



Cognitive analytic therapy: Comparing two measures of improvement

John Birtchnell^{1*}, Chess Denman² and Farouk Okhai²

¹Institute of Psychiatry, London, UK

²Addenbrooke's Hospital, Cambridge, UK

A sample group of 49 patients began a course of cognitive analytic therapy and completed two questionnaires at the end of the first session. The 32 patients who completed the course completed the same two questionnaires at the penultimate session, and the 30 patients who attended a 3-month follow-up completed the questionnaires again then. The two questionnaires were the Person's Relating to Others Questionnaire (PROQ2) and the Clinical Outcomes in Routine Evaluation (CORE). There was a high positive correlation between the total scores of these two measures and between the CORE and two, three and four of the PROQ2 scales at the start of therapy, the end of therapy and at follow-up respectively. By the end of therapy there were significant drops in scores on three of the eight PROQ2 scales and on the total score, and on the CORE. By follow-up there were significant drops on four PROQ2 scales and the CORE, but only two of these were the same as at the end of therapy. There was no obvious explanation why some patients registered a greater drop on one questionnaire than on the other.

Cognitive analytic therapy (CAT), is a pragmatic form of psychotherapy, usually of 16-sessions duration, that was developed over a 20-year period by Ryle (1979, 1985, 1990, 1995). From an early stage, elaborate training procedures were set in place to ensure a uniformity and minimum standard of practice. The aim of CAT is to enable patients to confront the ways in which they fail to revise certain manifest though unrecognized maladaptive forms of thinking and acting. It focuses upon a range of so-called self-limiting or self-harming procedures which are classified under the three headings of traps, dilemmas and snags. Traps are negative beliefs that promote a

* Correspondence should be addressed to Dr John Birtchnell, PO59, Institute of Psychiatry, De Crespigny Park, Denmark Hill, London SE24 9JL, UK.

response in others that leads only to their confirmation; dilemmas describe the vacillating between two extremes of opinion or behaviour both of which can be equally disabling; and snags are the patient's own self-imposed barriers to progress. According to CAT theory, interpersonal procedures rely upon stored knowledge of relationship structures called reciprocal roles. These specify the aims and range of activities of both parties in a relationship and are learned interpersonally. Like procedures, reciprocal roles may be deformed or maladaptive, and when they are, they may result in procedures which elicit confirmatory responses from the other in such a way as to confirm the expectations encoded in the reciprocal role template. In this way, interpersonal difficulties become 'locked in'.

Despite this clearly delineated theory, no measures have been developed for determining the presence or absence of maladaptive reciprocal roles. Consequently, the efficacy of CAT can be measured only indirectly by means of existing measures. Short accounts by Denman (1995) and Dunn, Golyukina, Ryle, and Watson (1997), both from the CAT Clinic at Guy's Hospital, London, reported significant improvements on the Beck Depression Inventory (BDI) (Beck, Ward, Mendelson, Mock, & Erbaugh, 1961), the Symptom Checklist (SCL) (Derogatis, 1977) and the Inventory of Interpersonal Problems (IIP) (Horowitz, Rozenberg, Baer, Ureno, & Villasenor (1988); but Margison (2000) observed that, 'Given the evidence of the impact of researcher allegiance, it may not be desirable to have all research on any new method of therapy restricted to its proponents' (p.146). In the present study, only the second author has any acknowledged allegiance to CAT.

Although the terminology of CAT is predominantly cognitive, in practice, most of the patient's problems can be defined in terms of deficiencies in relating to others. For this reason, the Person's Relating to Others Questionnaire (PROQ) (Birtchnell & Evans, 2004) would be an appropriate measure for the assessment of its therapeutic efficacy. The PROQ is a 96-item, computer-scored, eight-scale questionnaire, based upon Birtchnell's (1996) theory of relating. Since it cannot be assumed that all readers are aware of the scales of the PROQ2, or the theoretical model, the interpersonal octagon (Birtchnell, 1994), upon which they are based, a brief summary will now follow: The octagon is constructed around the intersecting of a vertical axis concerning upperness versus lowerness, and a horizontal one concerning closeness versus distance. Upperness is relating from above downwards, from a position of relative strength, and lowerness is relating from below upwards, from a position of relative weakness. Closeness is becoming involved with another and distance is moving into a position of separateness. These four polar positions cover a broad range of attitudes and forms of behaviour which are fully described by Birtchnell (1996, 2002). The theory proposes that everyone needs to become competent in relating in all four positions, and for each position, people who are, are said to relate positively; and people who fall short of complete competence, are said to relate negatively. The octagon is completed by inserting mid-way positions between each of the four main positions. Each octant has a two-word name, the first referring to the vertical axis, the second

