Objectivity in psychoanalytic assessment of couple relationships[†]

M. LANMAN, F. GRIER and C. EVANS

Background Clinicians claim that partners in a couple can be understood to share a mode of relating, at an unconscious level. Assessment of this depends on inference from observable data. This study tests the viability and reliability of a modification of the Personal Relatedness Profile (PRP) for this purpose.

Aims To test the interrater reliability and construct validity of a joint PRP score for couples.

Method Seven therapists independently rated couples' interactions using the 30-item PRP and segments of videotaped interviews with 19 couples.

Results Interrater reliability was good and correlations between items clearly supported the underlying Kleinian bipolar model used (paranoid–schizoid/ depressive positions).

Conclusions Psychoanalytic couple psychotherapists agree in independent judgements of the nature of couple functioning, these judgements being based on envisaging couples in terms of an unconsciously shared state of mind.

Declaration of interest The study was funded by the Lord Chancellor's Department.

Psychoanalytic couple psychotherapists are concerned with aspects of couples' functioning that the couple initially may be unaware of. This form of therapy aims to facilitate change in the relationship between the partners. It focuses not simply on partners as individuals and not only on the conscious and rational level, but also on the interaction between partners that operates unconsciously, which, if not engaged with, can interfere powerfully with the possibility of lasting change. The approach considers a couple's relationship in terms of how the functioning of the two individuals can be perceived as fitting together to form one predominant joint mode of relating. This paper describes the trial of a measure that assesses this shared underlying 'fit'. Such assessment requires that the assessor is trained in perceiving unconscious processes, both in themselves and in their patients, and also is accustomed to thinking of couples as a unit in this sense.

Background

It is increasingly recognised that couple relationships make an important contribution to patients' responses to a very wide range of physical and emotional problems, and that couple-focused interventions are helpful in many of these situations (Leff *et al*, 2000). For many couple and family therapies, a fundamental axiom is that the intervention is directed at, and works through, the couple or family system, and not the individuals. There is some evidence that the nature of the change sought by therapy is important in predicting the durability of that change (Snyder *et al*, 1991). Hence it becomes important to be able to measure different kinds of change sought by different therapies in order to test for a link between type of therapy and durability of change. Such information is important also in service development and training. Researching analytically informed couple

therapies requires measures of the couple relationship that detect unconscious as well as conscious changes; measures of symptomatic improvement alone are inadequate for this purpose. However, this area tends to be neglected because recognition and evaluation of psychological functioning at an unconscious level involves assessing a complex matrix of behaviours and feelings, using inference as well as overt evidence, whether assessing individuals or couples (Milton, 1997). But, to borrow a quotation from Slade & Priebe's recent editorial, 'the challenge is to make the important measurable, not the measurable important' (Robert McNamara, former US Secretary of State, quoted in Slade & Priebe, 2001).

Although measures of individual psychological functioning abound, few are psychoanalytically based and the contribution of psychoanalytic thinking to mental health has been controversial for many reasons, including the difficulty in providing evidence of the objectivity, reliability and validity of its judgements. It might be compared with the state of diagnosis in psychiatry before the series of studies that pioneered assessment of reliability and validity, including the use of operationalised ratings and videotaped interviews (e.g. Spitzer *et al*, 1967; Wing *et al*, 1974; Wing & Nixon, 1975). There has been some progress in this regard in psychoanalytic theory and therapies for individuals, but assessment of psychoanalytic couple therapy has lagged behind.

Particular problems arise for couple therapy in that there is no perfect formula for combining the individual 'scores' for each partner, to yield a 'couple score'. To capture and evaluate changes in a couple's patterns of relatedness, a measure is needed that looks at the couple as a unit, and we believe that there are currently no measures with established reliability and validity that assess the unconscious functioning of a couple. Hence it is necessary that such a measure be developed to complement measures of individuals, to provide an empirical test of the theoretical understanding on which psychoanalytic couple therapy is based.

Psychoanalytically informed couple therapy has a strong theoretical base and a strong body of anecdotal case reports and case series but traditionally it has not drawn on or developed nomothetic measures of the theory. Psychoanalytic couple therapists think of the patterns of interaction established between two

individuals making up a couple as being rooted in shared or similar aspects of their individual psychological states of mind, either conscious or unconscious, such as expectations, anxieties and defences. These interact via unconscious processes of mutual projection (Ruszczynski, 1993). What this means is that the partners are understood to deal with certain rejected or feared aspects of themselves by assuming them to be located in the other, and act accordingly without necessarily being aware consciously of doing so. It is thought that partners tend to choose each other partly because there is some unconscious 'fit' between them: the expectations and anxieties involved are, to some extent, similar for both and each has a way of coping with these that fits in with the projections from the other (Balint, 1993).

