

Cross-cultural implementation of a Chinese version of the Schedules for Clinical Assessment in Neuropsychiatry (SCAN) in Taiwan*

A. T. A. CHENG, A. Y. TIEN, C. J. CHANG, T. S. BRUGHA, J. E. COOPER, C. S. LEE, W. COMPTON, C. Y. LIU, W. Y. YU and H. M. CHEN

Background There are no published reports of cross-cultural equivalence and interrater reliability at the level of individual symptom items assessed by a semi-structured clinical interview employing operationalised clinician ratings.

Aims To assess the cross-cultural clinical equivalence and reliability of a Chinese version of the World Health Organization Schedules for Clinical Assessment in Neuropsychiatry (SCAN).

Method UK–US and Taiwanese groups of psychiatrists used Chinese and English transcripts of videotape interviews of Taiwanese patients to discuss cross-cultural issues and ratings of SCAN items. Item ratings were compared quantitatively individually and pooled by SCAN section.

Results Chinese equivalents were found for all SCAN items. No between-group differences were found for most individual items, but there were differences for some scaled items. Average agreement between the two groups was 69–100%.

Conclusions Cross-cultural implementation based on SCAN in Taiwan appears valid.

Declaration of interest This work was supported by a grant from the Taiwan National Health Research Institutes (DD01-86IX-MD-60IS).

This paper reports some of the work that is being done in preparation for a programme of epidemiological studies in Taiwan, in some of which the Schedules for Clinical Assessment in Neuropsychiatry (SCAN) (World Health Organization, 1994; Brugha, 1998) will be used. The SCAN have been developed under World Health Organization (WHO) auspices as a comprehensive instrument for the assessment and classification of psychiatric disorders in adults, incorporating the PSE-10 (Wing *et al*, 1990; Wing, 1996), which in turn has been the product of a long evaluation of the Present State Examination (PSE) (Wing *et al*, 1974). One of the strengths of SCAN is the PSE emphasis on phenomenology at the item level, a ‘bottom-up’ approach designed to provide relative independence from contemporary diagnostic constructs and associated potential assessment biases (Wing *et al*, 1998b).

The results of cross-cultural testing of an earlier version of SCAN at the level of diagnostic categories show satisfactory levels of agreement (Easton *et al*, 1997; Wing *et al*, 1998a), but at the important level of individual symptom items there are as yet no published reports of cross-cultural equivalence or interrater reliability.

The small-scale interrater reliability study reported here was based upon the use of interviews with Chinese patients that were conducted in Chinese by SCAN-trained Chinese psychiatrists. This allows a detailed examination of both translation and cross-cultural equivalence problems at the same time.

METHOD

In SCAN, the method of interviewing, although systematic and standardised, is flexible and based upon clinical expertise; it depends upon detailed questioning of the subject to enable an interviewer to

judge whether a recognisable symptom or sign is present during a specified period of time, and if so, with what severity. The SCAN is organised into 27 sections containing, in total, over 1400 items covering all major areas of psychopathology, which are described and defined in the SCAN Glossary. The interviewer is trained to rate the presence and degree of severity of these symptoms, each in their own right and independent of their relevance to any possible diagnoses that may or may not eventually apply to the subject being interviewed.

Preparation of the new Chinese SCAN; translation and back-translation

Translation and back-translation as recommended by Sartorius & Kukeny (1993) were carried out as follows. In 1993 translation of the English SCAN into Chinese (Mandarin) in Taiwan was started by a group of native Taiwanese psychiatrists. In the translation, Fukienese (the main native dialect in Taiwan) as well as Mandarin phrases and terms were used. Group discussions then were held by the translators, to compare in detail each section of the SCAN in English and Chinese. Further improvements were made to the text, giving conceptual equivalence priority over word-for-word linguistic equivalence. A back-translation to English was then carried out by bilingual mental health professionals who had not been trained in SCAN. This back-translation was reviewed for distortions from the original document, and areas of uncertainty or difficulty were discussed and resolved by the panel of psychiatrists.