referring to the horizontal one. Moving clockwise around the octagon, starting from the top, the scale names are: upper neutral, upper close, neutral close, lower close, lower neutral, lower distant, neutral distant and upper distant. For convenience, these are abbreviated to their initial letters: UN, UC, NC, LC, LN, LD and UD, and these abbreviations are adopted throughout the paper. A summary of the characteristics of the positive and negative forms of relating for each of the eight octants is provided in Figure 1. Throughout the paper, explanatory words, based upon these, will be added in parenthesis after the initials.

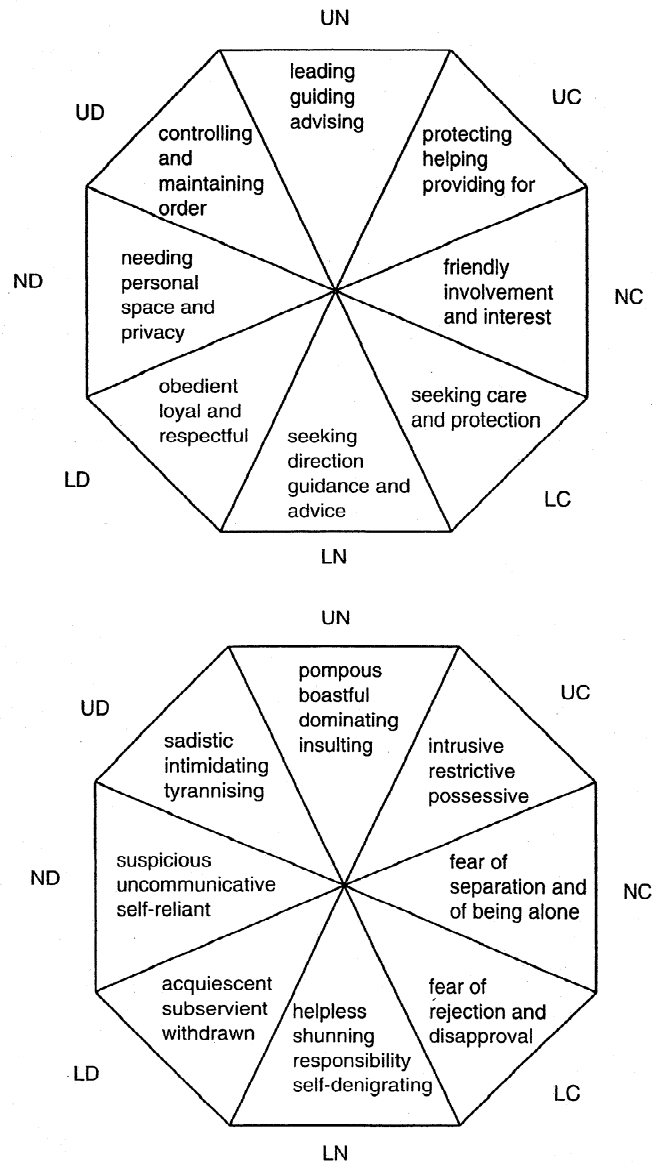


Figure 1. Positive (upper diagram) and negative (lower diagram) forms of relating. The pairs of initial letters are abbreviations for the full names of the octants given in the text. The diagrams first appeared in Birtchnell, J. The interpersonal octagon: An alternative to the interpersonal circle. *Human Relations*, 47, pp.518 and 524. Copyright The Tavistock Institute, 1994. Reproduced by permission.

The eight scales of the PROQ2 correspond to the eight octants of the interpersonal octagon; and scores are always presented in the clockwise sequence given above. Of the 96 items, only 80 refer to negative relating, i.e. 10 for each octant, and only these 80 items contribute to the eight PROQ2 scales. The 16 positive items are included only to reduce the negative quality of the questionnaire. Each item carries a score ranging from 0 to 3, depending upon the strength of agreement. Thus, for each scale, there is a maximum score of 30; and for the entire questionnaire, there is a maximum total score of 240.