This approach is influenced by psychoanalytic ideas about mental functioning derived from the work of Klein (1935, 1946) and widely used in contemporary psychoanalysis (see Britton, 1998: pp. 29–40). In particular, it draws on the ideas referring to two constellations of psychological functioning known as 'paranoid–schizoid' and 'depressive', and characteristic unconscious defensive structures are associated with each. Briefly, 'paranoid–schizoid' refers to a state of mind in which uncomfortable feelings tend to be denied in the self and experienced as located somewhere else, making the environment or the other person seem threatening; 'depressive' refers to a state of mind in which the self feels guilt and responsibility for damage to, or failings in, others or the environment. (It should be noted that there is some overlap between the psychoanalytic and psychiatric uses of the terms 'paranoid' and 'schizoid', but the use of 'depressive' in the two fields is different, psychiatric depression often having paranoid–schizoid rather than depressive aspects in Kleinian analytical assessment.)

Thinking about couples as tending to share a pattern of relating at an unconscious level does not imply that the two individuals necessarily appear or feel similar, or are consciously aware of what they have in common. They may appear to be like chalk and cheese and yet one may find that in the course of therapy they exchange roles at times. How rigidly or flexibly different psychological functions are distributed between the partners is regarded by couple psychotherapists

as a key factor in determining the contribution of the relationship to emotional and physical health.

A measure of this shared psychology could provide a joint couple 'score'. If reliable, it would provide support for the way in which psychoanalytic couple psychotherapists understand relationships, and be of use in the evaluation of relationship therapies.

This study aimed to develop and test an instrument for this purpose. The question was: could independent raters agree on their assessment if they were asked to rate the couple as a single unit, thinking in terms of a single state of mind or mode of psychological functioning being shared by both partners.

Hypotheses

Our hypotheses were as follows.

(a) Independent clinicians would agree in their clinical judgements about the patterns of relatedness in segments of videotaped interviews, using the couple form of the Personal Relatedness Profile (PRP; Hobson *et al*, 1998).

(b) The predominant states of mind would be bipolar, with each couple's shared state of mind being predominantly 'paranoid–schizoid' or predominantly 'depressive' in quality. Markers of the presence of one would be inversely related to the presence of the other, so that much of the variance in the ratings would form a single dimension of difference, with the items loading on that dimension as predicted. This would strongly support the construct validity of the scale.

METHOD

The PRP, a measure based on rating segments of videotaped interviews, was modified for use with couples. The original version of the PRP provided a psychoanalytically based instrument for the assessment of individuals that showed excellent interrater reliability and cross-validation against diagnostic categories. The modification involved altering the instructions to the raters, such that the raters were asked to consider the two partners in the couple 'as if they shared a single mode of psychological functioning' (see Appendix for modified instructions and some sample questions).

The raters were clinicians who were either psychoanalytic couple psychotherapists trained at the Tavistock Marital Studies Institute in London ($n=6$) or in training as such ($n=1$), and they were asked to use their clinical judgement in rating the states of mind and patterns of relating of the series of couples on the basis of the first 30 min of the couples' consultations with similarly qualified therapists. The authors were trained by Hobson and Patrick in the use of the PRP, and the raters had two and a half hours of guided practice in its use, which involved discussion and the rating of two brief extracts from videotapes of couple consultations not then used in the study.

Extracts from 19 videotaped consultations were rated. Several different therapists conducted the consultations. Out of a total of 26 available tapes, four were discarded because the sound quality was too poor and a further three on the grounds that the consultation got going so slowly that there was not enough material to rate within the first half-hour. These assessments had been conducted according to routine clinical practice in a specialist couple psychotherapy unit.

The first objective of the study was to assess interrater agreement. This was done as in the original study by the use of Kendall's coefficient of concordance W (Siegel, 1956: pp. 229–238), as calculated by SPSS version 10.07 (Norusis, 1992). Kendall's coefficient W lies in the range 0–1, where 1 indicates perfect agreement among the raters on the rank order of the videotapes. The type I error criterion α was set at 0.05. Formal statistical power was not calculated in advance (modelling power for multiple raters is complex) but the decision was made to use slightly more raters and videotapes than in Hobson *et al* (1998), to ensure that at least as much statistical power was available. Reliability was compared with that reported for rating individuals and differences were tested by assessing how many of the 30 items were rated more reliably in one study than the other, applying Wilcoxon's non-parametric test of ranked differences.

