

International Journal of Developmental Disabilities

Inter-rater reliability of Treatment Fidelity and Therapeutic Alliance Measures for Psychological Therapies for Anxiety in Young People with Autism Spectrum Disorders --Manuscript Draft--

Manuscript Number:	JDD166R2
Full Title:	Inter-rater reliability of Treatment Fidelity and Therapeutic Alliance Measures for Psychological Therapies for Anxiety in Young People with Autism Spectrum Disorders
Article Type:	Original Research Paper
Keywords:	Anxiety; Autism Spectrum Disorders; Inter-rater Reliability; Therapeutic Alliance; Treatment Fidelity.
Corresponding Author:	Suzanne Murphy, PhD University of Bedfordshire UNITED KINGDOM
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	University of Bedfordshire
Corresponding Author's Secondary Institution:	
First Author:	Rachel Brown, BSc
First Author Secondary Information:	
Order of Authors:	Rachel Brown, BSc Zeinab Iqbal Laura Reynolds, MSc Dee. A. Press Hadi Shaker-Naeeni Louise Scrivener Suzanne Murphy, PhD Uttom Chowdhury
Order of Authors Secondary Information:	
Abstract:	<p>Objectives: This article presents work undertaken to establish inter-rater reliability for a measure of treatment fidelity and a measure of therapeutic alliance for therapies for anxiety for young people with autism spectrum disorders. The discussion and decision-making processes behind achieving consensus of raters are rarely published. Margolin et al. (1998) have highlighted this issue and called for researchers to communicate the details of their observational and rating procedures. This article is a response to their call for greater transparency so that these methods are readily accessible for comparison with other studies.</p> <p>Methods: Participants were young people with autism spectrum disorders receiving treatment for anxiety, clinical staff treating these young people and the independent raters assessing the treatment sessions. We report: (a) the processes involved in establishing inter-rater reliability for two instruments (b) the results obtained with a sample of young people with autism spectrum disorders using these instruments.</p> <p>Results and conclusions Results demonstrate that it was possible to attain satisfactory inter-rater reliability with each of these two instruments with a client group with autism spectrum disorders, even though the instruments were originally designed for typically-developing populations.</p>

6. In cases where the article is a translation of a previously published article, the author(s) warrant(s) that full permission has been obtained to carry out and publish the translation of the article from its original language, and that it in no way infringes the rights of others (including the holders of copyright in material used within the original Article).
7. The author(s) warrant(s) that the article contains no plagiarised material and agree(s) that the article may be checked for plagiarism and other ethical questions (including checks using automated software packages). The Publisher reserves the right to withdraw or retract any article for which, in its reasonable opinion, valid concerns exist over the originality of the article, or its compliance with the Journal's ethical policy. Maney's plagiarism and ethical policy may be read in full at www.maneyonline.com/page/authors/publishingethics-general.

*If the author(s) cannot assign copyright for any or all the supplementary material, the clause in parentheses should be struck out and a licence to publish provided for this material. If copyright is held by a third party, permission to reproduce this material must be obtained from the copyright holder as laid out in paragraph 5 of this form and submitted to the publisher.

Copyright Assignment

I confirm that all the co-authors, named above in 1, know that the article has been submitted to the Journal.

I hereby sign this Assignment of Copyright with the full knowledge and agreement of all authors.

Signed: 

Date: 18th April 2014.

Printed name: DR SUZANNE MURPHY

Please note: it is essential for publication that Maney Publishing receive this signed Assignment of Copyright form. Without it, we will be unable to proceed to publication.

Please complete all specified fields and return this signed form to the **Production Editor** of the Journal.

Production Editor, JDD, for Maney Publishing
 jdd@maneypublishing.com; fax: +44 (0) 20 7451 7307
 1 Carlton House Terrace, London SW1Y 5AF, UK

International Journal of Developmental Disabilities ('the Journal')

ASSIGNMENT OF COPYRIGHT

In order for W.S Maney & Son Ltd, trading as Maney Publishing ('the Publisher'), to ensure the widest possible dissemination and protection of articles published in the Journal, we request authors to assign worldwide copyright in print, digital and other media in their article, including abstracts, to The British Society for Developmental Disabilities. This enables the Publisher to ensure maximum international copyright protection against infringement, and to disseminate your article, and the Journal, to the widest possible worldwide readership. Should the article be rejected from the Journal for any reason, all rights assigned within this document revert back to the author(s) of the article.

Please read the explanatory notes of this agreement on the following page.

1. In consideration of the undertaking set out in paragraph 2, the author(s) as beneficial owner(s) hereby assign(s) to The British Society for Developmental Disabilities all rights, title and interest in the copyright throughout the world in the article named below (**which shall be taken to include any or all supplementary material associated with the article***) for the full legal term of copyright.

Article title: *Inter-rater reliability of Treatment Fidelity and Therapeutic Alliance measures for psychological therapies for Anxiety in young people with Autism Spectrum disorders.*
 Name(s) of Author(s): *Rachel Brown, Zeinab Khal, Laura Reynolds, Suzanne Murphy, Hadi Shaker-Naeni, Louise Scrivener, Utham Chandhok.*
 Manuscript no./ref. (if known) *JDD166.*

If this paper is sponsored for Open Access by your grant funders, please follow the MORE OpenChoice route by signing and returning this copyright assignment to the address below, and find details of the policy and how to make payment at www.maneyonline.com/openaccess.

So that there should be no doubt, it is understood and agreed that this assignment includes the right to publish or adapt (subject to paragraphs 3 and 4) the material in the article for use in conjunction with computer systems, including networks. The article may be published in printed, CD-ROM, microfiche, on-line, or other machine-readable form, or in any other format or medium (electronic or otherwise) approved by The British Society for Developmental Disabilities; and subject to data mining.

2. Maney Publishing, as the Journal's publisher, hereby undertakes to prepare for publication and publish in the Journal, the Article named in paragraph 1.
3. The editor of the Journal, and the Publisher, are empowered to make such editorial changes as may be necessary to make the Article suitable for publication. Every effort will be made to consult the Author(s) if substantive changes are required.
4. The author(s) hereby assert(s) their moral rights under the terms of the Copyright Designs and Patents Act 1988 to be identified as the Author(s) of the Article.
5. The author(s) warrant(s) that the article is the author(s)' original work, has not been published before, either in part or whole, and is not currently under consideration for publication elsewhere. The author(s) also warrant(s) that (a) the article contains no libellous or unlawful statements, (b) it in no way infringes the rights of others (including the holders of copyright in material used within the article), (c) that permission has been obtained to reproduce any material or illustrations for which they do not hold copyright, and that (d) the author(s), as the owners of the copyright, is (are) entitled to make the assignment set out in this agreement. The author(s) hereby indemnifies/indemnify the Publisher against any claims for breach of the warranties given above.

