on transcripts, such as names of partners, should be deleted or blacked out).

- Where all identifying details of partially anonymised data have been deleted/erased, this data will be treated as anonymised data, and subjected to the same procedures as above.

- In instances where partially anonymised data can not be fully anonymised (for instance, audio recordings in which the participant may be identifiable from their voice), this data will be kept for ten years, and will be stored according to the protocols for non-anonymised data.

- Within this ten year period, partially anonymised data may be used for subsequent research projects and data analyses at the discretion of the Chief Investigator (provided that this is made explicit to participants in consent forms).

  Data that is required to be electronically transferred for example, audio recordings from handheld electronic devices to a computer for encryption, will be done so through a cloud based server with a two stage authentication program accessible only by the researchers.

**The eight general principles of the data protection act, 1998**

- Personal data shall be processed fairly and lawfully (with specific requirements regarding sensitive personal data).

- Personal data shall be obtained only for one or more specified and lawful purposes, and shall not be further processed in any manner incompatible with that purpose or those purposes.

- Personal data shall be adequate, relevant and not excessive in relation to the purpose or purposes for which they are processed.

- Personal data shall be accurate and, where necessary, kept up to date.

- Personal data processed for any purpose or purposes shall not be kept for longer than is necessary for that purpose or those purposes.

- Personal data shall be processed in accordance with the rights of data subjects.

- Appropriate technical and organisational measures shall be taken against unauthorised or unlawful processing of personal data and against loss or destruction of, or damage to, personal data.
• Personal data shall not be transferred to a country or territory outside the European Economic Area, unless that country or territory ensures an adequate level of protection for the rights and freedoms of data subjects in relation to the processing of personal data.
### Appendix P: Parcel Allocations using Factor Loading

<table>
<thead>
<tr>
<th>Level of Regard Item (FL)</th>
<th>Empathy Item (FL)</th>
<th>Unconditionality Item (FL)</th>
<th>Congruence Item (FL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>E1</td>
<td>U1</td>
<td>C1</td>
</tr>
<tr>
<td>BL21</td>
<td>BL 22</td>
<td>BL 23</td>
<td>BL 12</td>
</tr>
<tr>
<td>(0.80)</td>
<td>(0.89)</td>
<td>(0.62)</td>
<td>(0.80)</td>
</tr>
<tr>
<td>BL5</td>
<td>BL 18</td>
<td>BL 35</td>
<td>BL 40</td>
</tr>
<tr>
<td>(0.72)</td>
<td>(0.86)</td>
<td>(0.55)</td>
<td>(0.73)</td>
</tr>
<tr>
<td>BL13</td>
<td>BL 34</td>
<td>BL 11</td>
<td>BL 24</td>
</tr>
<tr>
<td>(0.70)</td>
<td>(0.84)</td>
<td>(0.40)</td>
<td>(0.72)</td>
</tr>
<tr>
<td>R2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BL37</td>
<td>BL 30</td>
<td>BL 27</td>
<td>BL 8</td>
</tr>
<tr>
<td>(0.67)</td>
<td>(0.83)</td>
<td>(0.35)</td>
<td>(0.67)</td>
</tr>
<tr>
<td>BL33</td>
<td>BL 2</td>
<td>BL 15</td>
<td>BL 32</td>
</tr>
<tr>
<td>(0.65)</td>
<td>(0.82)</td>
<td>(0.34)</td>
<td>(0.62)</td>
</tr>
<tr>
<td>BL17</td>
<td>BL 26</td>
<td>BL 7</td>
<td>BL 28</td>
</tr>
<tr>
<td>(0.59)</td>
<td>(0.82)</td>
<td>(0.32)</td>
<td>(0.60)</td>
</tr>
<tr>
<td>R3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BL9</td>
<td>BL 6</td>
<td>BL 3</td>
<td>BL 20</td>
</tr>
<tr>
<td>(0.50)</td>
<td>(0.80)</td>
<td>(0.30)</td>
<td>(0.58)</td>
</tr>
<tr>
<td>BL1</td>
<td>BL 10</td>
<td>BL 19</td>
<td>BL 4</td>
</tr>
<tr>
<td>(0.46)</td>
<td>(0.79)</td>
<td>(0.27)</td>
<td>(0.52)</td>
</tr>
<tr>
<td>BL29</td>
<td>BL 38</td>
<td>BL 39</td>
<td>BL 36</td>
</tr>
<tr>
<td>(0.42)</td>
<td>(0.65)</td>
<td>(0.27)</td>
<td>(0.45)</td>
</tr>
<tr>
<td>BL25</td>
<td>BL 14</td>
<td>BL 31</td>
<td>BL 16</td>
</tr>
<tr>
<td>(0.31)</td>
<td>(0.39)</td>
<td>(0.23)</td>
<td>(0.31)</td>
</tr>
</tbody>
</table>