We also report overall reliability, to bring the assessment more in line with diagnostic and other ratings that are made on the summation of ratings from multiple separate indicators or items. The parameter used to provide a direct comparison with many reliability studies of multi-item measures or interrater studies is Cronbach's

coefficient α (equivalent to the mixed-effect, consistency, intraclass correlation coefficient (ICC; see Bravo & Potvin, 1991; MacLennan, 1993).

The second objective of the study was to assess whether the ratings of the couples suggested that the paranoid–schizoid and depressive positions were inversely related, such that if a couple was likely to be rated higher on the 15 PRP depressive items it would be more likely to be rated lower on the paranoid–schizoid items. As previously, this was assessed with two separate tests on the mean ratings on each item across the seven raters.

The first test involved the formulation of two composites, allocating the first seven and the last eight items for each of the paranoid–schizoid and depressive types exactly as in Hobson *et al* (1998). These were subjected to maximum likelihood exploratory factor analysis. If a very large proportion of the variance across those four composite ratings is the first factor, this indicates that the paranoid–schizoid and depressive items are opposed. Formal tests comparing the proportion of variance in the first factor in the two studies are not readily available. However, a markedly lower proportion of variance in the first factor in this study would raise questions about the relative construct validity of the PRP when used to rate individuals and couples.

The second, more-fundamental test of the paranoid–schizoid/depressive dimensionality is to look at the exploratory principal component analysis of all 30 mean ratings (after reversing the paranoid–schizoid items). Items showing negative loadings on the first component would be failing to fit into this paranoid–schizoid/depressive dimensional model. As previously, items that showed loadings below 0.3 on the first component were censored as being unlikely to represent reliable variance on that dimension. The binomial distribution was used to test the likelihood that the items would have loaded as strongly as they have by chance alone.

RESULTS

Some raters considered the therapy extracts insufficient for certain ratings. This was true for 81 of the 3990 ratings (19 extracts, 7 raters, 30 items: $19 \times 7 \times 30 = 3990$), which was 2.03% – a rate equivalent to the 1.7% reported by Hobson *et al* (1998: p. 173). The 81 unrateable items were not

restricted to a few videotapes of the couples, to a few raters or to a few items; however, item 22, referring to the experience of solitude, was omitted the most often. We report parameters after replacing missing values with the mean that the rater gave the other videotapes on that item – a method of mean substitution that will not bias the interrater agreement unless there are very marked differences in rater means and omission of the same items by most raters, neither being the case here. Recalculation of all the following results on the complete data alone produced essentially similar findings.

The Kendall concordance coefficients for this study (W) for each of the 30 items are shown in Table 1. All were statistically

Table 1 Kendall concordance coefficients from Hobson *et al* (1998; W_1) and for this study (W)

Item	W_1	W	ICC
1	0.35	0.54	0.86
2	0.31	0.62	0.86
3	0.32	0.46	0.81
4	0.29	0.44	0.83
5	0.34	0.58	0.85
6	0.35	0.56	0.83
7	0.46	0.47	0.82
8	0.26	0.45	0.81
9	0.27	0.44	0.78
10	0.38	0.43	0.77
11	0.40	0.42	0.75
12	0.21	0.37	0.68
13	0.31	0.32	0.65
14	0.30	0.30	0.66
15	0.42	0.47	0.82
16	0.33	**0.27**	0.51
17	0.40	0.62	0.89
18	0.48	0.42	0.76
19	0.24	0.54	0.87
20	0.34	0.34	0.71
21	0.26	0.44	0.75
22	0.32	0.40	0.77
23	0.34	0.61	0.87
24	0.53	**0.39**	0.77
25	0.54	**0.38**	0.81
26	0.39	**0.28**	0.59
27	0.47	**0.31**	0.57
28	0.53	0.74	0.94
29	0.60	**0.37**	0.70
30	0.29	**0.24**	0.38

Figures in bold indicate interrater agreements that were lower in this study than in Hobson *et al* (1998).

significant at $P < 0.05$ (the lowest was for item 30: $W = 0.24$, $P = 0.04$). Concordance was moderately higher than in Hobson *et al* (1998: mean = 0.44 *v*. 0.37, median = 0.44 *v*. 0.34; binomial test, $P = 0.006$; Wilcoxon $P = 0.014$). Intraclass correlation coefficients, which are based on scores, not ranks, are shown for comparison with other reliability studies and are generally very acceptable for single-item reliability on seven raters.