The first version of the PROQ featured in a paper by Birtchnell, Falkowski, and Steffert (1992). A revised version, called the PROQ2, was introduced in 1995. Its psychometric properties are reported upon in Birtchnell and Evans (2004). For both the PROQ and the PROQ2, the α coefficients for all scales are very good. Most inter-scale correlations are low, but some, particularly those between the four polar scales and the neighbouring intermediate scales, are high. This is to be expected because the intermediate scale definitions are a blending of those of the polar scales to either side of them. A factor analysis provides strong support for five or six of the scales, depending on the sample examined. Samples of patients seeking psychotherapy have been shown to have significantly higher mean scores than samples of non-patients on the total score and six of the eight octant scores.

Two earlier, unpublished PROQ studies of CAT patients

The first version of the PROQ was administered to a series of 75 patients assessed for CAT at Guy's Hospital, London, but only 23 of these (30.7%) completed therapy and were re-tested at follow-up several months later. This disappointingly high rate of attrition was not unusual at this centre: Denman (1995) reported that 23% of patients booked-in for therapy failed to attend their first session and 33% failed to complete therapy. Additionally, Dunn *et al.* (1997) reported that around 50% did not attend follow-up. Within these considerable restrictions, there was a non-significant drop in the total mean score from 115.2 to 105.0, and drops of a low order of significance (p values both .04) were recorded on only two scales, LN and LD (helpless and acquiescent). The revised version, the PROQ2, was administered to a series of 57 patients, again assessed for CAT at Guy's Hospital, London. This time only 16 patients (28.1%) completed therapy and they were re-tested several months later. The drop in the total mean score was minimal (135.2 to 132.3) and there was no significant drop on any of the individual scales.

PROQ studies of patients receiving psychodynamic psychotherapy

In parallel with the two CAT studies described above, two studies were carried out in an NHS psychotherapy department in which the predominant form of therapy was

psychodynamic. These studies are reviewed here in order to demonstrate that the PROQ is sensitive to change in patients receiving another form of psychotherapy. This sets the CAT results in perspective. One study involved using the PROQ, the other involved using the PROQ2. The one difference between the two CAT studies and the two psychodynamic studies was that, in both the latter, the second measure was taken shortly after the end of therapy, and not at follow-up. In the first, involving 31 patients, the drop in total mean score (from 128.7 to 102.9) was highly significant and there were highly significant drops on five of the eight individual scales. In the second, involving 22 patients, the drop in mean score was even greater than in the first study (from 140.2 to 96.5) and there were highly significant drops on six of the eight individual scales. Although this might suggest that psychodynamic therapy is more effective, the effect of the difference in the timing of the second measure cannot be discounted.

In a recently published study (Birtchnell, 2002) the PROQ2 was administered to 219 patients receiving conventional, psychodynamic psychotherapy in three NHS departments, one of which was the one just described. The duration of therapy was not predetermined, and ranged from fewer than 20 sessions (29.7%) to 60 or more sessions (16.0%). Patients completed the PROQ2 shortly before the start of therapy and shortly after the end of therapy. The mean total score dropped from 132.7 to 104.2, and there were highly significant drops in the mean score on six of the eight scales. In one department, patients were tested at the time of assessment for treatment and at the start of therapy 9 months later, and there was no significant change in mean scores over this waiting period. In order to be more comparable with the two above-reported CAT studies, in the same department, 56 patients were re-tested 3-6 months after the end of therapy, and their mean scores were compared with their mean start of therapy scores. There remained a highly significant drop in the total mean score, from 123.0 to 102.5, and the drops on most of the scales were as great as at the end of therapy.

Conclusions from the previous studies

There is a striking contrast between the findings of the two published CAT studies; and the two unpublished CAT studies; the two published studies reporting significant improvement on the BDI, SCL and IIP, and the two unpublished studies reporting minimal or no improvement on the two versions of the PROQ. There is also a striking contrast between the two unpublished CAT PROQ studies and the two unpublished psychodynamic PROQ studies, with the CAT studies showing minimal or no improvement and the psychodynamic studies showing highly significant improvement. The published, larger, psychodynamic PROQ study confirmed the findings of the two, smaller, unpublished psychodynamic studies.

Why did the psychodynamic patients show greater improvement on the PROQ? First it should be stated that not all of the CAT patients did badly. In the two unpublished studies, 39% and 19%, respectively, achieved a drop in the total PROQ score of 20 or

more. The simplest explanation is that the CAT therapists were poorly trained. Another is that any improvement that occurred had become reversed by the end of the long follow-up period. A third is that while 16 sessions may be long enough to achieve symptomatic relief, it may not be long enough to correct basic relating deficits. Against this argument is the fact that, in two other 16-session therapies, the cognitive therapy of Beck, Rush, Shaw, & Emery (1979) and the interpersonal therapy of Klerman, Weissman, Rounsaville, & Chevron (1984), depressive symptoms were shown to be significantly reduced. Besides, in both the published CAT studies there was a significant improvement on the IIP, which is a measure of interpersonal difficulties.

