Proof Delivery Form

Please return this form with your proof

CUP reference:

Date of delivery:

Journal and Article number: BCP 339

Volume and Issue Number: 35(0)

Number of colour figures: Nil

Number of pages (not including this page): 11

There follows a proof of the article you have written for publication in **Behavioural and Cognitive Psychotherapy**. Please check the proofs carefully, make any corrections necessary and answer queries on the proofs. Queries raised by the sub-editor are listed below; the text to which the queries refer is flagged in the margins of the proof.

Please return the **corrected proof** together with the **offprint order form** as soon as possible (no later than 4 days after receipt) to:

Carole Hughes (Copyeditor)
21 Shandon Road
London
SW4 9HS
Tel: 020 8675 3719 (evenings) or 020 7403 7458 (day)
Email: randc.hughes@virgin.net **(for minor corrections this is the quickest way to respond)**

To avoid delay from overseas, please send the proof by air mail or courier.

- You are responsible for correcting your proofs! Errors not found may appear in the published journal.
- The proof is sent to you for correction of typographical errors only. Revision of the substance of the text is not permitted.
- Please answer carefully any queries raised from the sub-editor.
- A new copy of a figure must be provided if correction of anything other than a typographical error introduced by the printer is required.

Note:

If you have any queries, please telephone the Editorial Office on Tel: 020 7848 5023 Fax: 020 7848 5024

Author queries:

Typesetter queries:

Non-printed material:

Offprint order form

CAMBRIDGE UNIVERSITY PRESS

PLEASE COMPLETE AND RETURN THIS FORM. WE WILL BE UNABLE TO SEND OFFPRINTS (INCLUDING FREE OFFPRINTS) UNLESS A RETURN ADDRESS AND ARTICLE DETAILS ARE PROVIDED.

VAT REG NO. GB 823 8476 09

Behavioural and Cognitive Psychotherapy (BCP) *Volume:* ☐ *no:* ☐

Offprints

25 offprints of each article will be supplied free to each **first named author and sent to a single address**. Please complete this form and send it to **the publisher (address below)**. Please give the address to which your offprints should be sent. They will be despatched by surface mail within one month of publication. For an article by **more than one author this form is sent to you as the first named. All extra offprints should be ordered by you in consultation with your co-authors.**

Number of offprints required in addition to the 25 free copies: ☐

Email: .....

Offprints to be sent to (print in BLOCK CAPITALS): .....

.....

.....

.....

Post/Zip Code: .....

Telephone: ..... *Date (dd/mm/yy):* ..... / ..... / .....

Author(s): .....

.....

Article Title: .....

.....

All enquiries about offprints should be addressed to ***the publisher****: Journals Production Department, Cambridge University Press, The Edinburgh Building, Shaftesbury Road, Cambridge CB2 2RU, UK.*

Charges for extra offprints (excluding VAT) *Please circle the appropriate charge:*

Number of copies	25	50	100	150	200	per 50 extra
1-4 pages	£68	£109	£174	£239	£309	£68
5-8 pages	£109	£163	£239	£321	£399	£109
9-16 pages	£120	£181	£285	£381	£494	£120
17-24 pages	£131	£201	£331	£451	£599	£131
Each Additional 1-8 pages	£20	£31	£50	£70	£104	£20

Methods of payment

If you live in Belgium, France, Germany, Ireland, Italy, Portugal, Spain or Sweden and are not registered for VAT we are required to charge VAT at the rate applicable in your country of residence. If you live in any other country in the EU and are not registered for VAT you will be charged VAT at the UK rate. If registered, please quote your VAT number, or the VAT number of any agency paying on your behalf if it is registered. *VAT Number:* .....

Payment **must** be included with your order, please tick which method you are using:

☐ Cheques should be made out to Cambridge University Press.

☐ Payment by someone else. Please enclose the official order when returning this form and ensure that when the order is sent it mentions the name of the journal and the article title.

☐ Payment may be made by any credit card bearing the Interbank Symbol.

Card Number: ☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐

Expiry Date (mm/yy): ..... / ..... *Card Verification Number:* ☐☐☐☐

The card verification number is a 3 digit number printed on the **back** of your **Visa** or **Master card**, it appears after and to the right of your card number. For **American Express** the verification number is 4 digits, and printed on the **front** of your card, after and to the right of your card number.

Signature of card holder: .....

Amount (Including VAT if appropriate): £.....

Please advise if address registered with card company is different from above

CAMBRIDGE UNIVERSITY PRESS

transfer of copyright

Please read the notes overleaf and then complete, sign, and return this form to **Sue Perkins, Journals Publishing, Cambridge University Press, The Edinburgh Building, Shaftesbury Road, Cambridge, CB2 2RU, UK** as soon as possible.

Behavioural and Cognitive Psychotherapy

In consideration of the publication in **Behavioural and Cognitive Psychotherapy**

of the contribution entitled: .....

.....

by (all authors' names): .....

