

## **Measuring Adherence in CBT for Psychosis: A Psychometric Analysis of an Adherence Scale**

Rebecca Rollinson

*University of East Anglia, UK*

Ben Smith and Craig Steel

*University College London, UK*

Suzanne Jolley, Juliana Onwumere, Philippa A. Garety, Elizabeth Kuipers and  
Daniel Freeman

*Kings College London, UK*

Paul E. Bebbington

*University College London, UK*

Graham Dunn

*University of Manchester, UK*

Mike Startup

*University of Newcastle, Australia*

David Fowler

*University of East Anglia, UK*

---

Reprint requests to Rebecca Rollinson, Psychological Therapies and Primary Care Service, Community Resource Centre, Northgate Hospital, Great Yarmouth NR35 1BU, UK. E-mail: rebecca.rollinson@nwmhp.nhs.uk

© 2007 British Association for Behavioural and Cognitive Psychotherapies

**Abstract.** This paper reports the psychometric properties of a CBT for psychosis adherence scale, the Revised Cognitive Therapy for Psychosis Adherence Scale (R-CTPAS). The scale's factor structure, inter-rater reliability and concurrent validity were analysed in a sample of 67 audiotaped sessions of CBT for psychosis. The concurrent validity of the scale was examined through comparison with the Cognitive Therapy Scale (CTS, Young and Beck, 1980). Principal components analysis of the trial data suggested three factors: "engagement/assessment work", "relapse prevention work" and "formulation/schema work". Satisfactory levels of inter-rater reliability were established between rater dyads. Moderate correlations with the CTS provided an indication of concurrent validity. The R-CTPAS is concluded to be a reliable and useful instrument that can assess adherence to CBT for psychosis using the Fowler, Garety and Kuipers (1995) therapy model.

*Keywords:* Fidelity, adherence, competence, CBT, psychosis.

### Introduction

There is increasing evidence for the effectiveness of CBT for treatment-resistant positive symptoms of psychosis (Kuipers et al., 1997; Tarrier et al., 1998; Sensky et al., 2000; Gould, Mays, Goff, Mueer and Bolton, 2001; Rector and Beck, 2001; Pilling et al., 2002; Tarrier and Wykes, 2004). It is now included in UK and US national guidelines for best practice in the treatment of schizophrenia and is being disseminated across clinical services (National Institute for Clinical Excellence, 2002; American Psychiatric Association, 2004).

Manuals describing techniques of CBT for psychosis in detail have been available for over a decade. However, we lack adequate means to test the fidelity of the treatment delivered (Moncher and Prinz, 1991; Startup, Jackson and Pearce, 2002). Treatment fidelity entails both *therapist adherence*; "the extent to which a therapist used interventions and approaches prescribed by the treatment manual and avoided the use of interventions proscribed by the manual" (Waltz, Addis, Koerner and Jacobson, 1993), and *therapist competence*; "level of skill . . . [that is] . . . the extent to which the therapists took the relevant aspects of the therapeutic context into account and responded to these contextual variables appropriately" (Waltz et al., 1993). Formal tests of treatment fidelity provide the basis for research into processes of therapy, and are the starting point for attempts to understand the mechanisms of change underpinning complex interventions such as CBT for psychosis. They provide clearer differentiation between therapeutic approaches, facilitate the interpretation of varied outcomes in the same "brand" of therapy, and allow inferences about the influence of the manualized treatment upon outcome. Measures of treatment fidelity are equally important for dissemination, as they provide a clear definition of the treatment under study for programmes of training and supervision.

Randomized controlled trials of CBT have so far given priority to ensuring consistency of treatment within trials through ongoing supervision of the fidelity of the intervention. Rigorous and independent checks of fidelity are relatively recent. Kuipers et al. (1997) held peer supervision sessions between expert therapists, while Tarrier et al. (1998) had an independent rater judge whether a sample of sessions involved cognitive therapy or supportive counselling (the control condition) and Turkington and Kingdon (2000) published data to support their claim of treatment fidelity. The therapy provided in these trials was mainly carried out by the authors of the original manuals, and is therefore likely to be of good quality and consistent with the manual, but there is no way of quantifying adherence or competence. The study by Sensky et al. (2000) was more systematic. Independent raters assessed a representative sample

of therapy tapes from the control and experimental conditions with the Cognitive Therapy Rating Scale (CTRS) (Vallis, Shaw and Dobson, 1986). Competence was satisfactory, and significantly more cognitive techniques were used in the experimental condition. However, the CTRS is a general cognitive therapy competency scale, and CBT for psychosis involves specific additional considerations. For instance, clients may have little meta-cognitive awareness at the outset of therapy and may be extremely emotionally avoidant and distrustful of the therapist. Significant work may therefore be required on the therapists' part to establish a meaningful level of engagement and it may be difficult or even counter productive to attempt to adhere to a clear structure or implement specific techniques for some sessions. This type of work is not generally necessary with a help-seeking client wishing to understand their own thought processes better so would be less likely to be incorporated within a general cognitive therapy scale.

Haddock et al. (2001) adapted another general competency scale, the Cognitive Therapy Scale (CTS), to attend to many of these differences encountered in psychosis work. The adapted scale, (the CTS-Psy), alters primarily the scaling system, however, and the items retain the original focus on the use of key general cognitive skills. Both the original CTS and its adaptation, the CTS-Psy, provide important indications of the quality of general therapeutic skills, and their use, but they do not assess adherence to the specific procedures prescribed by a treatment manual.