A degree of overlap between symptom improvement and improvement in relating has to be acknowledged. The Birtchnell *et al.* (1992) study involved only the treatment of a major depressive episode with an antidepressant, and yet there were significant reductions in PROQ scores; and the more complete the recovery, the greater were the reductions. Furthermore, the psychotherapists in the Birtchnell (2002) study were simply pursuing their normal practices, and were not focusing specifically on relating difficulties.

Present study

The present study is a further attempt to determine the extent to which CAT can effect changes in PROQ2 scores. It aims to be more rigorous than those previously reported, and certainly to reduce the level of attrition. In addition to the PROQ2, the study includes the Clinical Outcomes in Routine Evaluation (CORE). This is a 34-item questionnaire designed specifically to measure change in psychotherapy, but directed primarily at more conventional clinical characteristics. Where the PROQ2 has a total score and eight scale-scores, the CORE has just a total score. Whilst the CORE items can be divided into domains (e.g. well-being, 4 items, problems 12 items, functioning, 12 items, and risk, 6 items) these are not normally scored separately and, besides, there are high (above 0.7) intercorrelations between the first three of them. The risk domain correlates less highly with the other three. Highly significant differences between the mean scores of non-clinical and clinical samples have been reported (Evans *et al.*, 2002).

The first author scored the PROQ2 and analysed the data, but is not a CAT therapist and took no part in the therapy. The second author, who was trained in CAT at Guy's Hospital, was not one of the therapists. The third author managed the project and was also one of the therapists.

Method

The therapy

The project received ethical approval from the Norfolk Mental Health Trust. It was carried out in the psychotherapy department at Addenbrooke's Hospital, Cambridge. Altogether 22 therapists were involved. All had received instruction in CAT but few had

extensive experience of using it. Each patient was to receive 16 sessions of CAT and to attend a follow-up session 3 months after the end of therapy. In fact, three patients who received 7, 11 and 24 sessions were also included.

Procedure

Patients were told that they did not have to participate in the study, and they were asked to sign a consent form. They were invited to complete both questionnaires at the end of the first therapy session, at the end of the penultimate session (returning it at the last session), and at 3-months follow-up. This strategy was adopted to maximize the take-up rate. The null hypothesis was that CAT would have no significant effect upon the PROQ2 or CORE scores between testing points.

A group of 49 patients (12 men, 37 women) were taken into therapy and completed the questionnaires at the start of therapy. Their ages ranged from 22 to 61 years (mean = 39.0, *SD* 10.1). Of these, 29 (6 men, 23 women) completed therapy, attended follow-up and completed all the questionnaires. Three completed therapy, completed the end-of-therapy questionnaires, but did not attend follow-up. Three completed therapy, did not complete the end-of-therapy questionnaires, but did the follow-up questionnaires. Three completed therapy but completed no questionnaires. Eleven (22.5%) did not complete therapy and did no questionnaires. Before being shown the test scores, the therapists were asked to provide a rating of 1–3 to indicate whether they considered the outcome of therapy to be poor, good or very good. Only one patient received no such rating.

Analyses

Statistical significance between sample mean scores at different stages was determined by a *t*-test for paired samples, and the two-tailed level of significance was adopted in all instances. The Pearson correlation coefficient was used to compare scores of the two instruments.

Patients who did better on one measure than the other

Two methods were used to identify those patients whose total PROQ2 and CORE score changes were either markedly similar or markedly different. In Method 1, the change in the PROQ2 score was plotted against that in the CORE score (Fig. 2). In Method 2, patients were listed in order of degree of improvement in total PROQ2 score and CORE score, and each patient's position on the one list was subtracted from that on the other. Of the nine patients who showed the greatest divergence (ranging from a difference of 24 places to one of 13 places), five did better on the PROQ2 and four did better on the CORE. Of the eight who showed the greatest convergence (ranging from an identical position to a difference of three places), three did very well on both, one did very badly on both and four were intermediate.