.....

1 To be filled in if copyright belongs to you

Transfer of copyright

I/we hereby assign to Cambridge University Press, full copyright in all formats and media in the said contribution.

I/we warrant that I am/we are the sole owner or co-owners of the material and have full power to make this agreement, and that the material does not contain any libellous matter or infringe any existing copyright.

I/we further warrant that permission has been obtained from the copyright holder for any material not in my/our copyright and the appropriate acknowledgement made to the original source. I/we attach copies of all permission correspondence.

I/we hereby assert my/our moral rights in accordance with the UK Copyrights Designs and Patents Act (1988).

Signed (tick one) ❑ the sole author(s)

❑ one author authorised to execute this transfer on behalf of all the authors of the above article

Name (block letters) .....

Institution/Company .....

Signature: ..... Date: .....

(Additional authors should provide this information on a separate sheet.)

2 To be filled in if copyright does not belong to you

a Name and address of copyright holder .....

.....

.....

.....

b The copyright holder hereby grants to Cambridge University Press the non-exclusive right to publish the contribution in the journal and to deal with requests from third parties in the manner specified in paragraphs 3 and 5 overleaf.

(Signature of copyright holder or authorised agent) .....

3 US Government exemption

I/we certify that the paper above was written in the course of employment by the United States Government so that no copyright exists.

Signature: ..... Name (Block letters): .....

4 Requests received by Cambridge University Press for permission to reprint this article should be sent to (see para. 4 overleaf)

Name and address (block letters) .....

.....

.....

Notes for contributors

1 The Journal's policy is to acquire copyright in all contributions. There are two reasons for this: *(a)* ownership of copyright by one central organisation tends to ensure maximum international protection against unauthorised use; *(b)* it also ensures that requests by third parties to reprint or reproduce a contribution, or part of it, are handled efficiently and in accordance with a general policy that is sensitive both to any relevant changes in international copyright legislation and to the general desirability of encouraging the dissemination of knowledge.

2 Two 'moral rights' were conferred on authors by the UK Copyright Act in 1988. In the UK an author's 'right of paternity', the right to be properly credited whenever the work is published (or performed or broadcast), requires that this right is asserted in writing.

3 Notwithstanding the assignment of copyright in their contribution, all contributors retain the following **non-transferable** rights:

- The right to (continue to) post a preprint of the contribution on their personal or departmental web page provided the first screen contains the statement that the paper has been accepted for publication in Behavioural and Cognitive Psychotherapy published by Cambridge University Press together with the appropriate copyright notice. On publication the full bibliographical details (volume: issue number (date), page numbers) must be inserted after the journal title.
- Subject to file availability, the right to post the contribution as published on their own or their departmental home page provided the first screen includes full bibliographical details and the appropriate copyright notice.
- The right to make hard copies of the contribution or an adapted version for their own purposes, including the right to make multiple copies for course use by their students, provided no sale is involved.
- The right to reproduce the paper or an adapted version of it in any volume of which they are editor or author. Permission will automatically be given to the publisher of such a volume, subject to normal acknowledgement.

4 We shall use our best endeavours to ensure that any direct request we receive to reproduce your contribution, or a substantial part of it, in another publication (which may be an electronic publication) is approved by you before permission is given.

5 Cambridge University Press co-operates in various licensing schemes that allow material to be photocopied within agreed restraints (e.g. the CCC in the USA and the CLA in the UK). Any proceeds received from such licenses, together with any proceeds from sales of subsidiary rights in the Journal, directly support its continuing publication.

6 It is understood that in some cases copyright will be held by the contributor's employer. If so, Cambridge University Press requires non-exclusive permission to deal with requests from third parties, on the understanding that any requests it receives from third parties will be handled in accordance with paragraphs 4 and 5 above (note that your approval and not that of your employer will be sought for the proposed use).

7 Permission to include material not in your copyright
If your contribution includes textual or illustrative material not in your copyright and not covered by fair use / fair dealing, permission must be obtained from the relevant copyright owner (usuallly the publisher or via the publisher) for the non-exclusive right to reproduce the material worldwide in all forms and media, including electronic publication. The relevant permission correspondence should be attached to this form.

If you are in doubt about whether or not permission is required, please consult the Permissions Controller, Cambridge University Press, The Edinburgh Building, Shaftesbury Road, Cambridge CB2 2RU, UK. Fax: +44 (0)1223 315052. Email: lnicol@cup.cam.ac.uk.