Startup et al. (2002) developed the Cognitive Therapy for Psychosis Adherence Scale (CTPAS) as a measure of adherence in CBT for psychosis as part of the North Wales trial (Startup, Jackson and Bendix, 2004). Specific items were based on the Fowler et al. (1995) therapy manual. The scale lists 12 therapeutic activities considered to be core components of the manual. The frequency of each item within a given session is rated on a 7-point scale. It allows both observer ratings of an audio-tape of the therapy session, and therapist ratings of their own session. The manual provides definitions of each item, and illustrative examples of both adherent and non-adherent therapy. Startup et al. (2002) analysed the psychometric properties of the original 12 item CTPAS on a sample of 29 sessions of CBT. The internal consistency of the total scale was moderate, with an alpha co-efficient of 0.47. Principal components analysis indicated two main factors "focus on delusions" ( $\alpha = 0.76$ ) and "focus on problems" ( $\alpha = 0.58$ ). Inter-rater reliability was adequate (intra class correlations  $>.7$ ) for 9 of the 12 items, with less agreement on items with very low in-session frequency.

The CTPAS was used, together with the CTS, to monitor levels of therapy adherence and competence in the Tayside-Fife clinical trial of CBT for psychosis (Durham et al., 2003). An expert rated the trial therapists as highly competent on the CTS, but only 4 of the 13 available tapes involved the use of specific CBT for psychosis techniques as defined on the CTPAS. As the scale was designed to reflect the Fowler et al. (1995) model of CBT for psychosis, this finding may reflect a difference in therapy models. It does, however, indicate both the importance of being able to distinguish the different therapeutic approaches comprising "CBT for psychosis", and the need for an adherence measure that can specify the differences in these approaches. The experience of using the CTPAS in this study suggested refinements that led to a revision (the Revised Cognitive Therapy for Psychosis Adherence Scale: R-CTPAS) incorporating significant extensions for use in the Psychological Prevention of Relapse in Psychosis (PRP) trial. Significant alterations were also made to the scaling system.

This paper reports on the psychometric properties of the R-CTPAS (reproduced in the Appendix). First, an exploratory principal components analysis was carried out on the PRP

trial data. Next, inter-rater reliabilities from three independent studies are reported. Finally, we examined the concurrent validity of the scale by assessing the association between R-CTPAS scores and ratings on the CTS, which are currently regarded as standard competence ratings for cognitive therapists.

## Method

### *Participants*

The main data for this study derived from taped sessions involving 67 people who were receiving CBT for non-affective psychosis as part of the PRP randomized controlled trial of CBT and family work (ISRCTN83557988). Participants were eligible for recruitment at the time of a relapse in positive symptoms.

CBT following the treatment manual of Fowler, Garety and Kuipers (1995) was offered for approximately 20 sessions over 9 to 12 months. Therapy was provided primarily by the 5 research clinical psychologists (RCPs) employed full-time on the trial (seeing 72% of patients) and also by 10 part-time local therapists within the research centres. All local therapists were trained to a criterion level of competence on the R-CTPAS, and received ongoing supervision from the full-time trial RCP in their research centre.

At the time of analysis, 103 trial participants had been allocated to CBT. Observer R-CTPAS and CTS ratings were collected throughout the trial. In order to ensure that a representative sample of therapy sessions was collated, the 20 planned sessions were divided into four “blocks” of five sessions. Therapists were asked to submit at least one tape from each five session block for rating by a research clinical psychologist (RCP) in one of the other research centres as part of the internal supervision processes within the trial. In order to minimize any shared variance between session scores, only one tape from each client was included in the analysis. Tapes were randomly selected after stratifying for block of therapy. This process produced a sample of 67 sessions, with 37.3% ( $n = 25$ ) taken from Block One, 25.4% ( $n = 17$ ) from Block Two, 19.4% ( $n = 13$ ) from Block Three, and 18% ( $n = 12$ ) from Block Four. The trial RCP’s rating the tapes (RR, JO, SJ, BS, CS) were all doctoral clinical psychologists with an average of 3 years experience of providing CBT for psychosis under expert supervision. They had received training on the use of the R-CTPAS from the scale authors and carried out regular (bi-monthly) calibration checks with each other.

### *Inter-rater reliability samples*

Tapes of CBT for psychosis from two other studies were used to provide further information on the inter-rater reliability of the scale. There were 12 tapes from a study of CBT for psychosis being used in routine clinical settings in Boulder, Colorado (Rollinson et al., in press), and 11 tapes from the Sensky et al. (2000) randomized controlled trial. The clients in both settings were experiencing distressing positive psychotic symptoms, while the therapists were drawn from a variety of professions and were receiving high levels of supervision and training. All tapes used for the inter-rater reliability analysis were rated by RR (an RCP on the PRP trial), and one of two external raters who had been trained on the R-CTPAS. These raters were trainee clinical psychologists with limited experience of providing CBT for psychosis. Eight of the

tapes from the Sensky trial were of CBT sessions, three were of a “befriending” intervention. All raters were blind to treatment allocation.