FL = Factor Loading
Appendix Q: Descriptive Statistics for Pilot Data

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Pilot Rater 1</th>
<th>Pilot Rater 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Regard</td>
<td>20.67</td>
<td>6.61</td>
</tr>
<tr>
<td>Empathy</td>
<td>14.20</td>
<td>14.64</td>
</tr>
<tr>
<td>Unconditionality</td>
<td>22.87</td>
<td>2.13</td>
</tr>
<tr>
<td>Congruence</td>
<td>12.60</td>
<td>9.49</td>
</tr>
<tr>
<td>BLRI Total</td>
<td>70.33</td>
<td>29.31</td>
</tr>
</tbody>
</table>
### Appendix R: Summary of Means and Standard Deviations of Sample 1: BLRI Items and Subscales

<table>
<thead>
<tr>
<th>Item</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.16</td>
<td>0.53</td>
</tr>
<tr>
<td>2</td>
<td>1.74</td>
<td>1.03</td>
</tr>
<tr>
<td>3</td>
<td>2.06</td>
<td>0.62</td>
</tr>
<tr>
<td>4</td>
<td>1.63</td>
<td>0.97</td>
</tr>
<tr>
<td>5</td>
<td>1.68</td>
<td>0.73</td>
</tr>
<tr>
<td>6</td>
<td>1.58</td>
<td>1.29</td>
</tr>
<tr>
<td>7</td>
<td>2.16</td>
<td>0.39</td>
</tr>
<tr>
<td>8</td>
<td>1.24</td>
<td>1.31</td>
</tr>
<tr>
<td>9</td>
<td>2.18</td>
<td>0.67</td>
</tr>
<tr>
<td>10</td>
<td>1.77</td>
<td>0.94</td>
</tr>
<tr>
<td>11</td>
<td>2.18</td>
<td>0.69</td>
</tr>
<tr>
<td>12</td>
<td>1.97</td>
<td>0.77</td>
</tr>
<tr>
<td>13</td>
<td>1.91</td>
<td>0.68</td>
</tr>
<tr>
<td>14</td>
<td>1.94</td>
<td>0.53</td>
</tr>
<tr>
<td>15</td>
<td>2.01</td>
<td>0.58</td>
</tr>
<tr>
<td>16</td>
<td>-0.85</td>
<td>1.46</td>
</tr>
<tr>
<td>17</td>
<td>2.01</td>
<td>0.67</td>
</tr>
<tr>
<td>18</td>
<td>1.63</td>
<td>1.02</td>
</tr>
<tr>
<td>19</td>
<td>2.29</td>
<td>0.49</td>
</tr>
<tr>
<td>20</td>
<td>0.57</td>
<td>1.57</td>
</tr>
<tr>
<td>21</td>
<td>1.78</td>
<td>0.89</td>
</tr>
<tr>
<td>22</td>
<td>1.70</td>
<td>1.31</td>
</tr>
<tr>
<td>23</td>
<td>2.24</td>
<td>0.43</td>
</tr>
<tr>
<td>24</td>
<td>1.52</td>
<td>1.02</td>
</tr>
<tr>
<td>25</td>
<td>2.71</td>
<td>0.45</td>
</tr>
<tr>
<td>26</td>
<td>1.75</td>
<td>0.85</td>
</tr>
<tr>
<td>27</td>
<td>2.24</td>
<td>0.76</td>
</tr>
<tr>
<td>28</td>
<td>1.35</td>
<td>1.21</td>
</tr>
<tr>
<td>29</td>
<td>2.01</td>
<td>0.77</td>
</tr>
<tr>
<td>30</td>
<td>1.72</td>
<td>1.09</td>
</tr>
<tr>
<td>31</td>
<td>2.57</td>
<td>0.51</td>
</tr>
<tr>
<td>32</td>
<td>1.71</td>
<td>0.68</td>
</tr>
<tr>
<td>33</td>
<td>2.05</td>
<td>0.74</td>
</tr>
<tr>
<td>34</td>
<td>1.75</td>
<td>1.18</td>
</tr>
<tr>
<td>35</td>
<td>2.22</td>
<td>0.57</td>
</tr>
<tr>
<td>36</td>
<td>1.35</td>
<td>0.52</td>
</tr>
<tr>
<td>37</td>
<td>1.93</td>
<td>0.70</td>
</tr>
<tr>
<td>38</td>
<td>2.07</td>
<td>0.80</td>
</tr>
<tr>
<td>39</td>
<td>2.11</td>
<td>0.36</td>
</tr>
<tr>
<td>40</td>
<td>1.65</td>
<td>1.00</td>
</tr>
<tr>
<td>R</td>
<td>20.44</td>
<td>4.46</td>
</tr>
<tr>
<td>E</td>
<td>17.65</td>
<td>8.20</td>
</tr>
<tr>
<td>U</td>
<td>22.09</td>
<td>2.56</td>
</tr>
<tr>
<td>C</td>
<td>12.63</td>
<td>6.91</td>
</tr>
<tr>
<td>BLRI Total</td>
<td>72.80</td>
<td>18.33</td>
</tr>
</tbody>
</table>
### Appendix S: Summary of Means and Standard Deviations of Sample 2: PCEPS-YP Items and BLRI Subscales