Please make a duplicate of this form for your own records

Behavioural and Cognitive Psychotherapy, 2006, 35, 1–11
Printed in the United Kingdom doi:10.1017/S1352465806003390

A Comparison of Two Versions of the Cognitive Therapy Scale

P. Kenneth Gordon

Hampshire Partnership Trust, Winchester, UK

Abstract. The Cognitive Therapy Scale is a well-established tool for assessing skills in delivering cognitive therapy, but has been subject to criticism. It has recently been updated by two groups, producing the Revised version or CTS-R (Blackburn, James, Milne and Reichelt, 2000) and a version designed for therapy of psychosis, the CTS-Psy (Haddock et al., 2001). The present study made a direct comparison of these scales to evaluate their inter-rater reliability, the extent to which they measure the same therapist qualities, and their utility for assessment of skills in trainee therapists working with different client-types. Twenty-six trainees submitted tapes of therapy with clients suffering either personality disorder or psychosis. Each tape was rated by two independent assessors on each of the two scales. Results suggest the scales are both fairly easy to use and produce highly similar estimates of student competence. Client diagnosis has no significant influence on the scores obtained by the therapist. However, inter-rater reliability is relatively low. It is concluded that safeguards are needed where these scales are used as a training outcome measure.

Keywords: CBT, Cognitive Therapy Scale, reliability, competence.

Introduction

The assessment of therapist performance is an important issue in a number of settings. In trials of therapy methods, the notion of experimental control requires us to prove that the therapy is being delivered as planned. Formal scales have been developed to assess this "treatment fidelity". For example, the Cognitive Therapy for Psychosis Adherence Scale (Startup, Jackson and Pearce, 2002) is designed to check adherence to a treatment manual written by Fowler, Garety and Kuipers (1995). In trials that involve a comparison between methods, the discrimination of one therapy from another becomes the crucial concern and this approach is exemplified by the Collaborative Study Psychotherapy Rating Scale (Evans, Piasecki, Kriss and Hollon, 1984), which distinguishes the characteristics of cognitive-behavioural and interpersonal therapy on the basis of observer ratings. In the fields of training and quality assurance, however, the issue becomes a little different as we are seeking a broader examination of therapist competence, which covers both general therapeutic qualities and technique-specific session activity. It was for this purpose that Beck's group originally introduced the Cognitive Therapy Scale (Young and Beck, 1980). This rapidly became the standard tool in cognitive therapy training, although it has been found to have limitations, and has been subject to several recent revisions.

Reprint requests to P. Kenneth Gordon, Psychology Services, Hampshire Partnership Trust, 59 Romsey Road, Winchester, Hants SO22 5DE, UK. E-mail: ken.gordon@hantspt-mid.nhs.uk

The Cognitive Therapy Scale (CTS) includes 11 items that are rated by an observer during a single therapy session, which may be audio or videotaped. The first group of ratings (general therapeutic skills) is made up of six items: agenda setting, feedback, understanding, interpersonal effectiveness, collaboration, and pacing/use of time. The second group of five items address specific cognitive therapy skills under the heading "conceptualization, strategy and technique". These ratings comprise: guided discovery, focus on key cognitions or behaviour, strategy for change, application of cognitive-behavioural techniques, and homework. The scale also has sections to note how the therapist dealt with problems in the session, and to make overall ratings of the difficulty of the case and the therapist competence. Over the years, there have been concerns about the reliability of the CTS, as different studies have yielded variable inter-judge reliability estimates. Details are given by Barber, Liese and Abrams (2003) and Reichelt, James and Blackburn, (2003). The former note that reliability is lower when CT experts (i.e. those using the scale) have not been trained together.

In more recent years, revised versions of the scale have been developed. The Cognitive Therapy Adherence and Competence Scale (Barber et al., 2003) was developed in the USA. This sought to improve the item selection, and also to separate the concept of adherence to CT procedures from the appropriateness and quality with which the work was carried out. The very high inter-correlations reported in the study suggested that these separate ratings may not be useful, but a final, 21-item version achieved good reliability over its four sections: cognitive therapy structure, development of a collaborative therapeutic relationship, development and application of the case conceptualization, and cognitive and behavioural techniques. Intra-class correlations ranged from .33 to .91 for individual scales, and .67 (adherence) or .73 (competence) for scale totals.

Two other adaptations of the CTS were independently researched within the UK. The Cognitive Therapy Scale-Revised (CTS-R) was introduced by Blackburn et al. (2000), primarily in response to their concerns about the reliability and conceptual basis of the original scale (Milne, Claydon, Blackburn, James and Sheikh, 2001; Blackburn et al., 2001; Reichelt et al., 2003). Their scale continues to use a 6-point rating, but this is tied to a clearer framework of skill levels, based on the Dreyfus "Levels of Competence" (Dreyfus, 1989; James, Blackburn, Milne and Reichelt, 2001b). The items themselves are also revised from the original, to better reflect the scope of current cognitive therapy (CT). They have been refined from an original group of 14 (Blackburn et al., 2000) to a 13- and 14-item version (Blackburn et al., 2001) and most recently to a final, 12-item scale (James, Blackburn, Milne and Reichelt, 2001a).