### Measures

*Revised Cognitive Therapy for Psychosis Adherence Scale (R-CTPAS).* This 21-item scale was developed from Startup et al.’s (2002) original 12-item version. The original scale was designed to reflect four of the six major components of the Fowler et al. (1995) therapy manual; “facilitating adaptive strategies to cope with psychotic symptoms”; “developing an understanding of psychosis in collaboration with the client”; “modifying delusional beliefs and beliefs about voices”, “relapse prevention and the management of social disability” (Startup et al., 2002). The scale was revised by David Fowler and the PRP trial therapists through an extended process of rating and discussion of different therapy tapes in the course of intensive training on the model of therapy used in the PRP trial. Five items were added to the scale. These items allowed the remaining two components of the Fowler et al. (1995) therapy manual (“Engagement” and “Strategies to assess dysfunctional assumptions” - termed “Schema work” in the revised scale) to be included. The relapse prevention item, “Keeping well”, was expanded to four items to reflect the focus of the PRP trial. The scoring sheet showing the individual items is reproduced in the appendix.

The scale can be completed as a self-report measure by the therapist immediately after the session, or as an observer rated measure, on the basis of an audiotape and rough transcription of a complete therapy session.

The scaling system used by Startup et al. (2002) was also revised to include a measure of therapist competence. In particular, it was revised to allow the identification of therapeutic activity that met general cognitive therapy competence criteria but was insufficiently matched to clients presenting with psychosis.

In the revised scale, each item is first rated for frequency within the session on a 1 to 7 scale. The rater then decides if the therapy activity reflected in the item is carried out in such a way that it meets the minimal competence criteria specified within the scale manual. If it does not, the same frequency scaling and anchor points are used, but with negative values assigned ranging from  $-1$  to  $-7$ . The competency criteria are defined in accordance with the Fowler et al. (1995) therapy manual and may therefore differ between different models of CBT for psychosis. The three main criteria used to define “competent” practice are that each therapeutic activity should be carried out in a manner that is clearly *matched* to the individual client’s presentation, *individualized* to their particular difficulties, and carried out *collaboratively*. These judgements are made on the basis of how the client and therapist present in that particular session, and the activities that actually occur, rather than a consideration of what the rater thinks the therapist *should* be doing at that point in therapy. The scale manual provides the rater with an outline of expected therapy activity at and above the minimal competence criteria for each item, as well as the type of therapy activity that would score below this criterion. As the scale incorporates measures of both competence and adherence, it is technically a “fidelity” scale, but in order to clarify the origins of the scale in Startup et al.’s (2002) version, the original name was retained.

The anchor points were also altered in the revised version of the scale. In the original Startup et al. (2002) scale, a score of 2 indicated that an item was present within the session. This was altered in the R-CTPAS so that a score of “1” indicated that the item was “sufficiently present

to be considered a dose". Subsequent anchor points were scores of 3 ("Quite frequent: one significant, or a few brief occurrences"), 5 ("Very frequent: characterized the session") and 7 ("Extremely frequent: present throughout the whole session").

*Cognitive Therapy Scale (Young and Beck, 1980).* The CTS is a widely used measure of general competency in cognitive behavioural therapy. The 11-item version of this scale was used. The first 6 items comprise the subscale "General Therapeutic Skills". This covers the use of structure within the session and non-specific relationship aspects such as "feedback" and "interpersonal effectiveness". The remaining 5 items comprise the subscale "Conceptualization, Strategy and Technique". These items refer to the use of general CBT principles such as "guided discovery" and "application of CBT techniques". Items are rated on a 6-point scale where 0 = Poor and 6 = Excellent. More detailed guidelines are provided for individual items. A score of two and above is generally regarded as a cut-off for competent cognitive therapy.

Vallis et al. (1986) examined the scale's psychometric properties, and reported strong internal reliability for the scale as a whole (Alpha coefficient 0.95) and moderate inter-rater reliability of .59 for the total score with individual items ranging from .27 to .59. The CTS is a pure competence measure and therefore does not provide a direct comparison with the R-CTPAS (which is a hybrid adherence/competence scale). It was selected for comparison with the R-CTPAS in the absence of other such hybrid measures at the time, and in order to explore whether the minimal levels of competent cognitive therapy ascribed within the R-CTPAS were comparable with those reflected in the CTS.

### *Statistical analyses*

Analyses were conducted using SPSS for Windows (version 10.0) (SPSS, 2000). Significance results are quoted as two-tailed probabilities. We first report descriptive statistics for the ratings on the PRP trial, followed by the principal components analysis (varimax rotation). Intra-class correlations (two-way mixed effects model, consistency definition) were used to examine inter-rater reliability between three rater dyads. The average measure correlation is reported to allow direct comparison with Startup et al. (2002). Finally, a non-parametric correlation analysis (Spearman's Rho, two tailed) is used to compare observer ratings on the R-CTPAS and the CTS.

## **Results**

### *PRP patient characteristics*

Most of the 67 participants were male ( $n=47$ ). The mean age was 40 years ( $SD=10.6$ ), and they were classed as: White ( $n=45$ ), Black African ( $n=6$ ), Black Caribbean ( $n=6$ ), Black other (1), Indian ( $n=3$ ), Other ( $n=6$ ). The diagnoses included schizophrenia ( $n=59$ ), schizoaffective disorder ( $n=5$ ) and delusional disorder ( $n=3$ ). The mean length of illness was 12.5 years ( $SD=10.6$ ).