<table>
<thead>
<tr>
<th>Item/Subscale</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PCEPS-YP</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Client Frame Of Reference</td>
<td>4.24</td>
<td>0.80</td>
</tr>
<tr>
<td>Tracking</td>
<td>4.84</td>
<td>0.68</td>
</tr>
<tr>
<td>Empathic Resonance</td>
<td>4.30</td>
<td>0.79</td>
</tr>
<tr>
<td>Accepting Presence</td>
<td>5.02</td>
<td>0.69</td>
</tr>
<tr>
<td>Genuineness</td>
<td>4.62</td>
<td>0.81</td>
</tr>
<tr>
<td>Emotion Focused</td>
<td>3.88</td>
<td>1.04</td>
</tr>
<tr>
<td>Emotion Symbolisation</td>
<td>3.26</td>
<td>0.80</td>
</tr>
<tr>
<td>Facilitation Of Client Self-Development</td>
<td>3.76</td>
<td>0.82</td>
</tr>
<tr>
<td>Developmental Responsiveness</td>
<td>5.20</td>
<td>0.67</td>
</tr>
<tr>
<td><strong>PCEPS-YP Total</strong></td>
<td>39.12</td>
<td>5.17</td>
</tr>
<tr>
<td><strong>BLRI</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level Of Regard</td>
<td>19.90</td>
<td>4.07</td>
</tr>
<tr>
<td>Empathy</td>
<td>17.58</td>
<td>6.25</td>
</tr>
<tr>
<td>Unconditionality</td>
<td>21.86</td>
<td>1.76</td>
</tr>
<tr>
<td>Congruence</td>
<td>13.38</td>
<td>7.03</td>
</tr>
<tr>
<td><strong>BLRI Total</strong></td>
<td>72.72</td>
<td>15.34</td>
</tr>
<tr>
<td>Item</td>
<td>T1: 50 BLRI</td>
<td>T2: 50 BLRI</td>
</tr>
<tr>
<td>------</td>
<td>------------</td>
<td>------------</td>
</tr>
<tr>
<td>1</td>
<td>2.10</td>
<td>2.18</td>
</tr>
<tr>
<td>2</td>
<td>1.72</td>
<td>1.86</td>
</tr>
<tr>
<td>3</td>
<td>2.06</td>
<td>2.12</td>
</tr>
<tr>
<td>4</td>
<td>1.76</td>
<td>1.74</td>
</tr>
<tr>
<td>5</td>
<td>1.62</td>
<td>1.94</td>
</tr>
<tr>
<td>6</td>
<td>1.42</td>
<td>1.66</td>
</tr>
<tr>
<td>7</td>
<td>2.12</td>
<td>2.04</td>
</tr>
<tr>
<td>8</td>
<td>1.18</td>
<td>1.56</td>
</tr>
<tr>
<td>9</td>
<td>2.12</td>
<td>1.98</td>
</tr>
<tr>
<td>10</td>
<td>1.80</td>
<td>1.90</td>
</tr>
<tr>
<td>11</td>
<td>2.16</td>