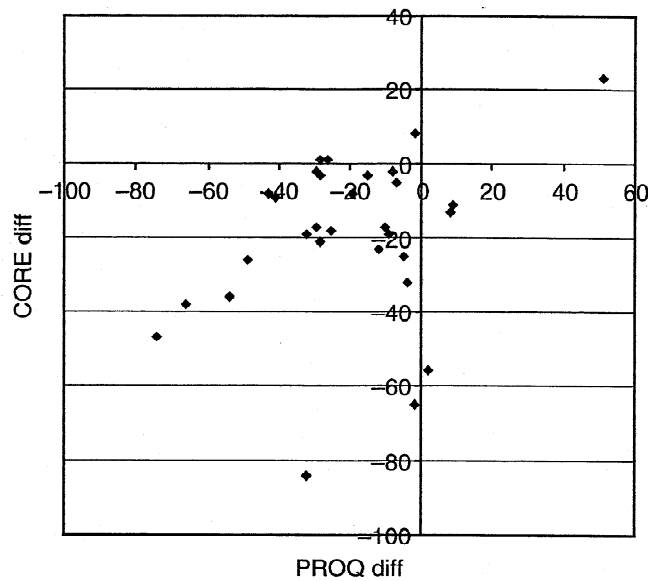


Figure 2. The drop in the total PROQ2 score (PROQ diff) plotted against the drop in total CORE score (CORE diff) from the start to the end of therapy ($N = 30$).

Results

The mean pre-therapy and post-therapy PROQ2 scores are given in Table 1. There was a significant drop in the mean total score from 122.3 ($SD\ 27.1$) to 104.4 ($SD\ 26.2$), $t = 3.92, p < .000$. There was also a significant drop on the three lower octant scales of LC, LN and LD (fear of rejection, helpless and acquiescent). The mean pre-therapy and follow-up PROQ2 scores are given in Table 2. The drop in the mean total score, from 126.7 ($SD\ 23.3$) to 98.4 ($SD\ 27.9$), $t = 5.83, p < .000$, was higher than the pre-to-post-therapy one. The significant drops on LC and LD were sustained, but that on LN was not. However, this time there were significant drops on UN and UC (dominating and possessive).

The mean pre- and post-therapy CORE scores ($N = 31$) were 53.4 ($SD\ 26.3$) and 33.6 ($SD\ 30.1$), $t = 5.02, p < .000$. The mean post-therapy and follow-up CORE scores ($N = 26$) were 32.2 ($SD\ 28.6$) and 33.3 ($SD\ 26.9$), $t = -0.24, ns$. The mean pre-therapy

Table 1. Mean PROQ2 pre-therapy and post-therapy scores, $N = 32$

	UN	UC	NC	LC	LN	LD	ND	UD	TOT
Pre	16.8	20.8	12.2	19.4	13.2	16.9	13.3	9.4	122.3
SD	5.9	6.4	7.8	6.8	7.8	7.2	9.2	5.8	27.1
Pos	16.2	18.9	10.5	14.5	10.0	12.0	12.4	10.0	104.4
SD	6.8	7.5	6.7	7.4	5.5	6.9	9.5	5.8	26.2
95%	-2.9	-4.1	-0.3	2.3	1.4	2.5	-1.2	-1.0	-27.1
CI	1.7	0.3	3.8	7.5	5.0	7.3	3.0	2.3	-8.6
r	0.56	1.75	1.75	3.83	3.71	4.15	0.84	0.79	3.92
p	.58	.90	.09	.001	.001	.000	.41	.43	.000

Table 2. Mean PROQ2 pre-therapy and follow-up scores, $N = 30$

	UN	UC	NC	LC	LN	LD	ND	UD	TOT
Pre	17.5	21.8	12.3	20.3	12.7	18.0	14.4	9.7	126.7
SD	5.6	6.0	7.5	5.8	7.1	6.1	9.4	5.8	23.3
Fol	14.6	17.8	10.6	13.7	10.4	11.7	12.0	9.5	98.4
SD	5.5	7.0	6.9	7.9	6.7	7.9	10.0	4.5	27.9
95%	-5.1	-6.1	-4.0	-9.5	-4.4	-8.9	-5.0	-1.9	-38.2
CI	-0.7	-1.9	0.8	-3.5	-0.3	-3.7	0.1	1.4	-18.4
<i>t</i>	2.71	3.86	1.40	4.44	2.35	5.00	1.94	0.29	5.83
<i>p</i>	.01	.001	.17	.000	0.26	.000	.63	.78	.000

and follow-up scores ($N = 29$) were 55.1 ($SD 27.7$) and 38.5 ($SD 30.1$), $t = 3.08$, $p < .005$).