### *Descriptives*

The mean, standard deviation and the percentage of sessions in which each item was endorsed are given for all R-CTPAS items in Table 1. Items are listed in descending order of mean score. There were no "non-competent" ratings (between  $-1$  and  $-7$ ) in the 67 sessions indicating a

**Table 1.** Descriptive statistics for the PRP sample ( $n = 67$ )

R-CTPAS item	Mean	Standard Deviation	% endorsed
Assessing psychotic experiences	1.25	1.42	36
Engagement	1.10	1.60	28
Developing a narrative perspective	1.10	1.80	25
Developing a model of psychosis	1.00	1.40	24
Enhancing self-regulatory strategies	1.00	1.52	21
Voices and other hallucinations	1.00	1.70	17
Recognizing problems	0.72	1.31	21
Resolving ambivalence	0.60	1.41	13
Schema work	0.60	1.22	16
Delusions	0.55	1.27	13
Developing a model of relapse	0.54	1.10	17
Depression work	0.43	1.24	9
Relapse prevention interventions	0.40	1.00	12
Normalizing	0.34	1.00	10
Assessment of previous relapse	0.34	1.00	12
Validity testing	0.30	1.00	7
Anxiety work	0.25	0.80	7
Columbo style	0.20	0.60	8
Verbal challenge of delusions	0.18	0.72	5
Evidence for delusional beliefs	0.10	0.40	4
Relapse cognitions	0.10	0.32	4
Total scale mean	0.56	1.03	

good level of therapy competence in the PRP trial. The mean number of items rated as present in any given session was 4.6, leaving, on average, 16 items rated as zero in each session. Given that the mean R-CTPAS score per session was 11.7 ( $SD = 5.1$ ) with a frequency score of seven indicating “present throughout the whole session”, it would appear that the PRP CBT sessions are characterized by a combination of several R-CTPAS items, rather than a single activity in isolation.

The high frequency of zero ratings may account for the low mean scores and the highly skewed distribution of the data set (skew 2.6,  $SD = 1.06$ ; kurtosis 7.66,  $SD = 7.0$ ). The low mean and standard deviation of many items (e.g. “evidence for delusional beliefs”, “validity testing”, and “relapse cognitions”) indicates low frequency and variance within the sample. Subsequent analyses of these items must therefore be treated with caution.

### *Principal components analysis*

The Cronbach’s alpha reliability co-efficient for the whole scale was  $-.14$ , indicating a very low level of internal reliability across the scale as a whole. This may suggest that the items are operating quite independently of each other rather than as a unitary construct. In order to explore this further, a principal components analysis (varimax rotation) was carried out. The items “delusions” and “voices and other hallucinations” were removed prior to the analysis, as they indicated only the topic of conversation and confounded with many other technical items that focus on these symptoms. Also, the items “anxiety work” and “depression work” were

excluded as these related specifically to CBT for non-psychotic phenomena. The stability of the results was assessed through the use of an informal, internal cross-validation or “jack-knife” procedure (Mosteller and Tukey, 1977; Krzanowski, 1987). The analysis was carried out on the entire sample ( $n = 67$ ) and repeated on nine systematically identified sub samples in order to check the stability of any factor structure identified. The sub samples were identified by assigning each participant a number from one to nine. Sub sample one then contained all participants other than those numbered “one”, sub sample two contained all participants other than those numbered “two”, and so on. Four sub samples contained 59 participants, five contained 60.

A scree plot on the total sample indicated three factors. Factor one (labelled, “engagement and assessment work”) was present in 9 of the 10 analyses, accounted for 13% of the total scale variance and had an internal reliability (Cronbach’s alpha) of .55. It comprised the items “assessing psychotic experiences”, “engagement”, “columbo style” and “evidence for delusional beliefs”. Factor two (“relapse prevention”) was present in 7 of the 10 analyses, accounted for 12% of the total scale variance and had an internal reliability of .5. It comprised the items “relapse prevention interventions”, “enhancing self-regulatory strategies” and “formulating a personal model of relapse”. Factor three (“formulation and schema work”) was present in 6 of the 10 analyses, accounted for 11.74% of the total scale variance and had an internal reliability of .35. It comprised the items “relapse cognitions”, “schema work” and “developing a model of psychosis”. All items loaded at .4 or above (range .4 to .81). All three factors had eigenvalues above 1 in the original extraction, and together accounted for 36.74% of the scale variance.

#### *Inter-rater reliability*

Tables 2 and 3 report the mean and standard deviation of each item in the PRP trial inter-rater reliability sample (Table 2) and the Boulder and Sensky trial samples (Table 3). The intra-class correlation coefficients between the principal rater (RR) and the second rater on the PRP sample (BS), the Boulder sample and the Sensky trial sample are given for each R-CTPAS item, each of the three sub-scales, and the total scale.

Tables 2 and 3 indicate high levels of rater agreement for most items across the three samples, and all report satisfactory correlations for the three subscales identified above. Within the PRP sample, the items “developing a narrative perspective” and “relapse prevention interventions” did not show significant inter-rater reliability, while the item “engagement” fell just short of statistical significance at  $p = .056$ . The low frequency of specific relapse prevention activities in the Boulder and Sensky samples makes it hard to draw conclusions regarding inter-rater reliability, other than that both sets of raters were agreed about its absence. The apparent variation in therapy content across different settings and therapy models is an interesting but complex topic. Unfortunately, the small sample sizes obtained for the current study are insufficient to provide a valid comparison of these approaches.