At the same time, a second adaptation of the scale was proposed by Haddock et al. (2001) because the practice of CT had broadened over recent years to cover interventions for psychosis, and the standard CTS did not seem well-suited to measuring the skills needed for this style of work. Their Cognitive Therapy Scale for Psychosis (CTS-Psy) not only altered some of the rated areas, but fundamentally changed the rating system used. Instead of a qualitative rating of skill in each domain on a 0–6 point scale, the CTS-Psy uses a check list of six micro-skills within each domain, each of which is allocated one point. Thus the marker simply rates each of these sub-items as being present, absent or appropriately omitted during the therapy session. This scoring method was anticipated to improve reliability by turning a subjective rating into a series of behavioural observations.

The two UK-based revisions of the CTS therefore differ in some of the therapist skills addressed, in the target client population for whom they are seen as suitable, and in the very different scoring system adopted. Table 1 shows the key differences between the scales. In

Table 1. Comparison of two revisions of the Cognitive Therapy Scale

Cognitive Therapy Scale – Revised (CTS-R)	Cognitive Therapy Scale – Psychosis (CTS-Psy)
1. Agenda setting and adherence	a. Agenda
2. Feedback	b. Feedback
3. Collaboration	e. Collaboration
4. Pacing and efficient use of time	–
5. Interpersonal effectiveness	c. Understanding
6. Eliciting of appropriate emotional expression	d. Interpersonal effectiveness
	–
7. Eliciting key cognitions	g. Focus on key cognitions
8. Eliciting and planning behaviours	–
	h. Choice of intervention
9. Guided discovery	f. Guided discovery
10. Conceptual integration	–
11. Application of change methods	–
	j. Quality of intervention
12. Homework setting.	i. Homework setting.
Scoring: 0–6 rating scale for each item; anchored by descriptions of each level.	Scoring: 0–6 points per item, each assigned for presence of specific behaviour (except item j, which is rated.

The items of the CTS-Psy have been re-ordered in some cases so as to correspond with their closest equivalent on the CTS-R. Initial letters indicate their original order. Blank spaces indicate items where no equivalent exists on the other scale.

terms of content, the CTS-R drops the relatively unreliable rating of "understanding", then adds two new items on "eliciting emotions", and "eliciting behaviour and plans". It also uses a rating of "conceptual integration" in place of "strategies for change". The CTS-Psy stays a little closer to the original scales, but drops "pacing and use of time" as inappropriate for clients with psychosis, and clarifies the two ratings on "strategies for change" and "application of CBT techniques" into ratings of the appropriate choice of intervention, and the quality with which it is applied.

The authors of each of these revised scales have produced data on their reliability. Blackburn et al. (2001) report inter-rater reliability across pairs of raters (4 in all), ranging from $r = 0.34$–0.79 (product-moment correlation) and 0.40 to 0.87 (intra-class correlations), all of which were significant. Haddock et al. (2001) report on the CTS-Psy using a group of four raters, whose intra-class correlation for total score was 0.94 and was highly significant.

As yet, however, no comparative data have been published. This is an important area to address for those who supervise and teach cognitive therapy skills. In choosing a measure of CT ability as a training aid or outcome evaluation, we need to be clear about the relative strength and weaknesses of a scale. It is also important to assess whether any single scale can be used to measure the range of skills needed by a therapist working with severe and complex cases. The CTS was originally designed for use in cognitive therapy for depression, and Haddock et al. (2001) argued that CBT for psychosis has a significant number of departures from that format, which affect the process as well as content of sessions. Thus the client's problems in regulation of attention and arousal, and in processing of social cues, may require

that a different approach is adopted by the therapist within the session. Judging the pace of the work, and maintaining rapport become particularly vital. The CTS-Psy attempts to allow for those differences. However, similar considerations may apply in work with other complex cases, such as borderline personality disorder, where the client's interpretation of therapist behaviour and their instability of mood and arousal may also impact on the session, and on the way the therapist proceeds. This could suggest the CTS-Psy might be useful for a more heterogeneous client group than psychosis alone. Turning to the CTS-R, as a generic scale it could seem less well tailored to work with psychosis, yet its format allows it to be adapted to the case at hand, because the rater judges the appropriateness of what they observe, not simply its presence. For example, an item such as "Elicitation of emotional expression" might be scored highly for a therapist who facilitated greater emotional expression during standard CBT, but also rated highly where the therapist working with a borderline client helped them moderate their emotion sufficiently to allow effective dialogue during the session. For these reasons, it seems important to compare the value of each of these scales across different groups of complex cases.

The aim of the current project was to make a direct comparison of the utility of the CTS-Psy and the CTS-R for assessing therapy involving clients with complex problems. In addition, we were interested in the reliability of the scales outside a research context, and when used by a larger group of experienced CT practitioners and trainers than those involved in the scale development. The following questions were investigated:

1 How reliable is each scale across pairs of markers?
2 What degree of agreement is there between the two scales when each is applied to the same therapy tapes?
3 What is the effect of client diagnosis on ratings and, specifically, does either scale tends to depress or elevate ratings of therapy where clients present with (a) personality disorder or (b) psychosis?
4 How do assessors rate their satisfaction with the two scales in terms of difficulty of use and perceived accuracy?