<td>2.08</td>
</tr>
<tr>
<td>12</td>
<td>1.96</td>
<td>2.04</td>
</tr>
<tr>
<td>13</td>
<td>1.84</td>
<td>1.98</td>
</tr>
<tr>
<td>14</td>
<td>1.94</td>
<td>1.74</td>
</tr>
<tr>
<td>15</td>
<td>2.02</td>
<td>1.90</td>
</tr>
<tr>
<td>16</td>
<td>0.68</td>
<td>0.38</td>
</tr>
<tr>
<td>17</td>
<td>2.02</td>
<td>1.92</td>
</tr>
<tr>
<td>18</td>
<td>1.72</td>
<td>1.90</td>
</tr>
<tr>
<td>19</td>
<td>2.28</td>
<td>2.12</td>
</tr>
<tr>
<td>20</td>
<td>0.62</td>
<td>0.32</td>
</tr>
<tr>
<td>21</td>
<td>1.82</td>
<td>1.82</td>
</tr>
<tr>
<td>22</td>
<td>1.74</td>
<td>1.92</td>
</tr>
<tr>
<td>23</td>
<td>2.20</td>
<td>2.00</td>
</tr>
<tr>
<td>24</td>
<td>1.50</td>
<td>1.66</td>
</tr>
<tr>
<td>25</td>
<td>2.64</td>
<td>2.62</td>
</tr>
<tr>
<td>26</td>
<td>1.84</td>
<td>1.86</td>
</tr>
<tr>
<td>27</td>
<td>2.16</td>
<td>1.98</td>
</tr>
<tr>
<td>28</td>
<td>1.52</td>
<td>1.60</td>
</tr>
<tr>
<td>29</td>
<td>1.92</td>
<td>2.02</td>
</tr>
<tr>
<td>30</td>
<td>1.62</td>
<td>1.84</td>
</tr>
<tr>
<td>31</td>
<td>2.56</td>
<td>2.42</td>
</tr>
<tr>
<td>32</td>
<td>1.76</td>
<td>1.92</td>
</tr>
<tr>
<td>33</td>
<td>1.92</td>
<td>2.06</td>
</tr>
<tr>
<td>34</td>
<td>1.58</td>
<td>1.86</td>
</tr>
<tr>
<td>35</td>
<td>2.26</td>
<td>2.04</td>
</tr>
<tr>
<td>36</td>
<td>2.00</td>
<td>2.08</td>
</tr>
<tr>
<td>37</td>
<td>1.90</td>
<td>1.96</td>
</tr>
<tr>
<td>38</td>
<td>2.20</td>
<td>2.00</td>
</tr>
<tr>
<td>39</td>
<td>2.04</td>
<td>1.98</td>
</tr>
<tr>
<td>40</td>
<td>1.76</td>
<td>2.06</td>
</tr>
<tr>
<td>R</td>
<td>19.90</td>
<td>20.48</td>
</tr>
<tr>
<td>E</td>
<td>17.58</td>
<td>18.54</td>
</tr>
<tr>
<td>U</td>
<td>21.86</td>
<td>20.68</td>
</tr>
<tr>
<td>C</td>
<td>13.38</td>
<td>14.60</td>
</tr>
<tr>
<td>BLRI Total</td>
<td>72.72</td>
<td>74.30</td>
</tr>
</tbody>
</table>
Appendix U: Summary of Means, Standard Deviations and Cronbach’s Alpha of Sample 4: BLRI Totals and Subscales for Raters 1, 2, and 3

