UNIVERSITY OF Southampton

University of Southampton Research Repository

Copyright © and Moral Rights for this thesis and, where applicable, any accompanying data are retained by the author and/or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This thesis and the accompanying data cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder/s. The content of the thesis and accompanying research data (where applicable) must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holder/s.

When referring to this thesis and any accompanying data, full bibliographic details must be given, e.g.

Thesis: Author (Year of Submission) "Full thesis title", University of Southampton, name of the University Faculty or School or Department, PhD Thesis, pagination.

Data: Author (Year) Title. URI [dataset]

University of Southampton

Faculty of Social, Human and Mathematical Sciences

School of Psychology

Brief BA for Depression Symptoms in Adolescents: Development of the Brief BA Fidelity Scale, Psychometric Evaluation, and Link to Outcome and Alliance

By

Elizabeth Jo Hodgson

Thesis for the degree of Doctorate in Clinical Psychology

May 2019

University of Southampton

Abstract

Faculty of Social, Human and Mathematical Sciences

School of Psychology

Thesis for the degree of Doctorate in Clinical Psychology

Brief BA for Depression Symptoms in Adolescents: Development of the Brief BA Fidelity Scale, Psychometric Evaluation, and Link to Outcome and Alliance

by

Elizabeth Jo Hodgson

Brief Behavioural Activation (Brief BA) is a manualised intervention for low mood and depression in adolescents (Pass & Reynolds, 2014) shown to improve depression symptoms and functioning from pre- to post- treatment. To draw conclusions about the effectiveness of Brief BA it is important to establish therapist adherence and competence (i.e. treatment fidelity). There are currently no published measures of treatment fidelity for Behavioural Activation. In this study, a measure of Brief BA fidelity was developed, and psychometric properties were tested with 30 Brief BA cases where treatment was delivered in schools. The scale evidenced good inter-rater reliability, internal consistency, and face validity and treatment fidelity was generally high. There was a significant reduction in client self-reported depression symptoms and an increase in client self-reported functioning from pre- to post- Brief BA treatment. The relationship between session-specific Brief BA fidelity and the therapeutic alliance was not significant at the beginning or middle of treatment but was significant at the end of treatment. There was no significant relationship between Brief BA fidelity and client outcome, which may be due to lack of variance given the high rates of both fidelity and client improvement in the sample. Results suggest the Brief BA fidelity scale is a reliable and valid measure, which can be used to inform future training and supervision.

Table of Contents

List of Tables ..... **i**

List of Figures ..... **i**

Declaration of Authorship ..... **iii**

Acknowledgements ..... **v**

Chapter 1 A Literature Review of Assessor-Rated Measures of Therapist Competence in Cognitive and/or Behavioural Therapies

1.1 Introduction 1

1.1.1 Defining therapist competence 1

1.1.2 Therapist adherence and therapist competence 2

1.1.3 Methods of assessing therapist competence 2

1.1.4 Aims of this literature review 3

1.2 Method 5

1.2.1 Inclusion criteria 5

1.2.2 Exclusion criteria 5

1.2.3 Location of the literature 5

1.2.4 Search strategy 6

1.2.5 Article selection 6

1.2.6 Methodological quality 8

1.2.7 Data extracted 8

1.3 Results 9

1.3.1 Overview 9

1.3.2 COSMIN risk of bias checklist 18

1.3.3 Interpretation of statistical findings 21

1.3.4 Adult transdiagnostic scales 21

1.3.5 Adult disorder specific scales 22

1.3.6 Child and young person transdiagnostic scales 24

1.3.7 Child and young person disorder specific scales 25

1.4 Discussion 27

1.4.1 Summary and critique of the competence scales 27

1.4.2 Recommendations 30

1.4.3 Limitations of this review 30

1.4.4 Conclusion 31

Chapter 2 The Brief BA Fidelity Scale: Development, Psychometric Evaluation, and Link to Outcome and Alliance

2.1 Introduction 33

2.1.1 Symptoms of depression and NICE guidelines 33

2.1.2 Behavioural activation 34

2.1.3 Measurement of treatment fidelity 35

2.1.4 Existing measures of treatment fidelity 35

2.1.5 Treatment fidelity and therapeutic alliance 36

2.1.6 Treatment fidelity and client outcome 37

2.1.7 Study aims and research questions 38

2.2. Method 41

2.2.1 Overarching study 41

2.2.2 Brief BA treatment, training, and supervision 41

2.2.3 Development of the Brief BA fidelity scale 42

2.2.4 Measures 43

2.2.5 Procedure for this research study 44

2.2.6 Inter-rater reliability 46

2.2.7 Analysis plan 46

2.3 Results 49

2.3.1 Psychometric properties 49

2.3.2 Brief BA treatment fidelity 50

2.3.3 Change in fidelity across treatment 51

2.3.4 Client outcomes from Brief BA 51

2.3.5 Brief BA fidelity and therapeutic alliance 52

2.3.6 Brief BA fidelity and client outcome 53

2.4 Discussion 55

2.4.1 Brief BA treatment fidelity 55

2.4.2 Change in fidelity across treatment 56

2.4.3 Client pre-treatment functioning and Brief BA fidelity 56

2.4.4 Client outcomes from Brief BA 57

2.4.5 Brief BA fidelity and therapeutic alliance 57

2.4.6 Brief BA fidelity and client outcome 57

2.4.7 Limitations and future research 58

2.4.8 Conclusion 59

Appendices 61

Appendix A- COSMIN Risk of Bias Checklist- adapted version 61

Appendix B- Definitions of Psychometric Properties 65

Appendix C- University of Southampton Ethical Approval 67

Appendix D- Brief BA Fidelity Scale 69

Appendix E- Brief BA Fidelity Scale Scoring Sheet 77

Appendix F- RCADS Depression Subscale 83

Appendix G- Outcome Rating Scale (ORS) 85

Appendix H- Session Rating Scale (SRS) 87

Appendix I- Brief BA Fidelity Scale Scoring Examples 89

Appendix J- Inter-Rater Reliability 99

List of References 101

List of Tables

Table 1. Summary of Study Characteristics 10
Table 2. Summary of Assessor-Rated Measures of Therapist Competence 14
Table 3. Methodological Quality Ratings (COSMIN) for the Psychometric Properties of Each Scale 20
Table 4. Participant Characteristics 45
Table 5. Brief BA Fidelity Scale Intraclass Correlation Coefficients, Means, and Standard Deviations 49
Table 6. Brief BA Fidelity Scale Means, Standard Deviations and Ranges 50
Table 7. Pre- and Post-Treatment RCADS Depression Subscale Raw Scores and ORS Scores 51
Table 8. Pre- and Post-Treatment Comparisons of RCADS Depression Subscale Raw Scores and ORS Scores 52
Table 9. Correlations Between Brief BA Fidelity Scores and Associated SRS Score at the Beginning, Middle, and End of Treatment 53

List of Figures

Figure 1. Flow Diagram of Literature Search Results 7

Research Thesis: Declaration of Authorship

Print name:	Elizabeth Jo Hodgson

Title of thesis:	Brief BA for Depression Symptoms in Adolescents: Development of the Brief BA Fidelity Scale, Psychometric Evaluation, and Link to Outcome and Alliance

I declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

This work was done wholly or mainly while in candidature for a research degree at this University;

Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;

Where I have consulted the published work of others, this is always clearly attributed;

Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;

I have acknowledged all main sources of help;

Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;

None of this work has been published before submission

Signature:		Date:	

Acknowledgements

I would like to thank my supervisors Dr Laura Pass and Dr Angharad Rudkin for their invaluable advice and guidance throughout the process of developing and writing my research thesis. I would also like to thank Laura for her help and patience with the inter-rater reliability process, and all those who gave their feedback on the Brief BA fidelity scale in the development stage.

I am extremely grateful to the young people who received Brief BA in schools and the therapists who delivered Brief BA, as without this it would not have been possible to conduct this research.

Finally, I would like to thank my partner, Oliver, and my parents, Carol and Tony, for their ongoing patience, encouragement, and support throughout the three years of training.

Chapter 1: A Literature Review of Assessor-Rated Measures of Therapist Competence in Cognitive and/or Behavioural Therapies

1.1 Introduction

Evidence suggests that Cognitive Behavioural Therapy (CBT) is effective for a range of mental health difficulties including adult and adolescent depression, generalised anxiety disorder, panic disorder, social phobia, post-traumatic stress disorder, and child anxiety and depression symptoms (Butler, Chapman, Forman, & Beck, 2006). In order to draw conclusions about the efficacy of a particular therapeutic approach, competence must be assessed to ensure that treatments were delivered as intended and in line with the evidence base. The measurement of therapist competence enables the assessment of treatment quality, which provides a framework for intervention if low competence is identified, and a method of assessing the success of this. It offers a structure for providing formative feedback to individual therapists and informing training by identifying aspects of treatment that are delivered with less skill. It is therefore essential that therapists, assessors, and researchers have access to valid, reliable, and coherent measures of therapist competence.

1.1.1 Defining therapist competence.

Therapist competence is defined as "the extent to which a therapist has the knowledge and skill required to deliver a treatment to the standard needed for it to achieve its expected effects" (Fairburn & Cooper, 2011, p. 374). Barber, Sharpless, Klostermann, and McCarthy (2007) describe competence as a sense of "appropriateness, responsiveness, good judgment, and clinical acumen" (p.494). Barber et al. (2007) propose that there are two meanings of competence: global competence and limited-domain competence. Global competence consists of general therapeutic skills demonstrated throughout the therapist's interventions and is applicable across treatment models, whereas limited-domain competence is only expressed within the context of a specific therapeutic intervention. Roth and Pilling (2007) present a competence framework specific to CBT, which consists of five domains required to deliver effective CBT. The first domain reflects generic therapeutic competences, such as the ability to engage the client. The other four domains are basic CBT competences (e.g. use of summaries and feedback to structure the session), specific CBT techniques (e.g. activity monitoring and scheduling), problem-specific competences (e.g. behavioural activation for depression), and metacompetences (e.g. capacity to select and apply the most appropriate CBT method). The generic CBT competences are applicable across

treatment models and therefore align with the idea of global competence proposed by Barber et al. (2007). The other four domains are specific to CBT and align with limited-domain competence.

1.1.2 Therapist adherence and therapist competence.

Fairburn and Cooper (2011) distinguish between therapist adherence and competence; adherence refers to whether an intervention was delivered as intended, and competence refers to the skill with which the intervention was delivered. The meaning of adherence within the context of a manualised treatment is the extent to which a therapist implements procedures outlined within the treatment manual (Barber et al., 2007). Adherence is independent of the context, whereas competence is dependent on contextual factors, such as the severity of client impairment, the client's life situation and stress, and the timing of interventions in a treatment session (Waltz, Addis, Koerner, & Jacobson, 1993). Competence is not sufficient without adherence (a therapist may be skilful at delivering the treatment but use techniques that are not outlined in the protocol). Furthermore, adherence is not sufficient without competence (a therapist may demonstrate a high level of adherence by rigidly following a treatment manual but being unresponsive to the client's behaviours, level of engagement and understanding). Although adherence and competence are separate concepts, there is overlap between them in practice and it is proposed that adherence presupposes competence, but adherence does not guarantee competence (Waltz et al., 1993). Treatment fidelity comprises both competence and adherence, and effective delivery of treatment requires both to ensure skilful delivery of techniques from a specific treatment model (Perepletchikova & Kazdin, 2005).

1.1.3 Methods of assessing therapist competence.

Muse and McManus (2013) provide a framework for the methods of assessing CBT therapist competence, based on Miller's (1990) clinical skills hierarchy. There are four levels: knowledge-based assessments (e.g. essays and multiple-choice questions), assessments of practical understanding (e.g. case reports and clinical vignettes), assessments of practical application of knowledge (e.g. standardised role-plays), and clinical practice assessments (e.g. therapist self-assessment, supervisory assessment, and assessor-rated assessment). Therapist competence involves not only knowledge and understanding, but the application of therapeutic skills in clinical practice (Roth & Pilling, 2007). Therefore, although knowledge-based assessments and assessments of practical understanding are useful, they are unlikely to be sufficient alone. Role-plays are useful for therapists practicing therapeutic skills but are not a standardised method of assessing therapist competence and may not be representative of clinical practice (Sharpless &

Barber, 2009). Clinical practice assessment is the highest level of competence assessment. Therapist self-assessment and supervisory assessment are often used in practice as formative measures. Assessor-rated competence is a formal method of competence assessment where an independent rater observes treatment sessions live or via recordings and rates the skill with which treatment was delivered using a standardised scale. This is considered the 'gold-standard' method of assessing therapist competence (Muse & McManus, 2013; Rapley & Loades, 2018). Although more time-consuming than alternative methods, it is a more objective method with less risk of bias.

1.1.4 Aims of this literature review.

The aim of this review is to describe and evaluate existing assessor-rated measures of therapist competence for cognitive and/or behavioural therapies. Muse and McManus (2013) previously evaluated various methods of assessing CBT competence, whereas this review focusses on the gold-standard method of assessing therapist competence using assessor-rated measures. This review builds upon the Muse and McManus (2013) review by including measures that have been designed to assess therapist competence in delivering treatment for children and young people, as well as measures that assess competence in delivering treatment for adults. As in the review by Muse and McManus (2013), measures assessing therapist competence only as well as measures assessing both competence and adherence (i.e. fidelity) are included. In this review articles from January 1980 to January 2019 are included, whereas Muse and McManus (2013) included articles from January 1980 to July 2012. Additional measures of therapist competence have been developed since Muse and McManus conducted their search in 2012; therefore, this review provides an up-to-date synthesis of the literature in this area. As evidence suggests that Behavioural Activation (BA) is as effective as CBT for adult depression (Richards et al., 2016), the search has been widened to include competence scales for behaviourally focussed therapies such as BA. The psychometric properties of each measure and the quality of analyses are evaluated. Each measure is described in terms of the subscales and individual items, and the feasibility of using each measure is considered.

1.2 Method

1.2.1 Inclusion criteria.

Articles were included if they introduced an assessor-rated scale of therapist competence or fidelity (competence and adherence) and investigated the psychometric properties of the scale. Articles were included if the scale measured therapist competence to a form of CBT (Gaudiano, 2008; e.g. CBT, Cognitive Therapy, or Cognitive Processing Therapy) or behavioural therapy (e.g. BA) and if the treatment was individual, face-to-face therapy. To ensure only high quality research was reviewed, articles were only included if they had been published in peer-reviewed journals. For practical reasons, articles needed to have been published in English.

1.2.2 Exclusion criteria.

Articles were excluded if they did not introduce an assessor-rated scale that measured therapist competence or fidelity. Where multiple studies relating to the same measure were obtained, only the initial development study was included. This was a necessary criterion to ensure a feasible number of studies for literature review and synthesis. It is good practice for the initial development study to include psychometric properties of the scale as it is the main article that will be cited. Articles were excluded if they related to a therapy delivered via telephone or internet or if the therapy was delivered in a group format, with couples or families. Articles were excluded if they related to any other therapy apart from cognitive and/or behavioural therapy (e.g. Mindfulness or Dynamic Psychotherapy). Articles were excluded if they had not been peer-reviewed, to ensure that only high quality research was reviewed. Grey literature (e.g. conference papers and dissertations), study protocols and systematic reviews were excluded from the review but were included for hand searching to check for additional relevant articles to include in this review. A language limiter was used, and articles were excluded if not published in English.

1.2.3 Location of the literature.

First, a search of internet-based bibliographic databases (PsycINFO and Web of Science Core Collection) was conducted, covering January 1980 to January 2019. These two databases were chosen because they contain peer reviewed journal articles relevant to psychology. The search was conducted from 1980 and onwards to ensure the inclusion of the Cognitive Therapy Scale (CTS; Young & Beck, 1980). Abstracts were screened and full-text articles were reviewed by the primary author for eligibility. For articles where eligibility was unclear, discussions were held with the research supervisors. Reference lists of retained articles were inspected for relevant

studies and databases were used to search for the abstracts and, if relevant, full text articles. Reference lists of literature reviews were also checked for relevant studies.

1.2.4 Search strategy.

The following search terms were used in both PsycINFO and Web of Science: ((assess* OR measure* OR scale) AND (therapist* OR clinic* OR psychologist* OR practitioner*) AND (competenc* OR skill* OR quality OR expertise OR fidelity) AND (“CBT” OR “cognitive behavio* therapy” OR “behavio* therapy” OR “cognitive therapy” OR "BA" OR "behavio* activation" OR "BATD")). BATD stands for Behavioural Activation for the Treatment of Depression in adults (Lejuez, Hopko, Acierno, Daughters, & Pagoto, 2011). A scoping search of the literature was conducted to identify search terms that had previously been used to retrieve information on this topic and search terms were developed through discussions with the supervisors of this research project. In PsycINFO, the terms were searched for within abstracts and in Web of Science, the terms were searched for within the topic (title, abstract and keywords). Results were limited to the following categories, based on relevance to this literature review: (a) psychology; (b) psychology clinical; (c) psychology developmental; (d) psychology MDT; (e) family studies; (f) psychology applied; and (g) behavioural sciences.

1.2.5 Article selection.

A flow chart detailing the article identification and selection process, following guidelines from PRISMA (Moher, Liberati, Tetzlaff, & Altman, 2009) is presented in figure 1. The PsycINFO and Web of Science searches retrieved 1092 results, of which 250 were duplicates. There were 842 unique citations and two additional citations were identified through inspecting reference lists of relevant articles and reviews. Abstract screening led to the exclusion of 776 citations. Full-text articles of the remaining 68 citations were retrieved and reviewed for eligibility, and 12 were identified as eligible. The reasons for non-eligibility are presented in Figure 1. The main reason was that they did not introduce an assessor-rated scale that measured therapist competence or fidelity.

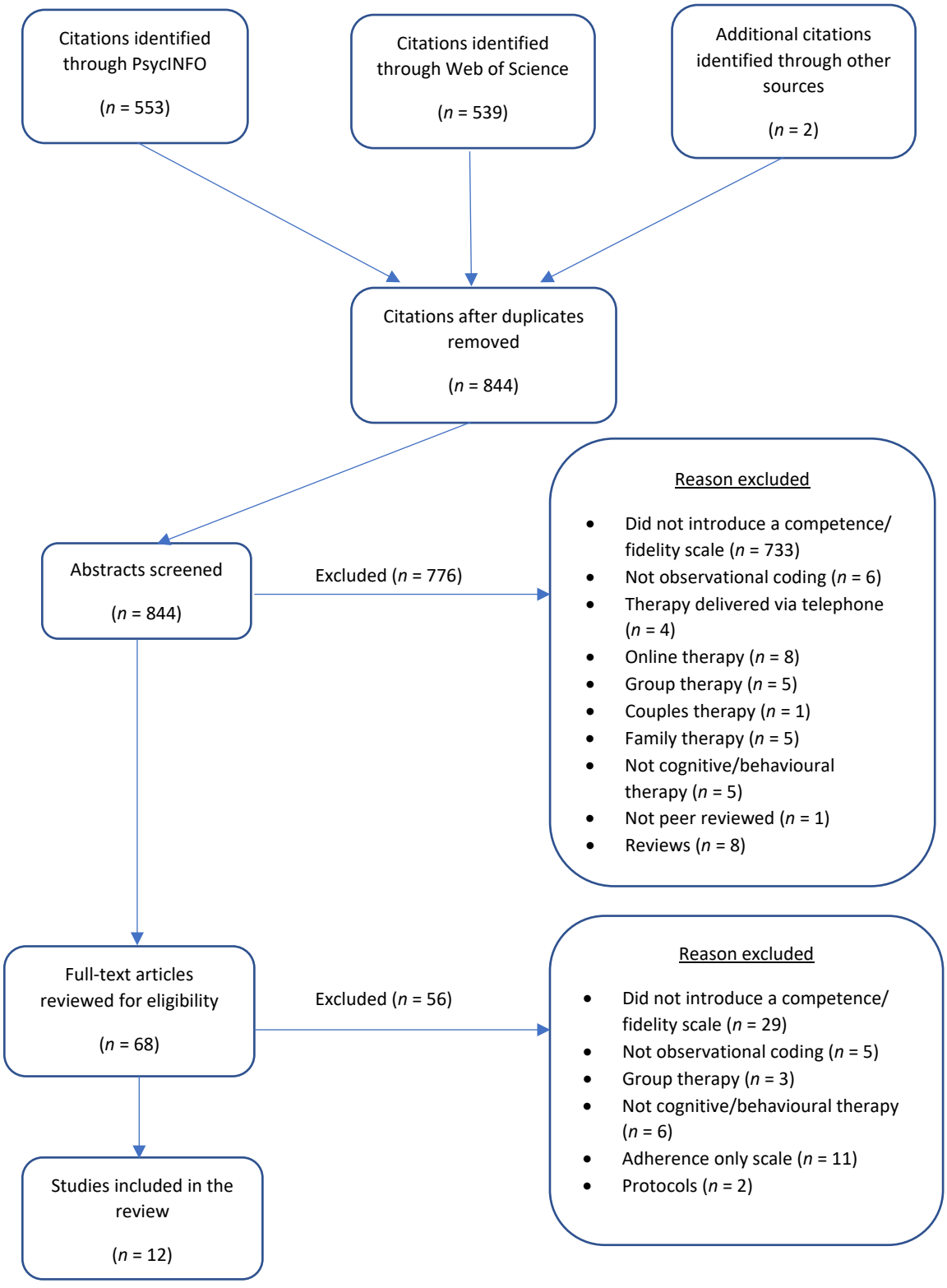


Figure 1. Flow Diagram of Literature Search Results

1.2.6 Methodological quality.

An adapted version of the COSMIN risk of bias checklist for systematic reviews of Patient-Reported Outcome Measures (PROMs; Mokkink et al., 2018a) was used to evaluate the methodological quality of studies (see Appendix A). This is a measure of whether the results are trustworthy based on the methodological quality of the study. The COSMIN user manual (Mokkink et al., 2018b) states that "the methodology can also be used for other types of measurement instruments (like clinician-reported outcome measures or performance-based outcome measures), or other applications (e.g. diagnostic or predictive applications), but the methodology may need to be adapted for these other purposes" (p. 6).

As the measures included in this review are assessor rated measures of therapist competence, some parts of the COSMIN tool were not relevant; therefore, it was adapted to include only the relevant standards. The edits were discussed and agreed with qualified research clinical psychologists. The adapted tool consists of the following seven categories (described in the COSMIN checklist as 'boxes'): measure development, content validity, structural validity, internal consistency, reliability, criterion validity, and construct validity. The following three categories were removed from the original COSMIN checklist because they were specific to PROMs and not relevant to this review: measurement error, cross-cultural validity, and responsiveness. There is a 4-point rating system where each standard within a category is rated as 'very good', 'adequate', 'doubtful', or 'inadequate'. For each study, the overall rating for each category is determined by taking the lowest rating of any standard in that category (i.e. the lowest score counts).

1.2.7 Data extracted.

The following information was extracted from each study: a) client age; b) client diagnosis; c) type of therapy; d) therapist sample size and information about their experience and qualifications; e) number of assessors and information about their experience and qualifications; f) assessor training; g) number of session recordings; h) full session recordings or clips; i) psychometric properties evaluated; j) the country in which the study was conducted; and k) information about the scale including subscales, items in the scale, whether it is transdiagnostic or disorder specific, response options, and time taken to complete the scale.

1.3 Results

1.3.1 Overview.

Twelve studies were identified, which are summarised in Table 1. Studies were categorised according to the nature of the scale. First, studies were classified into two groups based on whether the scale was transdiagnostic or disorder specific. Next, each of these two groups were classified into two groups based on whether the scale was specific to working with children and young people or adults. The characteristics of each scale are summarised in Table 2. The scales are listed in the following order: adult transdiagnostic (*n* = 3), adult disorder specific (*n* = 5), child transdiagnostic (*n* = 2), and child disorder-specific (*n* = 2).