Table 1. *Characteristics of client participants*

Client Number	Age	Gender	Trial Allocation
1	15	female	CBT
2	16	male	CBT
3	17	male	CBT
4	16	female	CBT
5	15	male	CBT
6	14	female	CBT
7	16	female	counselling
8	16	female	counselling
9	15	male	counselling
10	15	male	counselling
11	15	male	counselling
12	16	male	counselling
13	12	female	counselling

Table 2 – means and standard deviation for TPOCS-A scale items

TPOCS-A scale item	<i>McLeod & Weiss study2005 M (SD)</i>	<i>Current study M (SD)</i>
1. Client experienced therapist as supportive	2.96 (0.81)	3.43 (0.99)
2. Client shows hostility to therapist	4.49 (0.56)	4.81 (0.26)
3. Client demonstrates positive affect towards therapist	2.54 (0.73)	3.38 (0.55)
4. Client shares experience with therapist	3.08 (0.89)	2.57 (1.39)
5. Client appears uncomfortable with therapist	4.25 (0.65)	2.86 (2.04)
6. Degree client and therapist have difficulty interacting	4.31 (0.70)	4.24 (0.46)
7. Client used therapeutic tasks to make changes	0.85 (0.74)	1.81 (1.60)
8. Client did not comply with therapeutic tasks	4.49 (0.78)	4.15 (0.72)
9. Client and therapist work equally on tasks	2.87 (0.78)	3.05 (1.18)

Table 3 – means and standard deviation for CBT subscale items for the PCTPRS

PCTPRS CBT scale item	Counselling sessions		CBT sessions		<i>P</i> ^a
	Godfrey et al (2007)	Current study	Godfrey et al (2007)	Current study	
1. Rationale for behavioural procedures	1.4 (0.7)	1.2 (0.4)	5.1 (1.3)	4.8 (0.7)	<.01 **
2. Practise alternative behaviours	1.8 (1.2)	1.1 (0.2)	4.5 (1.3)	4.3 (2.4)	<.01 **
3. Homework assigned/reviewed	1.3 (0.7)	1.1 (0.1)	4.3 (1.3)	5.5 (1.0)	<.01 **
4. Rationale for cognitive procedures	1.1 (0.4)	1.1 (0.2)	3.7 (1.8)	4.9 (2.2)	<.01 **
5. Recognise cognitive errors	1.1 (2.7)	1.0 (0.1)	2.7 (1.5)	2.6 (2.2)	<.01 **
6. Search for alternative explanations	3.4 (1.2)	1.1 (2.1)	3.4 (0.9)	2.7 (0.3)	.14
7. Maintain gains	1.1 (0.4)	1.3 (0.5)	1.6 (0.6)	3.3 (1.7)	.03 *

^a = *p*-values for differences between counselling sessions and CBT sessions in the current study, * = *p* <.05, ** = *p* <.01.

Table 4 – means and standard deviation for counselling subscale items for the PCTPRS

PCTPRS counselling scale item	Counselling sessions		CBT sessions		<i>P</i> ^a
	Godfrey et al (2007)	Current study	Godfrey et al (2007)	Current study	
1. Negotiating style	4.5 (0.9)	5.4 (0.7)	3.7 (0.8)	5.0 (0.2)	.19
2. Relating change to therapy	1.4 (0.7)	1.9 (2.1)	1.1 (0.3)	2.7 (1.2)	.14
3. Understanding hypotheses	2.9 (1.1)	3.7 (0.9)	3.2 (1.1)	3.1 (0.9)	<.01**
4. Focusing	2.5 (1.5)	4.3 (1.1)	1.2 (1.2)	2.5 (0.4)	<.01**
5. Acknowledging affect	3.4 (1.3)	4.4 (1.3)	1.6 (0.7)	2.2 (0.9)	<.05*
6. Accepting affect	2.9 (1.2)	3.1 (1.2)	1.4 (0.5)	2.4 (1.0)	.12
7. Limitations	1.1 (0.4)	1.0 (0.1)	1.0 (0.2)	1.0 (0.1)	1.00

^a = *p*-values for differences between counselling sessions and CBT sessions in the current study, * = *p*<.05, ** = *p*<.01.

Inter-rater reliability of Treatment Fidelity and Therapeutic Alliance Measures for
Psychological Therapies for Anxiety in Young People with Autism Spectrum Disorders

Ms Rachel Brown¹, Dr Zeinab Iqbal², Ms Laura Farley², Ms DeeA. Press², Dr Hadi Shaker-
Naeeni², Ms Louise Scrivener², Dr Suzanne Murphy*³, and Dr Uttom Chowdhury²

1. Child & Adolescent Mental Health Service, Berkshire Healthcare NHS Foundation Trust,
Slough, Great Britain.

² Specialist Child and Adolescent Mental Health Services, South Essex Partnership
University NHS Foundation Trust, Dunstable, Great Britain.

³ Institute of Health Research, University of Bedfordshire, Great Britain.

*Corresponding author Suzanne.murphy@beds.ac.uk

This project was supported by a grant from the South Essex Partnership University NHS
Foundation Trust. There are no conflicts of interest for any of the authors.

Comments from the Editors and Reviewers:

Reviewer #1: Thank you for revising the manuscript in accordance with the feedback from the reviewers. The authors have adequately responded to the comments and have improved the manuscript significantly. This work in this manuscript should make a significant contribution to the studies of CBT in youth with ASD.

A few minor points need to be corrected in the references

1) the references should be given in the usual IJDD format e.g. Lindner, M., Ko Iker, S., Schulze, A., Christensen, E., Greenberg, C. R. and Hoffmann, G. F. 2004. Neonatal screening for glutaryl-CoA dehydrogenase deficiency, J. Inherit. Metab. Dis., 27, 851–859.

References have now been amended to this format.

2) please use "and" (not "&")

'And' has now been used throughout the reference section in place of '&'.

3) Lerner et al., 2001 is cited on page 9 but this reference is not listed in the reference section

Apologies, this was a typing error and should have read 'Lerner, 2011', this has now been amended.

4) Perepletchikova et al., 2007 is cited in the text (e.g. twice on page 1) but in the reference section the year is given as 2006

This should read 2007 and has now been corrected.