Training of Chinese psychiatrists in SCAN and first adjustments to the new version

A training course for 12 Taiwanese academic psychiatrists was conducted in Taipei in 1996 in English by two experienced SCAN trainers according to the established format developed by WHO. Eight of the Taiwan trainees formed a local ‘SCAN club’ and held monthly meetings to review the translation and to test SCAN. Each psychiatrist contributed interviews that were designed to determine whether patients understand the SCAN questions and concepts. Based on these interviews, minor changes were made to the interview and probes.

*Results from this study were presented at the XIII Congress of the International Federation of Psychiatric Epidemiology, March 1999, Taipei, Taiwan.

Preparation of interview videotapes

Videotape recordings of clinical interviews with 40 patients using the Chinese SCAN 2.1 version (World Health Organization, 1999) were made by the SCAN club members in Taiwan for this study. These were regarded as a pool of videotapes from which a manageable number of the best possible quality for translation and detailed discussion could be selected. The patients were selected as being representative of four areas of psychopathology – neurotic, affective, substance misuse or dependence, and psychotic – and were intended to be typical cases seen in hospital practice likely to provide ratings on a wide variety of SCAN items. About half of the recordings were excluded because of unsatisfactory technical quality (usually poor sound) or because of patients who gave many vague replies or who were very talkative; some interviews were excluded because of periods of unsatisfactory interviewing techniques. Selection in this way was appropriate, because the main properties required of the final group of interviews were simply that the clinical states of the patients should be reasonably typical and that the interviews should not set special problems for the translators. A final set of 16 Chinese videotapes was retained, with four patients in each of the four areas of psychopathology.

Transcripts in Chinese were prepared from the videotapes by local bureaux. These were reviewed and revised by each interviewer to produce a final Chinese version. The Chinese transcripts then were sent to another bureau for English translation. The English transcripts were reviewed by a bilingual clinician and grammatical errors were corrected. English subtitles were added to copies of the videotapes and time codes then were matched back to the printed transcripts for reference use.

Rating of the videotapes

Ratings were carried out by two groups of psychiatrists. The native English-speaking ('US-UK') group consisted of one British, one Irish and two American psychiatrists. All had had extensive experience teaching and using SCAN in their own centres and also had conducted SCAN courses in centres and countries other than their own. The native Chinese-speaking ('Taiwan') group consisted of six of the

original SCAN club group. It was decided that ratings would be based primarily on transcripts but that excerpts of the videos would be viewed so that raters could check for clinically relevant abnormalities of behaviour and gain a general sense of each patient. A special meeting of the two groups of psychiatrists, lasting 1 week, took place in Taipei in September 1998 to complete the comparison.

Symptom ratings

Ratings of one interview that was sent ahead to all the psychiatrists were compared first. Discrepancies were discussed but changes were not made to the ratings. For the remainder of the cases, videotapes with subtitles were shown to the group for 5–10 min, without discussion. Each rater then used the printed transcript in his or her native language to rate the SCAN items. One patient from each type of psychopathology was rated, and after the rating was completed it was discussed item by item. Again, no changes were made in the ratings. Subsequently, the remaining transcripts were rated without detailed item-by-item discussion. If conceptual issues arose during the rating, they were noted and presented after completion of the rating for group discussion. All ratings were entered into a spreadsheet for later comparison and quantitative analysis.

Clinical diagnosis

Although a subsidiary issue in this study, the raters recorded their opinion about the likely clinical diagnosis, using ICD-10, Chapter V (World Health Organization, 1993). Excellent agreement on the clinical diagnoses was found. This aspect of the study is not described further here, but details may be obtained from A.T.A.C. upon request (or seen at <http://www.mdlogix.com/id115.htm>).

Data management

Ratings made on paper recording forms were entered into a spreadsheet and the summarised numerical results within and across the two groups of raters (US-UK and Taiwan) were available for group discussion shortly after all the ratings were completed.