Table 1

Summary of Study Characteristics

Authors of the study	**Scale**	**Client age**	**Client diagnosis**	**Therapy**	**Therapist sample**	**Assessor sample**	**Assessor training**	**Number of session recordings**	**Full sessions or clips**	**Psychometric evaluation**	**Country**
Dobson, Shaw & Vallis, 1985	Cognitive Therapy Scale; CTS (Young & Beck, 1980)	NR (assumed to be adults)	Depression	Cognitive therapy	21 therapists (10 psychiatrists and 11 psychologists) Minimum 2 years post-doctoral experience	4 cognitive behavioural therapists, who had contributed to CT development	Trained in use of the CTS (no further details given)	21 (one submitted by each therapist)	Full sessions	Internal consistency, inter-rater reliability	United States
Blackburn et al., 2001	Cognitive Therapy Scale-Revised; CTS-R (Blackburn et al., 2001)	19-70 years (*M* = 37)	Depression, social phobia, panic disorder, OCD and GAD	Cognitive Therapy	21 trainee therapists (5 psychiatrists, 6 psychologists, 7 nurses, 1 trainee psychologist, 1 senior registrar, 1 senior nurse)	4 expert raters (no further information given)	NR	102 (3 from each of the 34 patients-one from beginning, middle and end)	Full sessions	Internal consistency, inter-rater reliability, face validity, discriminant validity	United Kingdom
Barber, Liese, & Abrams, 2003	Cognitive Therapy Adherence and Competence Scale; CTACS (Liese, Barber & Beck, 1995)	NR (assumed to be adults)	Cocaine dependence	Cognitive Therapy + group drug counselling	18 therapists with a doctoral degree, Master of Social Work degree, or medical degree and 6 months to 1-year experience	1 clinical psychologist and 1 master's-level psychiatric nurse. Both had training in CT and clinical and supervision responsibilities	20 hours of instruction in the CTACS use, followed by monthly telephone conferences	134 (92 of CT sessions with 88 clients)	Full sessions	Inter-rater reliability, internal consistency, discriminant validity	United States

Authors of the study	Scale	Client age	Client diagnosis	Therapy	Therapist sample	Assessor sample	Assessor training	Number of session recordings	Full sessions or clips	Psychometric evaluation	Country
Carroll et al., 2000	Yale Adherence and Competence Scale; YACS (Carroll et al., 2000)	*M* = 30 (*SD* = 5.5)	Cocaine dependence	Cognitive Behavioural Therapy	NR	Masters-level clinicians with experience in treating substance users and often have experience in one or more of the study treatments evaluated	Review of manual. 10 practice recordings rated, and reliability checked. Sessions to prevent rater drift	741 from 117 participants	Full sessions	Inter-rater reliability, factor analysis, internal consistency, criterion validity, discriminant validity	United States
Huppert et al., 2001	Multicenter Collaborative Study for the Treatment of Panic Disorder-Global Competence Item; MCSTPD-GCI (Huppert et al., 2001)	19-65 years (*M* = 36)	Panic disorder without agoraphobia or with low levels of agoraphobia	Cognitive Behavioural Therapy	14 therapists (13 psychologists and 1 [psychiatrist). CBT experience ranged from 1-18 years	NR	Raters were trained to a high level of reliability (no further details given)	526 (no further information given)	NR	No psychometric evaluation	United States
Haddock et al., 2001	Cognitive Therapy Scale for Psychosis; CTS-Psy (Haddock et al., 2001)	NR (assumed to be adults)	Psychosis	Cognitive Behavioural Therapy	21 trainee therapists on a diploma level course- CBT for psychosis, with at least 1-year post-qualification experience	2 clinical psychologists, 1 mental health nurse and 1 research fellow with a background in social work. All had undergone specialist training in CBT for Psychosis	Raters received intensive training on the CTS-Psy (no further details given)	NR	Full sessions	Inter-rater reliability, discriminant validity	United Kingdom

Authors of the study	Scale	Client age	Client diagnosis	Therapy	Therapist sample	Assessor sample	Assessor training	Number of session recordings	Full sessions or clips	Psychometric evaluation	Country
Davidson et al., 2004	Manual Assisted Cognitive Behaviour Therapy Rating Scale; MACT-RS (Davidson et al., 2004)	NR (assumed to be adults)	Recurrent deliberate self-harm	Manual-Assisted Cognitive-Behaviour Therapy (MACT)	21 therapists: 12 nurses, 1 psychiatrist, 4 psychologists, 2 social workers & 2 occupational therapists	2 experienced CBT trainers involved in the development of the scale	NR	49 (no further information given)	NR	Inter-rater reliability	United Kingdom
von Consbruch, Clark, & Stangier, 2012	Cognitive Therapy Competence Scale for Social Phobia; CTCS-SP (Clark, von Consbruch, Hinrichs, & Stangier, 2007)	NR (assumed to be adults)	Social Phobia	Cognitive Therapy	51 therapists trained in CT for social phobia	6 trainee clinical psychologists and 1 clinical psychotherapist	2 days of training in using the scale, 5 recordings rated and discrepancies discussed	161 (from 98 clients)	Full sessions	Inter-rater reliability, internal consistency, retest reliability	Germany
Stallard, Myles, & Branson, 2014	Cognitive Behaviour Therapy Scale for Children and Young People; CBTS-CYP (Stallard, Myles, & Branson, 2014)	9-17 years (*M* = 14.4)	Depression, separation anxiety, social anxiety, OCD, panic, GAD, specific phobia, PTSD	Cognitive Behavioural Therapy	18 CBT therapists on the CYP-IAPT course	12 assessors for the CYP-IAPT course	Training in using the CBTS-CYP (no further details given)	48 (from 18 therapists)	Full sessions	Face validity, internal consistency, convergent validity, discriminant validity	United Kingdom

Authors of the study	Scale	Client age	Client diagnosis	Therapy	Therapist sample	Assessor sample	Assessor training	Number of session recordings	Full sessions or clips	Psychometric evaluation	Country
Brown et al., 2018	Global Therapist Competence Scale for Youth Psychosocial Treatment; GCOMP (Brown et al., 2018)	RCT 1: *M* = 10.4, RCT 2: *M* = 11.3	GAD, separation anxiety, social phobia, or specific phobia	Coping Cat (Individual CBT; Kendall & Hedtke, 2006)	RCT 1: 16 trainee clinical psychologists and clinical psychologists. RCT 2: 13 social workers, psychologists, and 'other'	8 trainee clinical psychologists	Reading the manual, reviewing sessions with trainers, coding discussion at meetings, independent coding	744 (from 68 clients)	Full sessions	Inter-rater reliability, variance components analysis, construct validity	United States
Bjaastad et al., 2015	Competence and Adherence Scale for CBT; CAS-CBT (Bjaastad et al., 2015)	8-15 years (*M* = 11.5)	Separation anxiety, social phobia, or GAD	Manualised CBT (the FRIENDS program; Barrett, 2004, 2008)	10 clinical psychologists, 6 Masters of Education with 2 years clinical training, and 1 clinical social worker	2 CBT therapists/supervisors (considered expert raters) and two graduate psychology students (student raters)	Student raters completed training and discussed ratings with the expert raters	181 (from 173 clients)	Full sessions	Inter-rater reliability, rater stability, inter-item correlations, internal consistency, factor analysis	Norway
McLeod et al., 2018	Cognitive-Behavioural Treatment for Anxiety in Youth Competence Scale; CBAY-C (McLeod et al., 2018)	7–15 years (*M* = 10.6)	GAD, separation anxiety, social phobia, or specific phobia	Coping Cat (individual CBT; Kendall & Hedtke, 2006)	RCT 1: 16 trainee clinical psychologists and clinical psychologists. RCT 2: 13 social workers, psychologists, and 'other'	4 trainee clinical psychologists (all had training and clinical experience delivering CBT for anxiety)	Reading the manual, reviewing sessions with trainers, coding discussion at meetings, independent coding	744 (from 68 clients)	Full sessions	Inter-rater reliability, construct validity, variance components analysis	United States

Note. NR = Not Reported. OCD = Obsessive Compulsive Disorder. GAD = Generalised Anxiety Disorder. PTSD = Post Traumatic Stress Disorder. CYP-IAPT = Children and Young People's Improving Access to Psychological Therapies. RCT = Randomised Control Trial.

Table 2

Summary of Assessor-Rated Measures of Therapist Competence

Scale	Subscales	Items	Scale Category	Rating	Time to complete scale
CTS	General therapeutic skills	1. Agenda setting 2. Feedback 3. Understanding 4. Interpersonal effectiveness 5. Collaboration 6. Pacing and use of time	Adult transdiagnostic scale	7-point Likert scale (0 = *poor* and 6 = *excellent*) with item specific descriptors for even ratings. Range 0- 66	NR
	CBT specific skills	7. Empiricism 8. Focus on key cognitions 9. Strategy for change 10. Application of C B techniques 11. Homework			
CTS-R	General therapeutic skills	1. Agenda setting and adherence 2. Feedback 3. Collaboration 4. Pacing and efficient use of time 5. Interpersonal effectiveness	Adult transdiagnostic scale	7-point Likert scale (0 = *incompetent* (non-compliance) and 6 = *expert* (compliance and high skill). Range 0-72	NR
	CBT specific skills	6. Eliciting appropriate emotional expression 7. Eliciting key cognitions 8. Eliciting behaviours 9. Guided discovery 10. Conceptual integration 11. Application of change methods 12. Homework setting			
CTACS	CT Structure	1. Agenda 2. Mood check 3. Bridge from previous visit 4. Inquired about ongoing problem 5. Reviewing previous h/w 6. Assigning new h/w 7. Capsule summaries 8. Patient summary and feedback 9. Focus/structure	Adult transdiagnostic scale	7-point Likert scale: adherence rating (0 = *none* and 6 = *thorough*) and quality rating (0 = *poor* and 6 = *excellent*) with item specific descriptors for even ratings. Range 0–126	NR
	Development of a collaborative therapeutic relationship	10. Socialization to CT 11. Warmth/genuineness/ congruence 12. Collaboration			
	Development and application of the case conceptualization	13. Eliciting automatic thoughts 14. Eliciting core beliefs and schemas 15. Eliciting meaning/understanding 16. Addressing key issues 17. Case conceptualization			
	Cognitive and behavioural techniques	18. Guided discovery 19. Asking for evidence 20. Use of alternative techniques 21. Overall performance			

Scale	Subscales	Items	Scale Category	Rating	Time to complete scale
YACS	Substance use disorder general therapeutic skills a) assessment	1. Assess alcohol use 2. Assess cocaine use 3. Assess other substances 4. Assess psychopathology 5. Assess general functioning	Adult disorder specific (substance use disorder)	Each item is rated for adherence (0 = *not at all* and 5 = *extensively*). Where the item did occur, a competence rating is completed (0 = *not at all* and 5 = *extensively*). Range 0 - 105	NR
	b) general support	6. Praise patient efforts 7. Explore feelings 8. Explore level of family support 9. Optimistic reassurance 10. Show natural spontaneity			
	c) goals of treatment	11. Explore patient's treatment goals 12. Discrepancy- behaviour & goals 13. Commitment to abstinence 14. Reflective listening 15. Feedback about urine results			
	Substance use disorder CBT specific skills	16. Skills training 17. Debrief past high-risk situations 18. Cognitions 19. Plan future high-risk situations 20. Difference between slip v relapse 21. Conditioning			
MCSTPD -GCI	Global skill	1. Global competence	Adult disorder specific scale (panic disorder)	7-point Likert scale (1 = *clearly inadequate* and 7= *excellent*). Range 1–7	NR
CTS-Psy	General therapeutic skills (as applied in psychosis)	1. Agenda 2. Feedback 3. Understanding 4. Interpersonal Effectiveness 5. Collaboration	Adult disorder specific scale (Psychosis)	Six aspects for items 1 to 9 (1 = *present/ appropriately omitted* and 0 = *absent*). Item 10 is rated as 1 = *barely acceptable* to 6 = *excellent*. Range 0–60	NR
	CBT specific skills (as applied in psychosis)	6. Guided discovery 7. Focus on key cognitions 8. Choice of intervention 9. Homework 10. Quality of intervention (overall)			
MACT-RS	Skill in delivering self-harm treatment techniques	1. Structure 2. Pacing 3. Collaboration 4. Appropriate techniques 5. Skilful execution of techniques 6. Helpfulness of session 7. Empathy 8. Client problem/difficulty 9. Linking sessions 10. Using the manual 11. Homework assignments	Adult disorder specific scale (Self-Harm)	Each item is rated on a 1 to 7 scale with item-specific anchors provided at the low, mid, and high scale points. Range 11–77	NR

Scale	Subscales	Items	Scale Category	Rating	Time to complete scale
CTCS-SP	General therapeutic skills	1. Agenda 2. Dealing with questions 3. Clarity of communication 4. Pacing and efficient use of time 5. Interpersonal effectiveness 6. Resource orientation 7. Review of diary and questionnaires 8. Reviewing homework 9. Use of feedback and summaries	Adult disorder specific scale (Social Phobia)	7-point Likert scale (0 = *poor* and 6 = *excellent)*. Range 0- 66	NR
	CBT specific skills	10. Guided discovery 11. Focus on cognitive model 12. Rationale 13. Selection of appropriate strategies 14. Implementation of techniques 15. Integration of discussion and experiential techniques 16. Homework setting			
CBTS-CYP	Process	1. Partnership working 2. Right developmental level 3. Empathetic 4. Creative 5. Investigative 6. Self-efficacy 7. Enjoyable and engaging	Child and young person transdiagnostic scale	7-point Likert scale (0 = *incompetent* (non-compliance) and 6 = *expert* (compliance and high skill). Range 0-84	NR
	Method	8. Assessment and goals 9. Behavioural techniques 10. Cognitive techniques 11. Discovery experiments 12. Emotional techniques 13. Formulation 14. General skills			
GCOMP	Alliance building	1. Understanding 2. Positive regard 3. Client's perspective 4. Collaboration	Child and young person transdiagnostic scale	7-point Likert-scale (1 = *very poor* and 7 = *excellent*). Range 14-98	NR
	Positive expectancies	5. Treatment expectancies 6. Therapist credibility 7. Client self-efficacy			
	Focussing treatment	8. Structure 9. Continuity 10. Key themes			
	Instigating change	11. Change strategies 12. Active participation			
	Responsiveness	13. Motivation 14. Flexibility			

Scale	Subscales	Items	Scale Category	Rating	Time to complete scale
CAS-CBT	CBT structure	1. H/w review and setting new h/w tasks (adherence) 2. Structure and progress (adherence) 3. Parental involvement (adherence) 4. Cognitive behaviour therapy structure (competence score for items 1-3)	Child and young person disorder specific scale (anxiety)	7-point Likert scale: adherence (0 = *none* and 6 = *thorough* and competence (0 = *poor skills*, 6 = *excellent skills*). Range 0-66	NR
	Process and relational skills	5. Positive reinforcement (adherence) 6. Collaboration (adherence) 7. Process and relational skills (competence score for 5-6) 8. Flexibility (competence score)			
	Facilitating and completing session goals	9. Session goal 1 (adherence) 10. Session goal 2 (adherence) 11. Session goals (competence score for items 9-10)			
CBAY-C	Standard	1. Within session focus 2. Across session focus 3. Structure/phase 4. Homework review 5. Homework assignment	Child and young person disorder specific scale (anxiety)	7-point Likert-scale (0 = *not present* and 7 = *excellent*). Range 0-175	NR
	Model	6. Psychoeducation- anxiety 7. Emotion education 8. Fear ladder 9. Relaxation 10. Cognitive- anxiety 11. Problem solving 12. Self-reward 13. Coping plan 14. Exposure preparation 15. Exposure 16. Exposure de-brief 17. Maintenance			
	Delivery	18. Didactic teaching 19. Collaborative teaching 20. Modelling 21. Rehearsal 22. Coaching 23. Self-disclosure			
	Global	24. Skilfulness 25. Responsiveness			

Note. NR= Not Reported.

1.3.2 COSMIN risk of bias checklist.

The identified studies varied in the measurement properties that were reported, with an average of four properties per study. Definitions of each measurement property are outlined in Appendix B (adapted from the COSMIN user manual; Mokkink et al., 2018b). Quality ratings based on the COSMIN risk of bias checklist are summarised in Table 3, including a scoring system that was devised for this review to compare the measures in terms of the methodological quality. The measures with the lowest overall quality ratings were the CTS, MCSTPD-GCI, and MACT-RS, whereas the measures with the highest overall quality ratings were the CTS-R, CTACS, and YACS.

Measure development. All studies described the development of the scale. Measure development was rated as 'adequate' or 'very good' on the COSMIN checklist for all of the scales in this review apart from the MCSTPD-GCI. For this scale there was a lack of information about the construct to be measured (the theoretical ground on which the scale was developed was unclear) and the context of use (the intended application of the scale was not clearly described). It is possible that this information is provided elsewhere; however, it was not easily available.

Content validity. Content validity was reported for eight of the scales. The quality of the analysis was rated as 'adequate' on the COSMIN checklist for all scales where it was reported apart from the YACS, which was rated as 'doubtful'. For this scale it was not clear whether every item was considered in terms of relevance to the construct of interest. It is important for authors to provide details of the scale development including consideration of each item for inclusion in the scale (as recommended by the COSMIN checklist), so that it is clear how each item is relevant and representative of the construct the scale is measuring.

Structural validity. Structural validity was only reported for two of the scales in this review (YACS and CAS-CBT). For the YACS, confirmatory factor analysis was conducted to assess structural validity; however, the quality of this analysis was rated as 'inadequate' due to the small sample size. Exploratory factor analysis was conducted for the CAS-CBT and the quality of this analysis was rated as 'adequate' on the COSMIN checklist.

Internal consistency. Internal consistency was reported for six of the scales. This reliability analysis evaluates the interrelatedness of items and whether items on the scale measure the same construct. For five out of six scales (CTS, CTS-R, CTACS, CTS-SP, and CAS-CBT) the quality of the analyses was rated as 'very good' on the COSMIN checklist. This suggests that for these scales, items measure the same construct, and therapists who demonstrate competence in one area demonstrate competence in other areas. For the CBTS-CYP the quality of this analysis was rated

as 'doubtful' on the COSMIN Checklist due to not reporting Cronbach's alpha. Cronbach's alpha is recommended for reliability analysis when there are multiple items in a scale because it calculates item-total correlations and Cronbach's alpha if each item was deleted.

Inter-rater reliability. All studies apart from one (MCSTPD-GCI) reported inter-rater reliability. Inter-rater reliability is particularly important for measure development as assessors must be in agreement to ensure accurate ratings. The quality of the inter-rater reliability analysis was rated on the COSMIN checklist as 'adequate' or 'very good' for many of the measures but some were rated as 'doubtful' (CTS-Psy, CTS, MACT-RS and CBTS-CYP). The quality rating was marked down when ICCs were not calculated for individual items as well as the overall scale, or when Pearson's correlation coefficient was calculated instead of ICC, as Pearson's correlation does not take systematic error into account (Terwee et al., 2007).

Criterion validity. Criterion validity is an estimate of the extent to which a measure agrees with a gold standard. This analysis was only reported for one scale (YACS), which may be because it is not clear what the gold standard measure of competence or fidelity is. For the YACS, the quality of the criterion validity analysis was rated as 'very good' on the COSMIN checklist.

Construct validity. Construct validity was reported for seven of the scales. In particular, discriminant validity was measured to test whether concepts or measurements that are not supposed to be related are actually unrelated. The quality of the analysis was rated as 'very good' for six of the scales (CTS-R, CTACS, YACS, CBTS-CYP, GCOMP, and CBAY-C). The exception was the CTS-Psy, which was rated as 'doubtful' due to the small sample size.

Table 3

Methodological Quality Ratings (COSMIN) for the Psychometric Properties of Each Scale

Classification	Not reported	Inadequate	Doubtful	Adequate	Very good
Score	0	1	2	3	4

	Measure Development	Content Validity	Structural Validity	Internal Consistency	Inter-Rater Reliability	Criterion Validity	Construct Validity	Total Score
CTS								9
CTS-R								18
CTACS								18
YACS								19
MCSTP D-GCI								2
CTS-Psy								11
MACT-RS								5
CTCS-SP								10
CBTS-CYP								15
GCOMP								15
CAS-CBT								16
CBAY-C								15

1.3.3 Interpretation of statistical findings.

In addition to the methodological quality ratings on the COSMIN checklist, statistical findings were also considered. Intraclass Correlation Coefficients were interpreted following Cicchetti's (1994) guidelines; less than .40 reflects poor agreement, .40 to .59 reflects fair agreement, .60 to .74 reflects good agreement, and .75 and higher reflects excellent agreement. Cronbach's alpha was interpreted using George and Mallery's (2003) guidelines; > .90 is excellent, > .80 is good, > .70 is acceptable, > .60 is questionable, > .50 is poor, and < .50 is unacceptable.

1.3.4 Adult transdiagnostic scales.

Three transdiagnostic scales were identified that assess competence in delivering cognitive therapy with adults; the Cognitive Therapy Scale (CTS; Young & Beck, 1980), the Cognitive Therapy Scale- Revised (CTS-R; Blackburn et al., 2001), and the Cognitive Therapy Adherence and Competence Scale (CTACS; Liese, Barber, & Beck, 1995).

The CTS was developed to evaluate therapist competence in implementing the cognitive therapy protocol developed by Beck, Rush, Shaw, and Emery (1979). Criticism of the CTS included that rating points were not adequately defined (Whisman, 1993) and that there were aspects of competence not captured by the scale (Shaw et al., 1999). The CTS-R was subsequently developed and the main changes from the CTS were that interpersonal effectiveness items were combined into one item and facilitation of emotional expression was included as an additional item. The CTACS (Liese et al., 1995) was also based on the CTS but this measure was developed to assess therapist competence to CBT for cocaine abuse. The authors suggest the CTACS can be used in non-substance abuse populations; however, the psychometric properties of the scale have yet to be tested in other populations. The CTS measures competence only, whereas the CTS-R combines the measurement of adherence and competence in a single rating. The CTACS measures both adherence and competence separately, and Pearson's correlation coefficient indicated a strong correlation between adherence and competence (r = .96).

The scales vary in terms of their items, subscales, and response options. The CTS has 11 items rated on a 7-point scale (0 = *poor* and 6 = *excellent*) and there is a detailed scoring manual with behavioural descriptors. The CTS-R has 12 items and items are still scored on a 7-point scale (0 = *incompetent* and 6 = *expert)* but a clearer framework of the level of skill was developed based

on the Dreyfus Levels of Competence (Dreyfus, 1989) as described in the CTS-R manual (James, Blackburn, & Reichelt, 2001). The authors examined the psychometric properties of the CTS-R (Blackburn et al., 2001) and concluded that the scale may benefit from clearer definitions of items and guidelines for discriminating between different points on the scale. However, the scale has not since been updated. The CTS and the CTS-R both have two subscales (general therapeutic skills and CBT specific skills). The CTACS has 21 items separated into five subscales (cognitive therapy structure, development of a collaborative therapeutic relationship, development and application of the case conceptualization, cognitive and behavioural techniques, and overall performance). Each item is assessed for adherence and competence on a 7-point scale (for adherence 0 = *none* and 6 = *thorough*, and for competence 0 = *poor* and 6 = *excellent*).