Procedure

The University of Southampton offers a Postgraduate Diploma/MSc in Cognitive Therapy that focuses on the needs of clients with severe mental health problems. Over a period of a year, therapist competence is assessed in CT work with simpler cases such as mood disorders, and also with complex cases that span both personality disorder and psychosis. This range of cases matches those encountered in most secondary care settings, and provides an ideal opportunity to examine the two revised versions of the CTS in a naturalistic study.

A total of 26 students agreed to participate in the research, from amongst the 31 students who formed two consecutive cohorts of the diploma intake. As part of their training, each was routinely required to conduct CT with several clients, and submit audiotapes for assessment by their supervisor. At the end of the year, one such tape is chosen by the student and rated for examination purposes by two independent assessors, and this tape was the one used for the research evaluation. These tapes are subject to informed consent (i.e. the client agrees that the tape will be heard by a supervisor, and that it may also be rated by staff from the programme as part of the therapist's training). A total of nine assessors contributed to this rating process.

Because each student elects to follow a specialist track, working in the latter part of training with clients who are diagnosed as either having personality disorder (PD) or psychosis, we were able to assess 10 tapes of clients with PD and 16 involving psychosis. Typically, clients with psychosis are diagnosed with schizophrenia and receiving therapy related to auditory hallucinations or psychotic beliefs, whilst clients with PD are usually described as borderline, dependent or avoidant, and referred with a wide range of issues including self-harm, mood dysregulation, and social and relationship problems.

Given the double-marking process, a data set of 52 ratings was generated by the 26 tapes. Unfortunately, this applied only to the CTS-R, as two of the assessors were, for practical reasons, unable to complete the CTS-Psy ratings on the six tapes they examined. Therefore, for the inter-rater comparisons, only 20 sets of paired ratings were available on the CTS-Psy.

In all cases, tapes were rated by clinicians who specialize in work with the relevant client-type, as well as being experienced cognitive therapists and supervisors. Two were employed on the programme and the others were external contributors. Tapes were never rated by staff who had supervised the student in question, which was intended to ensure, as far as possible, that the rating was carried out "blind". All were familiar with one or more versions of the cognitive therapy scale, though with a greater level of experience on the CTS(R), as this had been the standard measure used on the programme for 3 years prior to the project. All raters were provided with detailed written manuals for each scale (James et al., 2001a; Haddock et al., 2004), and there were informal opportunities to discuss the scales at meetings of supervisors and with the course staff. In addition, half-day workshops were provided by the programme on both the CTS-Psy and the CTS-R. All raters attended at least one of these workshops, with seven attending the CTS-R and five the CTS-Psy training.

For each tape submitted, the marker listened to the whole tape (of approximately one hour duration) and then assessed it on each of the two rating scales. The order in which the two scales were rated was balanced across tapes and markers. They then completed a brief feedback form related to their opinion of the process. This form required percentage ratings of the difficulty of marking the tape on each scale, and of the marker's belief about the accuracy with which the score on each scale reflected the student's skill in CT.

Results

Table 2 summarises the scores awarded to students working with each client-type, and on each of the cognitive therapy scales. After summarizing the data, the analysis below follows the order of the four research questions presented earlier, with a final assessment of the practical implication of the reliability figures obtained.

Student performance

The 52 CTS-R scores ranged from 21 to 61, with a mean of 42.16 ($SD = 8.72$). Little normative data are available for comparison, but Blackburn et al. (2001) report a mean score on the 13-item version for 11 students at the end of their training. At 38.9 (*SD* 5.9), this suggests a score of around 36 would be typical for the current, 12-item scale, and this does not seem greatly dissimilar from our result, given our larger sample and greater variance.

The scores on the CTS-Psy ranged from 27 to 58 with a mean of 46.14 ($SD = 9.58$). For comparison, Haddock et al. (2001) report a mean total score of 33.5 in their experimental

Table 2. Scores awarded on two versions of the Cognitive Therapy Scale for students submitting taped sessions concerning either psychosis or personality disorder

Client type	Psychosis					Personality disorder			
Scale	CTS-R		CTS-Psy			CTS-R		CTS-Psy	
Marker/ student	1	2	1	2	Student	1	2	1	2
1	60.0	52.5	56.0	57.0	17	52.0	44.0	53.0	44.0
2	53.0	46.0	57.0	52.0	18	41.0	40.0	46.0	40.0
3	53.0	38.0	54.0	50.0	19	49.0	39.5	50.0	30.0
4	42.5	47.0	57.0	46.0	20	50.5	47.5	46.0	51.0
5	31.0	33.0	31.0	36.0	21	61.0	51.0	58.0	54.0
6	42.0	34.0	55.0	42.0	22	32.0	36.0	34.0	32.0
7	22.5	41.0	27.0	51.0	23	40.0	53.0	52.0	56.0
8	47.0	48.0	45.0	54.0	24	32.0	42.0	31.0	46.0
9	21.0	48.0	30.0	44.0	25	40.0	56.0	41.5	56.0
10	37.0	41.5	51.0	46.0	26	33.5	34.5	28.0	46.0
11	38.5	41.0	52.0	*					
12	29.0	33.0	*	*					
13	49.0	52.0	*	57.00					
14	47.0	37.5	58.0	*					
15	38.0	42.0	44.0	*					
16	36.0	37.0	30.0	*					
Total									
M	40.41	41.97	46.21	48.64		43.10	44.35	43.95	45.50
(*SD*)	(11.03)	(6.39)	(11.75)	(6.53)		(9.73)	(7.30)	(10.09)	(9.30)