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Rater 1 (KB)</th>
<th>Rater 2 (CH)</th>
<th>Rater 3 (RT)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>α</td>
</tr>
<tr>
<td>Regard</td>
<td>19.90</td>
<td>4.07</td>
<td>0.86</td>
</tr>
<tr>
<td>Empathy</td>
<td>17.58</td>
<td>6.25</td>
<td>0.88</td>
</tr>
<tr>
<td>Unconditionality</td>
<td>21.86</td>
<td>1.76</td>
<td>0.29</td>
</tr>
<tr>
<td>Congruence</td>
<td>13.38</td>
<td>7.03</td>
<td>0.85</td>
</tr>
<tr>
<td>BLRI Total</td>
<td>72.72</td>
<td>15.35</td>
<td>0.91</td>
</tr>
</tbody>
</table>
Appendix V: Percentage of Absolute Agreement Across Raters Per Item

<table>
<thead>
<tr>
<th>Item</th>
<th>% (N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>62 (31)</td>
</tr>
<tr>
<td>2</td>
<td>36 (18)</td>
</tr>
<tr>
<td>3</td>
<td>14 (7)</td>
</tr>
<tr>
<td>4</td>
<td>0 (0)</td>
</tr>
<tr>
<td>5</td>
<td>40 (20)</td>
</tr>
<tr>
<td>6</td>
<td>32 (16)</td>
</tr>
<tr>
<td>7</td>
<td>0 (0)</td>
</tr>
<tr>
<td>8</td>
<td>0 (0)</td>
</tr>
<tr>
<td>9</td>
<td>34 (17)</td>
</tr>
<tr>
<td>10</td>
<td>42 (21)</td>
</tr>
<tr>
<td>11</td>
<td>52 (26)</td>
</tr>
<tr>
<td>12</td>
<td>40 (20)</td>
</tr>
<tr>
<td>13</td>
<td>42 (21)</td>
</tr>
<tr>
<td>14</td>
<td>56 (28)</td>
</tr>
<tr>
<td>15</td>
<td>0 (0)</td>
</tr>
<tr>
<td>16</td>
<td>2 (1)</td>
</tr>
<tr>
<td>17</td>
<td>28 (14)</td>
</tr>
<tr>
<td>18</td>
<td>44 (22)</td>
</tr>
<tr>
<td>19</td>
<td>36 (18)</td>
</tr>
<tr>
<td>20</td>
<td>12 (6)</td>
</tr>
<tr>
<td>21</td>
<td>4 (2)</td>
</tr>
<tr>
<td>22</td>
<td>26 (13)</td>
</tr>
<tr>
<td>23</td>
<td>58 (29)</td>
</tr>
<tr>
<td>24</td>
<td>20 (10)</td>
</tr>
<tr>
<td>25</td>
<td>24 (12)</td>
</tr>
<tr>
<td>26</td>
<td>50 (25)</td>
</tr>
<tr>
<td>27</td>
<td>0 (0)</td>
</tr>
<tr>
<td>28</td>
<td>32 (16)</td>
</tr>
<tr>
<td>29</td>
<td>46 (23)</td>
</tr>
<tr>
<td>30</td>
<td>42 (21)</td>
</tr>
<tr>
<td>31</td>
<td>4 (2)</td>
</tr>
<tr>
<td>32</td>
<td>48 (24)</td>
</tr>
<tr>
<td>33</td>
<td>20 (10)</td>
</tr>
<tr>
<td>34</td>
<td>48 (24)</td>
</tr>
<tr>
<td>35</td>
<td>36 (18)</td>
</tr>
<tr>
<td>36</td>
<td>52 (26)</td>
</tr>
<tr>
<td>37</td>
<td>4 (2)</td>
</tr>
<tr>
<td>38</td>
<td>46 (23)</td>
</tr>
<tr>
<td>29</td>
<td>3 (2)</td>
</tr>
<tr>
<td>40</td>
<td>36 (18)</td>
</tr>
</tbody>
</table>
## Appendix W: Percentage of Absolute Agreement Across Raters Per Recording