Three measurement properties were reported for the CTS (Dobson et al., 1985), whereas five measurement properties were reported for the CTS-R (Blackburn et al., 2001) and the CTACS (Barber, Liese, and Abrams). All three scales demonstrated excellent internal consistency; Cronbach's alpha was .95 for the CTS, ranged from .92 to .97 for the CTS-R, and was .93 for the CTACS. The quality of each of these analyses were rated as 'very good' on the COSMIN checklist. For the CTS inter-rater reliability analysis, Pearson's correlation coefficient was .94 for the total score and individual items ranged from .54 (feedback) to .87 (application of cognitive behavioural techniques). The quality of this analysis was rated as 'doubtful' on the COSMIN checklist because Pearson's correlation coefficient was calculated instead of ICC. For the CTS-R and the CTACS, ICCs were conducted. The total ICC for pairs of raters on the CTS-R ranged from .40 to .86 (fair to excellent), and for individual items the lowest ICC was −.14 (Collaboration) and the highest was .84 (Guided discovery). The total ICC for the CTACS was .73 (good), and for individual items ranged from .34 (Addressing key issues) to .92 (Mood check). The quality of inter-rater reliability analysis was rated as 'adequate' on the COSMIN checklist for both the CTS-R and the CTACS. The CTS did not report the development of the scale whereas the CTS-R and the CTACS did provide this and the quality of the analysis was rated as 'adequate' on the COSMIN checklist. The CTS-R and the CTACS both demonstrated good construct validity and the quality of the analysis was rated as 'very good' for both scales on the COSMIN checklist.

1.3.5 Adult disorder specific scales.

Five disorder specific scales were identified that assess competence in delivering therapy with adults; the Yale Adherence and Competence Scale (YACS; Carroll et al., 2000), the Multicentre Collaborative Study for the Treatment of Panic Disorder- Global Competence Item (MCSTPD-GCI; Huppert et al., 2001), the Cognitive Therapy Scale for Psychosis (CTS-Psy; Haddock

et al., 2001), the Manual-Assisted Cognitive Behaviour Therapy Rating Scale (MACT-RS; Davidson et al., 2004), and the Cognitive Therapy Competence Scale for Social Phobia (CTCS-SP; Clark, von Consbruch, Hinrich, & Stangier, 2007).

These five scales all assess competence in delivering therapy for specific mental health disorders. The YACS is specific to the implementation of therapy for substance use disorders, the MCSTPD-GCI is for panic disorder, the CTS-PSY is for psychosis, the MACT-RS is for self-harm, and the CTCS-SP is for social phobia. The YACS measures both adherence and competence separately, and Pearson's correlation coefficient indicated a moderate correlation between adherence and competence for CBT (r = .38). The other four scales measure competence only.

The YACS has three general subscales (assessment, general support, and goals of treatment), and three treatment specific subscales (Clinical Management; CM, Twelve Step Facilitation; TSF, and CBT). There are 21 items rated on a 6-point scale and the scale has a detailed scoring manual to encourage a consistent and reliable approach. Each item is rated for adherence (0 = *not at all* and 5 = *extensively*). Where the item did occur, a competence rating is completed (0 = *not at all* and 5 = *extensively*). The MCSTPD-GCI has a single item (global competence) rated on a 7-point scale (1 = *clearly inadequate* and 7= *excellent)*. The CTS-Psy is a 10-item scale based on the CTS and includes two subscales (general therapeutic skills and CBT specific skills). The first nine items are each made up of six aspects rated on a 2-point scale (1 = *present/appropriately omitted* and 0 = *absent*). The final item relating to the quality of the intervention is rated on a 6-point scale (1 = *barely acceptable level of skill* and 6 = *excellent*). The MACT-RS has 11 items rated on a 7-point scale (response options are item specific) and all items fall under one subscale (skill in delivering self-harm treatment techniques). The CTCS-SP has 16 items rated on a 7-point scale (0 = *poor* and 6 = *excellent)* with two subscales (general therapeutic skills and CBT specific skills).

Six measurement properties were reported for the YACS (Carroll et al., 2000), one was reported for the MCSTPD-GCI (Huppert et al., 2001), four were reported for the CTS-Psy (Haddock et al., 2001), two were reported for the MACT-RS (Davidson et al., 2004), and three were reported for the CTCS-SP (von Consbruch, Clark, & Stangier, 2012). The YACS demonstrated excellent overall inter-rater reliability (ICC = .88) and ICCs for individual items ranged from .06 to .81. The CTS-Psy also demonstrated excellent overall inter-rater reliability (ICC = .94) and ICCs for individual items ranged from .41 (choice of intervention) to .95 (agenda). The MACT-RS demonstrated good inter-rater reliability (ICC = .66) and ICCs for individual items were not reported. The CTCS-SP demonstrated good to excellent overall inter-rater reliability (ICCs ranged from .73 to .88) and ICCs for individual items ranged from –.06 (feedback and summaries) to .98

(review of diary and questionnaires) for a pair of raters. The quality of the inter-rater reliability analysis on the COSMIN checklist was rated as 'very good' for the YACS, 'adequate' for the CTCS-SP, and 'doubtful' for the CTS-Psy and MACT-RS. The CTS-Psy was marked down due to the small sample size and the MACT-RS was due to not reporting ICCs for individual items. The CTCS-SP was the only adult disorder specific scale that measured internal consistency. This scale demonstrated good to excellent internal consistency (Cronbach's alpha coefficient ranged from .82 to .92) and the quality of this analysis was rated as 'very good' on the COSMIN checklist. Only the YACS and the CTS-Psy described the development of the scale; the CTS-Psy was rated as 'adequate' on the COSMIN checklist, whereas the YACS was rated as 'doubtful' because it was not clear if every item was considered for relevance. The YACS was the only adult disorder specific scale that measured structural validity; confirmatory factor analysis was conducted, which revealed that the subscales have good factor structure. However, as the sample size was less than five times the number of items on the scale, the quality of this analysis was rated as 'inadequate' on the COSMIN checklist. The YACS demonstrated good criterion validity and construct validity, and the quality was rated as 'very good' on the COSMIN checklist. The CTS-Psy also reported good construct validity, however the quality of this analysis was rated as 'doubtful' due to the small sample size.

1.3.6 Child and young person transdiagnostic scales.

Two transdiagnostic scales were identified that assess competence in delivering cognitive therapy with children and young people; the Cognitive Behaviour Therapy Scale for Children and Young People (CBTS-CYP; Stallard, Myles, & Branson, 2014) and the Global Therapist Competence Scale for Youth Psychosocial Treatment (GCOMP; Brown et al., 2018). Both of these scales assess therapist competence.

The CBTS-CYP was based on the CTS-R and adapted to reflect the use of CBT with children and young people. All skills are expected to be demonstrated in every session and are rated on a 7-point scale of therapist competence (0 = *incompetent* and 6 = *expert).* The first seven items comprise the process subscale, which includes the key aspects of CBT with children and young people, summarised by the acronym PRECISE (Partnership working, Right developmental level, Empathetic, Creative, Investigative, Self-efficacy, Enjoyable and engaging; Stallard, 2005). The remaining seven items comprise the method subscale, which includes assessment, formulation, and CBT specific skills. The GCOMP was designed to measure therapist global competence in psychological therapy with children and young people. Level 1 items consist of the five main domains (alliance building, positive expectancies, focussing treatment, instigating change, and responsiveness) and level 2 items are the 14 therapist behaviours that contribute to the level 1

domains. Items are rated on a 7-point Likert-scale (1 = *very poor* and 7 = *excellent*) and the scale is accompanied by a scoring manual.

Five measurement properties were reported for the CBTS-CYP (Stallard et al., 2014) and four properties were reported for the GCOMP (Brown et al., 2018). Both scales reported inter-rater reliability. The overall ICC for the CBTS-CYP is .96, indicating excellent inter-rater reliability. However, as reliability was based on 12 ratings of a single session and ICCs for individual items were not reported, the quality of reliability analysis was rated as 'doubtful' on the COSMIN checklist. Intraclass Correlation Coefficients on the GCOMP ranged from .61 (responsiveness) to .79 (instigating change), and the quality of the analysis was rated as 'very good' on the COSMIN checklist. Internal consistency was not reported by the GCOMP. The internal consistency of the CBTS-CYP was high, indicating considerable overlap between items. However, as Cronbach's alpha was not calculated, the quality of this analysis was rated as 'doubtful'. Content validity was rated as 'acceptable' on the COSMIN checklist for both scales. Construct validity was good for both scales and the quality of the analysis was rated as 'very good'. For the GCOMP inter-item corrections were conducted for the five domains, which ranged from .37 to .75, suggesting that therapists who are competent in one domain are likely to be competent in another.

1.3.7 Child and young person disorder specific scales.

Two disorder specific scales were identified that assess competence in delivering therapy with children and young people; the Competence and Adherence Scale for CBT (CAS-CBT; Bjaastad et al., 2015) and the Cognitive-Behavioural Treatment for Anxiety in Youth Competence Scale (CBAY-C; Mcleod et al., 2018). Both measures assess competence in delivering therapy for anxiety disorders. The CAS-CBT assessed therapist competence and adherence separately and Pearson's correlation coefficient indicated a strong correlation between adherence and competence (r = .79), whereas the CBAY-C measures therapist competence only.

The CAS-CBT was based on the structure of the CTACS. Seven items measure adherence on a 7-point scale (0 = *none* and 6 = *thorough*) and four items measure competence on a 7-point scale (0 = *poor skills* and 6 = *excellent skills*). There are supplementary items on global adherence, global competence, and how challenging the session was for the therapist. The CBAY-C has 25 items, each rated on a 7-point Likert-scale (0 = *not present* and 7 = *excellent*). Assessors are asked to consider the quality of delivery (skilfulness) as well as the timing and appropriateness of delivery for a given client and situation (responsiveness). The authors have developed a scoring manual to accompany the CBAY-C.

Five measurement properties were reported for the CAS-CBT (Bjaastad et al., 2015) and four properties were reported for the CBAY-C (McLeod et al., 2018). The CAS-CBT demonstrated good overall inter-rater reliability (ICC = .64) and ICCs for individual items ranged from .17 (session goal) to .90 (parental involvement). For the CBAY-C, the overall inter-rater reliability was also good (ICC = .67) and ICCs for individual items ranged from .37 (maintenance) to .80 (emotion education and self-reward). The inter-rater reliability analysis was rated as 'adequate' for the CAS-CBT and 'very good' for the CBAY-C on the COSMIN checklist. The CAS-CBT demonstrates good internal consistency (α = .87) and the quality of this analysis was rated as 'very good' on the COSMIN checklist. Internal consistency was not reported for the CBAY-C. Content validity was rated as 'acceptable' on the COSMIN checklist for both scales. Factor analysis was conducted for the CAS-CBT, which indicated a two-factor solution, and the quality of this analysis was rated as 'adequate' on the COSMIN checklist. The CBAY-C reported good construct validity and the quality of this analysis was rated as 'very good' on the COSMIN checklist, although scores on the two global items correlated highly with each other suggesting redundancy of these items.

1.4 Discussion

This review aimed to describe and evaluate published assessor-rated measures of therapist competence for cognitive and/or behavioural therapies. All twelve of the scales were developed to assess therapist competence in delivering CBT and there were none measuring competence in the implementation of purely behavioural therapies (e.g. BA). The COSMIN risk of bias checklist was used to evaluate methodological quality, and data was extracted from each study according to certain criteria.

1.4.1 Summary and critique of the competence scales.

All of the scales assessed therapist competence to CT or CBT, and there were no measures assessing therapist competence to a purely behavioural therapy. The majority of the scales had at least two subscales, which tended to be based on general therapeutic skills and CBT specific skills; this reflects the two categories of competences associated with treatment delivery (global competence and limited-domain competence) highlighted by Barber et al. (2007). For all scales, the items were not session specific, indicating that all items were expected to be seen within every session. The number of items on each scale varied. The MCSTPD-GCI (Huppert et al., 2001) includes just one item of global competence; the reliability and validity of this scale has not been examined and a single-item measure is unlikely to have the sensitivity to assess therapist competence. The scale with the largest number of items in this review is the CBAY-C (25 items). A scale of this length may be time-consuming to complete, depending on the time taken to score each item. For the CBAY-C, results supported the construct validity of the scale and the overall reliability was excellent (ICC = 0.75) but some of the items demonstrated lower reliability and global items were redundant. None of the scales reported time taken to complete the scale, which is a significant limitation for all scales in this review as researchers or clinicians are unable to assess feasibility. There is a need for comprehensive competence measures that are relatively quick and easy to use, and it is difficult to assess whether the current measures meet this need.

Many studies described the development of the scale including how items were identified but it was not always clear. It is important for potential users of the scale to be assured that the scale is based on theory and that items are relevant and representative of the overall construct. Scoring systems varied; most had a 7-point Likert scale (rated as 1-7 or 0-6) defined as either poor-excellent or incompetent-expert. The exceptions were the YACS, which had a 0-5 scale and the CTS-Psy which had 6 aspects for each item rated as either present or absent as well as the final item which was rated on a scale of 1-6. Most scales did not report a minimum clinical

standard, which makes it difficult for therapists and assessors to know what counts as an acceptable level of competence. For the CTS-R there is no validated competence threshold score, although a score of at least 2 on every individual item and a total score minimum score of 36 out of 72 (50%) is commonly implemented (James et al., 2001). For the MACT-RS (Davidson et al., 2004) therapists were classified into lower, mid, and high level of competence but it was unclear how these categories were set.

Some of the scales were accompanied by a scoring manual (CTS, CTS-R, YACS, GCOMP and CBAY-C), which included descriptions of items and guidelines for deciding between different points on the scale for each item. Although the CTCS-SP did not have a separate manual, the scale was detailed and included sufficient examples. The other scales were briefer and had item descriptions but no scoring examples. Having comprehensive and detailed scoring guidelines helps the assessor to distinguish between different points on the scale and can also be used in training to ensure consistency and reliability.

For many of the scales, ICCs for individual items were low. This may indicate a problem with reliability of the scale or insufficient assessor training in use of the scale. For example, on the CTACS, ICCs for individual items ranged from .34 to .92 with higher ICCs relating to therapeutic structure and lower ICCs related to collaboration and case conceptualization. On the CTS-R the lowest ICC for a pair of raters was -.14 (collaboration), and on the CTCS-SP, the lowest ICC was .06 (feedback and summaries). It may be that certain items (e.g. collaboration, case conceptualization, feedback and summaries) require more guidance than others and would benefit from detailed scoring instructions in the manual and/or an increased focus in training.

In two of the studies in this review, assessor training in using the scale was not reported, and in three studies training was mentioned but there was no description as to what this involved (see Table 1). The remaining studies provided descriptions of training, which included reading the manual, practice ratings, discussing discrepancies, attending training sessions and workshops, and ongoing discussions with the developers of the scale. Thorough training is essential to ensure a consistent approach, and should be reported so that those using the scale are aware of the training requirements and procedure. It would be helpful to have a recommendation of training requirements by the scale developers, so that researchers and clinicians are aware of what is involved and can assess feasibility.

Scales differed on focusing solely on competence, rating adherence and competence separately or combining adherence and competence into a single rating. Where adherence and

competence were rated separately (the CTACS, YACS, and CAS-CBT) the correlation between the two was examined. Authors examining the CTACS and the CAS-CBT found strong positive correlations between adherence and competence, and authors examining the YACS found a moderate positive correlation. It may be that therapists who are adherent to the particular approach are also able to deliver the treatment with a high level of skill or it could be that assessors have difficulty separating adherence from competence in their ratings. Waltz et al. (1993) propose that there is a considerable overlap between adherence and competence and differentiating between them may not be important because the main aim is to see if therapists are delivering the right treatment well. The CTS-R assesses adherence and competence in a single rating for each item by including a 7-point scale (0 = *non-adherence* and 6 = *adherence and very high skill*). This emphasises the importance of both adherence and competence and supports the idea that adherence is a necessary condition for competence, but adherence does not necessarily guarantee a high level of competence (Waltz et al., 1993).

Specific scales with developmentally appropriate criteria are required to measure therapist competence in treatment with young people (Fuggle, Dunsmuir, & Curry, 2013). There tend to be additional components compared to treatment with adults and it is not possible to capture fidelity to these components using adult scales (Bjaastad et al., 2015; Fjermestad, McLeod, Tully, & Liber, 2016). In this review, four scales were specific to young people. Some items on these scales are similar to CBT techniques with adults but have been adapted for young people (e.g. inclusion of a fear ladder instead of exposure). Some items are relevant to both age groups (e.g. positive reinforcement) and other items are unique to working with children and young people (e.g. parental involvement). The CBTS-CYP includes items defined by the PRECISE framework (Stallard, 2005), which comprises key aspects of CBT with children and young people.

McLeod et al. (2018) highlighted that competence may increase or decrease across treatment, and that measuring it at the beginning, middle and end of treatment is likely to be more representative than measuring it at one time point. In this review, some studies reported measuring competence at one time point only for each client, and for other studies it was unclear how many sessions were from the same course of therapy for a client. In the CTS-R the process of selecting three recordings from each of the clients (one at the beginning, middle, and end) was clearly described. In the study evaluating the CTS by Dobson et al. (1985) therapists selected a recording of their choice for analysis, which is likely to bias the results. In all studies (apart from two where it was not reported), full sessions were listened to rather than clips, which is likely to enable a more accurate rating of therapist competence.

1.4.2 Recommendations.

Based on the results of this review, the recommendation is to use the CTS-R as a transdiagnostic measure of therapist competence. This is because the scale demonstrated good internal consistency, discriminant validity, and overall inter-rater reliability, and the quality of each of the five psychometric properties evaluated was 'adequate' or 'very good'. The YACS would be recommended for substance abuse populations as it demonstrated good inter-rater reliability and discriminant validity and the quality of analysis was rated as 'very good'. The CTACS would also be recommended for substance abuse populations as it demonstrated good internal consistency, discriminant validity and inter-rater reliability and the quality of the analyses was rated as 'adequate' or 'very good'. The CTACS would not be recommended with nonsubstance abuse populations until the psychometric properties have been tested with this population. The MCSTPD-GCI and the MACT-RS would not be recommended until further psychometric evaluation has been conducted. The CTS-Psy and the CTCS-SP are promising but would benefit from further psychometric evaluation. In terms of scales developed specifically for measuring competence working with young people, all of the measures identified in this review evaluate at least four psychometric properties and the quality of all analyses was rated as 'adequate' or 'very good' for the GCOMP, CAS-CBT, and CBAY-C. However, it is suggested that these scales require refinement due to low reliability and redundancy of certain items or subscales. The CBTS-CYP requires further psychometric evaluation due to the quality of the internal consistency and inter-rater reliability analyses, which were rated as 'doubtful' on the COSMIN checklist.

1.4.3 Limitations of this review.

In the literature search, only studies that introduced an assessor-rated scale of therapist competence or fidelity and investigated the psychometric properties of the scale were included. It is good practice for researchers to evaluate and report the psychometric properties of a new measure when it is first included in a study; however, it is possible that subsequent studies examining the psychometric properties of the measure were missed. For example, subsequent studies may have evaluated structural or criterion validity, which were less commonly reported in this review. Furthermore, the authors who developed each scale reported on the psychometric properties; therefore, the results may be biased. It is possible that other important information was included in subsequent studies, such as time taken to complete the scale or minimum clinical standards. Widening the search to include all subsequent studies relating to the measures in this review was beyond the scope of the current review; however, it is important to acknowledge this as a possible limitation. Only peer-reviewed articles were included to ensure high quality and valid

research; however, it is important to consider possible publication bias whereby authors are more likely to publish studies that show significant findings.

The COSMIN risk of bias checklist was considered to be the most relevant quality assessment tool for this review, but it did not meet the requirements entirely. This tool was developed for patient-reported outcome measures, and thus it required adaptation to ensure relevance for assessor-rated measures. The COSMIN checklist assesses methodological quality, but does not take into account reported statistical findings. For example, internal consistency is scored as 'very good' on the COSMIN checklist if Cronbach's alpha is calculated, even if the reported statistic indicates poor internal consistency. In this review, reported statistics were considered in addition to the methodological quality rating in order to draw conclusions about the overall quality of each measure. It may have been preferable to develop a new tool based on the COSMIN checklist, and although outside the scope of the current project, future research could consider the development of an appropriate tool for this purpose. Additionally, as the COSMIN ratings were only completed by one reviewer, inter-rater reliability was not assessed and consequently the quality ratings are solely based on the judgement of one researcher.

1.4.4 Conclusion.

It is recommended that scales assessing competence in behavioural therapies are published. Time taken to complete the scale and training requirements should be reported in future scale development papers. Comprehensive assessor scoring instructions with clear item descriptions are important to ensure consistency and reliability. It is hoped that the findings of this review will provide clarity regarding which measures of therapist competence are valid and reliable, as well as highlighting priorities for future research.

Chapter 2: The Brief BA Fidelity Scale: Development, Psychometric Evaluation, and Link to Outcome and Alliance

2.1 Introduction

Anxiety and depression are common and serious mental health problems in young people. Sadler et al. (2018) conducted a national survey, which revealed that one in seven (14.4%) 11 to 16-year-olds and one in six (16.9%) 17 to 19-year-olds had a diagnosable mental health disorder. Emotional disorders (including anxiety and depression) were most common, present in 9.0% of 11 to 16-year-olds and 14.9% of 17 to 19-year-olds. If left untreated, emotional disorders in children and adolescents often persist into adulthood (Jones, 2013) and have a significant impact on well-being and development (Vogel, 2012). Depression is associated with a wide range of psychosocial difficulties in adulthood, including low income, unemployment, low perceived social support, and loneliness (Clayborne, Varin, & Colman, 2019). Early recognition and access to evidence-based interventions is essential, because providing effective treatment early on may prevent long-term negative outcomes.

2.1.1 Symptoms of depression and NICE guidelines.

The Diagnostic and Statistical Manual of Mental Disorders (5th ed.; DSM-5; American Psychiatric Association, 2013) outlines the symptoms of depression in young people. The main symptoms include depressed mood or irritability, and loss of interest or pleasure (anhedonia). Additional symptoms include sleep and appetite problems, psychomotor disturbances, fatigue or lack of energy, feelings of worthlessness or guilt, decreased concentration or indecisiveness, and recurrent thoughts of death or suicide. Five or more of these symptoms must be present during a two week period for a diagnosis of depression in children and adolescents.

National Institute for Health and Care Excellence (NICE, 2005) guidelines for the identification and management of depression in children and young people recommend a stepped care approach. Watchful waiting, guided self-help, and group Cognitive Behavioural Therapy (CBT) are recommended for mild depression, and psychological therapy is recommended for moderate to severe depression (individual CBT, interpersonal therapy, family therapy, or psychodynamic psychotherapy). Behavioural Activation (BA) is recommended as an intervention for adults with depression (NICE, 2009). The recently updated NICE guidelines for the treatment of depression in young people (2019) states that BA may “meet the needs of some children and young people with moderate to severe depression that are not already covered by the other

recommended psychological therapies" (p. 37) and a research recommendation has been made to investigate the effectiveness of BA in young people.

2.1.2 Behavioural activation.

Behavioural Activation (BA) is based on the behavioural theory of depression, which suggests that low mood is maintained by reinforcement for depressed behaviour and lack of reinforcement for non-depressed behaviour (Ferster, 1973; Lewinsohn, 1974). The aim of BA is to improve mood by increasing positive reinforcement for healthy behaviours. Jacobson et al. (1996) conducted a study examining the effective components in CBT for depression and found no significant differences in treatment outcome between CBT and BA alone. Subsequently, two contemporary BA approaches for adults with depression were developed. One approach consists of between 20-24 sessions and includes a detailed functional analysis as well as consideration of avoidance and approach behaviours (BA; Martell, Addis, & Jacobson, 2001). The other is a brief and more structured treatment (between 5-10 sessions) without functional analysis and with a focus on values (Behavioral Activation Treatment for Depression, BATD: Lejuez, Hopko and Hopko, 2001; Lejuez, Hopko, Acierno, Daughters and Pagoto, 2011).