The third author treated a quarter of the patients, but he was no more successful than the other therapists: his PROQ2 score dropped from 128.0 ($SD 20.07$) to 115.6 ($SD 20.5$). The equivalent scores for the others were 116.2 ($SD 33.6$) to 100.6 ($SD 20.5$). His CORE score dropped from 52.7 ($SD 26.1$) to 38.5 ($SD 17.9$). The equivalent scores for the others were 53.5 ($SD 27.0$) to 32.0 ($SD 29.4$).

Agreement between the PROQ2 and the CORE scores

The Pearson correlation coefficient between the two total scores was .40 at start of therapy ($N = 35$), .70 at end of therapy ($N = 31$) and .62 at follow-up ($N = 29$). There were positive correlations between CORE and the individual PROQ2 scales of LC (.53) and ND (.52) at the start of therapy, the individual scales of LC (.77), LD (.67) and ND (.77) (fear of rejection, acquiescence and suspicious) at the end of therapy, and the individual scales of LC (.74), LN (.59), LD (.62) and ND (.57) at follow-up. At the end of therapy and at follow-up there was a negative correlation between the CORE and UC (possessive) (-.33 and -.42, respectively).

Agreement between therapists' ratings and total scores

Only three patients received a therapist's rating of 3, i.e. the outcome of therapy was considered to be very good. On the PROQ2, order of degree of improvement these patients came first, second and 27. On the CORE, they came third, fourth and fifth. The patient who came 27 on the PROQ2 degree of improvement was the one with the biggest difference in outcome between the CORE (-56) and PROQ2 (+2). Equal numbers of patients had ratings of 2 and 1 (outcome good and poor) and they showed no clear relationship to measured degree of improvement. In fact, the patient who had the greatest improvement on the CORE (-84) received a rating of 1!

Patients who did better on one measure than the other

The individual scale scores of the five patients who did markedly better on the PROQ2 than on the CORE were examined. It was anticipated that these patients might have done better on those scales that did not correlate positively with the CORE; but no consistent pattern in score changes was found. The PROQ2 scale scores of the four patients who did markedly better on the CORE were also examined. Unexpectedly, all four patients had got worse on the UN and UD scales (pompous and sadistic) (mean increase of 6.8 and 7.0, respectively).

It had been anticipated that those patients who had done better on the CORE than on the PROQ2 would either have had problems that were not of an interpersonal nature, e.g. employment or financial difficulties, or would have had problems that were both expressed and treated within a symptomatic framework; but that those who had done better on the PROQ2 would have had more clearly demonstrable interpersonal difficulties and would have shown, through the course of therapy, more fundamental changes in their interpersonal functioning. In fact, there was no clear evidence of such a demarkation.

Discussion

Methodologically, the study represents an improvement over the two earlier, unpublished studies which used the PROQ and PROQ2 to measure change following CAT. In the earlier studies only 30.7 and 28.1%, respectively, of patients completed therapy and got to be re-tested at follow-up. In the present study 61.2% did. This attrition rate, though still high, is augmented by a number of factors along the way and is comparable to that recorded in a recent study using the PROQ2 to measure change following psychodynamic therapy (Birtchnell, 2002).

In terms of findings, the present study also demonstrates an improvement. Whereas, in the earlier studies, 39 and 19%, respectively, showed a drop of 20 points or more, from start to follow-up, in the present study, 70% did. Whereas, in the earlier studies, there was no significant change in mean total score from start to follow-up, in the present study there was a highly significant change; although the follow-up period was probably longer in the earlier studies. In them, the (first) author had no control over the timing of the follow-up, but in the present study, the (third) author was able to determine the timing with a fair degree of accuracy. Why then, was the present study more successful than the earlier two studies? The most likely explanation is that, although the present therapists were relatively inexperienced, they had been better trained.

Unlike the earlier studies, the present study had two time points for the assessment of change, at the end of the penultimate session and at the 3-month follow-up. Separate comparisons were made between start of therapy and end of therapy and between start of therapy and follow-up. Though slightly different numbers were involved, the drop in

mean total score was greater over the longer time period, which confirms that the improvement achieved by CAT is sustainable, at least over a 3-month period.

The Birtchnell (2002) psychodynamic psychotherapy study, had a larger N (219) and a longer therapy period (mean 36.3 sessions, SD 25.3), and the major part of the study was concerned only with a comparison of pre- and post-therapy scores, although, in a subsample of 56 patients, pre-therapy scores were compared with scores at 3-month follow-up. As in the present study, there was a highly significant drop in the mean total score from start of therapy to the end of therapy and also, in the subsample, from the start of therapy to follow-up. However, only 57% of the subsample patients showed a drop of 20 or more points, compared with 70% in the present study.