* indicates missing data.

group of "Thorn Course" trainees after training. As for the CTS-R, our trainees appear to score higher than the original normative sample, though in this case it may be explained by the fact that they are undertaking higher-level training and specifically in cognitive therapy (whereas the Thorn course is a broader introduction to psychosocial interventions).

Inter-rater reliability of each scale

Because a pool of nine assessors was used, the same two raters did not assess every tape. To make a reliability comparison between first and second marker, each rater was therefore randomly allocated to be designated "rater 1" or "rater 2" on each tape. Under these circumstances, Pearson's product moment correlation would produce a slightly different estimate of reliability for each different allocation of the raters to groups (Reichelt et al., 2003). Intra-class correlation (ICC) can, however, be used to provide a reliability estimate when assignment to pairs is arbitrary (Shrout and Fleiss, 1979; Howell, 2006), and is therefore used here. It also allows direct comparison with previous studies of the CTS that have used this method (Haddock et al., 2001; Blackburn et al., 2001). One-tailed tests are used for inter-rater comparisons, given the existing data attesting to the reliability of each scale.

The CTS-R total score shows a relatively low but significant ICC across pairs of markers ($\rho = .383$ ($df = 26$), $p < .05$, CI = .008 −.665). This is below the average ICC figure of 0.63 reported by Blackburn et al. (2001) for the four expert trainers who had developed the scale. Reichelt et al. (2003) also report product moment correlations for raters on the CTS-R before and after recent, specific training (0.44 rising to 0.67). A second analysis was therefore performed, excluding tapes where both markers had not attended the CTS-R training workshop. There were 10 remaining tapes, which had been assessed by a group of five markers. For this small sub-group, we obtained a much higher, and highly significant correlation ($\rho = 0.764$ ($df = 9$), $p < .001$, CI = .329 −.935).

The CTS-Psy total score shows a smaller intraclass correlation, which just fails to reach a significant level ($\rho = .284$ ($df = 19$), $p = .051$, CI = −.161 −.636). This is in contrast to the very high intra-class correlation of 0.94 reported by Haddock et al. (2001), using data from just four raters who had received intensive training on the scale, and who each rated five tapes that had been carefully selected to cover a range of skill level. Once again, data were therefore analysed only for assessors who had attended a training workshop on the CTS-Psy (12 tapes, heard by seven assessors) but this did not increase the ICC value obtained.

The inter-class correlations were also calculated for individual sub-scales. On the CTS-R, six out of twelve subscales produced significant ICCs (Agenda, Feedback, Pacing, Eliciting cognitions, Guided discovery, and Application of change). On the CTS-Psy, three of ten subscales had significant correlations (Feedback, Collaboration, and Guided discovery).

Agreement across scales

Because tapes were double-marked by two assessors, we used the average score obtained by each student on each scale as the data for comparison, a procedure suggested by Bland and Altman (1995). The relationship between the two scales (CTS-R and CTS-Psy) was tested with a Pearson's product-moment correlation, with significance being assessed on a two-tailed basis. There was a strong inter-scale agreement ($r = .794$, $n = 25$, $p < .001$) suggesting that the two scales are tapping broadly similar therapist skills. The correlation could be exaggerated a little by the procedure of making the two ratings immediately after one another, or if there is a global halo effect operating during rating.

Effect of diagnosis type on ratings

Students specializing in psychosis received marginally higher scores on the CTS-Psy ($M(25) = 47.28$ (SD 9.70)) than did the group working with PD cases ($M(20) = 44.72$ (9.48)). However, the psychosis group received lower scores on the CTS-R (M (32) = 41.19 (8.90) v. M (20) = 43.72 (8.39)). The direction of these differences is interesting, and could appear to suggest a slight tendency for the CTS-Psy to "favour" psychosis work styles and the CTS-R to favour work with PD clients, but a t-test showed that neither difference reaches statistically significance (t (43) = 0.887 for the CTS-Psy, and t (50) = 1.02 for the CTS-R).