Pass and Reynolds (2014) developed Brief Behavioural Activation (Brief BA), which is a manualised intervention for young people experiencing symptoms of depression. Brief BA was adapted from BATD to create a streamlined intervention that is appropriate and relevant for young people. The aim of Brief BA is to increase the young person's exposure to positive reinforcement by helping them to identify their personal values and engage in valued activities. It is based on behavioural theory and there are no cognitive techniques involved. Pass, Lejuez, and Reynolds (2017) conducted a pilot study of Brief BA for adolescents in a routine CAMHS setting, and found a significant decrease in depression symptoms and significant increase in functioning after receiving Brief BA. Reports from young people and their parents indicated high levels of satisfaction with the approach. Brief BA case studies provide evidence to suggest that Brief BA is effective for young people with complexity and risk (Pass, Whitney, & Reynolds, 2016) and it can be delivered by non-specialist clinicians (Pass, Hodgson, Whitney, & Reynolds, 2017) as training and delivery are relatively straightforward. Initial data from a feasibility study in schools (Pass, Sancho, Brett, Jones, & Reynolds, 2018) and a case study describing the outcomes for two young people with symptoms of depression (Totman, Reynolds, Brett, & Pass, 2019) suggest that Brief BA can be successfully delivered in schools. This is important in terms of increasing accessibility to evidence-based interventions.

2.1.3 Measurement of treatment fidelity.

In order to draw conclusions about the efficacy and effectiveness of a psychological treatment, it is important to understand whether the therapist delivered the approach with an acceptable level of fidelity. Treatment fidelity refers to "the extent to which a treatment was implemented as intended" (Fairburn & Cooper, 2011, p. 373). Fairburn and Cooper (2011) describe two aspects of treatment fidelity: therapist adherence and competence. Adherence is defined as "whether the right psychotherapeutic procedures were used" and competence is "how well the chosen procedures were implemented" (p. 373). A therapist who delivers an intervention with a high level of competence demonstrates skilful ability to take into account contextual factors, such as client engagement and understanding (Barber et al., 2007). National Institute for Health and Care Excellence (NICE) guidelines for the treatment of depression in young people state that interventions "should be provided by healthcare professionals who have been trained to an appropriate level of competence in the specific modality of psychological therapy being offered" (NICE, 2005, p. 15). Measuring treatment fidelity is a means of evaluating therapist training and ensuring that appropriate psychological treatments are provided in a skilful manner. There is also the potential to provide feedback to therapists and inform future training.

2.1.4 Existing measures of treatment fidelity.

Several assessor-rated measures of therapist competence or fidelity in CBT have been developed including transdiagnostic and disorder-specific scales (see Chapter 1 for a detailed review of these measures). The Cognitive Therapy Scale (CTS; Young & Beck, 1980) is a transdiagnostic scale designed to evaluate therapist competence in implementing general therapeutic skills and specific CBT skills. Limitations of the CTS were highlighted and the Cognitive Therapy Scale-Revised (CTS-R; Blackburn et al., 2001) was developed to improve the psychometric properties of the scale. Blackburn et al. (2001) assessed competence using the CTS-R in a sample of mental health professionals undergoing CBT training The mean competence score was 35.1 (*SD* = 7.2) at the earlier stage of training and 38.9 (*SD* = 5.9) at the later stage of training and there was an increase in competence across treatment. The CTS-R demonstrated high internal consistency and adequate inter-rater reliability. The CTS-R is commonly used in research and clinical practice, and as an assessment tool for CBT training.

Other transdiagnostic and disorder specific scales have also been developed. For example, the Cognitive Therapy Adherence and Competence Scale (CTACS; Barber et al., 2003) is a valid and reliable scale, however the psychometric properties have not yet been tested outside of the

substance abuse population. The Cognitive Therapy Scale for Psychosis (CTS-Psy; Haddock et al., 2001) and the Cognitive Therapy Competence Scale for Social Phobia (CTCS-SP; Clark et al., 2007) are promising disorder-specific scales; however, require further psychometric testing.

Specific measures have been designed to measure treatment fidelity in CBT when working with children and young people (see Chapter 1). For example, the Global Therapist Competence Scale for Youth Psychosocial Treatment (GCOMP; Brown et al., 2018), the Competence and Adherence Scale for CBT for anxiety disorders in youth (CAS-CBT; Bjaastad et al., 2015), and the Cognitive-Behavioural Treatment for Anxiety in Youth Competence Scale: CBAY-C (McLeod et al., 2018). These scales include developmentally appropriate criteria for measuring competence when working with young people and initial psychometric properties are promising. McLeod et al. (2018) highlight that competence is likely to change across treatment and it is important to measure it at more than one time point.

Whilst there are several measures of treatment competence for CBT, there are currently no published adult or young person measures of BA competence or fidelity. Research studies examining the effectiveness of BA for the treatment of depression in adults tend to measure adherence using a checklist (Dimidjian et al., 2006; Ekers, Richards, McMillan, Bland, & Gilbody, 2011) but do not measure competence.

2.1.5 Treatment fidelity and therapeutic alliance.

Therapist competence is hypothesised to promote a strong alliance but there is limited research in this area (Fjermestad et al., 2016). Ackerman and Hilsenroth (2003) conducted a review of therapist characteristics and techniques that positively influence the therapeutic alliance and concluded that therapist attributes such as being flexible, warm, and interested, and techniques such as exploration, reflection, and attending to the patient's experience contributed to the therapeutic alliance. These are aspects of global competence applicable across treatment models (Barber et al., 2007).

Creed and Kendall (2005) attempted to identify aspects of therapy associated with the alliance in CBT for child anxiety. An observational tool, the Therapist Alliance Building Behaviour Scale (TABBS) was developed to identify specific aspects of therapy. Youth and therapist rated alliance were rated using the Therapeutic Alliance Scale for Children (TASC; Shirk & Saiz, 1992). Higher scores for collaboration predicted higher youth-rated alliance early in treatment. Higher scores for collaboration and informality were linked to therapist-rated alliance later in treatment. Aspects of therapy that were not associated with child ratings of the therapeutic alliance included

the therapist being playful and providing hope and encouragement. The authors acknowledged that there was little variance in child ratings of the therapeutic alliance, which may mean it was difficult to detect significant relationships between therapist behaviours and child self-reported alliance. However, this study highlights that specific elements of therapy are likely to influence the client and therapist perspective of the therapeutic alliance.

The studies above include aspects of treatment fidelity included in the CTS-R (e.g. interpersonal effectiveness and collaboration) that are linked to alliance. Furthermore, in the CTS-R manual under collaboration it states that "in order to achieve a good therapeutic alliance, the therapist must assess the patient's needs, and particularly his/her preferred modes of learning" (James et al., 2001, p. 8). This suggests that a high score on the CTS-R may be associated with a good therapeutic alliance. However, there are no published studies investigating this.

2.1.6 Treatment fidelity and client outcome.

As therapist competence is measured to ensure high quality treatment provision with the aim of improving client mental health difficulties, a relationship between therapist competence and client outcome would be expected. Webb, DeRubeis and Barber (2010) conducted a meta-analysis examining the relationship between fidelity and outcome. The relationship between adherence and outcome was non-significant with a small effect size (r = .02), and the relationship between competence and outcome was non-significant with a small effect size (r = .07). When interventions specifically for depression were analysed, there was no significant relationship between adherence and outcome but a significant positive correlation between competence and outcome emerged. Zarafonitis-Müller, Kuhr, and Bechdolf (2014) found a small but significant effect of therapist competence on client outcome (r = .24) when the analysis was conducted across a range of disorders, and a moderate effect (r = .38) with depression interventions alone. There was no significant effect of adherence on client outcome. These meta-analyses suggest that adherence alone may not be associated with client outcome, but level of skill is associated with outcome and particularly within the context of treatment for depression.

Branson, Shafran, and Myles (2015) assessed competence in a sample of 43 therapists who were enrolled on an Improving Access to Psychological Therapies (IAPT) training programme. Each therapist submitted three recordings of CBT sessions (one from the beginning, middle, and end of training) and competence was assessed using the CTS-R. Results indicated that there was a significant increase in competence over the year of training. There was no significant association between overall competence score and improvement in symptoms of anxiety or depression.

However, significantly more clients of the most competent therapists (those scoring in the top 10%) demonstrated a reliable improvement in anxiety symptoms and significantly more clients treated by least competent therapists (those in the bottom 10%) demonstrated a reliable deterioration in symptoms of anxiety. The findings should be interpreted cautiously as competence was only rated at one time point for each client, which does not take into account change in competence across the course of therapy, thus the ratings may not be a true representation of competence. Furthermore, therapists selected the sessions themselves, which has the potential for bias because they are more likely to select sessions that they thought they did well in. However, competence scores at the beginning and end of treatment were comparable to the findings by Blackburn et al. (2001), indicating that the results are not biased. Kazantzis et al. (2018) compared the CTS and CTS-R as observational measures of therapist competence in a sample of 50 patients who received CBT for depression delivered in a Randomised Controlled Trial (RCT) context. Competence was measured by independent raters at two time points (early and late phase of treatment). Both the CTS and the CTS-R demonstrated acceptable reliability, and when assessed in the early phase of treatment, both predicted a statistically significant reduction in depression symptoms at the end of treatment. However, the relationship between competence and depression symptoms was not maintained at the 12-month or 24-month follow-up. Competence was significantly higher at the early phase of treatment than the late phase, and the authors discussed how it was unclear if the reduction in competence reflected a problematic decrease in level of skill or was a result of the therapist being less structured and shifting responsibility to the client as therapy progressed. The results of this study highlight that competence varies throughout treatment and is not a static construct.

It has been suggested that the mixed evidence in relation to the relationship between therapist competence and client outcome may be attributable to low reliability of competence measures (Perepletchikova & Kazdin, 2005; Webb et al., 2010). Furthermore, as research is often conducted in a controlled setting with highly trained and competent therapists, the resulting lack of variance in therapist competence may make it difficult to detect a relationship between competence and outcome (Hogue et al., 2008).

2.1.7 Study aims and research questions.

Development of the scale and psychometric evaluation. This study reports on the development and initial psychometric evaluation of the Brief BA fidelity scale, a measure designed to assess both therapist adherence and competence in the delivery of Brief BA for adolescent depression. The purpose was to develop a scale that is comprehensive enough to establish

adherence and competence but streamlined enough to be useable in routine clinical practice. The Brief BA fidelity scale was based on the CTS-R and adapted to be relevant and appropriate for Brief BA with young people. Inter-rater reliability, internal consistency, and face validity of the Brief BA fidelity scale were investigated to evaluate the extent to which the scale is an accurate representation of Brief BA fidelity.

Assessment of Brief BA fidelity. Once developed, the Brief BA fidelity scale was used to rate treatment sessions in the context of a small sample of therapists (*n* = 5) who delivered Brief BA to young people (*n* = 30) with depression symptoms in schools. The aim was to investigate whether therapists demonstrate an acceptable level of fidelity within this context. Fidelity scores at the beginning, middle, and end of therapy were measured to investigate whether there is a change in fidelity across therapy. Previous research suggests that competence changes over the course of therapy; however, the evidence is mixed with regards to whether it is an increase or decrease in fidelity.

Client outcomes from Brief BA. Change in client symptoms of depression and functioning from pre- to post- treatment were measured to explore the effectiveness of Brief BA in schools within this sample. Change in symptoms of depression and functioning were calculated by subtracting pre-treatment scores from post-treatment scores. Based on the results of a pilot study examining the effectiveness of Brief BA in a clinic setting (Pass et al., 2017) and a case study of Brief BA in schools (Totman et al., 2019), an improvement in client self-reported symptoms of depression and functioning following Brief BA was expected.

Brief BA fidelity and therapeutic alliance. The relationship between Brief BA fidelity and the young person's view of the therapeutic alliance at the beginning, middle and end stages of treatment were investigated. Research in this area is limited and it was not known whether there would be a relationship between Brief BA fidelity and the client-reported alliance.

Brief BA fidelity and client outcome. The relationship between overall Brief BA fidelity and change in client self-reported symptoms of depression and functioning was investigated. Evidence for a relationship between therapist competence and client outcome was mixed, so it was unclear whether there would be an association between fidelity and client outcome. The relationship between Brief BA fidelity at the beginning stage of therapy and change in client self-reported depression symptoms and self-reported functioning was also investigated. Based on the results of Kazantis et al. (2018), it was expected that fidelity at the beginning of treatment would be associated with a reduction in client depression symptoms. As functioning was not assessed by

Kazantis et al. (2018), the relationship between fidelity at the start of therapy and change in functioning was not known.

In summary, the research questions were as follows:

- What are the psychometric properties of the Brief BA fidelity scale (including inter-rater reliability, internal consistency, and face validity)?
- Is there an acceptable level of Brief BA fidelity in a small sample of therapists delivering Brief BA in schools, and does fidelity change across treatment?
- Is there an improvement in client self-reported symptoms of depression and functioning after receiving Brief BA in schools?
- Is there a relationship between Brief BA fidelity and the young person's view of the therapeutic alliance at the beginning, middle and end stage of treatment?
- Is there a relationship between overall Brief BA fidelity and change in client self-reported depression symptoms and functioning, and/or between Brief BA fidelity at the start of treatment and change in self-reported depression symptoms and functioning?

2.2 Method

2.2.1 Overarching study.

This study is part of a wider university research project looking at the effectiveness of Brief BA outreach work in schools as a way of increasing access to psychological therapy for adolescents with symptoms of depression. Inclusion criteria for the university research project were: elevated symptoms of depression measured by the Revised Child Anxiety and Depression Scale; RCADS (Chorpita, Yim, Moffitt, Umemoto, & Francis, 2000), Short Mood & Feelings Questionnaire; SMFQ (Angold et al., 1995) or the Kiddie Schedule for Affective Disorders and Schizophrenia diagnostic interview; K-SADS (Kaufman et al., 2013), help seeking (identified via a school survey of emotional health or school staff), and parental consent. Routine outcome measures to assess self-reported symptoms, functioning and experience of treatment were completed at every session. Ethical approval was obtained as part of the wider university research project, and this study received appropriate ethical approval from the University of Southampton (Study ID: 32114), see Appendix C.

2.2.2 Brief BA treatment, training, and supervision.

Brief BA involves the therapist and young person working together to identify the young person's values and scheduling valued activities to increase positive reinforcement for non-depressed behaviour. Brief BA is delivered using a session workbook (Pass & Reynolds, 2014), and a copy of the workbook is provided for the young person and their parent or carer. The Brief BA therapist guide and checklist are used (Pass, Brisco, Hodgson, & Reynolds, 2015); the therapist guide consists of an overview of each session, a list of outcome measures to be administered, and several 'frequently asked questions' and the therapist checklist contains a detailed list of what to cover in each session, including a reminder of the outcome measures to be completed. Brief BA in schools consists of between four to eight individual sessions with the young person, which last for up to an hour, and a review session approximately one month after the final session. The number of sessions is based on the school schedule and agreed with the young person and their parent or carer. The parent or carer does not attend the Brief BA sessions in schools, but are updated regularly by the therapist via telephone and email.

Training in Brief BA involves reviewing the workbook, therapist guide and checklist, and listening to a selection of audio recordings from treatment sessions. Therapists receive at least one full day of Brief BA training and additional training in working with young people and

specifically within the school setting. Brief BA supervision is typically weekly or fortnightly in a group setting and includes use of audio clips and case discussion.

2.2.3 Development of the Brief BA fidelity scale.

This study involved developing a measure of fidelity for Brief BA. The scale was based on the CTS-R as this is the most widely used scale in clinical practice and has good psychometric properties (Blackburn et al., 2001). The Brief BA fidelity scale has the same scoring system as the CTS-R and includes the five general therapeutic skills (agenda setting and adherence, feedback, collaboration, pacing and efficient use of time, and interpersonal effectiveness). Homework setting was also included, which is classed as a CBT specific skill in the CTS-R; however, it is also relevant to Brief BA. Item descriptions from the CTS-R were altered to reflect important aspects of therapy within the context of Brief BA with young people. For example, under collaboration it emphasises that the client should be involved in identifying their own values and the amount the client feels able to contribute may depend on their developmental level and the stage of therapy.

Four additional items were included in the scale: appropriate use of Brief BA techniques, positive reinforcement, clinical use of outcome measures, and risk assessment and management. Appropriate use of Brief BA techniques emphasises that Brief BA is about helping the young person do more of what matters to them, with techniques tailored to the individual, and the therapist should not engage in cognitive restructuring (Pass et al., 2015). Positive reinforcement is a key element of Brief BA and the therapist acknowledges small efforts or improvements the client has made in line with the behavioural model (Pass et al., 2015). Use of outcome monitoring is central to Children and Young People's Improving Access to Psychological Therapies (CYP-IAPT) to inform clinical practice and improve quality of care (Wolpert, Fugard, Deighton, & Görzig, 2012). This is integral to Brief BA as the CYP-IAPT principles were embedded into the approach. Consideration of risk is particularly important with this client group, as one of the key symptoms of depression in young people is suicidal ideation (Orchard, Pass, Marshall, & Reynolds, 2017). Therefore, to ensure the safety of the young person ongoing risk assessment and management is key.

The Brief BA fidelity scale was reviewed by four clinical psychologists, including the clinicians who developed Brief BA and deliver training and supervision in the approach. The scale was also reviewed by a Psychological Wellbeing Practitioner who has delivered Brief BA in schools, to ensure that the scale included aspects specific to delivering Brief BA in this context. Small additions and edits were made in response to feedback, including adding to the description

of items on the scale and editing the response options to provide further clarity for raters when deciding between different points of the scale (see Appendix D for the final version of the Brief BA fidelity scale, with additions highlighted).

The final scale consists of ten items, rated on a 7-point scale from 0 to 6 where a higher score indicates greater fidelity. It contains thorough descriptions of each item along with response options to help the rater make informed and reliable decisions. The same items are included regardless of the stage of treatment as all aspects are expected to be seen in every session. As Brief BA is purely a behavioural approach, evidence of cognitive techniques leads to a reduced score due to this going against the protocol. The maximum total score achievable is 60 and the suggested minimum standard is 30 out of 60 (50%) and a score of at least 2 on every item. This is based on the suggested minimum clinical standard for the CTS-R (the minimum score on the CTS-R is 36 out of 72, which is 50% of the total score). The time taken to listen to a full session recording and complete the scale is between one hour to one hour 15 minutes, depending on the length of the session. The Brief BA fidelity scale is accompanied by a scoring sheet (see Appendix E) for the assessor to provide comments as they listen to the recording.

2.2.4 Measures.

Revised Child Anxiety and Depression Scale (RCADS; Chorpita et al., 2000). The RCADS is a 47-item questionnaire measuring anxiety and depression in young people aged between 8-18 years (see Appendix F). Higher scores indicate more symptoms of anxiety and depression. The RCADS has good psychometric properties (Chorpita et al., 2000; Chorpita, Moffitt, & Gray, 2005). Raw scores are converted to t-scores according to the age and gender of the young person, which are used to determine whether their score is in the clinical range (t-scores of 0-64 are in the normal range, 65-69 are in the borderline range, and 70+ are in the clinical range). The full RCADS was completed at the assessment (or at the first Brief BA session) and review (or the final treatment session). The RCADS depression subscale was completed by the young person at the start of each session, which included an additional question to monitor risk ('I thought about killing myself').

Outcome Rating Scale (ORS; Miller & Duncan, 2000). The ORS is a self-report measure of functioning that is completed by the young person at the beginning of each Brief BA session (see Appendix G). The four areas of functioning are Individual, Interpersonal, Social, and Overall. Each of these areas are marked on a 10cm visual analog scale; they are combined to give a total score out of 40; higher scores indicate a higher level of functioning. The clinical cut-off score is 28

(scores below 28 are in the clinical range). Research indicates that the ORS has moderate to high reliability and moderate test-retest reliability (Bringhurst, Watson, Miller, & Duncan, 2006).

Session Rating Scale (SRS; Duncan et al., 2003). The SRS is a measure of the therapeutic alliance that is completed by the young person at the end of each session (see Appendix H). The four areas are Relationship, Goals and Topics, Approach or Method, and Overall Alliance. These areas are marked on four 10cm visual analog scales which are added together to give a total score; higher scores indicate higher client perception of therapeutic alliance. The SRS has good reliability, adequate validity, and high feasibility (Duncan et al., 2003).

2.2.5 Procedure for this research study.

Study design. Ninety full session recordings were rated by the primary author of this study (EH) using the Brief BA fidelity scale (three from each of the 30 participants). For each client, a beginning, middle and end Brief BA session was selected. For clients who received between 4 to 6 sessions, the first and last session were selected as well as a middle session from 2-5 (depending on how many there were in total). For clients who received 7 to 8 sessions, the beginning session was from 1-2, the end session was from 7-8, and the middle session was from 3-6. Sessions were selected to include a variety of sessions relating to the key concepts of Brief BA (e.g. activity scheduling, values, problem solving, and contracts). Selection was conducted blind to outcome measures. The content of the session was determined by listening to a short clip of the recording at the beginning, middle and end, which also allowed confirmation that the full session was recorded.

Participants. Participants were selected from a group of young people who received Brief BA in schools as part of a wider research project. Those who had an audio recording from a beginning, middle and end treatment session, and the required outcome measures were considered eligible. Thirty young people were selected to include a range of therapists, client ages, client genders, and schools. Proportions of these factors in the overall sample of young people who received Brief BA in schools were considered and clients for this study were selected to be representative. All clients gave their consent for sessions to be audio recorded and for their audio recordings to be used for research purposes. Participant Characteristics are presented in Table 4. The mean pre-treatment ORS score is in the clinical range (< 28) and the mean pre-treatment RCADS t-score is in the borderline range (65-69).

Table 4

Participant Characteristics

Participant Characteristic (*n* = 30)	Total
Gender	
Female	19 (63%)
Male	11 (37%)
Mean age in years (SD)	14.23 (1.46)
Age range in years	11-17
School year range	7-12
Ethnicity	
White British	21 (70%)
Mixed White and Black Caribbean	1 (3%)
Chinese	1 (3%)
Mixed White and Asian	4 (13%)
Asian British Indian	1 (3%)
Other mixed background	2 (7%)
Pre-treatment ORS	
Mean	19.93
Median	17.45
Mode	5.10
Pre-treatment RCADS t-scores	
Mean	68.31
Median	68.00
Mode	54

Therapists. There were five therapists in this study, who were trained Clinical Psychologists and Psychological Wellbeing Practitioners (PWPs). The sample included male and female therapists within the age range of 20-35. All therapists had received training in Brief BA and working with young people in the school setting. Therapists were supervised by one of the manual developers.

Raters. The author of this study was the primary rater (EH): A trainee Clinical Psychologist who was trained in Brief BA, had experience of implementing Brief BA with young people, and attended supervision sessions with one of the developers of the manual. The primary rater was involved in the development of the therapist guide and checklist for Brief BA and co-authored a published case study about the experience of delivering Brief BA as a non-specialist clinician. An

MSc student and Psychological Wellbeing Practitioner (LC) was recruited as a second rater for the inter-rater reliability analysis.

2.2.6 Inter-rater reliability.

The primary author and developer of the Brief BA fidelity scale (EH) provided training to the second rater (LC). Training involved becoming familiar with the Brief BA fidelity scale, Brief BA manual, therapist guide, and checklist. The learning phase involved EH and LC rating six Brief BA sessions independently (sessions were selected to include a range of therapists and stages of treatment). Discrepancies were discussed and ICCs were conducted for the six ratings. Scores were interpreted following the guidelines by Cicchetti (1994); less than .40 reflects poor agreement, .40 to .59 reflects fair agreement, .60 to .74 reflects good agreement, and .75 and higher reflects excellent agreement. As the total and individual items ICC indicated good to excellent inter-rater reliability, raters proceeded to the main reliability analysis.