<table>
<thead>
<tr>
<th>Recording</th>
<th>% (N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10 (4)</td>
</tr>
<tr>
<td>2</td>
<td>20 (8)</td>
</tr>
<tr>
<td>3</td>
<td>25 (10)</td>
</tr>
<tr>
<td>4</td>
<td>15 (6)</td>
</tr>
<tr>
<td>5</td>
<td>17.5 (7)</td>
</tr>
<tr>
<td>6</td>
<td>5 (2)</td>
</tr>
<tr>
<td>7</td>
<td>22.5 (9)</td>
</tr>
<tr>
<td>8</td>
<td>7.5 (3)</td>
</tr>
<tr>
<td>9</td>
<td>32.5 (13)</td>
</tr>
<tr>
<td>10</td>
<td>30 (12)</td>
</tr>
<tr>
<td>11</td>
<td>52.5 (21)</td>
</tr>
<tr>
<td>12</td>
<td>27.5 (11)</td>
</tr>
<tr>
<td>13</td>
<td>17.5 (7)</td>
</tr>
<tr>
<td>14</td>
<td>52.5 (21)</td>
</tr>
<tr>
<td>15</td>
<td>20 (8)</td>
</tr>
<tr>
<td>16</td>
<td>5 (2)</td>
</tr>
<tr>
<td>17</td>
<td>20 (8)</td>
</tr>
<tr>
<td>18</td>
<td>37.5 (15)</td>
</tr>
<tr>
<td>19</td>
<td>37.5 (15)</td>
</tr>
<tr>
<td>20</td>
<td>22.5 (9)</td>
</tr>
<tr>
<td>21</td>
<td>12.5 (5)</td>
</tr>
<tr>
<td>22</td>
<td>22.5 (9)</td>
</tr>
<tr>
<td>23</td>
<td>0 (0)</td>
</tr>
<tr>
<td>24</td>
<td>25 (10)</td>
</tr>
<tr>
<td>25</td>
<td>32.5 (13)</td>
</tr>
<tr>
<td>26</td>
<td>57.5 (23)</td>
</tr>
<tr>
<td>27</td>
<td>47.5 (19)</td>
</tr>
<tr>
<td>28</td>
<td>42.5 (17)</td>
</tr>
<tr>
<td>29</td>
<td>27.5 (11)</td>
</tr>
<tr>
<td>30</td>
<td>42.5 (17)</td>
</tr>
<tr>
<td>31</td>
<td>37.5 (15)</td>
</tr>
<tr>
<td>32</td>
<td>20 (8)</td>
</tr>
<tr>
<td>33</td>
<td>22.5 (9)</td>
</tr>
<tr>
<td>34</td>
<td>42.5 (17)</td>
</tr>
<tr>
<td>35</td>
<td>47.5 (19)</td>
</tr>
<tr>
<td>36</td>
<td>35 (14)</td>
</tr>
<tr>
<td>37</td>
<td>25 (10)</td>
</tr>
<tr>
<td>38</td>
<td>40 (16)</td>
</tr>
<tr>
<td>39</td>
<td>55 (22)</td>
</tr>
<tr>
<td>40</td>
<td>0 (0)</td>
</tr>
<tr>
<td>41</td>
<td>45 (18)</td>
</tr>
<tr>
<td>42</td>
<td>45 (18)</td>
</tr>
<tr>
<td>43</td>
<td>30 (12)</td>
</tr>
<tr>
<td>44</td>
<td>25 (10)</td>
</tr>
<tr>
<td>45</td>
<td>42.5 (17)</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>-----</td>
</tr>
<tr>
<td>46</td>
<td>25 (10)</td>
</tr>
<tr>
<td>47</td>
<td>37.5 (15)</td>
</tr>
<tr>
<td>48</td>
<td>45 (18)</td>
</tr>
<tr>
<td>49</td>
<td>15 (6)</td>
</tr>
<tr>
<td>50</td>
<td>42.5 (17)</td>
</tr>
</tbody>
</table>
Appendix X: Triad Participant Record Form (Barrett-Lennard, 2015)