Reviewer #2:

p. 2: "We focus here specifically on two of these neglected aspects: The..." - This section would be better suited after your discussion of alliance in the next paragraph. This way, you will have described treatment fidelity and it's importance, then alliance and it's importance, followed by your purpose - i.e., focusing on these two neglected aspects... This section might be better suited around p.5/6 where you refer to your study aims.

The first part of this section:

'We focus here specifically on two of these neglected aspects: The application of a measure of treatment fidelity to determine adherence to two kinds of psychological intervention and of a measure of therapeutic alliance with these interventions for young people with ASD' has been inserted into the middle of page three (with minor amendments to retain the sense of the sentence in a different section) just before a description of the two specific instruments used.

The second part of this section:

'McLeod and Weiss (2005) have stressed the importance of obtaining ratings from independent and blinded raters, as opposed to client's self-ratings, or clinician ratings, to avoid possible bias. Thus the measures used here are designed for blind observers, and, as for all observational measures, more than one observer should be used and inter-rater reliability between the observers reported' has now been inserted as the second paragraph on p6, after the descriptions detailing PCTPRS and TPOCS-A.

p. 4 - TOPCS-A should be "TPOCS-A"

This has now been changed

p. 9 - change "For the TPOCS-A, McLeod and Weiss (2005) found adequate inter-rater reliability ranging from .40 to .75 for the questionnaire items and internal consistency Cronbach's $\alpha = .95$ " to "For the TPOCS-A, McLeod and Weiss (2005) found adequate inter-rater reliability ranging from .40 to .75 for the questionnaire items and high internal consistency (Cronbach's $\alpha = .95$)"

-similarly at the bottom of p.9, use the symbol for alpha when in parentheses. If you are using the term in text, spell out the term 'alpha'

The above changes have been made on p9.

Watch APA formatting - e.g., p.9: "Godfrey et al., (2007) report..." - there should not be a "," in the sentence. Similarly, p. 19: "Sukhodolsky et al., (2013) identified..." - there are a couple of other places throughout the manuscript too.

Commas have now been removed from all 'at al.' references in the text.

Edit on p.10: I'm not sure that you need to provide details about evidence for the factor structure of the measure. The psychometric properties alone are fine.

Details of the factor structure for both the PCTPRS and the TPOCS-A have now been removed from pages 9 and 10 as requested.

Edit on bottom of p.11 - change to "inter-rater reliability"

This has now been changed.

I believe in one or two places you did not capitalize "kappa" - please edit to "Kappa"

This has now been changed.

Added reference: Sukhodolsky - confirm APA formatting

This reference, along with all the others, has now been amended to IJDD usual format as requested by reviewer 1.

Introduction

This article presents work undertaken to establish the use of measures of treatment fidelity and therapeutic alliance for young people with autism spectrum disorders (ASD). Measures of treatment fidelity and therapeutic alliance are more frequently developed for typically-developing populations; this article explores the potential for broadening their use to an ASD sample.

Treatment or intervention fidelity, also known as treatment/intervention integrity signifies the extent to which a treatment is implemented as intended (Vermilyea et al. 1984; Yeaton & Sechrest, 1981). Treatment fidelity measures are crucial to psychological interventions, both to ascertain their efficacy (i.e., how well they work under ‘ideal’ highly controlled settings such as clinical trials) and their effectiveness (i.e., how well they work in the ‘real-world’). Without them, it is difficult to interpret the outcomes of trials of psychosocial interventions.

Whilst acknowledging the importance of practitioner autonomy in clinical practice, and the need to tailor interventions to individuals’ unique needs; for the purposes of research and practice, it is important to ensure that the intervention offered is the intended one, and that clinicians have not ‘drifted’ from the treatment protocol.

Systematic reviews of clinical trials of psychological treatments (e.g., Perepletchikova et al. 2007, Weisz et al. 2005) frequently report that treatment fidelity measures are inconsistently applied or altogether lacking. Indeed, Weisz et al. (2005) found that only 32.2% of studies included used any form of fidelity check and Perepletchikova et al. (2007) claimed that only 3.5% of the psychosocial interventions that they reviewed addressed treatment fidelity adequately. Treatment fidelity is regarded as a multidimensional construct comprising a number of different components (McLeod et al. 2009). McArthur et al. (2012) report that some aspects of treatment fidelity are generally well-reported, such as specific, clear details of the intervention, number and duration of sessions and manualised versions of the

intervention to ensure consistency. However, these same authors also point out that assessments to check that clinicians adhere to the intervention and measures of non-specific intervention effects, such as aspects of the therapeutic relationship (e.g., therapeutic alliance), are rarely included.

The alliance between the therapist and client has for some time been considered as an important factor in the outcomes of interventions. The impact that the therapeutic alliance may have on therapy outcomes is clearly an important issue both for researchers and clinicians. A good therapeutic relationship is thought to facilitate a reduction in symptomology in children with anxiety by improving both the child's involvement in therapy (Chu et al. 2004) as well their participation in crucial exposure tasks (Kendall & Ollendick 2004).

Although a number of authors (e.g., McLeod et al. 2009, Perepletchikova 2011) have argued for the inclusion of more robust measures of fidelity in research studies evaluating treatments, a problem faced by researchers working with specialised client groups is that fully validated measures for these populations are rarely available. Researchers are thus faced with the choice of using these measures, with or without slight modification to allow for the characteristics of a particular population, or of undertaking major adaptations to an instrument with subsequent full evaluation of the psychometric properties. Whilst the latter course would seem preferable, this is not always practical or necessary for every instrument and every client group. Initial exploration of the properties of instruments may reveal that, with some minor amendments, such as more detailed instructions, the instrument is adequate for the purposes of a particular study and this is a course frequently adopted by researchers. The discussion and decision-making processes behind such revisions are rarely published. Margolin et al. (1998) have highlighted this issue and called for researchers to communicate the details of their observational coding and rating procedures so that these are readily

accessible for comparison with other studies. Thus, in response to Margolin et al. (1998), we report: (a) the processes involved in establishing inter-rater reliability for these two instruments (b) the results obtained with a sample of children with ASD using these instruments, and interpret what these findings indicate.

In reporting the work undertaken for the purposes of inter-rater reliability here, we hope that our methods will be available to other researchers for the purposes of implementing these and other observational measures, particularly with clients with ASD. We focus specifically on two of the neglected aspects of treatment fidelity (McArthur et al. 2012): The application of a measure of treatment fidelity to determine adherence to two kinds of psychological intervention and of a measure of therapeutic alliance with these interventions for young people with ASD.