Five participants from the total sample were selected at random for the main inter-rater reliability analysis and one recording was selected from the beginning, middle and end of therapy for each participant. Sessions were rated by EH and LC independently and ICCs were calculated, which indicated low reliability for some of the items. Several of the ratings were discussed and agreed between EH and the primary supervisor of this research project (one of the developers of the Brief BA manual). Discrepancies were then discussed between EH and LC, with the ratings by EH considered to be the 'gold standard'. An extra tool within the BA fidelity scale was subsequently developed, including examples of response options for each item to aid the second rater with the scoring process (see Appendix I). Next, five more participants from the total sample were selected at random and sessions were rated following the same procedure. Intraclass Correlation Coefficients from the learning stage, and the two subsequent stages are included in Appendix J.

2.2.7 Analysis plan.

The inter-rater reliability analysis was conducted with a subset of the total sample (15 session recordings from five participants). Inter-rater reliability was determined by calculating ICCs (3,1; Shrout & Fleiss, 1979), where ICC estimates and their 95% confidence intervals were calculated based on a single-measures, two-way mixed-effects model. Absolute consistency between raters was required.

All remaining analyses were conducted with the total sample (90 recordings from 30 participants). For correlational analyses where deviation from normality was evident in the

dataset, Pearson's correlation coefficient was conducted with bias corrected and accelerated (BCa) bootstrapped 95% confidence intervals. A significant result was shown by the fact that the bootstrapped confidence intervals did not cross zero.

The potential influence of client pre-treatment symptoms of depression and functioning on Brief BA fidelity was assessed. There was a slight negative correlation between client pre-treatment RCADS depression subscale score and Brief BA fidelity (a higher number of client depression symptoms at the start of treatment was associated with a lower level of therapist fidelity); however, this was not statistically significant (r = -.124, 95% CI [-4.86, .173]) with a small effect size (r value of .1 = small, .3 = medium, .5 = large; Cohen, 1992). There was a significant positive correlation between client pre-treatment ORS score and Brief BA fidelity (r = .371, 95% CI [.003, .675]), with a medium effect size (a higher level of client functioning at the start of treatment was associated with higher therapist fidelity). To assess the relationship between overall Brief BA fidelity and change in client self-reported symptoms of depression and functioning whilst controlling for the effect of client functioning at the start of treatment, partial correlations were conducted between Brief BA fidelity and change in RCADS depression subscale and ORS. For clarity of presentation, simple correlations are presented because the patterns of significance did not change after controlling for pre-treatment ORS.

2.3 Results

2.3.1 Psychometric properties.

Inter-rater reliability. Intraclass Correlation Coefficients from the final reliability analysis are presented in Table 5. The Brief BA fidelity scale demonstrated excellent inter-rater reliability (ICC = 0.93) for the total fidelity score and all individual items; ICCs ranged from 0.79 (collaboration) to 1.00 (risk).

Table 5

Brief BA Fidelity Scale Intraclass Correlation Coefficients, Means, and Standard Deviations

Item (*n* = 15)	ICC	95% Confidence Interval		*M*	*SD*
		Lower Bound	Upper Bound		
1. Agenda Setting and Keeping to the Agenda	.91	.76	.97	3.43	1.57
2. Feedback	.85	.62	.95	4.37	0.81
3. Collaboration	.79	.48	.92	4.50	0.68
4. Pacing and Efficient use of Time	.83	.57	.94	3.93	0.87
5. Interpersonal Effectiveness	.81	.54	.93	4.27	0.83
6. Positive Reinforcement	.94	.83	.98	4.33	1.03
7. Appropriate use of Brief BA Techniques	.80	.51	.93	4.30	0.70
8. Assigns Homework	.91	.75	.97	3.70	1.02
9. Clinical use of Outcome Measures	.91	.75	.97	4.37	1.03
10. Risk Assessment and Management	1.0	-	-	2.53	1.93
Total Fidelity Score	.93	.91	.95	3.97	1.24

Note. ICC = Intraclass Correlation Coefficient

Internal consistency. Cronbach's alpha was .81, indicating a good level of internal consistency of the scale. Removal of items 1-9 would result in a lower Cronbach's alpha, which indicates that these items should remain in the measure. Removal of item 10 (risk assessment and management) would lead to a small improvement in Cronbach's alpha (α = .89). However, the item was not removed because discussions with Brief BA experts and Clinical Psychologists experienced in working with adolescent depression indicated that it is an important aspect of fidelity. Furthermore, the scale had good internal consistency with the inclusion of this item. It is possible that risk assessment and management reflects a different aspect of fidelity to the other items.

Face validity. Feedback about the scale was very positive and suggestions were provided for additions to item descriptions, which were included in the final version of the scale. There was positive feedback about the inclusion of clinical use of outcome measures and risk assessment and management, which supports the inclusion of these items in the scale. The feedback suggests that the scale is an appropriate tool for assessing adherence and competence to Brief BA.

2.3.2 Brief BA treatment fidelity.

The total Brief BA fidelity score and individual item means, standard deviations, and ranges are presented in Table 6. The mean Brief BA fidelity score across all three sessions for all clients was 37.31 (*SD* = 5.65), which is above the suggested minimum standard of 30. The mean fidelity score across the three sessions was below the minimum standard for three of the 30 clients (10%). When looking at individual sessions, 77 out of 90 (86%) met the minimum total fidelity score of 30, and 60 out of 90 (67%) achieved a score of at least 2 on every item. The items that most commonly scored below 2 were agenda setting and risk assessment and management. Overall, 54 out of 90 sessions (60%) met both suggested minimum clinical standards (a minimum total fidelity score of 30 and a score of at least 2 on every item).

Table 6

Brief BA Fidelity Scale Means, Standard Deviations and Ranges

Item (*n* = 90)	*M*	*SD*	Range (0-6)
1. Agenda Setting and Keeping to the Agenda	3.17	1.38	0-5
2. Feedback	4.01	0.91	2-5
3. Collaboration	3.99	0.92	2-5
4. Pacing and Efficient use of Time	3.64	0.84	1-5
5. Interpersonal Effectiveness	3.93	0.76	3-5
6. Positive Reinforcement	3.77	1.07	2-6
7. Appropriate use of Brief BA Techniques	4.11	0.81	3-5
8. Assigns Homework	3.57	0.84	1-5
9. Clinical use of Outcome Measures	3.90	1.07	2-5
10. Risk Assessment and Management	3.22	1.72	0-5
Total Fidelity Score	37.31	5.65	27.66-47.33

Note. *n* = 90.

2.3.3 Change in fidelity across treatment.

Results of a one-way repeated measured ANOVA indicated that the mean fidelity scores differed significantly between time points, F (2, 58) = 3.329, p = .043, η^2 = .103. Post hoc tests using the Bonferroni correction revealed a slight decrease in fidelity from the beginning therapy session (M = 38.63, SD = 6.08) to the middle therapy session (M = 37.10, SD = 5.88); however, this was not statistically significant (p = .297). There was also a slight decrease in fidelity from the middle to end session (M = 36.20, SD = 7.18), which was not statistically significant (p = 1.00). The reduction in fidelity from the beginning to the end therapy session was also non-significant (p = .060).

2.3.4 Client outcomes from Brief BA.

T-tests revealed a significant reduction in client self-reported depression symptoms from pre- to post- treatment (see Tables 7 and 8), reflecting a large effect size (d value of 0.2 = small, 0.5 = medium, 0.8 = large; Cohen, 1992). The mean t-scores moved from the borderline range pre-treatment (M = 68.31, SD = 14.15), to the normal range at post-treatment (M = 55.83, SD = 17.42). There was also a significant increase in client self-reported functioning from pre- to post-treatment, with a large effect size.

Table 7

Pre- and Post-Treatment RCADS Depression Subscale Raw Scores and ORS Scores

Measure and time point	*n*	Mean raw score (*SD*)	Raw score range	Median raw score	Interquartile range
RCADS-Depression					
Pre-treatment	30	14.60 (5.44)	0-27	15	6.3
Post-treatment	30	9.80 (6.58)	0-27	9.5	10.5
ORS					
Pre-treatment	30	19.93 (9.25)	5.10-38.20	17.45	15.10
Post-treatment	30	29.68 (10.86)	6.10-40.00	34.60	18.95

Table 8

Pre- and Post-Treatment Comparisons of RCADS Depression Subscale Raw Scores and ORS Scores

Measure and comparison time points	*n*	*df*	*t*-value	BCa bootstrapped 95% CI	Cohen's *d*
RCADS-Depression					
Pre- to post-treatment	30	29	3.67	[2.34, 7.43]	0.80
ORS					
Pre- to post-treatment	30	29	5.63	[-12.92, -6.56]	0.97

df, degrees of freedom; CI, confidence interval.

Reliable and clinically significant change. Reliable change on the RCADS Depression subscale was shown by a change score greater than the published reliable change criterion based on the young person's age and gender (Law & Wolpert, 2014) and clinically significant change was defined as moving down a clinical category. Nine young people were already in the normal range on the RCADS Depression subscale pre-treatment, therefore clinically significant change was not possible. From the remaining 21, 10 (48%) showed reliable and clinically significant change. Three showed reliable improvement but did not change clinical category, and the others did not show reliable improvement. One young person appeared to show a reliable deterioration from pre- to post- treatment; however, inspection of the data revealed that the increase in depression symptoms occurred between the assessment and start of treatment.

Eight young people scored in the normal range (28+) on the ORS pre-treatment so clinically significant change was not possible. Twelve (55%) of the 22 young people who were able to evidence reliable and clinically significant change (an increase of more than five and move from clinical to normal range) demonstrated this from pre- to post-treatment. Another nine showed reliable improvement but their ORS scores stayed in the clinical range, and the rest did not show reliable improvement. None of the young people showed a reliable deterioration on the ORS.

2.3.5 Brief BA fidelity and therapeutic alliance.

Pearson's correlations indicated no significant correlation between SRS score at the beginning or middle of Brief BA with the same session Brief BA fidelity score (see Table 9). However, at the end of therapy there was a significant positive correlation between fidelity and SRS score with a medium effect size. This implies that in later sessions of Brief BA, a higher level of fidelity is associated with higher client-reported therapeutic alliance.

Table 9

Correlations Between Brief BA Fidelity Scores and Associated SRS Score at the Beginning, Middle and End of Treatment

Stage of Brief BA Therapy	*n*	*df*	*r* value	BCa bootstrapped 95% CI
Beginning	30	28	.256	[-.136, .173]
Middle	30	28	.317	[-.006, .602]
End	30	28	.422	[.134, .643]

df, degrees of freedom; CI, confidence interval.

2.3.6 Brief BA fidelity and client outcome.

Pearson's correlation indicated a slight positive correlation between Brief BA fidelity and improvement in self-reported depression symptoms; however, this was not statistically significant, r = .135, 95% CI [-.456, .233], and reflects a small effect size. There was a slight positive correlation between Brief BA fidelity and improvement in self-reported functioning; however, this was not statistically significant, r = .184, 95% CI [-.512, .207], and reflects a small effect size.

There was no significant relationship between Brief BA fidelity at the start of treatment and change in client self-reported depression symptoms, r = -.062, 95% CI [-3.92, .336]. There was no significant relationship between Brief BA fidelity at the start of treatment and change in client self-reported functioning, r = -.317, 95% CI [-.610, .046].

2.4 Discussion

This study reports on the development and initial evaluation of the Brief BA fidelity scale. The scale demonstrated excellent inter-rater reliability for the overall scale (ICC = 0.93) and individual items (ICCs ranged from 0.79 to 1.00). This is higher than the CTS-R (Blackburn et al., 2001) where the overall ICC ranged from 0.40 to 0.86 for pairs of raters, and for individual items ranged from -0.14 to 0.84. In the first stage of inter-rater reliability the overall Brief BA fidelity score was good (ICC = 0.71) but ICCs for individual items ranged from poor to excellent (0.21 to 0.96) and one item could not be calculated due to zero variance from the second rater. Inter-rater reliability for all individual items and the overall scale increased significantly from the first to the second stage following discussion of discrepancies and the addition of the Brief BA fidelity scale scoring examples document. This highlights the importance of having a detailed scale with specific examples to aid the rater with the scoring process.

Cronbach's alpha indicated a good level of internal consistency of the scale ($\alpha = 0.81$). Removal of item 10 (risk assessment and management) would lead to a small improvement in Cronbach's alpha ($\alpha = .89$). It may be that it measures a different aspect of fidelity to the other items. Given the clinical importance of risk assessment and management, particularly within the context of young people experiencing suicidal ideation, the decision was made for this item to remain in the scale. Feedback about the scale from Brief BA therapists and experts in Brief BA was very positive indicating good face validity and supporting the inclusion of all items.

2.4.1 Brief BA treatment fidelity.

The mean Brief BA fidelity score was 37.31 (*SD* = 5.65), which is above the suggested minimum standard of 30. For the CTS-R (Blackburn et al., 2001) the mean fidelity score at the earlier stage of training was 35.1 (*SD* = 7.2), which was just below the minimum standard of 36, and at the later stage of training was 38.9 (*SD* = 5.9), which was above the minimum standard. The mean Brief BA fidelity score is comparable to the mean score on the CTS-R after training, which is what would be expected given that the therapists were already trained in delivering Brief BA. This suggests that therapists are demonstrating an acceptable level of adherence and skill in delivering Brief BA in schools and supports the current training and supervision for therapists.

More than half of the individual sessions met both suggested minimum clinical standards (a total score of at least 30 and at least 2 for every item). The items that most commonly scored below 2 were agenda setting and risk assessment and management. Agenda setting was scored as 1 when significant items were missing from the agenda or 0 when an agenda was not set. It is

possible that an agenda was set before beginning the recording. Agenda setting is a core Brief BA skill that is included in the therapist guide and checklist. It sets expectations for the session and gives the opportunity for the client to consider what is most important for them to spend time on (Josefowitz & Myran, 2005). It is therefore recommended that the importance of agenda setting is emphasised in therapist training.

Risk assessment and management was sometimes absent and therefore was rated as 0. In Brief BA training therapists are advised that if no risk was identified at the initial assessment then it is not necessary to ask about risk verbally every session, as long as the additional question on the RCADS subscale is checked to ensure there has been no change. Therefore, this item may have been rated as 0 when it was appropriate for therapists not to ask about risk. It is also possible that risk assessment and management was discussed either before or after switching on the recorder. It is important to ensure that risk is considered, and the inclusion of risk assessment and management in the Brief BA fidelity scale emphasises this to clinicians and supervisors.

2.4.2 Change in fidelity across treatment.

There was a significant effect of stage of therapy on Brief BA fidelity, and mean fidelity scores indicated an overall decrease in Brief BA fidelity across treatment. This is consistent with Kazantis et al. (2018) who found a reduction in CTS and CTS-R scores from the early to late CBT treatment phase; however, it is contrary to Blackburn et al. (2001) who found an increase in fidelity across CBT treatment. It highlights that Brief BA fidelity varies across treatment, which emphasises the importance of rating fidelity at several sessions for each client to ensure a representative measurement. It may be that the decrease in Brief BA fidelity is a result of less structure and a shift in responsibility from therapist to client as sessions progress, as discussed by Kazantis et al. (2018). If this is the case, the decrease in fidelity is not problematic and the scale may need to be adapted to take this into account.

2.4.3 Client pre-treatment functioning and Brief BA fidelity.

There was a significant positive relationship between client pre-treatment ORS score and Brief BA fidelity score, indicating that higher client functioning is associated with higher treatment fidelity. This suggests that when the client's symptoms of depression are having a significant impact on their functioning, delivering Brief BA with a high level of skill may be more difficult. This emphasises the importance of considering client factors as well as therapist factors when rating treatment fidelity, as highlighted in the CTS-R (Blackburn et al., 2001). Flexibility and adjusting the approach according to the client's needs are emphasised in the Brief BA fidelity scale, and the

rater is reminded to take into account client difficulties (e.g. high levels of emotion and avoidance). Although it is possible for therapists to demonstrate high fidelity when client functioning is low, it might be more difficult for certain items, for example, those relating to pacing and therapy structure.

2.4.4 Client outcomes from Brief BA.

There was a significant improvement in client self-reported symptoms of depression and functioning following Brief BA. These outcomes are consistent with previous research examining the effectiveness of Brief BA in a routine CAMHS setting (Pass et al., 2017) suggesting that Brief BA is a promising intervention in the treatment of depression symptoms in young people. It also provides support for the suggestion that Brief BA can be delivered effectively in schools, which is important in terms of providing effective early support and improving access to specialist services for children and young people (Department of Health and Social Care and Department for Education, 2017). Further research is required to investigate the effectiveness of Brief BA, including a Randomised Controlled Trial (RCT) comparing Brief BA to CBT and a wait list control group.

2.4.5 Brief BA fidelity and therapeutic alliance.

There was no significant correlation between Brief BA fidelity and client-reported therapeutic alliance at the beginning or middle stage of therapy. However, at the end stage of therapy there was a significant positive correlation, with a higher level of Brief BA fidelity associated with higher client-reported alliance. The lack of variability in SRS scores and fidelity scores is likely to account for the non-significant correlations. As overall fidelity was lower at the end stage of treatment, it is likely that there was more variance resulting in a significant correlation. The relationship between Brief BA fidelity and alliance might have been significant at all stages of treatment if the variance was higher, although it is also possible that fidelity is particularly important for the client at the end of treatment.

2.4.6 Brief BA fidelity and client outcome.

Correlations between Brief BA fidelity and improvement in depression symptoms, and between Brief BA fidelity and improvement in self-reported functioning were non-significant. As therapists were working in the same research setting, received the same training, and had the same supervisor (one of the developers of Brief BA), fidelity was generally high and there was limited variance in the sample of therapists. Furthermore, there was little variance in client outcome due to the high number of young people demonstrating improvements. Although it is

positive that fidelity was high and client outcomes were good, the lack of variance in the sample may have resulted in non-significant correlations.

2.4.7 Limitations and future research.

This study was an important first step in evaluating the scale with a sample of therapists who were likely to demonstrate high Brief BA fidelity. Future research should assess fidelity, alliance, and client outcomes in a wider sample of Brief BA therapists, which may have greater variance and power to detect an effect if it is present. Possible moderators of the relationship between treatment fidelity and client outcomes should be explored, including the therapeutic alliance and the ability of the therapist to adapt to contextual factors, such as client engagement. It would also be interesting to investigate how fidelity changes over time, to see whether Brief BA training and supervision improves fidelity. It is possible that novice therapists may find it easier to show higher Brief BA fidelity than expert therapists who may be more aligned to other, more complex and integrative therapeutic styles. The Brief BA fidelity scale allows identification of aspects of Brief BA delivered with less skill to support the therapist in developing these skills through training and supervision. Therapists may require guidance in delivering the approach flexibly and identifying when deviation from the protocol is helpful or not. Future qualitative research could explore therapist experiences of delivering Brief BA.

It is important to consider the results within the context of young people who were experiencing depression symptoms but did not necessarily have a diagnosis of depression. The mean pre-treatment RCADS depression t-score was in the borderline range. The aim of the study was to examine fidelity in a small sample of therapists delivering Brief BA in schools with young people experiencing depression symptoms; therefore, the client sample was appropriate. Pass et al. (2017) examined the effectiveness of Brief BA in a routine CAMHS setting with young people experiencing depression symptoms, and in this study the mean pre-treatment RCADS depression t-score was in the clinical range. Future research may wish to examine the effectiveness of Brief BA in a sample of young people with a diagnosis of depression.

Brief BA is delivered in both clinic settings and schools. In the clinic setting, parents are involved in certain sessions. The fidelity scale would need to be adapted slightly for use in a clinic setting, by considering how therapists should include parents and how to manage having both the client and parent in the room. Future research could consider adapting the scale so that it can be used to assess fidelity in both settings, and conducting a pilot study to evaluate use of the Brief BA fidelity scale in the clinic setting with parental involvement in sessions.

One young person appeared to show a reliable deterioration in client-reported depression symptoms, which occurred between the assessment and start of treatment. Reasons for this are unknown, but one hypothesis is that the young person experienced a difficult life event between the assessment and start of treatment. The data also show that for some young people there was a reliable improvement in depression symptoms between the assessment and session 1. For these young people, talking about their difficulties at the assessment may have resulted in an improvement in depression symptoms. Although it was outside the scope of the current study, future studies evaluating the effectiveness of Brief BA may wish to consider the stage at which symptom change occurs and reasons for this.

Items on the Brief BA fidelity scale did not demonstrate the full range of scores for every item. Positive reinforcement was the only item rated as a 6, and this was for just two treatment sessions. On the Brief BA fidelity scale, it states that ratings of 6 are only provided when therapists demonstrate exceptional skills, particularly in the face of client difficulties. In the CTS-R manual (James et al., 2001) assessors are advised that the scoring profile should approximate a normal distribution with relatively few therapists scoring at the extremes. It is possible that raters were hesitant to give scores of 6 even when it was warranted. In the Brief BA fidelity scale scoring examples document, examples are given for the most commonly rated scores to help raters to distinguish between ratings. As no items were rated as 6 in the development stage, there were no examples of what this would look like, therefore raters may have been less likely to score items as 6 in the main coding stage. Further work is required to provide examples of 6 for each item, for inclusion in the scoring examples document. It is important that raters receive adequate training in use of the Brief BA fidelity scale; this would include reviewing the scale and scoring examples document, receiving instruction in use of the scale by expert raters, rating sessions independently, and discussing discrepancies with expert raters. The aim of this study was to develop a comprehensive measure of Brief BA fidelity that was streamlined enough to be useable in supervision and clinical work. Although listening to one-hour sessions was time-consuming, the scale itself took approximately 15 minutes to complete. It was not possible to compare the time taken to complete this scale with other scales, as it is not commonly reported. Future research could investigate whether reliable and valid ratings can be obtained by assessing clips of sessions, to reduce the time required for evaluation.

2.4.8 Conclusion.

The development of a Brief BA fidelity scale has made a novel contribution to the research, and initial exploration of the psychometric properties indicates that the scale is a

reliable and valid measure. Results of this study suggest that therapists are demonstrating an acceptable level of fidelity and provide evidence to support the effectiveness of Brief BA in the school setting, which is important in terms of early intervention and improving access to interventions for young people with depression symptoms. It is hoped that the Brief BA fidelity scale will be useful in informing Brief BA training and supervision, and maximising clinical effectiveness to make a positive difference to young people's lives.