Considering now the individual scales of the PROQ2: in the pre- versus post-therapy CAT comparisons, there were significant drops on the three lower scales, while in the pre- versus post-therapy (Birtchnell, 2002) psychodynamic comparisons, there were significant drops on seven out of the eight scales. Judging by the actual means, it seems unlikely that this smaller number of scale differences was due to the smaller size of the CAT sample; so it is possible that psychodynamic therapy is effective over a broader range of scales.

It might be considered that any possible superiority of psychodynamic psychotherapy over CAT was simply due to the larger number of sessions the psychodynamic patients received. Whilst the mean duration of the psychodynamic therapy was over twice that of the CAT, 16% of psychodynamic patients received 60 or more sessions, and a smaller proportion received over 100 sessions. It may be, of course, that the psychodynamic patients received more sessions than were necessary, but one might ask whether this relatively small gain in therapeutic output could be justified on economic grounds.

In the pre-therapy versus follow-up comparisons, the CAT patients lost one significant drop (LN - helpless) but gained two more (UN and UC - pompous and possessive); so, as with the total score findings, this suggests that CAT improvement is sustainable. Here, it is difficult to make a comparison with the psychodynamic study because only a subsample of the main sample was followed up. In this subsample, there were significant drops on only four of the eight scales, compared with seven of the eight in the pre- versus post-therapy comparison of the entire sample. Whilst this result may indicate a degree of falling back in the effects of the psychodynamic therapy; it could also be explained by the smaller N of the psychodynamic subsample.

Since the PROQ2 demonstrates improvement as effectively in psychodynamic psychotherapy as it does in CAT, can it be argued that the PROQ2 has particular relevance for the assessment of patients receiving CAT? There are obvious similarities between relating theory (Birtchnell, 1996, 2002), upon which the PROQ2 is based, and CAT theory. There is also a form of psychotherapy that is called relating therapy (Birtchnell, 2002) that bears a resemblance to CAT; though, unlike CAT, it proposes a set of relating objectives, and stresses the need for each person to become competent in their attainment. Both theories and both forms of therapy focus upon the correction of

maladaptive interpersonal strategies. Both forms of therapy, just like most other forms of therapy, proceed first by highlighting these maladaptive strategies and then encouraging and enabling the patient to correct them. It is possible that sometimes a patient's end-of-therapy PROQ2 scores are increased because s/he has become more aware of these strategies, but has not yet succeeded in eliminating them; and this may give a false impression of deterioration.

What then of the CORE? Unlike the PROQ2, it is a single variable measure. As such, it provides no guide to the direction the therapy might take or indication of specific areas of improvement and non-improvement. The correlation between the PROQ2 total score and the CORE was strikingly high (.40), and became even higher (.70) by the end of therapy. The overlap between the two measures may be explained partly by the fact that 10 of the CORE's 34 items refer to relating to others. Eight of these 10 items are included in the 12-item domain that is called Functioning. It is possible that the correlation between the PROQ2 total score and the Functioning domain would have been higher, but the domains are not normally treated as scales and they do show high intercorrelations. Another explanation for the overlap between the two measures is that, when people feel generally better, they are likely to feel more positive about their relating to others. This must be the explanation for the drop in PROQ scores with improvement in depression treated by antidepressant medication, observed in the Birtchnell *et al.* (1992) study; so does the relating of patients really improve or is it simply the patients' view of their relating? The fact that patients improve on some PROQ2 scales and not on others suggests the former.

The correlations between the CORE and the individual scales of the PROQ2 are interesting: from the start of therapy, through the end of therapy to follow-up, the number of scales with which the CORE showed positive correlations increased from two to three to four. With the exception of ND, these were the (three lower) scales on which CAT produced significant improvements. The negative correlation between the CORE and the UC scale confirms a long-held view (e.g. Birtchnell & Evans, 2004) that the UC scale is the one that is the least aligned with psychopathology. In the Birtchnell and Shine study (2000) it was the only scale to show no correlation with the PDQ-IV measure of personality disorders (Hyler, 1994).