The two instruments in question, the Therapy Process Observational Coding System for Child Psychotherapy-Alliance scale (TPOCS-A, McLeod & Weiss 2005) and the Primary Care Therapy Process Rating Scale (PCTPRS, Godfrey et al. 2007) are currently in use in a randomised controlled trial comparing cognitive behavioural therapy (CBT) with supportive counselling for adolescents with co-morbid anxiety and Autism Spectrum Disorder (ASD). This study, by the authors of this article, is on-going and results will be reported at a later date. Here, we report the processes involved in reaching inter-rater reliability for the two instruments.

The choice of TPOCS-A and PCTPRS was based on trying to find instruments as closely suited to our purposes as possible. Therapeutic alliance can be measured using the child's own self-report, parent report if they are included in the therapy, the treating clinician's report or finally, an outside independent observer's report. Each of these methods has their advantages and disadvantages. The child's self-report has the advantage of directly assessing the perspective of the person to whom the therapy is aimed. However, demand characteristics

may make the child feel obliged to give a positive report of the therapy. In our study, parent-report would be less than ideal, as parents were involved in some, but not all therapeutic sessions and hence would give only a partial report. Clinician report may be subject to a biased view of the success of the therapy. Therefore, we wished to use an independent observer's perspective. The TPOCS-A was designed to address limitations of previous instruments by using independent observer's report; most previous instruments rely on self-report from the child, parent or therapist. Furthermore, previous literature has identified three aspects of therapeutic alliance; bond, task and goal. The majority of previous measures are concerned with only one or two of these three measures whereas TPOCS-A encompasses all three (McLeod & Weiss 2005) and has been used successfully on youth with internalising disorders and ADHD. Therapeutic alliance and changes in alliance over the course of treatment as measured by the TPOCS-A were positively associated with treatment outcomes for individual CBT for children with anxiety (Chiu et al, 2009), parent friendship training for children with ADHD (Lerner et al. 2011) but showed mixed outcomes for a comparison of group and individual CBT for children with anxiety (Liber et al, 2010). Many earlier measures of therapeutic alliance were downward revisions of adult instruments, whereas the TPOCS-A was designed specifically for youth (McLeod & Weisz 2005). Taking into consideration all of these factors, TPOCS-A was our instrument of choice for the study. For selection of a measure to distinguish between CBT and counselling, the available range was extremely limited. For many trials of therapy, treatment fidelity is commonly assessed using a checklist compiled by the authors of the treatment manual. The checklist ensures that various points are included in therapy sessions, giving an indication that the treatment manual is being adhered to. However, as we were comparing CBT against counselling (as opposed to more typical comparisons of therapy versus treatment as usual or waiting list) we wished to apply a more stringent fidelity test to ensure that none of the counselling therapists were

using CBT techniques. Comparison of CBT against counselling, or of two types of psychotherapy with each other is rarely conducted, despite calls for this type of comparison (e.g., Sukhodolsky et al. 2013). An extensive literature search identified only one fidelity measure designed *specifically* to compare CBT and counselling in either adults or children; the PCTPRS. Although this was originally designed for use with adults, the focus of the scale is on assessing therapist techniques rather than client responses, thus it was decided to trial this measure with our sample to explore its possible use.

Godfrey et al. (2007) found that the PCTPRS distinguished between the CBT and counselling sessions well, the different sessions attained significantly different scores on the CBT scale for all items bar one, and on all the counselling items bar three. The authors propose that the overlap seen on the counselling scale reflects areas of common ground between CBT and counselling techniques. Godfrey et al. (2007) found that the ‘emotional experiencing’ factor of the counselling subscale was negatively correlated with fatigue at 6 months such that the more a patient engaged in the process of therapy acknowledged and processed their distress, the lower was his/her final fatigue score.

When results of fidelity and therapeutic alliance measures are given in published reports of psychological interventions, the processes for achieving inter-rater reliability for such measures are only briefly outlined. We therefore hope to detail some of the issues we encountered and describe the manner in which they were resolved. The aim of this article is to provide a full description of the processes and decisions involved in achieving inter-rater reliability for these two measures, with regard to the present study. Achieving inter-rater reliability for measures such as these involves making decisions and reaching consensus on a number of issues, some of which are not readily apparent at outset. A particular issue we will address in this article is that we were working with young people with ASD, whereas the measures we were using, the TPOCS-A and the PCTPRS had previously been used with

typically-developing young people and adults and not with special populations. More specifically, the TPOCS-A has been used with young people with anxiety and the PCTPRS in a study of adults with chronic fatigue.

McLeod and Weiss (2005) have stressed the importance of obtaining ratings from independent and blinded raters, as opposed to client's self-ratings, or clinician ratings, to avoid possible bias. Thus the measures used here are designed for blind observers, and, as for all observational measures, more than one observer should be used and inter-rater reliability between the observers reported.

For a client group with ASD, the issue of therapeutic alliance is of particular interest, as making meaningful connections with others is one of the obvious difficulties that this client group face. ASD is often defined in terms of deficits in social communication, social relationships, and social imagination. Both DSM-5 and ICD-10 criteria include symptoms relating to social interaction, communication, restrictive interests and repetitive behaviours, and delays or abnormal functioning. We therefore took these impairments into account when rating alliance: for example within a neuro-typical population, a clear demonstration of a good therapeutic alliance might be positive affect shown towards the therapist such as smiling or showing an interest in the therapist, however individuals with ASD tend not to see the need for such behaviours and thus are less likely to exhibit them during social interaction. The lack of obvious positive affect directed towards the therapist does not, however, indicate that positive therapeutic relationships are impossible to achieve with this client group. Signs of a positive therapeutic alliance may be more subtle within this client group, for example if the client feels able to speak openly with the therapist about their current difficulties. Furthermore, in a neuro-typical population, certain behaviours which may be perceived as indicating anxiety may, in a client with ASD, be a repetitive motor mannerism, thus not necessarily indicating that they are feeling anxious. Part of the task we set ourselves was to

incorporate our knowledge of young people with ASD into the decisions to be made for interrater reliability.

Method

Participants

Participants were (a) young people with autism spectrum disorders receiving treatment for anxiety (b) clinical staff treating these young people (c) the blind raters assessing the treatment sessions.