Appendix A COSMIN Risk of Bias Checklist- adapted version

Box 1. Measure development

1a. Measure design

General design requirements

1. Is a clear description provided of the construct to be measured?
   Very good or Inadequate
2. Is the origin of the construct clear: was a theory, conceptual framework, or disease model used or clear rationale provided to define the construct to be measured?
   Very good or Doubtful
3. Is a clear description provided of the target population for which the measure was developed?
   Very good or Inadequate
4. Is a clear description provided of the context of use?
   Very good or Doubtful
5. Was the measure development study performed in a sample representing the target population for which the measure was developed?
   Very good, Adequate, Doubtful, or Inadequate

Concept elicitation

6. Was an appropriate method used to identify relevant items for a new measure?
   Very good, Adequate, Doubtful, or Inadequate
7. Were skilled group moderators/interviewers used?
   Very good, Adequate, Doubtful, or N/a
8. Were the group meetings or interviews based on an appropriate topic or interview guide?
   Very good, Adequate, Doubtful, or N/a
9. Were the group meetings or interviews recorded and transcribed verbatim?
   Very good, Adequate, Doubtful, Inadequate, or N/a
10. Was an appropriate approach used to analyse the data?
    Very good, Adequate, Doubtful, or Inadequate
11. Was at least part of the data coded independently?
    Very good, Adequate, Doubtful, Inadequate, or N/a
12. Was data collection continued until saturation was reached?
    Very good, Adequate, Doubtful, Inadequate, or N/a
13. For quantitative studies (surveys): was the sample size appropriate?
    Very good, Adequate, Doubtful, Inadequate, or N/a

Box 2. Content validity

2d. Asking professionals about relevance

Design requirement

22. Was an appropriate method used to ask professionals whether each item is relevant for the construct of interest?

Very good, Adequate, Doubtful, or Inadequate

2e. Asking professionals about comprehensiveness

Design requirement

27. Was an appropriate method used for assessing the comprehensiveness of the measure?
    Very good, Adequate, Doubtful, or Inadequate

2f. Asking professionals about comprehensibility

Design requirement

Was an appropriate qualitative method used for assessing the comprehensibility of the measure?
Very good, Adequate, Doubtful, or Inadequate

Box 3. Structural validity

Statistical methods

1. For CTT: was exploratory or confirmatory factor analysis performed?
   Very good, Adequate, Inadequate or N/a
2. For IRT/Rasch: does the chosen model fit to the research question?
   Very good, Adequate, Doubtful, Inadequate or N/a
3. Was the sample size included in the analysis adequate?
   Very good, Adequate, Doubtful, or Inadequate
4. Were there any other important flaws in the design or statistical methods of the study?
   Very good, Doubtful, or Inadequate

Box 4. Internal consistency

Design requirements

1. Was an internal consistency statistic calculated for each unidimensional scale or subscale separately?
   Very good, Doubtful, or Inadequate

Statistical methods

2. For continuous scores: was Cronbach's alpha or omega calculated?
   Very good, Doubtful, Inadequate or N/a
3. For dichotomous scores: Was Cronbach's alpha or KR20 calculated?
   Very good, Doubtful, Inadequate or N/a
4. For IRT-based scores: was standard error of the theta (SE (θ)) or reliability coefficient of estimated latent trait value (index of (subject or item) separation) calculated?
   Very good, Inadequate or N/a
5. Were there any other important flaws in the design or statistical methods of the study?
   Very good, Doubtful, or Inadequate

Box 6. Reliability

Design requirements

3. Were the test conditions similar for the measurements? e.g. type of administration, environment, instructions (rater guidelines/consider if they rated separately)
   Very good, Adequate, Doubtful, or Inadequate

Statistical methods

4. For continuous scores: Was an intraclass correlation coefficient (ICC) calculated? (if calculated but no details on the ICC model go for adequate)
   Very good, Adequate, Doubtful, Inadequate or N/a
5. For dichotomous/nominal/ordinal scores: Was kappa calculated?
   Very good, Inadequate or N/a
6. For ordinal scores: Was a weighted kappa calculated?
   Very good, Doubtful or N/a
7. For ordinal scores: Was the weighting scheme described? e.g. linear, quadratic
   Very good, adequate or N/a

Other

8. Were there any other important flaws in the design or statistical methods of the study?
   Very good, Doubtful, or Inadequate

Box 8. Criterion validity

Statistical methods

1. For continuous scores: Were correlations, or the area under the receiver operating curve calculated? (AUC- ROC curve analysis)
   Very good, Inadequate, or N/a
2. For dichotomous scores: Were sensitivity and specificity determined?
   Very good, Inadequate, or N/a

Other

3. Were there any other important flaws in the design or statistical methods of the study?
   Very good, Doubtful, or Inadequate

Box 9. Hypothesis testing for construct validity

9a. Comparison with other outcome measurement instruments (convergent validity)
Design requirements

1. Is it clear what the comparator instrument(s) measures?
   Very good or Inadequate
2. Were the measurement properties of the comparator instrument(s) sufficient?
   Very good, Adequate, Doubtful, or Inadequate

Statistical methods

3. Was the statistical method appropriate for the hypotheses to be tested?
   Very good, Adequate, Doubtful, or Inadequate

Other

4. Were there any other important flaws in the design or statistical methods of the study?
   Very good, Doubtful, or Inadequate

9b. Comparison between subgroups (discriminative or known-groups validity)

Design requirements

5. Was an adequate description provided of important characteristics of the subgroups?
   Very good, Adequate, or Doubtful

Statistical methods

6. Was the statistical method appropriate for the hypotheses to be tested?
   Very good, Adequate, Doubtful, or Inadequate

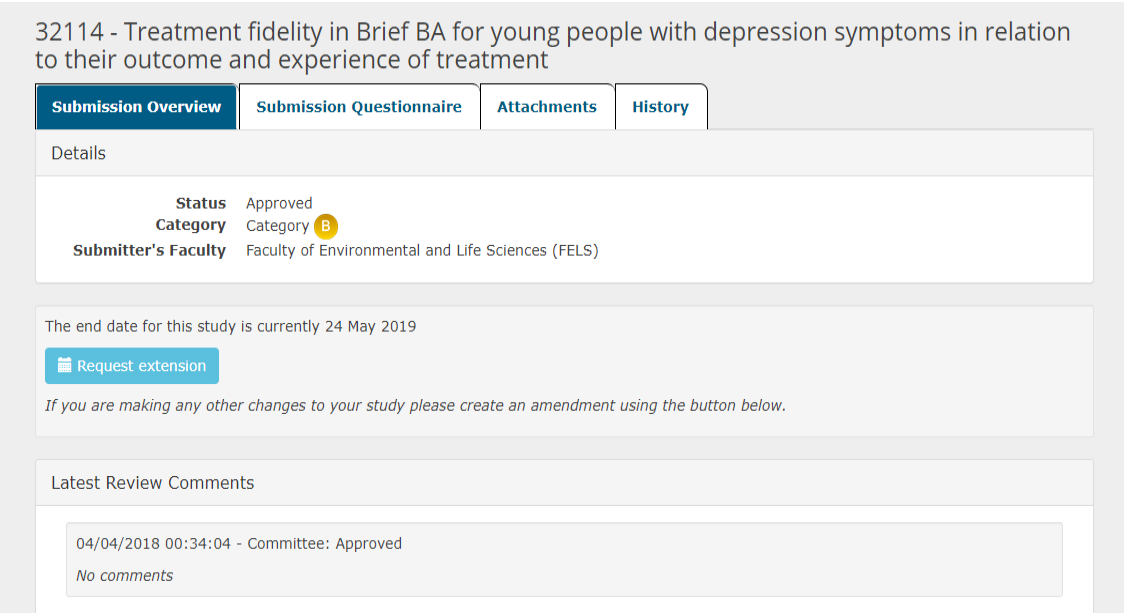
Other

7. Were there any other important flaws in the design or statistical methods of the study?
   Very good, Doubtful, or Inadequate

Appendix B Definitions of Psychometric Properties

Psychometric Property	Definition
Inter-rater reliability	The extent to which scores for patients are the same for repeated measurement by different persons on the same occasion
Internal consistency	The degree of the interrelatedness among the items
Content validity	The degree to which the content of a measure is an adequate reflection of the construct to be measured
Structural validity	The degree to which the scores of a measure are an adequate reflection of the dimensionality of the construct to be measured
Criterion validity	The degree to which the scores of a measure are an adequate reflection of a 'gold standard'
Construct validity	The degree to which the scores of a measure are consistent with hypotheses (for instance with regard to internal relationships, relationships to scores of other instruments, or differences between relevant groups)

Appendix C University of Southampton Ethical Approval



32114 - Treatment fidelity in Brief BA for young people with depression symptoms in relation to their outcome and experience of treatment

Submission Overview | Submission Questionnaire | Attachments | History

Details

Status Approved
Category Category B
Submitter's Faculty Faculty of Environmental and Life Sciences (FELS)

The end date for this study is currently 24 May 2019

Request extension

If you are making any other changes to your study please create an amendment using the button below.

Latest Review Comments

04/04/2018 00:34:04 - Committee: Approved

No comments

Appendix D Brief BA Fidelity Scale

Brief BA Fidelity Scale

This scale is designed to measure treatment fidelity to Brief BA, which is a manualised intervention for young people with depression or low mood (Pass & Reynolds, 2014). The scale measures therapist adherence to the specific treatment approach as well as therapist competence (the skilfulness of treatment delivery). The 7-point scale extends from (0) where the therapist did not adhere to that aspect of therapy to (6) where there is full adherence and a very high level of skill. The same scale can be used for any session of Brief BA. It is based on the Cognitive Therapy Rating Scale-Revised (CTS-R; Blackburn et al., 2000) and has been adapted for Behavioural Activation.

There are 10 items. Each item is rated on a Likert scale, ranging from 0-6. A general explanation of the scoring system is provided below and a description for each level is provided to help guide your decision. The scoring system is the same as the CTS-R.

Fidelity Level	Description
0	Absence of feature or highly inappropriate
1	Inappropriate with major problems evident
2	Evidence of competence, but numerous problems and lack of consistency
3	Competent, but some problems and/or inconsistencies
4	Good features, but minor problems and/or inconsistencies
5	Very good features, minimal problems and/or inconsistencies
6	Excellent, or very good even in the face of client difficulties

For all items, focus on the skill of the therapist, taking into account client difficulties (such as high levels of emotion, aggression, and avoidance) as well as therapist flexibility (the degree to which the therapist adjusts/tailors the intervention to the individual, taking into account the client's mood, the stage of therapy, level of engagement, and their developmental level).

Please note that ratings of 6 are only provided when therapists demonstrate exceptional skills, particularly in the face of client difficulties. As detailed on the scale, competent practice is considered to start at a 3 on the scale.

Maximum score on the scale is 60 (10 x 6). The minimum standard is 30 (50%), which is an average of 3 marks per item.

Please use the Brief BA Fidelity Scale- Scoring Sheet to record your ratings.

1. Agenda Setting and Keeping to the Agenda

The therapist establishes issues to discuss jointly and the client is asked what they would like to cover. The agenda should be clear, and the therapist should be specific about what will be covered. The agenda should include a review of items from the previous session(s), review of the homework assignment (will not apply to initial session), and 1 or 2 appropriate items to discuss in the current session. The agenda should also include time to review the session learning, set new homework and take feedback on the session but these may be included as a single "summing up" item. The homework review should be covered in sufficient detail. It should include a discussion of the client's activities during the week and may include a discussion re what was helpful or unhelpful about an activity, ACE-I ratings, how it affected their mood, and how it met their values. The homework review should inform the current session. Problem solving can be used to overcome barriers to completing homework. Once set, the agenda should be adhered to, unless deviation is important due to a new issue arising (for example, in an instance of new risk related information coming to light this should be added to the agenda and other items may be therefore omitted). Collaboration during agenda setting should be scored here, but collaboration during any other stage of the session should be scored under item 3 (collaboration).

Fidelity Level	Description
0	No agenda set, highly inappropriate agenda or agenda not adhered to
1	Inappropriate agenda set or significant items missing/not added
2	An attempt at an agenda made, but major difficulties evident. Poor adherence
3	Appropriate agenda set well, but some difficulties evident. Some adherence
4	Appropriate agenda, minor difficulties. Moderate adherence
5	Highly appropriate agenda, minimal difficulties. Agenda adhered to
6	Excellent agenda, or highly effective agenda in the face of difficulties

2. Feedback

The client's and the therapist's understanding of key issues should be helped using two-way feedback. The therapist provides and elicits feedback throughout, with summaries at the beginning (review of the week) at the end (session summary), and topic reviews throughout the session. The therapist checks in with the client before moving from one topic to another. The therapist should elicit regular feedback from the client on their understanding (e.g. asking if anything didn't make sense or asking the client to briefly summarise the content). The therapist should also elicit feedback regarding their experience (e.g. "how are you finding today's session?") as appropriate. When giving feedback, the therapist should give it in a manner that is constructive and helps move therapy forwards. The therapist should be sensitive to how feedback may be received by the client. The client should have the opportunity to express their thoughts about therapy sessions (this might involve asking the client how they found the session or a particular technique, what was the most helpful or important part for them, what they could do differently, or how they found a homework task). The SRS can be used as a prompt for this discussion.

Fidelity Level	Description
0	Absence of feedback or highly inappropriate (insensitive or not constructive)
1	Minimal appropriate feedback
2	Appropriate feedback, but not given and elicited frequently enough or too vague
3	Appropriate feedback given and elicited frequently, but some difficulties in content or method
4	Appropriate feedback given and elicited frequently, minor problems evident
5	Highly appropriate feedback given and elicited frequently, minimal problems
6	Excellent use of feedback, or highly effective feedback in the face of difficulties

3. Collaboration

The client is encouraged to be an active member in the session and there should be opportunity for both the therapist and client to contribute. The therapist and client should work towards shared goals. The client should have an active role in identifying their values and deciding which to focus on. The therapist might ask the client for ideas of homework tasks. The therapist should recognise that the amount the client feels able to contribute depends on factors such as their developmental level and the stage of therapy. The client's needs are considered (for example, when to listen, offer suggestions, or wait for the client to devise their own). The manual is viewed together during the session to aid collaboration through shared activity. The therapist should be open about the process of therapy (rationale for the intervention, involvement of parents, and risk discussions). The client should be aware that risk information may need to be shared with their parents, the school, or crisis services, depending on the severity of risk and safeguarding procedures in each setting. This should be discussed with the client in a collaborative way (what needs to be shared, with whom, and how).

Fidelity Level	Description
0	Client is actively prevented or discouraged from being collaborative
1	Therapist is too controlling, dominating, or passive
2	Some attempt at collaboration but style causes problems
3	Teamwork evident but some problems
4	Effective teamwork is evident but not consistent, minor problems
5	Effective teamwork throughout most of the session, minimal problems
6	Excellent teamwork, or highly effective teamwork in the face of difficulties

4. Pacing and Efficient use of Time

The session should be well time managed in relation to the agenda, with clear beginning, middle and end phases. The therapist should maintain the pace of the session appropriately. They may need to limit or interrupt discussion at times to do this. There should be sufficient time for homework review and homework setting at the end of the session. The therapist should not go over time without good reason (e.g. necessary risk assessment/management input). There should be sufficient time for the Brief BA concepts, and the pace should not be too slow or so fast that the client doesn't have time to grasp the concepts. The client's individual needs and speed of learning should be taken into account. In certain situations, the pace of the session may need to change according to the client's needs (e.g. to adapt to discussion of risk or if a client struggles to understand a topic). If the session is shorter in length than usual, the therapist should agree with the client how much time they have and consider how best to use the time.

Fidelity Level	Description
0	Poor time management leads either to an aimless or overly rigid session
1	The session is too slow or fast for the client's needs, session overruns significantly without due cause
2	Reasonable pacing but digression or repetitions leads to an inefficient use of time
3	Good pacing some of the time, some problems
4	Balanced allocation with start middle and end, minor problems
5	Good time management and flow to the session, minimal problems
6	Excellent time management, or highly effective management in face of difficulties

5. Interpersonal Effectiveness

The ability for the therapist to form a good relationship with the client is very important. The therapist should create a relaxed and comfortable atmosphere. The therapist may have informal discussions with the client, for example by asking more about their interests. The client should feel warmth, genuineness, empathy and understanding from the therapist. There should be trust in the therapeutic relationship so that the client is able to be open with the therapist. The therapist is non-judgemental and validating towards the client and the client should feel at ease, understood, and accepted. The therapist should listen to the client and attend to issues the client brings to sessions that are important to them instead of dismissing them. The therapist is effective in validating the client's presenting difficulties as part of the nature of depression as appropriate and helps the client to consider how Brief BA can break the cycle of depression.

Fidelity Level	Description
0	Therapist may be dismissive, and the client disengages, is distrustful and/or hostile
1	Therapist has difficulty in showing empathy, genuineness, and warmth
2	Therapist style at times impedes empathetic understanding
3	Therapist understands explicit meaning of client communication resulting in some trust, some inconsistencies
4	Therapist understands implicit and explicit meaning, minor problems
5	Very good interpersonal effectiveness, client is understood, minimal problems
6	High interpersonal effectiveness, even in the face of difficulties

6. Positive Reinforcement

In line with the behavioural model, the client should receive praise and encouragement for their progress, efforts in attending or contributing in sessions, and attempts to complete the homework. Even if the client has not completed the homework, they should be praised for any efforts they have made. The therapist should acknowledge seemingly small improvements or efforts even in the face of obstacles. When the client attempts a scheduled activity, they should receive praise. The therapist actively looks out for times when the client mentions something they have achieved or a skill and expands upon this by asking questions. The therapist's tone of voice is encouraging, and they are interested in what the client says. As the therapist is likely to serve a reinforcing function for the client, they should consider other ways the client will be able to get this reinforcement from their environment once the sessions have finished. This may be discussed separately with parents or school, so the therapist should not be marked down if it is not mentioned during the session.

Fidelity Level	Description
0	Therapist criticises or tells off the client
1	Therapist does not give positive reinforcement
2	Therapist misses many opportunities to praise and encourage the client
3	Therapist provides some praise and encouragement, some inconsistencies
4	Therapist gives praise and encouragement, minor problems or inconsistencies
5	Therapist gives a lot of praise and encouragement throughout, minimal problems
6	Excellent positive reinforcement, or very good even in the face of difficulties

7. Appropriate use of Brief BA Techniques

The essence of Brief BA is about helping the young person do more of what matters to them, and this message should be clear in all sessions. The therapist should use the manual as a workbook in the sessions to introduce the Brief BA techniques and provide examples. The therapist should help the client to understand how the concepts relate to them. The level of therapist skill in using the Brief BA techniques and clarity in describing the techniques should be taken into account (they should be jargon-free and tailored to the individual).

The focus of the sessions must be on behaviour. If the client identifies their thoughts in a certain situation, it is acceptable for the therapist to label the thoughts and ask how it made the client feel. The therapist may then explore how the thought would look behaviourally, ask the client what they did in the situation, and the effect on their mood. They could also help the client to consider if they would do anything different next time. The therapist should not engage in thought challenging or other cognitive techniques (e.g. setting up activities as behavioural experiments to test cognitions, suggesting using mindfulness for managing thoughts) as this is not part of the Brief BA approach.

Although you may expect to see certain Brief BA techniques in specific sessions as per the manual and checklist at the end of this scale, there may be times where techniques are introduced earlier or later based on the therapist's judgement and number of sessions available to cover the techniques. This is acceptable and should not influence the scoring for this item. The therapist should be flexible with the Brief-BA techniques. Some clients may complete their activity log on the handout, whereas others prefer to use their mobile phone or the computer; the therapist should encourage the client to record their activities in a way that suits them.

The client is given a copy of the workbook for them and a copy for their parents. Parents do not usually attend Brief BA sessions in schools; however, there may be discussion about how it can be useful for parents to also have an understanding of Brief BA and how they may be able to support the client with doing more of what matters.

Fidelity Level	Description
0	Therapist uses inappropriate techniques e.g. cognitive restructuring
1	Therapist does not use any of the Brief BA techniques outlined above
2	Some attempt to use Brief BA techniques but very rigid or unclear
3	Therapist uses Brief BA techniques, but some problems evident
4	Therapist uses appropriate Brief BA techniques, minor problems
5	Therapist uses appropriate Brief BA techniques flexibly, minimal problems
6	Therapist uses appropriate techniques skilfully, even in the face of difficulties

8. Assigns Homework

An appropriate homework task is set with the client, and the client understands the rationale for the task. The homework task is clearly linked to the content of the current session. It should be set jointly and negotiated with the client. The therapist may elicit reactions to the homework task and ask if the assignment is clear and sounds manageable. The therapist is flexible about the way in which the client records their homework so that it works for them. There should be sufficient time for the homework task to be explained clearly. The client might be asked to consider if there may be any potential obstacles to completing the task and if so how the obstacles could be overcome. At the final session there may not be a homework task set, however there should be a discussion about how the client can continue with the progress they have made independently.

Fidelity Level	Description
0	Therapist fails to set homework or sets inappropriate homework
1	Therapist does not negotiate homework, insufficient time allocated to discuss
2	Therapist negotiates homework in a routine way without explaining rationale
3	Therapist sets an appropriate homework task, but some problems evident
4	Appropriate homework jointly negotiated, clear rationale, minor problems
5	Appropriate homework jointly negotiated with clear rationale, obstacles explored, minimal problems
6	Excellent homework negotiated, or appropriate one set in the face of difficulties

9. Clinical use of Outcome Measures

Routine Outcome Monitoring (ROMs) is used in Brief BA. The client should be asked to complete the ORS and RCADS Depression Subscale at the start of each session. The therapist may refer to why ROMs are used and what each measure is looking at. The therapist should ask for feedback about what the scores reflect and whether it fits with how they have been feeling. The therapist might refer to a graph to discuss the change in scores across sessions, and they should explain what an increase or decrease means. At the end of the session the SRS is completed, and the therapist should ask if there is anything big or small the client would like to do differently. Goals should be elicited at the start of therapy and reviewed during the sessions (the client is asked how often they would like to review their goals, so it may not be every session). The therapist might ask what would be different if their scores were higher.

Fidelity Level	Description
0	Therapist fails to use ROMs or discuss them with the client
1	Minimal use of ROMs and does not discuss them with the client
2	Therapist uses ROMs but does not discuss them with the client
3	Therapist uses ROMs and discusses with the client, but some problems evident
4	Appropriate use of ROMs, minor problems
5	Appropriate and clinically meaningful use of ROMs, minimal problems
6	Excellent clinical use of ROMs, or appropriate use of ROMs in the face of difficulties

10. Risk Assessment and Management

The therapist should ask the client about risk at the start of therapy and check in on it throughout. The therapist may ask whether the client would feel able to disclose information about self-harm or suicidal ideation and agree a way to make this easier. The therapist may normalise that self-harm and suicidal ideation are common in young people with depression and explain that it is not the focus of Brief BA but will need to be monitored to keep the young person safe. The RCADS depression subscale has been edited to include a question about self-harm, which can be useful to prompt risk discussion. The therapist should notice any changes in their response to this question and ask the client about it. They may have a risk discussion with the client, for example asking about the severity, frequency, and intent. They might also refer to the safety plan, ask the client for feedback about how they have found using it, and make any changes. Extra sessions may be added if the therapist needs to take significant time to cover risk management. If the client does not report any risk at the beginning of therapy, this should still be checked in on briefly every session to ensure there has been no change.

Fidelity Level	Description
0	Therapist fails to ask the client about risk
1	An attempt at risk discussion, but inappropriate e.g. client feels judged
2	An attempt at risk discussion, but major difficulties evident
3	Therapist asks the client about risk, but some problems evident
4	Appropriate discussion about risk, minor problems
5	Appropriate discussion about risk, minimal problems
6	Excellent risk discussion, or appropriate in the face of difficulties

Appendix E Brief BA Fidelity Scale Scoring Sheet

Brief BA Fidelity Scale - Scoring Sheet

Rater: **Date of Rating:**

Therapist: **Client ID:**

Session Number:

Please highlight the fidelity level for each item and add comments which informed the score below each item. The full Brief BA Fidelity Scale should be referred to for more details about each item.

Once this has been completed, please transfer the fidelity level for each item to the table below and enter the total fidelity level (this the fidelity level for each item combined).

Please ensure you have completed the details about the session above.