Despite the close relationship between recovery as demonstrated by the two measures, there was a small number of patients (nine) who fared better on the one measure than on the other. It was hoped that a closer inspection of these patients might provide an explanation for this difference. First, the individual PROQ2 scale changes were examined; and while no clear pattern emerged from the five who fared better on the PROQ2, it was noted that the four who fared better on the CORE had all got worse on UN and UD. Whilst no confident conclusions can be drawn from such a small number of patients, it may be that these patients experienced a sense of well-being because they had become more dominant. Next, the patients' individual clinical descriptions were examined, but there was no convincing evidence that the difference in score changes reflected different degrees of clinical improvement.

Given the relatively small sample size of the present study, the shorter duration of CAT and the inexperience of the therapists, this is a promising result, on the basis of which, further studies involving larger samples with more experienced therapists are justified. Useful gains could be made from a study in which CAT patients were compared with patients receiving psychodynamic therapy but tested after only 16 sessions, or one in which CAT patients were tested after longer periods of therapy. At present, we do not know, for either treatment mode, the minimum number of sessions required to achieve an adequate level of improvement.

A more detailed clinical study of psychotherapy patients, though not necessarily CAT patients, who did markedly better on the one questionnaire than on the other would provide a clearer understanding of the clinical changes to which each of the two measures is most sensitive. In such a study, the therapists, who were blind to the questionnaire scores, should make written observations concerning whether the patients had improved only symptomatically or at a more fundamental relating level; and such observations should be assessed by a blind rater.

Acknowledgement

This paper is derived from one read at the Joint Meeting of the UK and European Chapters of the Society for Psychotherapy Research at the University of Leiden, March, 2001.

References

- Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). *Cognitive therapy of depression*. New York: Wiley.
- Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, 4, 561-571.
- Birtchnell, J. (1994). The interpersonal octagon: An alternative to the interpersonal circle. *Human Relations*, 47, 511-529.
- Birtchnell, J. (1996). *How humans relate: A new interpersonal theory*. Hove, East Sussex, UK: Psychology Press.
- Birtchnell, J. (2002). *Relating in psychotherapy: The application of a new theory*. Hove, East Sussex, UK: Brunner-Routledge.
- Birtchnell, J., & Evans, C. (2004). The Person's Relating to Others Questionnaire (PROQ2). *Personality and Individual Differences*, 36, 125-140.
- Birtchnell, J., Falkowski, J., & Steffert, B. (1992). The negative relating of depressed patients: A new approach. *Journal of Affective Disorders*, 24, 165-176.
- Birtchnell, J., & Shine, J. (2000). Personality disorders and the interpersonal octagon. *British Journal of Medical Psychology*, 73, 433-448.
- Denman, F. (1995). Auditing CAT. In A. Ryle (Ed.), *Cognitive analytic therapy: Developments in theory and practice* (pp. 000-000). Chichester: Wiley.
- Derogatis, L. R. (1977). SCL-90-R: Administration, scoring and procedures manual-II. Towson, MD: Clinical Psychometric Research.

- Dunn, M., Golyukina, K., Ryle, A., & Watson, J. P. (1997). *Psychiatric Bulletin*, 21, 165-168.
- Evans, C., Connell, J., Barkham, M., Margison, F., McGrath, G., Mellor-Clark, J., & Audin, K. (2002). Towards a standardised brief outcome measure: Psychometric properties and utility. *British Journal of Psychiatry*, 180, 51-60.
- Horowitz, L. M., Rosenberg, S. E., Baer, B. A., Ureno, G., & Villasenor, V. S. (1988). Inventory of Interpersonal Problems: Psychometric properties and clinical applications. *Journal of Consulting and Clinical Psychology*, 56, 885-892.
- Hyer, S. E. (1994). *Personality diagnostic questionnaire, fourth edition (PDQ-4)*. New York: New York State Psychiatric Institute.
- Klerman, G. L., Weissman, M. M., Rounsaville, B. J., & Chevron, E. (1984). *Interpersonal psychotherapy of depression*. New York: Basic Books.
- Margison, F. (2000). Cognitive analytic therapy: A case study in treatment development. *British Journal of Medical Psychology*, 73, 145-150.
- Ryle, A. (1979). The focus in brief interpretive psychotherapy: Dilemmas, traps and snags as target problems. *British Journal of Psychiatry*, 134, 46-54.
- Ryle, A. (1985). Cognitive theory, object relations and the self. *British Journal of Medical Psychology*, 58, 1-7.
- Ryle, A. (1990). *Cognitive analytic therapy: Active participation in change*. Chichester: Wiley.
- Ryle, A. (1995). *Cognitive analytic therapy: Developments in theory and practice*. Chichester: Wiley.

Received 11 June 2003; revised version received 23 March 2004