(a) The young people receiving treatment were aged 12 - 18 years-old with diagnoses of high-functioning autism spectrum disorders and anxiety which had been made by clinicians working within the three recruiting clinics (below). All diagnoses were confirmed and detailed using the Autism Diagnostic Interview (revised) (ADI-R, Le Couteur et al.2003) , the Autism Diagnostic Observation Schedule (ADOS, Lord et al. 2002) and the Anxiety Disorders Interview Schedule, child and parent versions(ADIS C/P, Silverman & Albano 1996) administered by staff who had completed recognised training for these instruments. Young people were randomly allocated to the CBT and counselling arms of the trial (a randomisation protocol was used and treatment allocation was ensured with sealed envelopes). Client gender and age characteristics for the sub-sample of 13 clients used for the inter-rater reliability calculations are given in Table 1. Our sample was somewhat unusual in containing a relatively high proportion of females (46%). This reflects gender distribution of the larger sample of the main study. Females were not purposely over-sampled; this is a chance feature of our sample which was randomly selected (consecutively-recruited clients at three child and adolescent clinics).

Insert table 1 here.

(b) Clinical staff working within three clinics of the National Health Service (NHS) in Great Britain. The clinics were all child and adolescent mental health centres, whose remit is to

provide treatment for moderate to severe cases. Participating staff providing the treatment for this study comprised three consultant child psychiatrists, one clinical psychologist and one counsellor. All were professionally trained in either CBT or counselling.

(c) The three raters were researchers independent of the treating clinicians. All were female, two aged 26 and one aged 53, all were qualified in psychology to at least graduate level, all with experience of using coding and rating systems for the assessment of children and adults with ASD and one of the raters had over 20 years' experience of use of rating instruments within psychological research contexts. Importantly, all of the raters had prior knowledge of ASD and experience of working with young people and adults with ASD. The three raters were blind, independent assessors; none were responsible for delivering the treatments, all were blind as to whether CBT or counselling was being delivered in each viewed session and all were blind as to the outcomes of the treatments.

Video-recording of treatment sessions

CBT and counselling intervention sessions were video-recorded by the treating clinicians. All sessions were video-recorded, even though only a subsample was rated, so that clinicians would be blind as to which sessions were assessed for treatment fidelity and therapeutic alliance. Recordings were taken such that both the clinician and the client were in full view thus enabling rating of facial expressions and body language as well as spoken communication.

Ethical approval

Written consent for video-recording sessions was obtained from the young people participating before treatment commenced. The full protocol for the study was reviewed by the National Research Ethics Service Committee (East of England) and approval granted.

Treatment Fidelity and Therapeutic Alliance Measures

The TPCOS-A was selected as a measure of therapeutic alliance as it was originally

developed for use in young people with anxiety (McLeod 2005). It is comprised of nine items which explore the two dimensions of therapeutic alliance – ‘bond’, which refers to the affective aspects of the alliance and ‘task’ which explores to the extent to which the client engages in the therapeutic tasks (McLeod & Weisz 2005). Five previous studies (Chiu et al. 2009, Fjermestad et al. 2012, Langer et al. 2011, Lerner et al. 2011, Liber et al. 2010, McLeod & Weisz 2005) have used TPOCS–A in clients groups with ADHD, anxiety, and depression.

For the TPOCS-A, McLeod and Weiss (2005) found adequate inter-rater reliability ranging from .40 to .75 for the questionnaire items and high internal consistency (Cronbach’s $\alpha = .95$).

The PCTPRS was originally designed as a process measure to examine adherence to CBT versus supportive counselling in an RCT investigating treatment for chronic fatigue and to provide a measure of therapeutic alliance (Godfrey et al. 2007). Thus, the PCTPRS comprises three sub-scales, one assessing the extent to which the therapist uses CBT techniques, one assessing the use of counselling and one to provide a measure of therapeutic alliance. For our main study, the comparison of CBT and counselling for young people with ASD, we used the first two of these subscales but used the TPOCS-A to measure therapeutic alliance instead of the PCTPRS alliance subscale. The CBT and counselling subscales comprise 7 items each. Godfrey et al. (2007) report mostly adequate inter-rater reliability for the different scales of the PCTPRS using weighted Kappa, with agreement above 90% and values of $k = 0.5$ (range 0.2–0.8). The internal reliability of all sub-scales was tested using Cronbach’s alpha, the subscales had good reliability, with a coefficients above 0.8.

Procedure

A random sample of the video-recorded treatment sessions were then viewed by the three authors and rated for treatment fidelity and therapeutic alliance using the TPOCS-A and

PCTPRS. Sessions were viewed in their entirety from start to finish and were typically 50 minutes long.

The two scales of the PCTPRS relating to CBT and counselling were used to provide a measure of treatment fidelity to their practised therapy by the clinicians taking part in the study. The PCTPRS subscales aim to distinguish between these two therapies, but also to indicate areas of overlap and treatment similarity, thus, every video-recording was rated on both the CBT and the counselling scales regardless of the intervention being delivered. There are 7 items relating to CBT and 7 relating to counselling, raters are required to score each item on a 7-point Likert rating scale (from 1- not at all, to 7-extensively). The TPOCS-A is broken down into two scales; one relating to the therapeutic 'bond' (6 items), and one relating to 'therapeutic tasks' (3 items). Raters are required to score each item on a 6-point Likert rating scale (from 0 - not at all, to 5 - a great deal). Recommended standards for rater training were adhered to as follows:

- (a) Raters read both manuals for these instruments alone before training commenced;
- (b) Discussion took place to clarify each rater's understanding of the scale items, from the outset and all the way through the process of establishing inter-rater reliability, the characteristics of our ASD sample were borne in mind in relation to the ratings ;
- (c) All three raters viewed and rated four sample recordings together, whilst discussing how they would apply the scale items;
- (c) All three raters then independently rated three sample recordings and then met afterwards to discuss the results. These sample recordings were selected to be therapy sessions taken from the main study, but were not to be used in the final analysis;
- (d) Ratings were then compared, and where discrepancies had arisen, they were discussed until a consensus was reached. This process was then repeated four times, with each rater

independently ratings 4-5 recordings. For each iteration, new recordings were used to prevent raters remembering previous recordings and discussions. Each time, recordings for this practice stage were used that were from the main study but that were not used in the final analysis for the inter-rater reliability (inter-rater reliability for all four iteration given in results section). After four repetitions, it was felt that agreement was good (i.e., $k > .40$) between the raters and should be tested for the final time for the study;

(e) In order to test the quality of the agreement for the final time, the three raters rated 13 recordings (6 CBT sessions and 7 counselling sessions) independently, and the inter-rater reliability was calculated for these 13 video-recordings using weighted Kappa.

Results

We provide here a description of the process of achieving inter-rater reliability, and also the means, standard deviations and inter-rater reliability ratings obtained for 13 video-recordings. Prior to obtaining these final inter-reliability ratings, we obtained ratings for Kappa for the first iteration of .40, .62 and .31 for TPOCS-A, the CBT scale and the counselling scale respectively, followed by ratings of .59, .66 and .29 for the second iteration, .61, .71. and .45 for the third and finally .74, .79 and .50 for the fourth iteration.