Inspection revealed that items 16, 24, 25, 26, 27, 29 and 30 showed lower values than in Hobson *et al* (1998). The finding that the majority of these items were 'general affect' items suggested that a *post hoc* analysis might throw more light on this because the 30 items of the PRP fall into three groups of three. The mean reliability for the first ten items in this study was 0.50 (cf. 0.33 in Hobson *et al*, 1998); for the second group of items the comparison was 0.41 *v*. 0.34; and for the last group of ten the comparison was 0.42 *v*. 0.43.

The dimensionality check showed that the first-factor eigenvalue of 3.47 accounted for 87% of the variance. This is higher than the values of 3.24 and 76%, respectively, found by Hobson *et al* (1998), indicating an even larger first dimension of variation across the couples rated.

Finally, the test of whether or not the 30 items displayed a bipolar structure in which the paranoid–schizoid items correlated negatively with the depressive items showed all but one item (item 25) loading above 0.3, in contrast to the finding of six low-loading items in Hobson *et al* (1998). The low-loading item had a negative loading of -0.25; hence 29 of the 30 items showed loading in the predicted direction. The probability of this happening by chance alone is vanishingly small ($P = 9 \times 10^{-10}$).

In light of the strong support for the first major dimension of variation, ICCs (equivalent to Cronbach's α) for the 30 items were calculated. The overall α for all 210 ratings (7 raters, 30 items) was 0.98 and for each rater it was 0.87, 0.94, 0.96, 0.90, 0.74, 0.96 and 0.87. The overall interrater reliability on the summing of the 30 items was 0.92.

DISCUSSION

Rater agreement

The first finding was that the raters reached a greater degree of agreement than was achieved in the original study.

This provides a clear and positive answer to our first hypothesis (i.e. whether or not raters can agree on rating couples) and shows that, at least for these questions and for trained raters, there are reliably observable phenomena that appear to fit the theoretical model. Not only are the reliabilities statistically significant but they are also strong overall. Only one rater showed an internal consistency below 0.8 (rater 5, ICC=0.74) and only seven of the 30 items showed interrater reliability below 0.7, a stringent criterion if applied at item level rather than at overall rating level. When items were summed to get a closer approximation to reliability checking of a composite rating (e.g. that of the Hamilton Rating Scale for Depression (Hamilton, 1967), the Beck Depression Inventory (Beck *et al*, 1961) or the multiple markers in an operationalised diagnostic system such as the DSM (American Psychiatric Association, 1994)), the reliability was excellent, at 0.92.

Construct validity

The second finding was that the data showed a clear first dimension of variation on which the paranoid–schizoid items correlated negatively with the depressive position items, with only one item not fitting the predicted pattern. This suggests that the Kleinian contrast of paranoid–schizoid and depressive may have strong construct validity as rated by the PRP. If the items showed no such empirical construct validity, the finding that 29 of the 30 loaded as expected would happen in about one in a billion such experiments.

What is being tested

It is important to be clear what is, and what is not, tested by the study. The question of whether raters are rating a 'shared state of mind' in the couples is not addressed directly by any one parameter in the analyses. Equally, whether a paranoid–schizoid or depressive unconscious state of mind is shared by the members of each couple is not tested directly either. What is tested is whether there are some shared qualities within each couple that can be rated by the majority of the raters on the majority of videotape extracts for the majority of the items (98% overall). The couples are seen to differ on these items and, if there were not some recognisable shared qualities of the couples, neither

the interrater reliability nor the strong validation of the single dimension of paranoid–schizoid/depressive would have been seen. The differences that were found reliably and that associated items as predicted with a bipolar dimension of difference between the couples do not prove the existence of Kleinian positions. Similarly, the reliable association of Schneiderian symptoms with each other, and separate from the major symptoms of anxiety, does not prove the existence of schizophrenia or anxiety as useful diagnostic categories. However, finding either unreliable ratings or no association of ratings as predicted by analytical theory would have supported rejection of either the PRP as a measure or the analytical theory of couple therapy, or both. The finding of reliability and dimensional opposition for the couple data supports the idea that trained raters can infer a 'couple mind'. Their ratings are congruent with theory. In this way these findings support construct validity.