Items rated 8 or 9 (SCAN codes for 'uncertain' and 'missing') were treated as missing values in mean score analysis. At the item level, a binary representation of

the presence or absence of a clinically significant symptom was of interest. This is based on the explicit rules of SCAN interviewing in which 2 and 3 (in Part 1) and 1, 2 and 3 (in Part 2) are considered 'clinically significant'. The same data transformation principles are used in diagnostic algorithms. Thus, the definite presence of a symptom meeting the glossary definition would be indicated by a value of 1, and anything less indicated by a value of 0. Specifically, item scores using Rating Scale I in Part 1 of the SCAN were transformed into dichotomous (0, 1) values by mapping 0 and 1 scores to 0 ($0, 1 \geq 0$) and 2 and 3 scores to 1 ($2, 3 \geq 1$). In Part 2 of the SCAN, items using Rating Scale II were similarly converted into dichotomous values, except that 1, 2 and 3 values were all mapped to 1.

Clinical equivalence and qualitative analyses

Following each rating session, discussion took place between the two groups of interviewers and detailed notes of these qualitative findings were kept. The purpose of these discussions was to enhance the clinical validity of the Chinese SCAN symptom items cross-culturally by making sure that the Chinese items addressed the same concept as the English. In addition, the possible effects of cultural differences in social desirability upon the responses of the patients to questions were discussed.

Quantitative analyses; interrater reliability

Owing to the small sample size, the statistical methods employed are considered as being primarily descriptive. However, two hypotheses were discussed and agreed upon at the start of the meeting, and these guided some aspects of the data analyses:

- (a) We expected that there would be a large measure of agreement between the item ratings of the two groups of psychiatrists.
- (b) Within the overall agreement, it was expected that some disagreements would be found, the most likely being that the Taiwanese psychiatrists would rate anxiety symptoms higher and depressive symptoms lower than the visiting US-UK group.

Comparison of scaled item ratings

For each patient, single-item group means were calculated for the US–UK and Taiwan groups of raters. A number of items from the various SCAN sections were omitted because they were not rated at all or there were too few ratings to make comparisons. Overall mean and standard deviations for items in each SCAN section were calculated across patients within each diagnostic group (e.g. Affective). The means and standard deviations were compared between the two rater groups using paired two-tailed *t*-test statistics for individual items and for items grouped by SCAN section.

Comparison of binary-transformed item ratings

Three methods were used to examine this aspect of items present or absent:

- (a) The overall presence or absence of ratings of clinical significance was summarised by the mean of the binary ratings within groups and converted into a percentage value.
- (b) Between-group agreement was calculated in three steps: the US–UK and Taiwan group means for the binary item ratings for an individual subject were calculated; these means were rounded to either 0 or 1; and the number of patients for which the rounded group mean of the binary ratings was in agreement between the US–UK and Taiwan groups was counted and converted to a percentage. Because there were only four patients in a diagnostic group, for the individual items the between-group percentage agreement could be only 0, 25, 50, 75 or 100. These between-group percentage agreements for SCAN item binary ratings were also averaged by SCAN section.
- (c) Within-group agreement was also examined. Because this follows the same general pattern as the more important between-group agreement, the results are not given here in detail; these may be obtained from A.T.A.C. upon request (or seen at <http://www.mdlogix.com/id115.htm>).

RESULTS

Psycholinguistic equivalents

To improve the way in which the concepts underlying the items were conveyed in

Chinese, it was decided that several items needed modification in the Chinese draft. First, the English question about being unable to relax in a general sense needed to be differentiated into questions about ‘being unable to relax your feelings’ and ‘being unable to relax your muscles’. This distinction is important because the concept of ‘relaxation’ in Chinese can apply to either the emotions or the physical body.

Second, in translating the item for ‘depressed mood’, the Chinese version also required slight modification. Just as English has several vernacular terms for depression (e.g. ‘down in the dumps’, ‘blue’, etc.), so does Chinese (e.g. ‘fallen into the valley’, ‘heart not clear up’, ‘sour heart’). Similarly, the English word ‘guilt’ has no exact Chinese equivalent. The original SCAN includes ‘blamed yourself’ and ‘ashamed of yourself’, which were judged to be relevant to the Chinese. In addition, ‘loss of face’, a common term in Chinese, was also incorporated into the Chinese SCAN.