#### *Concurrent validity*

In order to explore the concurrent validity of the R-CTPAS, scores were compared with CTS ratings for the same sessions. The PRP sample used in the principal components analysis was subjected to a correlation analysis comparing scores on these two scales. Of these 67 sessions,

**Table 2.** Item mean, standard deviation (*SD*) and intra-class correlation coefficient (ICC) for the PRP (*n* = 21) sample

R-CTPAS item	Mean	<i>SD</i>	ICC
Assessing psychotic experiences	1.81	0.81	.75
Engagement	0.95	0.03	.52*
Developing a narrative perspective	1.00	0.81	.40*
Developing a model of psychosis	1.12	0.64	.76
Enhancing self-regulatory strategies	1.21	0.57	.72
Schema work	0.31	0.23	.80
Voices and other hallucinations	1.17	0.17	.97
Delusions	1.21	0.37	.91
Formulating a personal model of relapse	0.33	0.20	.66
Relapse prevention interventions	0.43	0.40	.23*
Normalizing	0.07	0.10	n/p
Depression work	0.26	0.10	.93
Anxiety work	0.40	0.10	.97
Resolving ambivalence	0.31	0.17	.60
Assessment of previous relapse	0.62	0.47	.57
Columbo style	0.02	0.03	n/p
Recognizing problems	0.05	0.07	n/p
Verbal challenge of delusions	0.24	0.07	.90
Relapse cognitions	0.00	0.00	n/p
Validity testing	0.17	0.10	.83
Evidence for delusional beliefs	0.14	0.13	.56
Factor One – Engagement and assessment	0.68	1.35	.78
Factor Two – Relapse prevention	0.60	1.13	.66
Factor Three – Formulation and schema work	0.48	1.03	.80
Scale total	0.53	1.19	.80

\*ICC not significantly greater than zero. For all other ICC's  $p < .05$ .

n/p: Agreement between raters that the item was not present.

56 had observer CTS ratings available. The PRP therapists scored a mean CTS score of 33.20 ( $SD = 8$ ) overall and received significantly higher ratings on the General Therapeutic Skills subscale than on the Conceptualization, Strategy and Technique subscale (respective means 4 ( $SD = 1$ ) and 2.27 ( $SD = 1.04$ );  $t(55) = 10.83$  (two tailed)). A Spearman's rho reported moderate correlations between the R-CTPAS total score and the General Therapeutic Skills subscale ( $r = .5$ ,  $p < .001$ ), the Conceptualization, Strategy and Technique subscale ( $r = .36$ ,  $p < .001$ ), and the CTS total score ( $r = .5$ ,  $p < .001$ ).

## Discussion

### Summary of findings

Three factors were identified that remained relatively stable across the nine sub-samples tested. These appeared to distinguish between engagement and assessment work, relapse prevention work and formulation and schema level work. The internal reliability of the scale overall was

**Table 3.** Item mean, standard deviation (*SD*) and intra-class correlation coefficient (*ICC*) for the Boulder and Sensky trial inter-rater reliability samples

R-CTPAS item	Boulder <i>n</i> = 12			Sensky <i>n</i> = 11		
	Mean	<i>SD</i>	<i>ICC</i>	Mean	<i>SD</i>	<i>ICC</i>
Assessing psychotic experiences	2.42	1.44	.92	0.81	1.47	.96
Engagement	1.00	1.28	.95	0.81	1.16	.80
Developing a narrative perspective	0.58	1.16	.83	0.40	1.06	.98
Developing a model of psychosis	0.50	0.80	.54*	0.40	1.20	.93
Enhancing self-regulatory strategies	1.08	1.00	.85	0.00	0.00	n/p
Schema work	0.25	0.62	.91	0.00	0.00	n/p
Voices and other hallucinations	1.20	1.80	.95	0.27	0.90	1.00
Delusions	1.33	1.92	.93	1.18	1.61	.90
Formulating a personal model of relapse	0.00	0.00	n/p	0.00	0.00	n/p
Relapse prevention interventions	0.00	0.00	n/p	0.00	0.00	n/p
Normalizing	0.00	0.43	1.00	0.40	0.76	.90
Depression work	0.42	1.00	.98	0.00	0.00	n/p
Anxiety work	0.25	0.62	.80	0.00	0.00	n/p
Resolving ambivalence	0.17	0.57	.96	0.45	1.35	.97
Assessment of previous relapse	0.00	0.00	n/p	0.00	0.00	n/p
Columbo style	0.08	0.28	1.00	0.00	0.00	n/p
Recognizing problems	0.00	0.00	n/p	0.00	0.00	n/p
Verbal challenge of delusions	0.00	0.00	n/p	0.22	1.19	.73
Relapse cognitions	0.00	0.00	n/p	0.00	0.00	n/p
Validity testing	0.00	0.00	n/p	0.00	0.00	n/p
Evidence for delusional beliefs	0.17	0.40	.76	0.77	1.19	.61
Factor One – Engagement and assessment	1.02	1.55	.96	.54	1.13	.83
Factor Two – Relapse prevention	0.20	0.75	.87	n/p	n/p	n/p
Factor Three – Formulation and schema work	.5	1.20	.76	.27	1.39	.92
Scale total	0.45	1.14	.94	0.25	.90	.89

\**ICC* not significantly greater than zero. For all other *ICC*'s  $p < .05$ .

n/p: Agreement between raters that the item was not present.

very low, and these three factors accounted for 36.74% of the scale variance. Taken together, these findings suggest that, whilst 10 of the 21 items may be analysed in meaningful clusters, most of the individual scale items tend to reflect distinct therapeutic activities. The scale should therefore continue to be scored using all 21 items.