The circularity question

A further question is whether the ratings followed from training of the raters in such a way as to make the correlations between the ratings follow from theory to rating rather than allowing rating to test theory. There are well-recognised ways in which spurious construct validity can be shown. For example, this can be seen in relation to historical stereotypes of the 'epileptic personality'. A formal rating study of the concept some decades ago might have shown apparent construct validity with all the items loading as predicted, but if epilepsy were not observable so that the other characteristic could not be rated by the halo effect there would have been no interrater reliability. Similar, prevalent American and Russian definitions of schizophrenia before the 1980s might have shown apparent reliability, and validity might have been shown, because the descriptive process was circular (Wing & Nixon, 1975).

CLINICAL IMPLICATIONS

■ Clinicians can agree when assessing a couple as a unit, which supports an approach that is claimed to have clinical utility.

■ This approach offers some access to one partner's unconscious through what is articulated by the other.

■ It is reasonable to conceptualise states of mind in couples in terms of the concepts of paranoid–schizoid and depressive positions.

LIMITATIONS

■ The level of rater agreement in assessing couples as one unit does not, in itself, prove the 'existence' of a shared couple psychology. The experience of clinical utility may be the best evidence available, but is not tested here.

■ The training in working with unconscious factors required to work at the level of inference on which this instrument operates is a seemingly unavoidable limitation.

■ Despite clear discrimination of a bipolar factor (paranoid–schizoid/depressive) this would be more strongly supported by a study where separate groups of raters rated the two factors independently.

M. LANMAN, TQAP, MSPMP, F. GRIER, MBPAS, MSPMP, Tavistock Marital Studies Institute, London; C. EVANS, MRCPsych, Tavistock and Portman NHS Trust and Rampton Hospital, London

Correspondence: M. Lanman, Tavistock Marital Studies Institute, Tavistock Centre, 120 Belsize Lane, London NW3 5BA, UK. Tel +44 (0)20 7447 3724; e-mail: monica@tmsi.org.uk

APPENDIX

Instructions for application of the Personal Relatedness Profile (PRP) to couples, together with some sample questions from the schedule (for the full, original version of the PRP, see Hobson *et al*, 1998)

Rater's name:

Couple code number:

General instructions

Please circle a score for the couple against each question. Please note carefully what the scale represents, i.e. '1' is Very uncharacteristic, and '5' is Very characteristic. We want you to consider the couple as a unit, as if they were one person. Or, to put it another way, rate them in terms of their shared state of mind, thinking in terms of splitting and mutual projection. We want you to make a *clinical judgement* in answering the questions, that is, a judgement using all the data available to you as a clinician, including what the couple say and do, which could be overtly mostly about some third party such as a child, *and also including your countertransference response* and a degree of clinical inference about the internal object relations you are observing. But this would *not* include a deeply unconscious structure that the rater could only infer theoretically. You need to have some clinical evidence for your judgement.

Section 1: Personal relatedness

On this first part of the scale, we would like you to consider the quality of what the couple experiences to happen between them, or between either of them and other people, or between other people (as reported), and to make judgements on the extent to which each of the following characterise the couple's overall functioning (considered as two sides of a whole). The quality of relatedness between couple and interviewer also should be considered in making a judgement.

Characteristic 'relatedness patterns' involve:

	Very uncharacteristic			Very characteristic		Unclassified
4. Lack of concern, use of people as things	1	2	3	4	5	U
8. A capacity for ambivalence, in which the participant(s) grapple with the complexities of relationships	1	2	3	4	5	U

Section 2: Characteristics of people ('objects')

In this second part, we would like you to consider the nature of the people that the couple feel they encounter (possibly reflecting internal objects). The characteristics may be inferred from behaviour during the interview, and from the couple's own descriptions. The picture may contain apparent contradictions (i.e. objects of very differing natures, e.g. very good *and* very bad figures). Ratings may also apply to a couple's experience of themselves as well as of others. Once more, we would like you to judge the extent to which the following characterise the couple's overall experiences of people.

The figures are experienced as:

	Very uncharacteristic			Very characteristic		Unclassified
13. Emotionally available and caring, with recognition of the needs and wishes of others	1	2	3	4	5	U
19. Betraying, untrustworthy, abandoning, deserting	1	2	3	4	5	U

Section 3: Predominant affective states

Please rate the degree to which the following characterise or underlie the couple's conscious predominant affective state. We would encourage you again to use your intuitive and clinical skills in judging what the material expresses about overall functioning, in addition to basing ratings on explicit evidence. But you should have *some* evidence (which can be your countertransference, or a clear sense that what you are seeing is a defence against something) for your judgement, other than theoretical assumption.