Agreement was reached on similar modifications to a number of other items in order to improve the conceptual equivalence of the new Chinese version.

Social desirability set

Another conceptual area had to do with social desirability. For example, it was thought to be difficult for subjects in Taiwan to answer the question, “Would you say you were more calm and collected, less prone to irritability, than most people?” People in Taiwan have difficulty in answering a question worded in a positive manner because they are taught to be humble. In the Taiwan version the polarity of the probes was reversed to avoid this response bias problem. Thus the interview text reads “Would you say that you are more prone to be nervous and tense than most people?” This may also be a problem in countries such as Japan.

Similarly, in asking people in Taiwan about loss of enjoyment (anhedonia), the concept of ‘enjoying’ life may not apply because there is a cultural bias against admitting that activities are enjoyed. The cultural concern is that such an admission may be interpreted as boasting. To deal with this point, the SCAN item was modified by substituting a list of personal activities, and the respondent is asked about any changes in the level of participation in them.

A complete list of the SCAN items and the associated issues that were

discussed has been compiled and is available from A.T.A.C. upon request.

The SCAN item interrater reliability

Table 1 shows the quantitative results for selected representative individual SCAN items that differed significantly and the overall section summary ratings (‘grand means’) for both scaled ratings and binary-transformed ratings, comparing the US–UK and Taiwan groups of raters. (Tables containing the larger set of items can be obtained from A.T.A.C. upon request, or seen at <http://www.mdlogix.com/id115.htm>)

Seven out of 12 sections differed significantly in the grand mean ratings for the section between the Taiwan and US–UK groups when using scaled data, but only three of these differences were significant when using binary ratings. For all the sections, the group percentage agreement between the Taiwan group and the US–UK group was good, ranging from a low of 69% for Section 8 to 100% for Section 11.

In summary, the Taiwan group rated some items in both Section 3 (worry and tension) and Section 4 (anxiety) higher than the US–UK group. The Taiwan group also rated some of the affective symptoms higher (Sections 6, 7 and 10). In Section 11 (alcohol) some scaled items were rated higher by the Taiwan group and some by the US–UK group, but there were no differences in Section 12 (other substances). In Sections 16, 17 and 18 (perceptual changes, auditory hallucinations and thought disorder) there were several items rated higher by the Taiwanese group, but none in Section 19 (delusions).

DISCUSSION

Cross-cultural studies

This study illustrates how psychiatrists from different cultures can collaborate in improving the cross-cultural equivalence of a comprehensive measure of psychopathology. The study was conducted in Taiwan with experienced English-speaking SCAN trainers who rated interviews that were first conducted in Chinese and then translated to English. This was in contrast to previous studies of cross-cultural reliability where local SCAN assessors were trained and conducted reliability exercises in English at English-speaking sites. It was apparent to the experienced English-speaking SCAN trainers that this new

Table 1 Comparison of scaled and binary ratings of selected Schedules for Clinical Assessment in Neuropsychiatry (SCAN) items and section summary by rater group (US–UK and Taiwan)