Startup et al. (2002) comment on the considerable difficulties inherent in establishing inter-rater reliability in a scale differentiating therapy activities *within* rather than between therapy models. Nonetheless, the satisfactory levels of inter-rater reliability they reported were replicated in this analysis.

#### *Content of therapy sessions*

The low mean item scores for the PRP sample (Table 1), together with the high average session rating of 11.7 ( $SD = 5.1$ ) and the presence of an average of 4.6 items per session,

suggests that individual PRP therapy sessions comprise relatively few therapy activities, all occurring with moderate frequency. As the 21 scale items are not an indication of what *should* be occurring within each session, the low endorsement of many scale items in each session appears reasonable. The means still appear, however, much lower than the Startup et al. (2002) sample. This may be partly attributable to scaling differences as a score of 2 indicated presence in the Startup sample, whereas presence is scored as 1 in the PRP sample.

### *Principal components analysis*

The reported factor structure appears to reflect aspects of Startup et al.'s (2002) findings. They reported a bipolar factor with insight oriented work at one pole and more active interventions (“keeping well”, “developing a model of psychosis”, “enhancing self-regulatory strategies” and “normalizing”) at the other. These two poles appear to be represented within the PRP sample as two separate factors; “engagement and assessment work” and “relapse prevention”. The Startup et al. factor “focus on delusions” is not replicated in the current study. The high loading items on this factor (“verbal challenge of delusions” and “evidence for delusional beliefs”) were very low frequency items in the PRP trial (Table 1). This difference therefore probably reflects the different emphases within the therapy provided across the two trials.

It is possible that these factor based subscales are reflecting the particular emphases of different therapy models rather than stable properties of the scale. This raises implications as to whether the scale reflects adherence exclusively to the Fowler et al. (1995) model. As the scale is designed without a cut-off score to reflect “adherent” therapy, the intention is that a judgement of adherence to a particular therapy model can be made by comparing the scoring profile with the therapy model as stated a priori. Indeed, the variation in scoring profiles observed across different therapy models and service settings (Tables 1 and 3) suggests that the R-CTPAS may be useful in distinguishing between different types of CBT for psychosis. This area requires further investigation with larger sample sizes and more varied therapy models.

### *Inter-rater reliability*

The inter-rater reliability co-efficients were moderately high across the three samples studied. In the PRP sample, 10 of the 17 items rated as present achieved intra class correlation coefficients (ICC) of .7 or above. There was 100% rater agreement on the absence of the remaining four items. In the smaller Boulder and Sensky samples respectively, 13 of 14 and 9 of 10 items rated as present achieved an ICC of .7 or above. These findings are comparable with those reported for the original CTPAS (9 of 12 items with ICC of .7 or above ranging from .43 to .89) and the CTS-Psy (6 of 10 items with ICC of .7 or above ranging from .43 to .95). As no “non-competent” therapy was identified in any of the samples, we cannot quantify the inter-rater reliability in therapy samples where the level of competence may be more varied.

The current findings do not replicate Startup et al.'s (2002) pattern of lower reliability for lower frequency items. Much of the disagreement in ratings can instead be attributed to significant overlap between items. “Enhancing self-regulatory strategies”, for instance, can be difficult to distinguish from “relapse prevention interventions” as they both refer to work on coping strategies. Similarly, “developing a narrative perspective” was a high frequency item that overlaps considerably with “developing a model of psychosis”, as the formulation work frequently entailed an autobiographical account of symptom development.

Finally, “engagement” was just short of reaching a significant level of inter-rater reliability within the PRP sample ( $p = .056$ ).

Of note is the greater inter-rater reliability between the index rater and the non-expert raters in the Boulder and Sensky samples. It is possible that non-expert raters adhere more closely to manualized guidelines rather than their own expert “sense” of what is occurring in therapy. This phenomenon requires further exploration in a larger sample with more varied levels of competence.

### *Concurrent validity*

The moderate correlations between the R-CTPAS and the CTS provide support for the concurrent validity of the R-CTPAS, but also suggest the two scales are reflecting different aspects of therapy. This is perhaps to be expected given the function of the CTS as a pure competence scale, and R-CTPAS as a hybrid competence and adherence scale.

The CTS subscale scores indicate that the CBT for psychosis carried out within the PRP trial sample was competent cognitive therapy as defined by the CTS rating manual (Young and Beck, 1980) as therapists scored an average of 2 or above on each subscale, despite very low levels of homework setting. The PRP trial therapists did score significantly higher on the “General Therapeutic Skills” subscale than the “Conceptualization, Strategy and Technique” subscale. This may reflect a greater emphasis upon assessment and flexible engagement in PRP CBT for psychosis, and the less structured use of traditional CBT techniques, such as “homework” and “validity testing” (Table 1).