Triad Participant Record Form R-U
Developed by Godfrey T. Barrett-Lennard, PhD

Please fill in the next line first
Today’s date ...................................... Your first name ..............................................
Partners’ names ........................................ and ..............................................................

Below are ways that one person might feel and respond to another. Please consider each numbered statement, noting names in the first item of each section. Answer Part A immediately after your turn as listener, Part B right after being the speaker, and Part C after being the observer.

Mark each statement in the main answer column on the right, according to how strongly you feel that it is true, or not true, in this interaction. Please be sure to mark every one. Write in a positive number (+3, +2, or +1) for each ‘yes’ answer, and negative numbers (−1, −2, or −3) to stand for ‘no’ answers. Here is the meaning of each answer number:

+3: Yes (!), I strongly feel this is true.  
+2: Yes, I feel that this is true.  
+1: (Yes) I feel that this is probably more true than untrue.  
−3: No(!), I strongly feel that this is not true  
−2: No, I feel that this is not true.  
−1: (No) I feel that this is probably untrue, or more untrue than true.

<table>
<thead>
<tr>
<th><strong>A. How I found myself responding when I was the listener:</strong></th>
<th>LISTENER</th>
<th>Speaker</th>
<th>Observer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I felt a personal respect for _________ (name) as s/he spoke on in the session.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. I found myself evaluating or sizing up ________’s problem, and how s/he came across.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Whether _______ was self-critical or shared positive feelings about her/himself didn’t and wouldn’t affect my own feeling toward her/him</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. I felt an inner warmth and reaching out as _______ shared and expressed her/himself.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. I couldn’t help expressing some discomfort or disapproval in relation to _______.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Some things s/he expressed triggered feelings I have and was feeling in myself.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. I did feel some impatience with ______, and maybe a sense of ‘so what’.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. My own reaction to ______ held steady; I didn’t feel “turned off” toward her/his process at some moments, and more responsive to her (him) during other sharing.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. I wanted to help ______ get a better perspective or handle on her/his problem and develop a solution.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
B. **How I experienced the listener when I was the client/speaker:**

1. I felt respected by [name] as a feeling person with my own view.
2. I sensed that [name] was evaluating me or my world, and how I came across.
3. Whether I was expressing painful or troubling things about myself/my life, or more positive ones, did not seem to effect [name]'s attitude and response to me.
4. I felt a warmth and caring interest from [name], as I spoke about myself.
5. I couldn’t help feeling at moments that [name] was disapproving of me.
6. I felt that [name]'s own attitudes and feeling toward the issues I spoke about were affecting his/her response to me.
7. I sensed [name] was impatient with me, maybe uninterested or critical.
8. I sensed that I could say anything, or show different sides of myself, without it changing [name]'s attitude or feeling toward me.
9. [name] tried to help me to get a better perspective or handle on my problem and figure out a solution.

C. **My view of the listener’s (Ls) response, from the position of observer:**

1. L was visibly respecting of [name] as a person with his/her own view.
2. I sensed that L was appraising and trying to see through [name]'s style and messages.
3. L's attitude and responsiveness seemed to me to be the same, whether I was expressing conflicted negative feelings or positive things about self.
4. I sensed a warmth and caring from L as [name] spoke on about him/herself.
5. I felt that in some way L disapproved of [name].
6. I thought that L's response tended to come from his/her own attitudes rather than from what [name] was trying to say.
7. I sensed that L was impatient with [name], maybe uninterested or critical.
8. It seemed [name] could express and show anything s/he experienced without disturbing or changing L's attitude in response to him/her.
9. I felt that L wanted [name] to get a fresh grip and perspective on his/her problem and then come to a solution.