Item	Description	Scaled SCAN ratings			Binary-transformed ratings		
		Mean		P-value (t-test)	% of presence		Between-group agreement (%)
		US–UK	Taiwan		US–UK	Taiwan	
3.001	Worry	1.19	1.85	0.16	50	65	75
3.002	Nervous tension	1.00	1.65	0.03	38	55	100
3.006	Restlessness	1.67	2.25	0.21	60	95	75
4.001	Presence of anxiety	0.94	0.90	0.39	94	90	100
4.003	Cannot get breath	0.89	1.03	0.26	31	45	75
4.004	Heart pounding	1.67	1.22	0.21	67	50	75
6.001	Depressed mood	1.81	2.03	0.51	69	75	100
6.004	Anhedonia	1.21	1.41	0.20	42	44	100
6.006	Loss of hope	0.56	0.75	0.39	33	33	75
7.002	Concentration	1.75	2.17	0.42	75	63	75
7.004	Loss of interest	1.77	1.93	0.62	67	50	100
7.006	Loss of energy	2.06	1.95	0.34	67	70	100
8.005	Change in appetite	1.94	2.19	0.07	75	75	100
8.009	Sleep problem with depressed mood	0.75	0.75	N/A	75	75	100
8.011	Delayed sleep	1.25	1.75	0.23	38	56	75
10.001	Expansive mood	1.00	1.25	0.39	50	50	100
10.002	Irritable mood	0.94	1.14	0.40	44	40	100
10.005	Overtalkative	0.69	0.98	0.32	25	40	75
11.002	Frequency of drinking	7.00	6.63	0.06	N/A	N/A	N/A
11.004	Amount during heavier periods	5.63	4.96	0.65	N/A	N/A	N/A
11.014	Risk-taking behaviour with alcohol	1.13	1.28	0.03	50	45	100
12.024	Impaired capacity to control	1.92	1.20	0.53	42	45	100
12.033	Tolerance to drugs	1.00	1.00	N/A	25	25	100
12.040	Drug withdrawal problems	1.50	1.50	1.00	50	50	100
16.001	Unusual sensations	0.42	0.46	0.39	25	25	100
16.005	Changed time perception	1.15	1.33	0.39	25	25	100
16.006	Derealisation (things)	0.67	1.33	0.42	19	25	100
17.001	Probe for hallucinations	1.00	1.00	N/A	100	100	100
17.006	Quality of auditory hallucinations	2.00	2.00	N/A	0	0	100
17.008	Voice commenting	2.40	2.48	0.18	54	60	100
18.003	Thoughts being read	1.06	0.21	0.20	19	6	75
18.007	Thought broadcast	0.00	0.81	0.20	0	13	75
18.012	Replacement of will	0.75	1.04	0.57	25	50	75
19.001	Probe delusions of reference	0.88	1.00	0.39	83	100	100
19.003	Being spied upon	1.94	2.25	0.19	50	50	100
19.009	Delusional perception	0.56	0.88	0.58	13	36	75
Section summary							
3		0.79	1.15	0.01	27	43	88
4		0.74	0.87	0.02	36	37	88
6		1.09	1.31	0.01	42	49	91
7		1.64	2.09	0.03	58	63	84
8		1.10	1.34	0.24	45	50	69
10		0.91	1.14	0.01	38	40	89
11		1.80	1.48	0.03	45	44	100
12		1.30	1.35	0.64	41	40	97
16		0.61	0.76	0.18	28	30	86
17		1.14	1.27	0.09	42	45	90
18		0.37	0.81	0.01	17	36	72
19		1.02	1.11	0.15	47	47	85

N/A, not applicable.

method gave appropriately more prominence to the opinions and experiences of the local psychiatrists.

Translation of symptoms

The emphasis on the equivalence of experiences of the patients at the level of symptoms ensures as much cross-cultural equivalence as possible and also avoids the potential problem that is sometimes called the 'category fallacy'. This term refers to the warning given by some medical and social anthropologists that psychiatrists will be likely to arrive at misleading conclusions if they automatically apply the diagnostic concepts that they have become familiar with in their own culture to patients from different cultures (Kleinman & Good, 1985). Cross-cultural studies in which PSE or SCAN have been used (e.g. many of those coordinated by WHO) therefore minimise this problem.

It was apparent in the discussions that the major area in which the refinements were needed was not so much in the formal translation of the clinical definitions contained in the SCAN items and Glossary, but more in the vernacular terms used to operationalise communication of clinical concepts to the patients. It may be useful to divide the cross-cultural validity and equivalence issue into two parts: the terms, concepts and knowledge of languages needed to maximise communication between the professional research staff; and the terms and more vernacular style of language used to communicate these clinical concepts to the subjects.

Differences between Taiwanese and US–UK raters

Turning next to our quantitative findings, we found that although few individual items had group ratings that were significantly different, there were significant differences between items grouped by SCAN section. As expected, the Taiwan scaled ratings of anxiety symptoms were higher than the US–UK ratings. Although this difference disappeared for the section (Section 4) that covers panic and phobias when the ratings were transformed from scaled SCAN ratings to binary presence or absence of clinically significant symptoms, the difference persisted for non-specific symptoms (Section 3), consistent with the *a priori* hypothesis. It is possible that the difference could have an effect on studies of neurotic anxiety disorders.