### *Limitations of the findings*

We have built upon previous research by examining the psychometric properties of the R-CTPAS in a larger sample. It is still a small sample for a principal components analysis, however, and falls short of the recommended minimum sample size (Nunnally and Bernstein, 1994). Further replication should involve a larger therapy sample and employ independent raters, as the raters in the current study also provided therapy on the trial. The study is also limited by the lack of “non-competent” ratings, making it difficult to draw conclusions regarding the level of inter-rater reliability to be expected when rating therapy of a more varied level of competence.

The findings are further limited by the fact that the therapists chose tapes to be rated, rather than having these selected at random. Tape rating also formed part of an internal supervision process, so the tapes selected may have been more challenging sessions, providing a slightly distorted profile of therapy activity.

### *Potential applications of the scale*

While its primary function was to allow adherence to the Fowler et al. (1995) model to be monitored and maintained within the PRP trial, the R-CTPAS has other potential uses. The detailed description it provides of CBT for psychosis can inform the interpretation of outcome findings, allow comparison between research trials, and assist in the generalization of these findings to clinical settings. It has also shown itself to be a useful aid in the development

and evaluation of training programmes, and can facilitate reflective practice when used in supervision.

### Conclusions

The results of this analysis build upon Startup et al.'s (2002) findings. They suggest an "engagement and assessment" factor specific to PRP therapy, a "relapse prevention" factor common to both samples and an additional, less stable, factor reflecting formulation and schema based work. They replicate the finding that satisfactory inter-rater reliability can be established on the R-CTPAS and suggest that this level of reliability can be established across varied therapy settings and models, although further work is required to investigate the impact of variations in level of competence. The moderate correlation with the CTS supports the concurrent validity of the scale and suggests that the two scales require similar levels of competence from CBT therapists. The detailed description of therapeutic activity offered by the scale means that it can contribute to future clinical effectiveness trials, process research and clinical practice.

### Acknowledgements

The authors would like to thank all members of the PRP trial research group, the therapists and clients from the Boulder Mental Health Center, Colorado and the authors of the Sensky trial. The study was funded by the Wellcome Trust (program grant 062452).

### References

- American Psychiatric Association** (2004). Practice guidelines for the treatment of patients with schizophrenia. (2nd ed.). Arlington, VA: Author.
- Durham, R. C., Guthrie, M., Morton, R. V., Reid, D. A., Treliving, L., Fowler, D. and Macdonald, R.** (2003). Tayside-Fife clinical trial of cognitive-behavioural therapy for medication-resistant psychotic symptoms: results to 3-month follow up. *British Journal of Psychiatry*, 182, 303–311.
- Fowler, D., Garety, P. A. and Kuipers, E.** (1995). *Cognitive Behavioural Therapy for Psychosis: theory and practice*. Chichester: Wiley.
- Gould, R. A., Mays, V., Goff, D., Mueser, K. T. and Bolton, E.** (2001). Cognitive therapy for psychosis in schizophrenia: an effect size analysis. *Schizophrenia Research*, 48, 335–342.
- Haddock, G., Devane, S., Bradshaw, T., MCGovern, J., TARRIER, N., KINDERMAN, P., BAQULEY, 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.
- Krzanowski, W. J.** (1987). Cross-validation in principal component analysis. *Biometrics*, 43, 575–584.
- Kuipers, E. K., Garety, P. A., Fowler, D., Dunn, G., Bebbington, P., Freeman, D. and Hadley, C.** (1997). London-East Anglia randomized controlled trial of cognitive-behavioural therapy for psychosis. I: effects of the treatment phase. *British Journal of Psychiatry*, 171, 319–327.
- Moncher, F. J., and Prinz, R. J.** (1991). Treatment fidelity in outcome studies. *Clinical Psychology Review*, 11, 246–266.
- Mosteller, F. and Tukey, J. W.** (1977). *Data Analysis and Regression*. Boston, Mass: Addison Wesley.

- National Institute for Clinical Excellence (NICE)** (2002). *Schizophrenia: core interventions in the treatment and management of schizophrenia in primary and secondary care*. London: Author.
- Nunnally, J. C. and Bernstein, I. H.** (1994). *Psychometric Theory* (3rd ed.) New York: McGraw-Hill.
- Pilling, S. P., Bebbington, P., Kuipers, E., Garety, P. A., Geddes, J., Orbach, G. and Morgan, C.** (2002). Psychological treatments in schizophrenia: I. meta-analysis of family intervention and cognitive behaviour therapy. *Psychological Medicine*, 32, 763–782.
- Rector, N. A. and Beck, A. T.** (2001). Cognitive behavioural therapy for schizophrenia: an empirical review. *Journal of Nervous and Mental Disorder*, 189, 278–287.
- Rollinson, R., Haig, C., Warner, R., Garety, P. A., Kuipers, E., Bebbington, P., Freeman, D., Dunn, G. and Fowler, D.** (in press). The application of CBT for psychosis in clinical and research settings. *Psychiatric Services*.
- Sensky, T., Scott, J. L., O'Carroll, M., Scott, J., Turkington, D. and Kingdon, D.** (2000). A randomized controlled trial of cognitive-behavioural therapy for persistent symptoms in schizophrenia resistant to medication. *Archives of General Psychiatry*, 57, 165–172.
- Startup, M., Jackson, M. and Bendix, S.** (2004). North Wales randomized controlled trial of cognitive behaviour therapy for acute schizophrenia spectrum disorders: outcomes at 6 and 12 months. *Psychological Medicine*, 34, 413–422.
- Startup, M., Jackson, M. and Pearce, E.** (2002). Assessing therapist adherence to cognitive-behaviour therapy for psychosis. *Behavioural and Cognitive Psychotherapy*, 30, 329–339.
- Tarrier, N. and Wykes, T.** (2004). Is there evidence that cognitive behaviour therapy is an effective treatment for schizophrenia? A cautious or cautionary tale? *Behaviour Research and Therapy*, 42, 1377–1401.
- Tarrier, N., Yusupoff, N. L., Kinney, C., McCarthy, E., Gledhill, A., Haddock, G. and Morris, J.** (1998) Randomized controlled trial of intensive cognitive behaviour therapy for patients with chronic schizophrenia. *British Medical Journal*, 317, 303–307.
- Turkington, D. and Kingdon, D.** (2000). Cognitive-behavioural techniques for general psychiatrists in the management of patients with psychoses. *British Journal of Psychiatry*, 177, 101–106.
- Vallis, T., Shaw, B. F. and Dobson, K.** (1986). The Cognitive Therapy Scale: psychometric properties. *Journal of Consulting and Clinical Psychology*, 54, 381–385.
- Waltz, J., Addis, M. E., Koerner, K. and Jacobson, N. S.** (1993). Testing the integrity of a psychotherapy protocol: assessment of adherence and competence. *Journal of Consulting and Clinical Psychology*, 61, 620–630.
- Young, J. and Beck, A. T.** (1980). *Cognitive Therapy Scale Rating Manual*. Unpublished manuscript.