Markers' opinions of the two scales

These views were reflected in the markers' percentage ratings of the difficulty of use and perceived accuracy of each scale. There was no significant difference between the two scales

regarding difficulty, with both appearing to be relatively straightforward to apply (mean percentage difficulty for CTS-Psy = 34.03; CTS-R = 31.79, t (38) = .863, *ns*). Accuracy was seen as relatively high for both scales. However, the CTS-R was seen as giving the more accurate picture of the therapist's skills (mean for CTS-Psy = 63.08; CTS-R = 72.95, t (38) = −3.92, $p < .001$).

Implications of measurement error in practice

One way to look at the implications of variability of scores across markers is to look at the discrepancy in total score assigned by each pair of raters. Thus if one marker assigns a score of 32 and the second marker only 27, the discrepancy between them is 5 points. This calculation shows there are fairly substantial inter-rater disagreements. For the CTS-R the mean discrepancy is 7.48 (*SD* 6.33) and for the CTS-Psy the mean is 9.42 (*SD* 6.47). Thirteen out of 26 pairs of markers on the CTS-R (50%) have a discrepancy of more than 5 points, and 11 out of 20 (55%) for the CTS-Psy.

These variations mean that there will be a degree of error when interpreting the test score awarded to an individual by any one assessor. This can be quantified using the standard error of measurement (SE), a statistic that assumes an individual's score on any one testing is an estimate of their "true" level of performance. Using the data reported here for the CTS-R, we obtain an SE of 6.68 and a 95% confidence interval of 13.09. Given that a score of 36 on the CTS-R is the level usually defined as "competence" in cognitive therapy, to be 95% sure that a student was less than competent after a single assessor marked one taped therapy session, they would have to obtain a score of below 23 points. From the same single assessment, 95% confidence about achieving competence would require a score of over 49. For the CTS-Psy the SEM is 8.10, but because there is no published "competence" threshold, the full calculation cannot be repeated.

Discussion

Perhaps the most striking message from these data is that therapy rating scales such as the CTS-R and CTS-Psy have a disappointing level of inter-rater reliability amongst a large staff team, even where they have experience of assessing therapy skills and a good working knowledge of the scales. There seems to be a particular problem with the CTS-Psy in this respect, as reliability levels were significant for only three sub-scales, and not for the total score. For the CTS-R, reliability was significant for total score and for half the sub-scales, but was at a level that still allows considerable variation between markers of individual tapes.

It is important to say that this is a naturalistic study, and inevitably suffers from several shortcomings. The numbers were relatively low, and only complex cases were assessed. As these cases require a flexible style of therapy, this might affect the reliability of the scales. Training of raters was, despite all efforts, incomplete and the different results for the two scales could be attributed in part to our group of markers using the CTS-Psy less skilfully than the CTS-R, as they generally had greater experience of the latter scale. Nevertheless, the study provides an indication of the problems with rating cognitive therapy in a practice setting.

The reliability levels obtained here are below those of the small group of highly trained raters described in the standardization study for the CTS-Psy (Haddock et al., 2001) or the raters with extended training on the CTS-R (Reichelt et al., 2003). In that sense, we can endorse the conclusions of Reichelt et al. about the importance of training. This conclusion is also supported by the higher correlations obtained for the CTS-R when we re-analysed data from only those raters who had attended a training workshop. At the same time, it has to be recognized that our sample of judges were experienced, had been given detailed written manuals for the scales, and in all cases had attended short workshops on one or other of the scales. It seems possible that many current supervisors around the country could be using versions of the CTS with more limited experience than that. Therefore we would tend to endorse the conclusions from a review of the original version of the CTS, where Barber et al. (2003) say research results "indicate that the CTS reliability is low when CT experts are not trained together . . . it is quite clear that, even with training, inter-judge reliability is not high."

How can this reliability problem be overcome in practice? Our current procedure in Southampton ensures every tape is rated by two judges, and their mean score adopted as the final mark. We have also used a relatively lenient pass mark for parts of the programme (30 points on the CTS-R, rather than the "competence" level of 36). Given the earlier calculations on standard error of assessment, and taking the double-marking into account, this means we have reduced the possibility of false positives (i.e. failing a student in error) to an acceptable level. However, it does open the possibility of type two error (i.e. passing a non-competent student). We intend therefore to revert to a higher pass-mark, but with more than one tape counting towards the student's assessment. On the basis of both the existing literature and the data reported here, it is recommended that assessment using versions of the CTS should incorporate thorough rater training, double-marking, and multiple sampling of student performance.

In the longer term, further research on reliability is clearly needed. There might also be benefits in linking this to research on the views of cognitive therapy trainers about what constitute core CT skills, as trainers who have widely different views about the most important ingredients of therapy, or who are trained in different "schools" of CT, might be likely to rate students differently, particularly where sub-scales are broadly defined and open to interpretation. Some of the inter-rater variation found here may indeed come from genuinely different approaches to CT amongst our raters. Ratings of competence in a broad style of work (CT) will never be as reliable as ratings of adherence to one specific protocol.