Issues encountered around achievement of inter-rater reliability

We describe below some of the issues encountered at stage (d) above, and how these issues were resolved, and which were specific to an ASD client group and which apply more generally to all rating systems of this kind.

1. Global impressions ratings

The rating schemes for both the treatment fidelity and therapeutic alliance were molar rather than microanalytic (Bull 2002). This means that raters worked without transcripts, relying instead on memory, notes and an overall, global impression of the therapeutic session. Due to the complexity of the data i.e., the patterns of interaction between the therapist and the client,

raters sometimes attended to different aspects of the communication. Typically, this involved the observation of facial expressions, where one rater may have been observing the face of either the therapist or the client, whereas at the same point in the recording, the other rater may be have been observing the other person. With facial expressions and other non-verbal communication changing rapidly, this could be a source of occasional disagreement.

Normally, however, the salient and significant incidents of the session were noted by all raters. This was one of the most difficult issues to resolve however, and could only really be addressed by ensuring that sufficient numbers of recordings were viewed by the rating team.

2. Clarifying definitions

There were individual items of the rating scales which presented a greater source of uncertainty and required more extensive discussion to reach consensus. Sometimes reaching consensus involved further clarifying and defining guidelines given in the manuals to provide greater specificity. Calculating reliability as early as possible provides important feedback regarding which items are more problematic, and so inter-rater reliability was calculated at regular intervals for this reason. When discussing rating decisions, it is important to describe the thinking processes behind the decision, as this is reveals where the discrepancy lies. The following examples illustrate this process:

Item 9 of the TPOCS-A ‘To what extent did the therapist and client work together equally on tasks?’ which requires raters to make a decision concerning the extent to which client and therapist worked together on therapeutic tasks. This was an item that demanded careful consideration during training, owing to the characteristics of ASD clients. A not uncommonly observed type of conversation in this study, for both CBT and counselling interventions, was one in which the therapist interviewed the client and received a series of single-word responses. With a neuro-typical client group, one would be inclined to take this as an indication that the therapist and client were not working together collaboratively on

therapeutic tasks. With an ASD client group, however, possible literal interpretation of questions should be taken into account, and thus if the question asked could be answered with yes or no response they may give this and not be instinctively aware that more information is required until they are asked directly.

The ‘alternative overt behaviours’ specified on item 2 of the CBT scale of the PCTPRS (‘Did the therapist work with the client to plan, or to practice overt behaviours for the client to utilise outside of therapy?’) also prompted discussion. Although in supportive counselling alternative behaviours may be suggested/discussed, we found the crucial word in this question to be ‘overt’; during CBT, the discussion of alternative actions forms a major basis of the intervention and thus will be much more explicitly discussed than in supportive counselling. Similarly, Item 4 of the CBT scale ‘did the therapist provide a rationale which emphasised the importance of evaluating the accuracy of the client’s beliefs and changing inaccurate beliefs in order to alleviate the client’s anxiety?’ needed clarification. Again, supportive counselling may indeed include some discussion around a client’s beliefs, however identifying inaccurate beliefs and working to shift these is one of the main components of CBT, therefore it will play a prominent part in therapy. We therefore found that in this item the word ‘rationale’ was the most relevant, as in CBT the therapist will take a considerable amount of time explaining the theory of CBT and how our beliefs can impact on thoughts, feeling and behaviour. Again in item 5 ‘Did the therapist help the client to identify specific types of cognitive distortions or errors (e.g., all-or-nothing thinking, overgeneralisation) that were present in the client’s thinking?’ the discussions around cognitive distortions will be much more frequent than in supportive counselling, and the therapist is likely to explicitly name these unhelpful thinking habits. Some supportive counsellors may less-overtly discuss thinking errors, so the important word to focus on when rating this item was ‘specific’. On discussion of items 7 of both the CBT subscale and the

counselling subscale of the PCTPRS, we again agreed that in both interventions, the therapist may discuss the end of therapy and how the client might manage certain situations. The differentiation here, however, is that in CBT the therapist is much more likely to talk about these as overt, definable skills, whereas supportive counsellors may talk about this in more general terms.

In the counselling scale, item 1 generated some discussion when establishing inter-rater reliability. One particular session involved a client who had experienced a particularly stressful week, and thus we rated this scale as lower than may have been usually expected for a supportive counselling session. We therefore note that in some sessions where a client has a lot of information to relay about very challenging situations, the counsellor may take more of a supportive listening role.

One general pattern which was noted was that if during a previously viewed session there was no evidence of a particular item, warranting a score of '0', raters tended to then overcompensate by rating too highly on a latter session if there was some evidence of the item, so a general awareness of this pattern aids consistent and accurate rating.

Although definitions were clarified for the three raters as discussed above, it did not prove necessary to change questionnaire items on any of the instruments for this study.

3. Use of entire range of rating

Some authors have suggested that restricted variability in alliance scores account for past null findings i.e., a ceiling effect (Chu et al 2004). Although Lerner et al. (2011) discussed this point specifically and found no evidence of restriction in the range of scores using the TPOCS-A in their study, we were alert to this possibility, and were careful to ensure use of the entire scale during training.

4. Maintenance of training

Once all the raters have achieved inter-rater reliability, it is advisable to continue

maintenance sessions at regular intervals to check for rater ‘drift’, i.e., a tendency to move away from the agreed definitions and ratings without continued comparison. The main study is on-going, and meetings to maintain reliability are taking place.

Scores Obtained for the Video-Recordings

TPOCS-A Scores.

Table 2 provides scores obtained for the TPOCS-A for the 13 video-recordings used for interrater reliability and also provides a comparison with the original study using TPOCS-A, McLeod & Weiss (2005). It is interesting to note that, despite differences in the client groups used (McLeod & Weiss 2005 used a neuro-typical sample of 22 children aged 8 – 14 years-old, diagnosed as meeting DSM IV criteria for anxiety), the scores obtained are largely similar, suggesting that the clinicians involved in our study with young people with ASD were equally successful in establishing positive relationships with their clients. Only one difference between the studies exceeded 20%; item 5 ‘client appears uncomfortable with the therapist’. Item 5 is a reverse-scored item, hence, this suggests that client in our study felt more uncomfortable in sessions that did those in McLeod & Weiss’s (2005) study. Possible reasons for this difference may lie with the sampling of video-recordings. McLeod & Weiss (2005) sampled video-recordings across the course of therapy, whilst our recordings tended to come from the beginning of therapy, thus, it is possible that participants in our study had not yet reached the same level of ease with their clinicians.