Item	**Fidelity Level**
1. Agenda Setting and Keeping to the Agenda	
2. Feedback	
3. Collaboration	
4. Pacing and Efficient Use of Time	
5. Interpersonal Effectiveness	
6. Positive Reinforcement	
7. Appropriate Use of Brief BA Techniques	
8. Assigns Homework	
9. Clinical Use of Outcome Measures	
10. Risk Assessment and Management	
Total Fidelity Level	

1. Agenda Setting and Keeping to the Agenda

Fidelity Level	Description
0	No agenda set, highly inappropriate agenda or agenda not adhered to
1	Inappropriate agenda set or significant items missing/not added
2	An attempt at an agenda made, but major difficulties evident. Poor adherence
3	Appropriate agenda set well, but some difficulties evident. Some adherence
4	Appropriate agenda, minor difficulties. Moderate adherence
5	Highly appropriate agenda, minimal difficulties. Agenda adhered to
6	Excellent agenda, or highly effective agenda in the face of difficulties

Comments:

2. Feedback

Fidelity Level	Description
0	Absence of feedback or highly inappropriate (insensitive or not constructive)
1	Minimal appropriate feedback
2	Appropriate feedback, but not given and elicited frequently enough or too vague
3	Appropriate feedback given and elicited frequently, but some difficulties in content or method
4	Appropriate feedback given and elicited frequently, minor problems evident
5	Highly appropriate feedback given and elicited frequently, minimal problems
6	Excellent use of feedback, or highly effective feedback in the face of difficulties

Comments:

3. Collaboration

Fidelity Level	Description
0	Client is actively prevented or discouraged from being collaborative
1	Therapist is too controlling, dominating, or passive
2	Some attempt at collaboration but style causes problems
3	Teamwork evident but some problems
4	Effective teamwork is evident but not consistent, minor problems
5	Effective teamwork throughout most of the session, minimal problems
6	Excellent teamwork, or highly effective teamwork in the face of difficulties

Comments:

4. Pacing and Efficient Use of Time

Fidelity Level	Description
0	Poor time management leads either to an aimless or overly rigid session
1	The session is too slow or too fast for the client's needs, session overruns without due cause
2	Reasonable pacing but digression or repetitions leads to an inefficient use of time
3	Good pacing some of the time, some problems
4	Balanced allocation with start middle and end, minor problems
5	Good time management and flow to the session, minimal problems
6	Excellent time management, or highly effective management in face of difficulties

Comments:

5. Interpersonal Effectiveness

Fidelity Level	Description
0	Therapist may be dismissive, and the client disengages, is distrustful and/or hostile
1	Therapist has difficulty in showing empathy, genuineness, and warmth
2	Therapist style at times impedes empathetic understanding
3	Therapist understands explicit meaning of client communication resulting in some trust, some inconsistencies
4	Therapist understands implicit and explicit meaning, minor problems
5	Very good interpersonal effectiveness, client is understood, minimal problems
6	High interpersonal effectiveness, even in the face of difficulties

Comments:

6. Positive Reinforcement

Fidelity Level	Description
0	Therapist criticises or tells off the client
1	Therapist does not give positive reinforcement
2	Therapist misses many opportunities to praise and encourage the client
3	Therapist provides some praise and encouragement, some inconsistencies
4	Therapist gives praise and encouragement, minor problems or inconsistencies
5	Therapist gives a lot of praise and encouragement throughout, minimal problems
6	Excellent positive reinforcement, or very good even in the face of difficulties

Comments:

7. Appropriate Use of Brief BA Techniques

Fidelity Level	Description
0	Therapist uses inappropriate techniques e.g. cognitive restructuring
1	Therapist does not use any of the Brief BA techniques outlined above
2	Some attempt to use Brief BA techniques but very rigid or unclear
3	Therapist uses Brief BA techniques, but some problems evident
4	Therapist uses appropriate Brief BA techniques, minor problems
5	Therapist uses appropriate Brief BA techniques flexibly, minimal problems
6	Therapist uses appropriate techniques skilfully, even in the face of difficulties

Comments:

8. Assigns Homework

Fidelity Level	Description
0	Therapist fails to set homework or sets inappropriate homework
1	Therapist does not negotiate homework, insufficient time allocated to discuss
2	Therapist negotiates homework in a routine way without explaining rationale
3	Therapist sets an appropriate homework task, but some problems evident
4	Appropriate homework jointly negotiated, clear rationale, minor problems
5	Appropriate homework jointly negotiated with clear rationale, obstacles explored, minimal problems
6	Excellent homework negotiated, or appropriate one set in the face of difficulties

Comments:

9. Clinical Use of Outcome Measures

Fidelity Level	Description
0	Therapist fails to use ROMs or discuss them with the client
1	Minimal use of ROMs and does not discuss them with the client
2	Therapist uses ROMs but does not discuss them with the client
3	Therapist uses ROMs and discusses with the client, but some problems evident
4	Appropriate use of ROMs, minor problems
5	Highly appropriate and clinically meaningful use of ROMs, minimal problems
6	Excellent clinical use of ROMs, or appropriate use of ROMs in the face of difficulties

Comments:

10. Risk Assessment and Management

Fidelity Level	Description
0	Therapist fails to ask the client about risk
1	An attempt at risk discussion, but inappropriate e.g. client feels judged
2	An attempt at risk discussion, but major difficulties evident
3	Therapist asks the client about risk, but some problems evident
4	Appropriate discussion about risk, minor problems
5	Appropriate discussion about risk, minimal problems
6	Excellent risk discussion, or appropriate in the face of difficulties

Comments:

Appendix F RCADS Depression Subscale

Self-report weekly mood assessment:

RCADS Depression subscale

Please put a circle around the word that shows how often each of these things happen to you. There are no right or wrong answers.

	0	1	2	3
1. I feel sad or empty	Never	Sometimes	Often	Always
2. Nothing is much fun anymore	Never	Sometimes	Often	Always
3. I have trouble sleeping	Never	Sometimes	Often	Always
4. I have problems with my appetite	Never	Sometimes	Often	Always
5. I have no energy for things	Never	Sometimes	Often	Always
6. I am tired a lot	Never	Sometimes	Often	Always
7. I cannot think clearly	Never	Sometimes	Often	Always
8. I feel worthless	Never	Sometimes	Often	Always
9. I feel like I don't want to move	Never	Sometimes	Often	Always
10. I feel restless	Never	Sometimes	Often	Always
10a. I thought about killing myself	Never	Sometimes	Often	Always

RCADS depression subscale total:

Appendix G Outcome Rating Scale (ORS)

Outcome Rating Scale (ORS)

Name _____

Session # ____ Date: _____

Who is filling out this form? Please check one: Self_____ Other_____

If other, what is your relationship to this person? _____

Looking back over the last week, including today, help us understand how you have been feeling by rating how well you have been doing in the following areas of your life, where marks to the left represent low levels and marks to the right indicate high levels. *If you are filling out this form for another person, please fill out according to how you think he or she is doing.*

Individually
(Personal well-being)

|----------|

Interpersonally
(Family, close relationships)

|----------|

Socially
(School, friendships)

|----------|

Overall
(General sense of well-being)

|----------|

International Center for Clinical Excellence

www.scottdmiller.com

© 2000, Scott D. Miller and Barry L. Duncan

Appendix H Session Rating Scale (SRS)

Session Rating Scale (SRS)

Name _____

Session # ____ Date: _____

Who is filling out this form? Please check one: Self_____ Other_____

If other, what is your relationship to this person? _____

Please rate today's session by placing a mark on the line nearest to the description that best fits your experience.

Relationship

I did not feel heard, understood, and respected.	I----------I	I felt heard, understood, and respected.

Goals and Topics

We did *not* work on or talk about what I wanted to work on and talk about.	I----------I	We worked on and talked about what I wanted to work on and talk about.

Approach or Method

The therapist's approach is not a good fit for me.	I----------I	The therapist's approach is a good fit for me.

Overall

There was something missing in the session today.	I----------I	Overall, today's session was right for me.

International Center for Clinical Excellence

_____

www.scottdmiller.com

© 2002, Scott D. Miller, Barry L. Duncan, & Lynn Johnson

Appendix I Brief BA Fidelity Scale Scoring Examples

Brief BA Fidelity Scale Examples

1. Agenda Setting

Examples of a 3

+ The agenda included a review of content from previous sessions, and appropriate items for the session today, which were covered during the session.

− Lack of collaboration within the agenda setting. Homework review and summing up item was not included in the agenda.

+ Asks the client what they would like to cover in the session. Bridge to previous session.

− No clear agenda with specific items at the start- goes into discussion after setting items, rather than an overview of items at the start. Homework review and summing up item not present.

Examples of a 4

+ The therapist adds appropriate and specific items for the session and asks if there is anything the client would like to add. Recognised no time to review goal and added to agenda for next time.

− One of the items could have been clearer. Did not have time to review goal- could have checked at the start what client would like to prioritise.

+ The therapist adds appropriate items for the session and asks if there is anything the client would like to add. The agenda included a review from previous session.

− Homework review present but brief and could have been more specific in agenda.

Examples of a 5

+ The therapist adds appropriate and specific items for the session and includes homework review. Link to previous session. Asks if anything the client would like to add and where to start. Agenda is adhered to.

− No summing up item in agenda

A 2 would be an attempt at an agenda but very unclear and/or poor adherence for no reason. Do not score higher than 3 if there is a lack of collaboration with the agenda setting or if the therapist starts setting the agenda and then moves onto discussion of the items instead of completing the

agenda first. When scoring, consider if the homework review was included in the session as well as whether it was added to the agenda at the start. Consider adherence to the agenda as well as agenda setting. For a 4 the agenda should be collaborative with minor difficulties. If there are minor difficulties but the agenda is adhered to, rate it as a 4 rather than a 5 (even though the agenda is adhered to, it might not be clearly set from the start or may be missing items). A 5 is when there are only very minor things that could be changed.

2. Feedback

Example of a 3

+ Asked the client how they found completing the activity log and if there is anything they noticed. Checked if Brief BA techniques made sense.

− Did not summarise topics or provide a summary at the end. SRS completed but not discussed.

Examples of a 4

+ Checks in if what is being discussed makes sense and asks what the client has learnt from doing one of the Brief BA techniques. Asks for feedback on how they found the session.

− Could have used topic reviews to summarise what was learnt.

+ Lots of examples of eliciting feedback and provides appropriate summaries at the beginning and end of session. Asks how the client is finding the Brief BA techniques, and if they have any questions. Asks the client to clarify when unclear about something.

− Asks to complete the SRS but does not discuss.

Example of a 5

+ Asks how the client found the assessment and what they know about Brief BA so far- asks the client to explain Brief BA to them. Provides summaries and asks if the client has any questions. Asks if certain Brief BA techniques makes sense, checks if anything want to go over. Asks for feedback re SRS and if there is anything that could be done differently.

− Could have used open-ended questions e.g. what the client thought of a particular technique

A 2 would be very limited examples of feedback or too vague. A 3 may be when the feedback is appropriate but some difficulties (e.g. there is a lack of summaries and reviews between topics, and the therapist does not ask for feedback about the session). If the SRS is not discussed but there are lots of other examples of feedback, a 4 may be appropriate. A 5 is when it is very difficult to think of anything that could be done differently (the therapist provides and elicits feedback

throughout and may ask the client to summarise one of the Brief BA concepts to check their understanding).

3. Collaboration

Example of a 3

+ Gave suggestions to the client and asked for their input at times e.g. what would the client like help with when discussing contracts, and what activities they might like to try during the week. − The therapist was not very collaborative when discussing problems and did not encourage the client to think of their own solutions.

Examples of a 4

+ The therapist and client worked together throughout most of the session e.g. the client was asked for their ideas of values and which life area to focus on. − On one occasion the therapist gave their idea rather than asking the client

+ The therapist asked for ideas and made suggestions where appropriate. The therapist asked questions such as 'shall we write that down?' and 'shall we pick another value in this life area?' − The therapist did a lot of talking and client could have been more active and engaged during the later part of the session

Example of a 5

+ The therapist asks the client if they want to tick their symptoms on the manual. There were collaborative discussions throughout and the therapist involves the client by asking questions. The client is encouraged to be active and have choice e.g. which day to fill out on the activity log and which version to use. − The therapist may not have asked the client for ideas of homework tasks; however, they had a collaborative discussion around it.

A 2 would be if there is some attempt at collaboration but the therapist is directive and misses many opportunities to involve the client. A 3 is when there are examples of the therapist asking the client for their ideas and suggestions, but they miss many opportunities. A 4 is when there is effective teamwork most of the time and only one or two exceptions to this. A 5 is when there is effective teamwork throughout and there is not much that could have been done differently.

4. Pacing

Examples of a 3

+ There was sufficient time spent on the Brief BA techniques and the client seemed to grasp them.

− The session was shorter, but the therapist didn't refer to this and did not think how best to use the time. The therapist spoke very fast and there could have been more time spent on the beginning and end parts of the session.

+ Good amount of time reviewing the progress the client has made and what they have learnt and did finish the session on time.

− Possibly more time reviewing the homework. It seemed like quite a lot to cover and perhaps too much for the client to take in; there was the option for another session and this may have been more beneficial.

Examples of a 4

+ Discussed how best to make use of the time, and spent time discussing issues that were important to the client. Clear beginning middle and end phases.

− Perhaps could have spent more time on the session content and homework setting

+ Good amount of time discussing values and activities, good pace- the client seems able to grasp the concepts. Sufficient time on homework setting.

− Could have spent more time on homework review.

Examples of a 5

+ Appropriate amount of time discussing ROMs and risk. When going through an example on activity log, the therapist interrupted the discussion to encourage the client to jot their ideas down which helped to direct the focus back to the task and keep it on track. Sufficient time on the activity log and explaining homework. Good flow to the session

− Perhaps could have considered how best to use the time with the client

A 2 would be if there is an inefficient use of time throughout (e.g. due to digressions or repetitions). A 3 is when there is sufficient time on parts of the session (e.g. homework review or Brief BA concepts) but there are some problems (e.g. too much to cover, and the pace was rushed). A 4 is when there is clear beginning, middle and end phases and the client seems able to grasp the concepts but there are some minor problems (e.g. could have spent more time on one part). A 5 is when there is a good flow to the session throughout and there is not much at all that could have improved it.

5. Interpersonal effectiveness

Examples of a 3

+ Informal discussion and discussion of topics that seemed important to the client.

− The therapist shared a lot of their own experiences and missed following up on what the client said. Missed opportunities for validation.

+ The therapist helps the client to explore what is getting in the way of their revision and seems to be invested in helping the client.

− A lot of the session was about activities the client needs to do (even though were linked to their values)- possibly felt a bit overwhelming for the client. Missed following up on one point that was more about what the client wants to do.

Examples of a 4

+ Some informal discussion about half-term which seemed to put the client at ease. The therapist listens to the client and their tone of voice is encouraging.

− Perhaps came across as slightly informal or rehearsed at times.

+ Introduces self as had not met before. When the client is unable to answer one of the questions, the therapist validated that it was a difficult thing to think about and tried to explain it. Offers to write down client's goal.

− A couple of times when the client said something during the session, the therapist says 'that is something we can talk about'- but in terms of the relationship may have been useful to show some interest at the time by asking questions.

Example of a 5

+ The therapist is genuine, warm, and validating. The client seems to feel understood and is able to be open and honest. There is some informal discussion and particularly around something that was important to the client. The therapist offered to speak to parents, which the client seemed to appreciate.

− Perhaps more opportunities for informal discussion

A 2 would be if the therapist does not come across as particularly genuine, empathetic, or warm. A 3 is when the therapist listens to the client and seems genuine but misses opportunities for building the therapeutic relationship (e.g. with validation or noticing implicit meaning). A 4 is when the therapist seems to understand the client and is warm and genuine and any problems are minor (e.g. could have followed on from what the client was saying rather than sticking 'to the

book'). A 5 is when there is very good interpersonal effectiveness throughout and not much that could be improved.

6. Positive Reinforcement

Example of a 3

➕ The therapist provides some praise and encouragement (e.g. 'I think that's a really good goal'/ 'that's a good question')

➖ There could have been more examples of positive reinforcement and there were several missed opportunities

Examples of a 4

➕ There are several examples of where the therapist gives praise and encouragement ('That sounds like a brilliant thing to do', 'that's a great idea').

➖ Missed opportunity to reinforce on the client's efforts (their attempt to explain BA).

➕ The therapist provides praise and encouragement ("That's brilliant" / "that sounds really good"/ "that's really interesting")

➖ Missed opportunity for praise and encouragement (independence with resolving situation with homework difficulties). Tone of voice could be slightly more upbeat.

Example of a 5

➕ Considered how the client would feel about continuing with the Brief BA techniques after the sessions have finished. Praised the client many times and was encouraging throughout (e.g. 'that's brilliant', 'well done', 'you should be very proud of yourself').

➖ Could have praised the client specifically for completing their activity log over half term. However, later said, 'you did everything we planned, which is great'

A 2 is given when the therapist misses many opportunities to praise and encourage the client (there are only a couple of examples of this). A 3 is when there are at least a few examples of praise and encouragement, but some inconsistencies (the therapist misses some opportunities to praise their efforts or small improvements). A 4 is when there are many examples of praise and encouragement, but perhaps one or two missed opportunities. A 5 is when there is praise and encouragement throughout and not much that could be added.

7. Appropriate Use of Brief BA Techniques

Examples of a 3

+ The therapist linked activities to values and emphasised getting a balance of activities from different life areas. Linked job interview discussion to the client's values. Related to client (e.g. asking what the client might need help with).

− Stuck very closely to the manual. Explanation of positive reinforcement was not clear ("more opportunities where you can be positive in relation to these things"). Said how avoidance reinforces belief there is something to be afraid of- not strictly Brief BA.

+ The therapist looked through activity log and ACE-I ratings and referred to values. Tips for completing the activity log and flexibility. Individualises to the client and emphasises it is what they find important. Client talks about "positive attitude" – asks what it looks like, how do they get it.

− Towards the end of the session went into thought challenging with situation with friends the client was anxious about- "so no evidence to suggest… "But then checked if enjoys it- more BA.

Examples of a 4

+ The therapist asks about ACE-I and refers to values by asking what it means to the young person. An activity linked to values is discussed. The positive cycle of activity was referred to.

− The activity log wasn't discussed during this session, which may have been helpful.

+ Refers to the cycle of low mood and positive cycle of activity, symptoms of low mood and asks which relate to the client and triggers. Explains focus will be on the present. Refers to values and valued activities. Explains how to complete activity log, ACE-I ratings, daily mood ratings.

− Don't complete a day of the activity log together, which would have been helpful to enable them to do it at home. Top tips for completing activity log were referred to only briefly.

Examples of a 5

+ The therapist was flexible with using the manual. They discussed what depression is, referred to the cycle of depression and discussed how it related to the client. There was a focus on doing more things you enjoy/that are important to you, not just doing more. The activity log was discussed and ACE-I ratings. At one point asks what could have done that would have made it better- BA focus.

A 2 is given if there is an attempt to use Brief BA techniques but very unclear or rigid. A 3 is given if Brief BA techniques are used but there are some problems (e.g. skilful and flexible use of Brief BA techniques through most of the session but at one point goes very briefly into exposure or thought challenging and perhaps is unclear in explaining a particular technique). A 4 is given if appropriate techniques are used throughout and there are only minor problems (e.g. the therapist does not go through an example of the activity log specific to the client together which would have helped). A 5 is given if the therapist is flexible in their delivery of Brief BA and very minimal problems.

8. Assigns Homework

Example of a 3

+ The therapist set up what the client will try this week, when and what will help remind them to do so? Checked had enough copies of activity log. The therapist gave a helpful suggestion with the activity log ("What would be good is if you put it on there and if you don't do it, cross it out and re-plan it").

− Could have summarised homework task at the end, checked if it sounded manageable and if there were any obstacles.

Examples of a 4

+ After going through an example together the therapist asks the client to do another activity log during the week and gave them a spare copy. This was clearly linked to the session content and they discussed how this would help (the rationale). Checked this sounds okay.

− Possible obstacles weren't discussed

+ The client is helped to consider which activities they could try this week and details (e.g. when, where etc). Clearly linked to the session. Asks how the client feels about doing those and whether they seem manageable. Also asked them to fill in any parts of the manual about activities they are doing already, want to do, and need help with.

− Obstacles weren't discussed, and it felt slightly rushed at the end

Example of a 5

+ Asks the client which activities they would like to focus on this week – when, what? Checked it seemed manageable. Anything that might get in the way? Feel quite motivated? Asked how many days would be manageable to complete on the activity log, checked have enough copies. Wrote down the plan for homework. Clearly linked to the session.

− Could have reminded of rationale for completing activity log, but presume covered in a previous session

A 2 is given when the homework is not linked to the session and the client is unclear about the rationale. A 3 is given when an appropriate homework task is set but there are some problems (e.g. the homework task is set earlier in the session but not summarised at the end, the client is not asked if it seems manageable and potential obstacles are not discussed). A 4 is given when an appropriate task is set jointly with a clear rationale, but obstacles were not discussed, and perhaps more time could have been spent on it. A 5 is given when an appropriate homework task is set with clear rationale, possible obstacles explored and very minimal problems.

9. Clinical use of ROMs

Examples of a 4

+ Compared RCADS depression subscale score to the previous session and picked out an item that had changed to ask about. Explained that this level is a 'normal level of low mood'. Gave ORS score and compared score to previous sessions- explained this means has improved. Rated progress towards goal. Completed the SRS

− Did not discuss the SRS and no reminder of why the questionnaires were completed (but was a later session so client could know this by now)

+ There is discussion of why the questionnaires are completed and how they are used to track progress. Each scale is discussed and what a higher or lower score means.

− "That fits with what you are saying"- assumption and could have checked in instead.

Example of a 5

+ Compared measures to the previous weeks and referred to graph. Gave exact scores on the ORS and how they have changed- "gone up by X. Does that fit with how you have been feeling?"- "anything that impacted on that?" Explained what a decrease means on the RCADS and checked it fits with how client has been feeling. Also discussed goal in detail and asked about SRS.

− Perhaps could have varied questions around whether it fits with how the client is feeling

A 2 is given when the ROMs are completed but not used in a clinically meaningful way (not discussed or explained at all in the session). A 3 is given when the client is asked to complete the ROMs and there is some discussion about scores on one or two of the measures, but some aren't discussed and could have been discussed in more detail. A 4 is given when ROMs are completed and discussed but the therapist could have spent more time on it (e.g. asked for feedback on the SRS, checked in if the scores fit with how the client is feeling, and given exact scores). A 5 is given when all of the ROMs are completed and discussed in detail with the client in a clinically meaningful way.

10. Risk Assessment and Management

Example of a 3

+ The therapist said they just needed to check in on sensitive question- thoughts about self-harm or ending your life- have you had any of those thoughts? The client says no, and the therapist thanks the client

− Did not remind the client they would check in on this each time/remind them why, ask if they would feel able to disclose information about this and what would make it easier.

Example of a 4

+ The therapist mentions sensitive questions they will go through every session to check everything is going okay for the client- any thoughts about wanting to harm yourself or ending your life? The therapist asks about intention to harm themselves, plans and actions. They explain it is their duty of care, so they need to check in on this every time.

− Does not ask if they would feel able to disclose information about this and what would make it easier. Does not refer to a safety plan.

Example of a 5

+ The therapist checked in on risk- "how have things been since last time?" "Are you still having those thoughts?" "What was it you thought?" "Have you made any plans?" "Intentions to act on the thoughts?" Any thoughts about self-harm? Referred to safety plan and checked still happy using it, any changes to make. Reminded will check on it each time. Checked if anything else wanted to talk about.

− Perhaps could have normalised that it is common in young people with low mood and reminded why check in on it

A 2 is given when the therapist asks one unclear question regarding risk (e.g. 'is there anything I need to be aware of in regard to yours or other people's safety?') A 3 is given when the therapist asks a question about self-harm or suicide but does not expand upon this and perhaps could have checked in about a safety plan. A 4 is given when this is expanded upon by asking about intentions, plans or actions, and the client may be reminded that this will be checked on each session. A 5 is given when the client is asked about risk in sufficient detail, the therapist may remind them why they are asking, and refer to their safety plan. They are given the opportunity to discuss anything else relating to this before moving on.