The second aim of this project was to compare the utility of the two scales. Although they have been developed for rather different purposes, and use different scoring methods, the scores assigned on the two scales are highly inter-correlated in this fairly large group of markers and session tapes. There is also no significant difference in the results they produce for sub-groups of clients, despite the CTS-Psy being designed to measure therapeutic skills for work with psychosis. However, it would be premature to suggest that the two measures are interchangeable, and further data on their use with different client groups will be needed before any definite conclusion can be reached on the value of diagnosis-specific CT scales.

The views of participants were that both scales were relatively easy to use. The CTS-R was judged to be significantly more accurate in portraying the student's skill level, possibly because its competence rating scale allows a judge to feel they are making a comprehensive

judgement of ability. In contrast, the CTS-Psy simply checks for the presence of a number of specific therapist behaviours. Whilst this format should reduce subjectivity, it may be seen by assessors as a constraint. It must also be acknowledged that the sample of markers in this study had used the CTS-R for a longer period than the CTS-Psy, and this may have coloured their opinions.

At this stage, we cannot therefore make a definitive recommendation of either scale above the other. Both offer a useful way of assessing and giving feedback on CT skills, provided that their numerical results are interpreted with a degree of caution.

Acknowledgements

Thanks are due to the supervisors and teachers on the University of Southampton Diploma/MSc in Cognitive Therapy, who completed the ratings described here and contributed to discussions on the project and its findings, and to Ian James and Gillian Haddock, who provided essential information and materials related to the two scales being investigated.

References

Barber, J. P., **Liese, B. S. and Abrams, M. J.** (2003). Development of the Cognitive Therapy Adherence and Competence Scale. *Psychotherapy Research, 13*, 205–221.

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

Blackburn, I.-M., **James, I. A.**, **Milne, D. L. and Reichelt, F. K., with Garland, A.**, **Baker, C.**, **Standart, S. H. and Claydon, A.** (2000). Cognitive Therapy Scale- Revised (CTS-R). Unpublished manuscript.

Bland, J. M. and Altman, D. G. (1995). Calculating correlation coefficients with repeated observations: Part 1 – correlation within subjects. *British Medical Journal, 310*, 446.

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

Evans, M. D., **Piasecki, J. M.**, **Kriss, M. R. and Hollon, S. D.** (1984). *Raters Manual for the Collaborative Study Psychotherapy Rating Scale, Form 6*. University of Minnesota and St. Paul – Ramsey Medical Centre.

Fowler, D., **Garety, P. and Kuipers, E.** (1995). *Cognitive Behaviour Therapy for Psychosis: theory and practice*. Chichester: John Wiley and Sons.

James, I. A., **Blackburn, I.-M.**, **Milne, D. L. and Reichelt, F. K.** (2001). Manual of the Revised Cognitive Therapy Scale. Unpublished manuscript, Newcastle Cognitive and Behavioural Therapies Centre, Newcastle, UK.

James, I. A., **Blackburn, I.-M.**, **Milne, D. L. and Reichelt, F. K.** (2001). Moderators of trainee therapists' competence in cognitive therapy. *British Journal of Clinical Psychology, 40*, 131–140.

Haddock, G., **Devane, S.**, **Bradshaw, T.**, **McGovern, J.**, **Tarrier, N.**, **Kinderman, P.**, **Baguley, I.**, **Lancashire, S. and Harris, N.** (2001). An investigation into the psychometric properties of the Cognitive Therapy Scale for Psychosis (CTS-Psy). *Behavioural and Cognitive Psychotherapy, 29*, 221–233.

Haddock, G., **Devane, S.**, **Bradshaw, T.**, **McGovern, J.**, **Tarrier, N.**, **Kinderman, P.**, **Baguley, I.**, **Lancashire, S. and Harris, N.** (2004). Cognitive Therapy Scale for Psychosis: guidelines for raters. Unpublished Manuscript, University of Manchester, Manchester, UK.

Howell, D. C. (2002). Intraclass correlation: for unordered pairs. *Retrieved 18 May, 2006, from http://www.uvm.edu/~dhowell/StatPages/More_Stuff/icc/icc.html.*

Milne, D., **Claydon, T.**, **Blackburn, I.-M.**, **James, I. and Sheikh, A.** (2001). Rationale for a new measure of competence in therapy. *Behavioural and Cognitive Psychotherapy, 29*, 21–34.

Reichelt, F. K., **James, I. A. and Blackburn, I.-M.** (2003). Impact of training on rating competence in cognitive therapy. *Journal of Behavior Therapy and Experimental Psychiatry, 34*, 87–99.

Shrout, P. E. and Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin, 2*, 420–428.

Startup, M., **Jackson, M. and Pearce, E.** (2002). Assessing therapist adherence to cognitive-behavioural therapy for psychosis. *Behavioural and Cognitive Psychotherapy, 30*, 329–339.

Young, J. and Beck, A. T. (1980). The development of the Cognitive Therapy Scale. Unpublished manuscript, Center for Cognitive Therapy, Philadelphia, PA.