CLINICAL IMPLICATIONS

- In converting the clinical concepts of the items in the Schedules for Clinical Assessment in Neuropsychiatry (SCAN) from English to Chinese, no barrier appeared that could not be overcome.
- It is recommended that the training plan of the SCAN should be extended to include detailed follow-up work, and one form for this is a reliability study as reported here.
- In translating the questions to be asked and the instructions in the SCAN Glossary that indicate how symptoms should be rated, consideration should be given to local vernacular terms in addition to the more formal language that may be adequate for the mental health professionals.

LIMITATIONS

- Imperfections in some of the interview recordings used in this study may have been due to the artificial situation created by making a time-limited recording of an interview, knowing that it will be used for a reliability study.
- The small sample size has limited the statistical power to detect item-level differences.
- This study was limited to the cross-cultural equivalence of symptoms, and the cross-cultural reliability of diagnosis was not examined in detail.

A. T. A. CHENG, FRCPsych, Division of Epidemiology, Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan; A. Y. TIEN, MD, Department of Psychiatry, West Virginia University, Morgantown, West Virginia, USA; C. J. CHANG, MD, Department of Psychiatry, Cathay General Hospital, Taipei, Taiwan; T. S. BRUGHHA, MRCPsych, Section of Social and Epidemiological Psychiatry, Department of Psychiatry, University of Leicester, Leicester, UK; J. E. COOPER, FRCPsych, Department of Psychiatry, University of Nottingham, UK; C. S. LEE, MD, Department of Psychiatry, Lo-Tung Poh-Ai Hospital, Lo-Tung, I-Lan, Taiwan; W. COMPTON, MD, Department of Psychiatry, Washington University, St Louis, Missouri, USA; C. Y. LIU, MD, Department of Psychiatry, Chang Gung University College of Medicine and Chang Gung Memorial Hospital, Tao-Yuan, Taiwan; W. Y. YU, MD, Department of Neuropsychiatry, Kai-Suan Psychiatric Hospital, Kaohsiung, Taiwan; H. M. CHEN, PhD, Division of Epidemiology, Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan

Correspondence: Professor Andrew T. A. Cheng, Division of Epidemiology, Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan. Fax: 8862 2782 3047; e-mail: bmandrew@ccvax.sinica.edu.tw

(First received 16 August 2000, final revision 5 December 2000, accepted 12 December 2000)

On the other hand, the expected higher level of ratings of depressive symptoms by US–UK raters did not occur, and in fact the opposite result was observed, with the mean Taiwan rating for items in the mood sections (Sections 6, 7 and 10) being higher. This prediction was based upon the results of studies that found comparatively low rates for depressive disorders in Taiwan (Compton *et al*, 1991; Weissman *et al*, 1996). However, because the differences between the Taiwan and US–UK groups were decreased when the items were dichotomised, the effect on diagnosis

of depressive disorders may not be large. As with Section 3 and neurotic anxiety disorders, this needs to be kept in mind in future studies.

The higher US–UK rating for alcohol use disorders (Section 11) was not expected and did not persist after the scaled ratings were dichotomised. As with the results for other sections, this result suggests that it is important to distinguish scaled ratings from the basic presence or absence of clinically significant symptoms. When diagnosis is the main focus of a study, binary ratings should be sufficient.

Within the psychotic symptoms, Section 18 (thought disorder) stood out as being rated significantly higher by the Taiwan group than by the US–UK group, with both scaled and binary ratings. In SCAN training, particular attention is paid to these items, emphasising the need to avoid false-positive ratings. The group discussions clearly indicated that the Taiwan group rated too liberally with these items, and based the decision too much upon the first reply of the patient to the structured probe rather than on the clinical judgement of the interviewers after further questions and answers (a possible contributing cause for this has been noted already). The fact that ratings of the other psychotic symptom sections (Sections 16, 17 and 19) by the Taiwan group were not significantly different from the ratings by the US–UK group is partial evidence for this possibility, because Section 18 items require a higher standard of clinical evidence than items in the other sections. Alternatively, it is possible that the US–UK group rated too low on the Section 18 items because of information being changed or lost in translation.