Appendix J Inter-Rater Reliability

Brief BA Fidelity Scale Items	Learning stage			Main stage 1			Main stage 2		
	ICC	*M*	*SD*	ICC	*M*	*SD*	ICC	*M*	*SD*
1. Agenda Setting and Keeping to the Agenda	.85	4.17	1.03	.80	3.70	1.34	.91	3.43	1.57
2. Feedback	.83	3.92	1.17	.60	4.63	0.49	.85	4.37	0.81
3. Collaboration	.94	4.25	1.14	.21	4.77	0.43	.79	4.50	0.68
4. Pacing and Efficient use of Time	.83	4.17	0.94	.22	4.27	0.79	.83	3.93	0.87
5. Interpersonal Effectiveness	1.00	4.17	0.72	.31	4.60	0.56	.81	4.27	0.83
6. Positive Reinforcement	.84	4.33	0.99	.81	4.17	0.91	.94	4.33	1.03
7. Appropriate Use of Brief BA Techniques	.62	4.50	0.91	NC	4.77	0.50	.80	4.30	0.70
8. Assigns Homework	.71	3.83	0.72	.48	4.10	0.66	.91	3.70	1.02
9. Clinical use of Outcome Measures	.96	3.75	1.36	.44	4.80	0.48	.91	4.37	1.03
10. Risk Assessment and Management	1.00	2.83	2.37	.96	3.93	1.34	1.00	2.53	1.93
Total Fidelity Score	.95	3.99	1.25	.71	4.37	0.89	.93	3.97	1.24

Note. For means and standard deviations, items were averaged over the two raters. ICC = Intraclass Correlation Coefficient. *M* = Mean. *SD* = Standard Deviation. NC = not calculated due to zero variance from one rater.

List of References

Ackerman, S. J., & Hilsenroth, M. J. (2003). A review of therapist characteristics and techniques positively impacting the therapeutic alliance. *Clinical Psychology Review, 23*, 1–33. doi: 10.1016/S0272-7358(02)00146-0

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.

Angold, A., Costello, E. J., Messer, S. C., Pickles, A., Winder, F., & Silver, D. (1995). Development of a short questionnaire for use in epidemiological studies of depression in children and adolescents. *International Journal of Methods in Psychiatric Research, 5*, 237–249.

Barber, J. P., Liese, B. S., & Abrams, M. J. (2003). Development of the Cognitive Therapy Adherence and Competence Scale. *Psychotherapy Research, 13*(2), 205-221. doi: 10.1093/ptr/kpg019

Barber, J. P., Sharpless, B. A., Klostermann, S., & McCarthy, K. S. (2007). Assessing intervention competence and its relation to therapy outcome: A selected review derived from the outcome literature. *Professional Psychology: Research and Practice, 38*(5), 493-500. doi: 10.1037/0735-7028.38.5.493

Barrett, P. M. (2004). *Friends for life—Group leader's manual* (4th ed.). Brisbane, Australia: Australian Academic Press.

Barrett, P. M. (2008). *Friends for life* (Norwegian version). Bergen, Norway: Innovest AS.

Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). *Cognitive therapy of depression.* New York: Guilford.

Bjaastad, J. F., Haugland, B. S. M., Fjermestad, K. W., Torsheim, T., Havik, O. E., Heiervang, E. R., & Ost, L. G. (2015). Competence and Adherence Scale for Cognitive Behavioral Therapy (CAS-CBT) for anxiety disorders in youth: Psychometric properties. *Psychological Assessment, 28*(8), 908-916. doi: 10.1037/pas0000230

Blackburn, I. M., James, I. A., Milne, D. L., Baker, C., Standart, S., Garland, A., & Reichelt, F. K. (2001). The Revised Cognitive Therapy Scale (CTS-R): Psychometric properties. *Behavioural and Cognitive Psychotherapy, 29*, 431-446. doi: 10.1017/S1352465801004040

Branson, A., Shafran, R., & Myles, P. (2015). Investigating the relationship between competence and patient outcome with CBT. *Behaviour Research and Therapy, 68*, 19-26. doi:

10.1016/j.brat.2015.03.002

Bringhurst, D. L., Watson, C. W., Miller, S. D., Duncan, B. L. (2006). The reliability and validity of the Outcome Rating Scale: a replication study of a brief clinical measure. *Journal of Brief Therapy, 5,* 23-30.

Brown, R. C., Southam-Gerow, M. A., McLeod, B. D., Wheat, E. B., Tully, C. B., Reise, S. P., . . . Weisz., J. R. (2018). The global therapist competence scale for youth psychosocial treatment: Development and initial validation. *J. Clin. Psychol., 74*, 649-664. doi: 10.1002/jclp.22537

Butler, A. C., Chapman, J. E., Forman, E. M., & Beck, A. (2006). The empirical status of cognitive-behavioral therapy: A review of meta-analyses. *Clinical Psychology Review, 26*, 17-31. doi: 10.1016/j.cpr.2005.07.003

Carroll, K. M., Charla, N., Sifry, R. L., Nuro, K. F., Frankforter, T. L., Ball, S. A., . . . Rounsaville., B. J. (2000). A general system for evaluating therapist adherence and competence in psychotherapy research in the addictions. *Drug and Alcohol Dependence, 57*(3), 225-238. doi: 10.1016/S0376-8716(99)00049-6

Chorpita, B. F., Moffitt, C. E., & Gray, J. A. (2005). Psychometric properties of the revised child anxiety and depression scale in a clinical sample*. Behaviour Research and Therapy, 43*, 309-322. doi: 10.1016/j.brat.2004.02.004

Chorpita, B. F., Yim, L., Moffitt, C., Umemoto, L. A., & Francis, S. E. (2000). Assessment of symptoms of DSM-IV anxiety and depression in children: a revised child anxiety and depression scale. *Behaviour Research and Therapy*, *38*(8), 835–855. doi: 10.1016/S0005-7967(99)00130-8

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6,* 284–290. doi: 10.1037/1040-3590.6.4.284

Clark, D.M., Ehlers, A., Hackmann, A., McManus, F., Fennell, M., & Grey, N. (2006). Cognitive therapy versus exposure and applied relaxation in social phobia: a randomized controlled trial. *Journal of Consulting and Clinical Psychology, 74,* 568-578. doi: 10.1037/0022-006X.74.3.568

Clark, D. M., Ehlers, A., McManus, F., Hackmann, A., Fennell, M. J. V., Campbell, H. (2003). Cognitive therapy versus fluoxetine in generalized social phobia: a randomized placebo-

controlled trial. *Journal of Consulting and Clinical Psychology, 71*, 1058-1067. doi: 10.1037/0022-006X.71.6.1058

Clark, D. M., von Consbruch K., Hinrichs, J., & Stangier, U. (2007). *Cognitive Therapy of Social Phobia: checklist of therapist competency.* Unpublished manuscript, London: Kings College London.

Clark, D. M., & Wells, A. (1995). A cognitive model of social phobia. In R. Heinberg, M. Liebowitz, D. A. Hope, & F. R. Schneier (Eds.), *Social Phobia: diagnosis, assessment and treatment.* (pp. 69-93). New York: Guilford Press.

Clayborne, Z. M., Varin, M., & Colman, I. (2019). Systematic Review and Meta-Analysis: Adolescent Depression and Long-Term Psychosocial Outcomes. *Journal of the American Academy of Child & Adolescent Psychiatry, 58*(1), 72-77. doi: 10.1016/j.jaac.2018.07.896

Cohen, J. (1992). A power primer. *Psychological Bulletin, 122*, 155-159. doi: 10.1037/0033-2909.112.1.155

Creed, T., & Kendall, P. C. (2003). Empirically supported therapist relationship-building behavior within a cognitive–behavioral treatment of anxiety in youth. Manuscript submitted for publication.

Davidson, K., Scott, J., Schmidt, U., Tata, P., Thornton, S., & Tyrer, P. (2004). Therapist competence and clinical outcome in the Prevention of Parasuicide by Manual Assisted Cognitive Behaviour Therapy Trial: the POPMACT study. *Psychol. Med, 34*(5), 855–863. doi: 10.1017/S0033291703001855

Department of Health and Social Care and Department for Education (2017). *Transforming children and young people's mental health provision: a green paper.* Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/664855/Transforming_children_and_young_people_s_mental_health_provision.pdf

Dimidjian, S., Hollon, S. D., Dobson, K. S., Schmaling, K. B., Kohlenberg, R. J., Addis, M. E., . . . Jacobson, N. S. (2006). Randomized trial of behavioral activation, cognitive therapy, and antidepressant medication in the acute treatment of adults with major depression. *Journal of Consulting and Clinical Psychology, 74*(4), 658-670. doi: 10.1037/0022-006X.74.4.658

Duncan, B. L., Miller, S. D., Sparks, J. A., Claud, D. A., Reynolds, L. R., Brown, J., & Johnson, L. D.

(2003). The Session Rating Scale: Preliminary psychometric properties of a "working" alliance measure. *Journal of Brief Therapy, 3,* 3-12.

Dobson, K. S., Shaw, B. F., & Vallis, T. (1985). Reliability of a measure of the quality of cognitive therapy. *The British Journal of Clinical Psychology, 24*(4), 295-300. doi: 10.1111/j.2044-8260.1985.tb00662.x

Dreyfus, H. L. (1989). The Dreyfus model of skill acquisition. In J. Burke (Ed.), *Competency based education and training*. London: Falmer Press.

Ekers, D., Richards, D., McMillan, D., Bland, J. M., & Gilbody, S. (2011). Behavioural activation delivered by the non-specialist: phase II randomised controlled trial. *Br J Psychiatry, 198*, 66-72. doi: 10.1192/bjp.bp.110.079111

Fairburn, C. G., & Cooper, Z. (2011). Therapist competence, therapy quality, and therapist training. *Behaviour Research and Therapy, 49*, 373-378. doi: 10.1016/j.brat.2011.03.005

Ferster, C. B. (1973). A functional analysis of depression. *American psychologist*, *28*(10), 857.Fjermestad, K. W., McLeod, B. D., Tully, C. B., & Liber, J. M. (2016). Therapist characteristics and interventions: Enhancing alliance and involvement with youth. In S. Maltzman (Ed.), *Oxford handbook of treatment processes and outcomes in counselling psychology*. (pp. 97-116). New York: Oxford University Press.

Fuggle, P., Dunsmuir, S., & Curry, V. (2013). *CBT with children, young people, and families.* London: Sage.

Gaudiano, B. A. (2008). Cognitive-Behavioural Therapies: Achievements and Challenges. *Evidence-Based Mental Health, 11*(1), 5-7. doi: 10.1136/ebmh.11.1.5

George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference. 11.0 update (4th ed.).* Boston: Allyn & Bacon.

Haddock, G., Devane, S., Bradshaw, T., McGovern, J., Tarrier, N., Kinderman, P., . . . Harris, N. (2001). An investigation of the psychometric properties of the cognitive therapy scale for psychosis (CTS-Psy). *Behavioural and Cognitive Psychotherapy, 29*, 221–233. doi: 10.1017/S1352465801002089

Hogue, A., Henderson, C. E., Dauber, S., Barajas, P. S., Fried, A., & Liddle, H. A. (2008). Treatment adherence, competence, and outcome in individual and family therapy for adolescent behavior problems. *Journal of Consulting and Clinical Psychology, 76*, 544–555. doi: 10.1037/0022-006X.76.4.544

Huppert, J. D., Bufka, L. F., Barlow, D. H., Gorman, J. M., Shear, M. K., & Woods, S. W. (2001). Therapists, therapist variables and cognitive behavioral therapy outcome in a multicenter trial for panic disorder. *Journal of Consulting and Clinical Psychology, 69*, 747–755. doi: 10.1037/0022-006X.69.5.747

Jacobson, N. S., Dobson, K. S., Truax, P. A., Addis, M. E., Koerner, K., Gollan, J. K., . . . Prince, S. E. (1996). A component analysis of cognitive-behavioral treatment for depression. *Journal of Consulting and Clinical Psychology*, *64*, 295-304. doi: 10.1037/0022-006X.64.2.295

James, I. A., Blackburn, I. M., & Reichelt, F, K. (2001). *Manual of the Revised Cognitive Therapy Scale (CTS-R).* Unpublished Manuscript, Newcastle Cognitive and Behavioural Therapies Centre, Newcastle.

Jones, P. B. (2013). Adult mental health disorders and their age at onset. *The British Journal of Psychiatry, 54,* 5-10. doi: 10.1192/bjp.bp.112.119164

Josefowitz, N., & Myran, D. (2005). Towards a person-centred cognitive behaviour therapy. *Counselling Psychology Quarterly, 18*(4), 329-336. doi: 10.1080/09515070500473600

Kaufman, J., Birmaher, B., Axelson, D., Perepletchikova, F., Brent, D., & Ryan, N. (2013). *Schedule for Affective Disorders and Schizophrenia for school-age children – Present and Lifetime version (KSADS-PL): Working draft*. Unpublished manuscript.

Kazantis, N., Clayton, X., Cronin, T. J., Farchione, D., Limburg, K., & Dobson, K. S. (2018). The Cognitive Therapy Scale and Cognitive Therapy Scale-Revised as measures of therapist competence in cognitive behavior therapy for depression: Relations with short and long term outcome. *Cognitive Therapy and Research, 42*(4), 385-397.

Kendall, P. C., & Hedtke, K. A. (2006). *Cognitive-behavioral therapy for anxious children: Therapist manual* (3rd ed.). Ardmore, PA: Workbook.

Law, D., & Wolpert, M. (2014). *Guide to using outcomes and feedback tools with children, young people and families*. London, England: CAMHS Press.

Lejuez, C. W., Hopko, D. R., Acierno, R., Daughters, S. B., & Pagoto, S. L. (2011). Ten Year Revision of the Brief Behavioral Activation Treatment for Depression: Revised Treatment Manual. *Behavior Modification*, *35*(2), 111–161. doi: 10.1177/0145445510390929

Lejuez, C. W., Hopko, D. R., & Hopko, S. D. (2001). A brief behavioral activation treatment for depression: Treatment manual. *Behavior Modification*, *25*(2), 255-286.

Lewinsohn, P. M. (1974). A behavioral approach to depression. In R. J. Friedman, & M. M. Katz (Eds.), *The psychology of depression: Contemporary theory and research* (pp. 157-178). New York: John Wiley & Sons.

Liese, B. S., Barber, J. P., & Beck, A. T. (1995). *The cognitive therapy adherence and competence scale.* Unpublished manuscript.

Martell, C. R., Addis, M. E., & Jacobson, N. S. (2001). *Depression in context: Strategies for guided action.* New York: Norton

McLeod, B. D., Southam-Gerow, M. A., Rodriguez, A., Quinoy, A. M., Arnold, C. C., Kendall, P. C., & Weisz, J. R. (2018). Development and Initial Psychometrics for a Therapist Competence Instrument for CBT for Youth Anxiety. *Journal of Clinical Child and Adolescent Psychology, 47*(1), 47-60. doi: 10.1080/15374416.2016.1253018

Miller, G. E. (1990). The assessment of clinical skills/competence/performance. *Academic Medicine, 65*, 63-67. doi: 10.1097/00001888-199009000-00045

Miller, S. D., & Duncan, B. L. (2000). *The outcome rating scale*. Chicago: Author.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine, 151*(4), 264-269. doi: 10.7326/0003-4819-151-4-200908180-00135

Mokkink, L. B., Prinsen, C. A., Patrick, D. L., Alonso, J., Bouter, L. M., de Vet, H. C., & Terwee, C. B. (2018a). *COSMIN Methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs).* User manual.

Mokkink, L. B., Vet, H. C. W., Prinsen, C. A. C., Patrick, D. L., Alonso, J., Bouter, L. M., Terwee, C. B. (2018b). COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. *Quality of Life Research, 27*(5), 1171-1179. doi: 10.1007/s11136-017-1765-4

Muse, K., & McManus, F. (2013). A systematic review of methods for assessing competence in cognitive-behavioural therapy. *Clinical Psychology Review, 33*, 484-499. doi: 10.1016/j.cpr.2013.01.010

NICE (2005). CG28 *Depression in children and young people: identification and management.* NICE clinical guideline 28, issued September 2005.

NICE (2009). CG90 *Depression in adults: The treatment and management of depression in adults*

(update). NICE clinical guideline 90, issued October 2009.

NICE (2019). *Depression in children and young people: identification and management: Draft for consultation.* Retrieved from https://www.nice.org.uk/guidance/gid-ng10106/documents/draft-guideline

Orchard, F., Pass, L., Marshall, T., & Reynolds, S. (2017). Clinical characteristics of adolescents referred for treatment of depressive disorders. *Child and Adolescent Mental Health, 22*(2), 61-68. doi: 10.1111/camh.12178

Pass, L., Brisco, G., Hodgson, E., & Reynolds, S. (2015). *Brief Behavioural Activation for depressed adolescents (Brief BA) therapist guide.* Unpublished manuscript.

Pass, L., Hodgson, E., Whitney, H., & Reynolds, S. (2017). Brief Behavioral Activation Treatment for Depressed Adolescents Delivered by Nonspecialist Clinicians: A Case Illustration. *Cognitive and Behavioral Practice, 25*(2), 208-224. doi: 10.1016/j.cbpra.2017.05.003

Pass, L., Lejuez, C. W., & Reynolds, S. (2017). Brief Behavioural Activation (Brief BA) for Adolescent Depression: A Pilot Study. *Behavioural and Cognitive Psychotherapy, 46*(2), 182-194. doi: 10.1017/S1352465817000443

Pass, L., & Reynolds, S. (2014). *Treatment manual for brief behavioural activation (Brief BA) for depressed adolescents.* Unpublished manuscript.

Pass, L., Sancho, M., Brett, S., Jones, M., & Reynolds, S. (2018). Brief Behavioural Activation (Brief BA) in secondary schools: a feasibility study examining acceptability and practical considerations*. Educational and Child Psychology, 35*(2), 1-25.

Pass, L., Whitney, H., & Reynolds, S. (2016). Brief Behavioral Activation for Adolescent Depression: Working with Complexity and Risk. *Clinical Case Studies*, *15*(5), 360–375. 10.1177/1534650116645402

Perepletchikova, F., & Kazdin, A. E. (2005). Treatment integrity and therapeutic change: Issues and research recommendations. *Clinical Psychology: Science and Practice, 12*(4), 365-383. doi: 10.1093/clipsy.bpi045

Rapley, H. A., & Loades, M. E. (2018). A systematic review exploring therapist competence, adherence, and therapy outcomes in individual CBT for children and young people. *Psychotherapy Research, 22*, 1-10*.* doi: 10.1080/10503307.2018.1464681

Richards, D. A., Ekers, D., McMillan, D., Taylor, R. S., Byford, S., Warren, F. C., . . . Finning, K. (2016). Cost and Outcome of Behavioural Activation versus Cognitive Behavioural Therapy for Depression (COBRA): a randomised, controlled, non-inferiority trial. *The Lancet, 388*, 871-880. doi: 10.1016/S0140-6736(16)31140-0

Roth, A. D., & Pilling, S. (2007). *The competences required to deliver effective cognitive and behavioural therapy for people with depression and with anxiety disorders*. London: Department of Health.

Sadler, K., Vizard, T., Ford, T., Goodman, A., Goodman, R., & McManus, S. (2018). *Mental Health of Children and Young People in England, 2017.* Retrieved from https://www.gov.uk/government/statistics/mental-health-of-children-and-young-people-in-england-2017-pas

Sharpless, B. A., & Barber, J. P. (2009). A conceptual and empirical review of the meaning, measurement, development, and teaching of intervention competence in clinical psychology. *Clinical Psychology Review, 29*(1), 47-56. doi: 10.1016/j.cpr.2008.09.008

Shaw, B. F., Elkin, I., Yamaguchi, J., Olmsted, M., Vallis, T. M., Dobson, K. S., . . . Imber, S. D. (1999). Therapist competence ratings in relation to clinical outcome in cognitive therapy of depression. *Journal of Consulting and Clinical Psychology, 67*, 837–846. doi: 10.1037/0022-006X.67.6.837

Shirk, S. R., & Saiz, C. (1992). Clinical, empirical, and developmental perspectives on the therapeutic relationship in child psychotherapy. *Development and Psychopathology, 4*, 713–728. doi: 10.1017/S0954579400004946

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing reliability. *Psychological Bulletin, 86*, 420-428. doi: 10.1037/0033-2909.86.2.420

Southam-Gerow, M.A., McLeod, B.D., Arnold, C.A., Rodriguez, A., Cox, J.R., Reise, S.P., . . . Kendall, P. C. (2016). Initial development of a treatment adherence measure for cognitive behavioral therapy for child anxiety. *Psychological Assessment, 28(1),* 70–80. doi: 10.1037/pas0000141

Stallard, P. (2005). *A Clinicians Guide to Think Good Feel Good: using CBT with children and young people*. Winchester: John Wiley.

Stallard, P., Myles, P. & Branson, A. (2014) The Cognitive Behaviour Therapy Scale for Children and Young People (CBTS-CYP): development and psychometric properties.

Behavioural and Cognitive Psychotherapy, 42(3), 269-282. doi: 10.1017/S135246581300115X

Terwee, C. B., Bot, S. D., de Boer, M. R., van der Windt, D. A., Knol, D. L., Dekker, J., . . . de Vet, H. C. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol, 60*(1), 34-42. doi: 10.1016/j.jclinepi.2006.03.012

Totman, J., Reynolds, S., Brett, S., & Pass, L. (2019). *Brief Behavioural Activation Therapy for Adolescent Depression in Schools: Clinical and Educational Benefits.* Manuscript in Preparation.

Vogel, W. (2012). Depression in Children and Adolescents. *Continuing Medical Education, 30*(4), 114-117.

von Consbruch, K., Clark, D. M., & Stangier, U. (2012). Assessing therapeutic competence in cognitive therapy for social phobia: Psychometric properties of the cognitive therapy competence scale for social phobia (CTCS-SP). *Behavioural and Cognitive Psychotherapy, 40*, 149-161. doi: 10.1017/S1352465811000622

Waltz, J., Addis, M. E., Koerner, K., & Jacobson, N. S. (1993). Testing the integrity of a psychotherapy protocol: Assessment of adherence and competence. *Journal of Consulting and Clinical Psychology, 61*(4), 620-630. doi: 10.1037/0022-006X.61.4.620

Webb, C. A., DeRubeis, R. J., & Barber, J. P. (2010). Therapist adherence/ competence and treatment outcome: A meta-analytic review. *Journal of Consulting and Clinical Psychology, 78*, 200-211. doi: 10.1037/a0018912

Whisman, M. A. (1993). Mediators and moderators of change in cognitive therapy of depression. *Psychological Bulletin, 114,* 248–265. doi: 10.1037/0033-2909.114.2.248

Wolpert, M., Fugard, A. J. B., Deighton, J., & Gorzig, A. (2012). Routine outcomes monitoring as part of children and young people's improving Access to Psychological Therapies (CYP IAPT) – Improving care or unhelpful burden? *Child and Adolescent Mental Health, 17*(3), 129–130.

Wu, S. M., Whiteside, U., & Neighbors, C. (2007). Differences in inter-rater reliability and accuracy for a treatment adherence scale. *Cognitive Behaviour Therapy, 36*, 230-239. doi: 10.1080/16506070701584367

Young, J., & Beck, A. T. (1980). *Cognitive Therapy Scale: Rating manual.* Unpublished manuscript, University of Pennsylvania, Philadelphia.

Zarafonitis-Müller, S., Kuhr, K., & Bechdolf, A. (2014). The relationship between therapist's competence and adherence to outcome in cognitive-behavioural therapy: Results of a meta-analysis. *Fortschritte der Neurologie-Psychiatrie, 82*(9), 502–510.