	Very uncharacteristic			Very characteristic		Unclassified
23. Intolerable frustration or sense of deprivation and/or extreme emotional 'hunger'	1	2	3	4	5	U
26. Feeling gratified, enriched, satisfied or nourished	1	2	3	4	5	U

However, the 30 items used here covered three distinct domains and none of them individually is specific to the framework from which psychoanalytic formulations are derived. The raters were not instructed to formulate the couples they saw in the Kleinian positional spectrum; rather, they were asked to make the ratings without formulation. Hence, vulnerability to the charge of spurious construct validity appears to be minimised here. All construct validation must be a process of survival of a long series of empirical tests, not merely of one, and none is definitive. The final test that changes the theory is rare in the human and social sciences.

REFERENCES

American Psychiatric Association (1994) *Diagnostic and Statistical Manual of Mental Disorders* (4th edn) (DSM–IV). Washington, DC: APA.

Balint, E. (1993) Unconscious communications between husband and wife. In *Before I was I: Psychoanalysis and the Imagination* (eds J. Mitchell & M. Parsons), pp. 207–218. London: Free Association Books.

Beck, A. T., Ward, C. H., Mendelson, M., et al (1961) An inventory for measuring depression. *Archives of General Psychiatry*, **4**, 561–571.

Bravo, G. & Potvin, L. (1991) Estimating the reliability of continuous measures with Cronbach's alpha or the intraclass correlation coefficient: toward the integration of two traditions. *Journal of Clinical Epidemiology*, **44**, 381–390.

Britton, R. (1998) *Belief and Imagination*. London: Routledge.

Hamilton, M. (1967) Development of a rating scale for primary depressive illness. *British Journal of Social and Clinical Psychology*, **6**, 278–296.

Hobson, P., Patrick, M. & Valentine, J. (1998) Objectivity in psychoanalytic judgements. *British Journal of Psychiatry*, **173**, 172–177.

Klein, M. (1935) A contribution to the psychogenesis of manic–depressive states. *International Journal of Psycho-Analysis*, **16**, 282–310.

___ (1946) Notes on some schizoid mechanisms. *International Journal of Psycho-Analysis*, **27**, 99–110.

Leff, J., Vearnals, S., Wolff, G., et al (2000) The London Depression Intervention Trial. Randomised controlled trial of antidepressants v. couple therapy in the treatment and maintenance of people with depression living with a partner: clinical outcome and costs. *British Journal of Psychiatry*, **177**, 95–100.

MacLennan, R. N. (1993) Interrater reliability with SPSS for Windows 5.0. *American Statistician*, **47**, 292–296.

Milton, J. (1997) Why assess? Psychoanalytic assessment in the NHS. *Psychoanalytic Psychotherapy*, **11**, 47–58.

Norusis, M. J. (1992) *SPSS for Windows. Advanced Statistics.* Chicago, IL: SPSS Inc.

Ruszczynski, S. (1993) *Psychotherapy with Couples.* London: Karnac Books.

Siegel, S. (1956) *Nonparametric Statistics for the Behavioral Sciences.* Tokyo: McGraw-Hill International.

Slade, M. & Priebe, S. (2001) Are randomised controlled trials the only gold that glitters? *British Journal of Psychiatry,* **179**, 286–287.

Snyder, K. K., Wills, R. M. & Grady-Fletcher, A. (1991) Long-term effectiveness of behavioural versus insight-orientated marital therapy: a four-year follow-up study. *Journal of Counselling and Clinical Psychology,* **59**, 138–141.

Spitzer, R. L., Cohen, J., Fliess, J. L., *et al* **(1967)** Quantification of agreement in psychiatric diagnosis, a new approach. *Archives of General Psychiatry,* **17**, 83–87.

Wing, J. & Nixon, J. (1975) Discriminating symptoms in schizophrenia – a report from the International Pilot study of Schizophrenia. *Archives of General Psychiatry,* **32**, 853–859.

__ , **Cooper, J. E. & Sartorius, N. (1974)** *Measurement and Classification of Psychiatric Symptoms. An Instruction Manual for the PSE and Catego Program.* Cambridge: Cambridge University Press.