Insert Table 2 here

Primary Care Therapy Process Rating Scale (PCTPRS) scores.

Scores for the CBT subscale of the PCTPRS are given in table 3, and scores for the counselling subscale are given in table 4. These scores are divided between the counselling sessions and the CBT sessions and for comparison purposes, scores from Godfrey et al. (2007) who used the PCTPRS in a study comparing CBT and counselling for fatigue are also

given. Statistical analysis was applied to the differences for our own sample between scores for the CBT sessions and the counselling sessions on both the CBT and counselling subscales and is given in the last column of tables 3 and 4. Due to our sample being not normally distributed, non-parametric Mann-Whitney tests were used.

Godfrey et al. (2007) report that their raters, who were blind to which sessions were CBT and counselling, were able to correctly identify the actual therapy given using this scale in all but one case (98% time). Similarly, raters in the present study were able to identify which sessions were CBT or counselling 100% of the time.

Scores obtained by our sample were similar to those of the previous study, Godfrey et al. (2007) and for both the CBT subscale and the counselling subscale, differences were in the expected directions i.e., CBT sessions scored higher on the CBT scale, and counselling items scored higher on the counselling scale. For the CBT scale, item 6 was the only item on which we failed to reach a significant difference between CBT and counselling sessions. This replicates the results of Godfrey et al's (2007) study, where again, item 6 similarly did not show significant differences between the two types of intervention. It is likely that this reflects areas of overlap between the two types of therapy.

For the counselling scale, there were fewer significant differences in the scores between the CBT sessions and the counselling sessions. This also is consistent with the results reported by Godfrey et al. (2007) who found significant differences for items 1, 4, 5, 6, but not items 2, 3 and 7. In the case of our sample (see table 4), we found significant differences for items 3, 4 and 5, but not the others. It is likely that both our results and Godfrey et al's (2007) can be explained by overlap between the therapies. Godfrey et al. (2007) propose that the two therapies have some elements unique to each and some in common, and that there are more elements unique to CBT than there are to counselling.

Insert tables 3 and 4 here

Inter-rater reliability was calculated between the three raters using Cohen's Kappa (weighted) and was $k = .75$ for TPOCS-A, $k = .77$ for the CBT scale of the PCTPRS and $k = .46$ for the counselling scale of the PCTPRS. The two most frequently cited authors giving acceptability values for Cohen's Kappa are Fliess (1981) and Landis and Koch (1977). Fleiss's (1981) guidelines characterize Kappas over 0.75 as excellent, 0.40 to 0.75 as fair to good, and below 0.40 as poor. Landis and Koch (1977) regard values of 0–0.20 as slight, 0.21–0.40 as fair, 0.41–0.60 as moderate, 0.61–0.80 as substantial, and 0.81–1 as almost perfect agreement. Therefore, inter-rater reliabilities all exceed 0.4, thus fulfilling the criteria for 'fair' reliability. However, the TPOCS-A and CBT scale of the PCTPRS achieved reliabilities considered to be 'excellent' or 'substantial' whereas the counselling scale only just achieves 'fair' levels.

Discussion

In response to calls for greater transparency of the issues encountered when applying observational measures to specialised client groups we report our findings here.

It is interesting to note that scores on both the TPOCS-A and the PCTPRS differ little between neuro-typical samples and our ASD sample. This probably reflects a broadly consistent approach taken to therapy by clinicians working within the field of mental health and also that the clinicians participating in this study were experienced with ASD clients groups and were therefore able to establish good working relationships. We suggest also that a knowledge of ASD was useful when using the rating scales. For example, in a typically developing client group, a client's reluctance to discuss their emotions with the therapist may indicate that they are somewhat uncomfortable with the therapist and don't wish to disclose the information, however with an ASD client group this may indicate that they are not able to describe their emotions rather than reluctance.

The PCTPRS was developed to assess adherence as an adjunct to trials and therefore was designed to mirror, as closely as possible, methods already used in clinical practice, such as

listening to audiotapes in supervision. Generally, this proved feasible, although training did have to be quite extensive to reach inter-rater reliability. The results revealed that the new measure could reliably distinguish between the two therapies. The PCTPRS can be used to check that clinicians are delivering treatments as intended but also to measure the extent that there are common factors across treatments. As with the original study (Godfrey et al. 2007) where the development of the PCTPRS is reported, we found commonality between the therapies in terms of content. CBT appears to possess distinctive attributes from counselling, but the two therapies have many elements in common. The TPOCS-A and CBT scale of the PCTPRS achieved substantially higher reliabilities as measured by the Kappa coefficient than the counselling scale. The nature of counselling is such that it is a broader, flexible patient-driven therapy, and this, together with the overlap with CBT, may possibly account for the somewhat less reliable recording of clinician adherence. However, the lower reliability of the counselling scale is similar to that found by the authors of the scale Godfrey et al. (2007) who also reported lower reliability for this than for the CBT scale for a sample of adults with chronic fatigue. They suggest that some of the scale items are difficult to quantify and may need further refinement.

Research into the use of CBT or other psychological therapies for anxiety in children and young people with ASD is scant and consists of preliminary pilot studies. A recent systematic review and meta-analysis of this subject by Sukhodolsky et al. (2013) identified eight randomised controlled trials, with sample sizes ranging from 22-71. These authors highlighted the form of control conditions for these studies, reporting that five studies used waitlist, two used treatment as usual and one an attentional control (that is; one that controls for clinician time and attention by also providing this in for the control group). In conclusion they stressed the limitations of studies using waitlist or treatment as usual, and called for future studies to evaluate CBT for anxiety for children with ASD against further attention

controls. The main study associated with the results reported here, when completed will thus be one of the first studies in this area to use an attention control (in this case supportive counselling). Ensuring that there is no overlap in the use of CBT techniques between the therapists using CBT and those practicing counselling is thus an important consideration for this study and evaluation of these two types of therapy will rest substantively on the ability to distinguish these. The use of the PCTPRS, if successful in this case, could support future research and larger, definitive studies in this area. Similarly, the exploration of therapeutic alliance in ASD samples is in its infancy, but again the importance of measuring this aspect, though it may be difficult for this group, is evident when one considers CBT for anxiety is a child-focused intervention that requires high levels of motivation and cooperation.

Our main study is on-going, so we cannot yet report whether we found CBT or counselling to be more efficacious for our sample of this specialised client group. However, the findings of Godfrey et al. (2007) using the PCTPRS suggested that the specific techniques associated with particular 'brand names' of therapy (such as CBT or counselling) are not necessarily those that produce the greatest changes and that it may in fact be features that they have in common that are those to which future research should attend. We have described here the processes used to obtain inter-rater reliability as a first step in supporting future research investigating the differences of a variety of psychological therapies against each other.