In summary, the cross-cultural differences between the two groups were lessened when items were converted from scaled ratings to group percentage agreements for binary ratings, which were generally high.

Issues for future SCAN development and training

Even though the first draft of the Chinese translation of SCAN with which this

study started had been the result of a great deal of careful translation work, the novel methods used in this study showed that further improvements were still possible. It seems that to capitalise on the efforts made in standard SCAN training, follow-up training and intermittent joint discussions can be of great value. Cross-training in different centres is stipulated by the SCAN training material, but there has not been a positive requirement for detailed follow-up monitoring of SCAN performance by local raters using the native language of a new SCAN centre. To avoid drift of rating and some of the problems encountered in this study, translation and reliability exercises similar to the one described here are recommended. This should be balanced with an awareness of the amount of work required, but for implementation of the SCAN in a major language the work is justified.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the excellent performance of our research assistants C. H. Ling, T. H. Chang, S. T. Pan and C. H. Chiu, and all the SCAN club members.

REFERENCES

- Brugha, T. S. (1998)** Development of SCAN Version 2. In *Diagnosis and Clinical Measurement in Psychiatry: A Reference Manual for SCAN/PSE-10* (eds J. Wing, N. Sartorius & T. Ustun), pp. 116–127. Cambridge: Cambridge University Press.
- Compton, W., Helzer, J., Hwu, H., et al (1991)** New methods in cross-cultural psychiatry: psychiatric illness

in Taiwan and the United States. *American Journal of Psychiatry*, **148**, 1697–1704.

Easton, C., Meza, E., Mager, D., et al (1997) Test–retest reliability of the alcohol and drug use disorder sections of the Schedules for Clinical Assessment in Neuropsychiatry (SCAN). *Drug and Alcohol Dependence*, **47**, 187–194.

Kleinman, A. & Good, B. (1985) *Culture and Depression: Studies in the Anthropology and Cross-Cultural Psychiatry of Affect and Disorder*. Los Angeles, CA: University of California Press.

Sartorius, N. & Kukyen, W. (1993) Translation of health status instruments. In *Quality of Life Assessments in Health Care Settings*. Berlin: Springer-Verlag.

Weissman, M. M., Bland, R. C., Canino, G., et al (1996) Cross-national epidemiology of major depression and bipolar disorder. *Journal of the American Medical Association*, **276**, 293–299.

Wing, J. K. (1996) SCAN and the PSE tradition. *Social Psychiatry and Psychiatric Epidemiology*, **31**, 50–54.

—, **Babor, T., Brugha, T., et al (1990)** SCAN: Schedules for Clinical Assessment in Neuropsychiatry. *Archives of General Psychiatry*, **47**, 589–593.

—, **Cooper, J. & Sartorius, N. (1974)** *Measurement and Classification of Psychiatric Symptoms*. New York: Cambridge University Press.

—, **Sartorius, N. & Der, J. (1998a)** International field trials: SCAN–0. In *Diagnosis and Clinical Measurement in Psychiatry: A Reference Manual for SCAN/PSE–10* (eds J. Wing, N. Sartorius & T. Ustun), pp. 86–109. Cambridge: Cambridge University Press.

—, —, & **Ustun, T. B. (eds) (1998b)** *Diagnosis and Clinical Measurement in Psychiatry, a Reference Manual for SCAN*. Cambridge: Cambridge University Press.

World Health Organization (1993) *The ICD–10 Classification of Mental and Behavioural Disorders: Diagnostic Criteria for Research*. Geneva: WHO.

— (1994) *SCAN: Schedules for Clinical Assessment in Neuropsychiatry*. Geneva: WHO.

— (1999) *SCAN 2.1: Schedules for Clinical Assessment in Neuropsychiatry*. Cambridge: Cambridge University Press.