Appendix 1

**PRP** Psychological Prevention  
of Relapse in Psychosis  
Funded by The Wellcome Trust

Session no: Self/Observer  
Date:  
Therapist initial: Client initials:

Revised-Cognitive Therapy for Psychosis Adherence Scale

1.	<b>Recognizing problems</b> Did the therapist help the client to recognize, or encourage the client to acknowledge that they were experiencing problems?	
2.	<b>Engagement</b> Did the therapist pay particular attention to issues that would facilitate the engagement of the client?	
3.	<b>Columbo style</b> Did the therapist help the client to explain their reasons for holding a belief by apologizing for being confused but then carefully questioning to gain the details?	
4.	<b>Normalizing</b> Did the therapist help the client to recognize that their psychotic symptoms are similar to experiences of many people who do not have mental illness?	
5.	<b>Resolving ambivalence</b> Did the therapist help the client to resolve their ambivalence about a possible course of action?	
6.	<b>Assessing psychotic experiences</b> Did the therapist assess the antecedents, consequences, quality and impact of the client's psychotic experience?	
7.	<b>Developing a narrative perspective</b> Did the therapist help the client to construct a narrative account of his/her experiences as a meaningful sequence of events, and to develop and explore this narrative?	
8.	<b>Developing a model of psychosis</b> Did the therapist work with the client to develop an individualized, shared understanding of the nature of the client's psychotic disorder?	
9.	<b>Evidence for delusional beliefs</b> Did the therapist assess the evidence that the client uses to support his/her delusional beliefs?	
10.	<b>Verbal challenge of delusions</b> Did the therapist challenge the client's beliefs through discussion?	
11.	<b>Validity testing</b> Did the therapist encourage the client to 1) engage in specific behaviours for the purpose of testing the validity of their beliefs, OR 2) make explicit predictions about external events so that the outcomes of those events could serve as tests of those predictions OR 3) review the outcome of previous validity tests?	
12.	<b>Enhancing self-regulatory strategies</b> Did the therapist help the client to improve their self-regulatory strategies OR review the effectiveness of strategies previously discussed or practised?	
13.	<b>Schema work</b> Did the therapist assess and formulate underlying schemas and dysfunctional assumptions OR intervene on the basis of previous assessment of such schemas?	
14.	<b>Anxiety work</b> Did the therapist assess the client's anxiety experience, develop an individualized explanatory framework and/or facilitate the client's management of their anxiety?	
15.	<b>Depression work</b> Did the therapist work collaboratively with the client using cognitive techniques to address depression?	
16.	<b>Voices and other hallucinations</b> Did the therapist work with the client on issues concerning their experience of hallucinations?	
17.	<b>Delusions</b> Did the therapist work with the client on issues concerning their delusion(s)?	
18.	<b>Assessment of previous relapse</b> Did the therapist assess the antecedents, quality, impact and appraisal of the client's previous experience(s) of relapse?	
19.	<b>Formulating a personal model of relapse</b> Did the therapist work with the client to develop a shared, individualized understanding of the client's pattern of relapse?	
20.	<b>Relapse prevention interventions</b> Did the therapist work with the client on developing ways in which a relapse prevention plan could be implemented OR review the implementation of a previously developed relapse prevention plan?	
21.	<b>Relapse cognitions</b> Did the therapist assess and work with specific cognitions relevant to relapse?	

**Scoring guide**

1. Was the item considered present and meeting minimal competence criteria defined in the scale manual?

a) Not present	Rate as zero
b) Present and competent	Rate for frequency using +1 to +7 scale
c) Present but not competent	Rate for frequency using -1 to -7 scale

2. If present, how frequently did the item occur? (-1 to -7 if not competent)

1	2	3	4	5	6	7
Present		Quite frequent		Very frequent		Extremely frequent
Sufficiently present to be considered a "dose"		One significant or few brief occurrences		Characterized the session		Present throughout the whole session