Conclusion

Both the TPOCS-A and the PCTPRS appears to work well with our client group of young people with ASD, a somewhat more specialised population than the ones for which they were originally designed. However, the TPOCS-A has also been shown to be effectual with an ADHD sample, so this may reflect its general flexibility and potential for use with other client groups. The PCTPRS, although previously used only with adult samples, also performed well, clearly differentiating CBT and counselling sessions but also illuminating

similarities. It is reassuring to see that similar scores to earlier studies were obtained with our sample of young people with ASD, suggesting that the inter-rater reliability work has been carried out effectively. This suggests that future researchers could use these tools both with ASD and other special groups in future. However, we would stress that the raters in this study were highly experienced, both in terms of their knowledge of ASD and in terms of rating behavioural and communicative data. We would therefore recommend that others wishing to use these instruments note particularly the issues described above that we encountered during inter-rater reliability training, and take steps to ensure all potential raters remain aware of these throughout future studies.

References

- Bull, P. 2002. *Communication under the microscope: The theory and practice of microanalysis*. Hove: Routledge.
- Chiu, A., McLeod, B.D., Har, K., and Wood, J.J. 2009. Child–therapist alliance and clinical outcomes in cognitive behavioral therapy for child anxiety disorders. *J Child Psychol Psysc.*, **50**, 751–758.
- Chu, B.C., Choudhury, M.S., Shortt, A.L., Pincus, D.B., Creed, T.A., and Kendall, P.C.2004. Alliance, technology and outcome in the treatment of anxious youth. *Cogn Behav Pract.*, **11**, 44–55.
- Fjermestad, K.W., McLeod, B.D., Heiervang, E.R., Havik, O.E., Öst, L., and Haugland, B.S.M.2012. Factor structure and validity of the Therapy Process Observational Coding System for Child Psychotherapy–Alliance Scale. *J Clin Child Psychol.*, **41**, 246-254.
- Fleiss, J.L. 1981. *Statistical methods for rates and proportions* (2nd ed.). New York: John Wiley.
- Godfrey, E., Chalder, T., Ridsdale, L., Seed, P., and Ogden, J. 2007. Investigating the ‘active ingredients’ of cognitive behaviour therapy and counselling for patients with chronic fatigue in primary care: Developing a new process measure to assess treatment fidelity and predict outcome. *Brit J Clin Psychol.*,**46**, 253-272.
- Kendall, P.C. and Ollendick, T. 2004. Setting the research and practice agenda for anxiety in children and adolescence: A topic comes of age. *Cogn Behav Pract.*, **11**, 65–74.
- Landis, J.R. and Koch, G.G. 1977. The measurement of observer agreement for categorical data. *Biometrics.*, **33**, 159–174.
- Langer, D.A., McLeod, B.D., and Weisz, J.R. 2011. Do treatment manuals undermine youth–therapist alliance in community clinical practice? *J Consult Clin Psych.*, **79**, 427–432.

- Le Couteur ,A., Lord, C., and Rutter, M. 2003. The Autism Diagnostic Interview-Revised. Los Angeles: Western Psychological Services.
- Lerner, M., Mikami, A., and McLeod, B. D. 2011. The alliance in a friendship coaching intervention for parents of children with ADHD. *Behav Ther.*, **42**, 449–461.
- Liber, J.M., McLeod, B.D., Van Widenfel, B.M., Goedhart, A.W., Van der Leeden, A.J. M. and Utens E, M,W, J. et al. 2010. Examining the relation between the therapeutic alliance, treatment adherence, and outcome of cognitive behavioral therapy for children with anxiety disorders. *Behav Ther.*, **41**, 172–186.
- Lord, C., Rutter, M., DiLavore, P.C. and Risi, S. 2002. Autism Diagnostic Observation S chedule. Los Angeles: Western Psychological Services.
- Margolin, G., Oliver, P., Gordis, E., O’Hearn, H., Medin, A., Ghosh, C., and Morland, L. 1998. The nuts and bolts of behavioral observation of marital and family interaction. *Clin Child Fam Psych Rev.*, **1**, 195–213.
- McArthur, B.A., Riosa, P.B., and Preyde, M. 2012. Review: Treatment fidelity in psychosocial intervention for children and adolescents with comorbid problems. *Child Adolesc Ment Health.*, **17**, 139–145.
- McLeod, B.D. 2005. The Therapy Process Observational Coding System for Child Psychotherapy–Alliance Scales. Richmond: Virginia Commonwealth University.
- McLeod, B.D., Southam-Gerow, M.A., and Weisz, J.R. 2009. Conceptual and methodological issues in treatment integrity measurement. *School Psychol Rev.*, **38**, 541–546.
- McLeod, B.D. and Weisz, J.R. 2005. The Therapy Process Observational Coding System–Alliance Scale: Measure characteristics and prediction of outcome in usual clinical practice. *J Consult Clin Psych.*,**73**, 323–333.

- Perepletchikova, F., Treat, T.A. and Kazdin, A.E. 2007. Treatment Integrity in Psychotherapy Research: Analysis of the Studies and Examination of the Associated Factors. *J Consult Clin Psych.*, 2007; **75**, 892–841.
- Perepletchikova, F. 2011. On the Topic of Treatment Integrity. *Clin Psychol Sci Pract* **18**, 148–153.
- Silverman, W.K. and Albano, A.M. 1996. The Anxiety Disorders Interview Schedule for DSM-IV–Child and Parent Versions. San Antonio, TX: Graywind.
- Sukhodolsky, D.G., Bloch, M.H., Panza, K.E. and Reichow, B. 2013. Cognitive-Behavioral Therapy for Anxiety in Children With High-Functioning Autism: A Meta-analysis. *Pediatrics*, **132**, e1341-e1350.
- Vermilyea, B.B., Barlow, D.H. and O'Brien, G.T. 1984. The importance of assessing treatment integrity: An example in the anxiety disorders. *J Behav Assess.*, **6**, 1–11.
- Weisz, J.R., Doss, A.J., and Hawley, K.M. 2005. Youth psychotherapy outcome research: A review and critique of the evidence base. *Ann Rev Psych.*, **56**, 337–63.
- Yeaton, W. and Sechrest, L. 1981. Critical dimensions in the choice and maintenance of successful treatments: Strength, integrity, and effectiveness. *J Consult Clin Psych.*, **49**, 156–167.