

Preliminary psychometric evaluation of a standardized clinical family interview

Warren Kinston* and Peter Loader†

Though multiple task interviews are often preferable for family interaction research, clinical-style interviews have a rôle to play. The 'Standardized Clinical Family Interview', whose development, structure and mode of administration have been described in an earlier paper, is designed to elicit family interaction in a clinically relevant fashion. A series of studies which evaluated the reliability, sensitivity, validity and acceptability of this instrument is now provided. The results suggest that the SCFI is useful and deserves further study; however, interviewer training is a major task.

Introduction

Direct observation of family interaction is essential in family therapy research. The description and measurement of interaction depends, therefore, on the availability of a suitable method for its elicitation. Cromwell *et al.* (1976) have identified and evaluated four approaches to elicitation: naturalistic observation, use of single complex tasks, multiple task interviews and clinical-style interviews. To this list we may add elicitation occurring during diagnostic and therapeutic sessions (e.g. Kinston *et al.*, 1982; Furniss *et al.*, 1983).

Clinical-style interviews of the whole family, specifically designed for research purposes, are conspicuous by their rarity. Those that have been constructed (Walker *et al.*, 1984; Epstein and Bishop, 1981; Minuchin *et al.*, 1978; Wells and Rabiner, 1973; Riskin, 1976) are not geared to non-labelled families, are minimally described and have been subjected to little if any formal psychometric evaluation. The Family Studies Group at the Hospital for Sick Children (London, U.K.) therefore developed a research interview called the 'Standardized Clinical Family Interview' (SCFI) to meet these basic requirements.

Accepted version received June 1986.

*All correspondence to Dr Kinston, Family Research Program, Institute of Organization of Social Studies, Brunel University, Uxbridge, Middlesex UB8 3PH, U.K.

†The Hospital for Sick Children and the Hornsey Rise Child Guidance Clinic, London.

Description of the SCFI

A detailed description of the development, structure and mode of administration of the SCFI, together with the protocol, has been published in a companion paper (Kinston and Loader, 1984). The main points will be summarized here to set the methodological studies to follow in context.

The SCFI is designed to be used with a wide range of labelled and non-labelled families, and takes no more than one hour to administer. It resembles other forms of systematic clinical examination in that it must be performed by a professional who is trained to observe and test or probe the object of investigation in a clearly defined, repeatable manner. The procedure is not immediately linked to gathering any specific information but rather serves as an opportunity for the family to get involved in discussing and thinking about its own activities and for family members to interact. The interaction resulting from the SCFI can be assessed from a variety of perspectives and with differing instruments depending on the purpose of the research. In this it differs from most other methods in which the elicitation procedure is directly linked to the research interest.

The interviewer carries the family through a semi-structured protocol, and as he does so he follows various rules of procedure which promote standardization across families. These rules concern his attitude to the family, his focus during the interviewing process, his involvement of the children and the way he adheres to or modifies the given protocol. These rules in effect mean that the interviewer has to be an experienced family therapist. The interviewer must be able to get close to the family but avoid becoming enmeshed with them; and he must be able to adopt a neutral and non-therapeutic stance in relation to any problems which may emerge. The family must therefore be prepared for the interview.

The SCFI protocol consists of four phases. In phase 1, the introduction which lasts five to ten minutes, the interviewer meets the family and explains the interview rules. In phase 2, which can last from two to fifteen minutes, the reason for attendance is discussed with emphasis on the family perspective. Phase 3, System Properties, is the heart of the interview and lasts thirty to forty minutes. It consists of a series of questions, probes and statements about family life. The topics covered include the way the family sees itself, the degree of togetherness, who does what with whom, how families are alike and how they are different, views about stages of the family life cycle, the issue of rôles and

responsibilities, areas of conflict and ways of handling these, issues of discipline and ways of making decisions. Finally, the family's relationship with the families of origin and with the community is raised. Phase 4, the conclusion, brings the interview to a close.

The researcher using this, or any other method of eliciting family interaction, will wish to know whether it is indeed standardized, whether the interaction it elicits bears any relation to the interaction with which he is concerned, whether the interaction is elicited reliably or is just an artefact of the moment or of the personality of the interviewer, and whether it is acceptable to a wide range of families. The studies which are described in this paper were devised to provide evidence on these important matters.

Materials and methods

Source of families

All families were recruited through their attendance at a large children's hospital. They comprised two different groups.

- (a) Families of children with coeliac disease attending a coeliac clinic for surveillance. Coeliac disease usually manifests itself in infancy with bowel complaints. The children become and remain healthy as long as they adhere to a special diet. These children had been healthy for a variable number of years.
- (b) Families with children attending the dermatology clinic with atopic eczema. Eczema is a common condition, but the children at this clinic have usually been affected particularly severely.

Measures of the family

Families were assessed using two instruments which have been developed by the Family Studies Group in the Department of Psychological Medicine. The Current Family State Assessment (CFSA) consists of thirty scales (third edition) or thirty-three scales (fourth edition) and has been examined for reliability (Kinston *et al.*, 1979). The other instrument is the Family Health Scale (FHS) which provides a reliable and valid clinical assessment of family interaction along a poorly-functioning/healthy-functioning axis (Kinston *et al.*, in press). The FHS in the edition mainly used (third) draws on the assessment of communication, atmosphere, handling of affect, boundaries, alliances, parenting and relationship to the environment so as to generate an

overall rating for the family. The fourth edition includes more explicit assessments of stability-adaptability, conflict resolution and problem-solving. The FHS is designed to give a score between one and seven, where seven reflects optimal family functioning, five reflects adequate function and lower scores indicate progressive overall dysfunction.

Studies

The SCFI was investigated in a number of studies: no. 1, content-analytic study of check standardization; no. 2, reliability and sensitivity to interviewer change; no. 3, validation against another method of eliciting interaction; no. 4, validation against individual member psychopathology; no. 5, acceptability to families.

The nature of these studies and their results will be described consecutively below. Because the psychometric work was usually performed as a part of other substantive studies, details of the studies are curtailed to what is required for understanding and evaluating the SCFI.

Data handling and analysis

Data was coded and punched onto computer cards. Analysis was performed on the University of London Computer either using SPSS (Statistical Package for the Social Sciences) or specially written programs. Inter-rater reliability and correlations were checked using either weighted kappa (K_w ; Cohen, 1968; Hall, 1974) or Spearman's rank correlation coefficient (r_s ; Siegel, 1956). Significant differences between groups rated in two ways or on two occasions have been looked for with the Wilcoxon Matched Pair Signed Ranks Test (Siegel, 1956).

Procedures and results

Study no. 1: standardization

It is one thing to prepare a protocol and procedural rules and it is another for the interviewer to follow them. The interviewer was therefore asked to check off each standard item on his protocol as it was covered during the interview, and at the end of the interview the protocol was handed in. To confirm coverage, an independent observer watched videotapes of thirteen interviews and content-analysed the interviewer's vocal behaviour. A full account of this study was provided in Evans (1982) and it will be only briefly reported here.

The observer checked for any deviation from the required wording or order of administration and assessed the appropriateness of deviations in the light of the established principles of administration described in the companion paper (Kinston and Loader, 1984). Ancillary non-protocol interventions were also content-analysed and compared with the rules laid down.

Tables 1 and 2 summarize some of the quantitative findings. The protocol was adhered to flexibly, but 14% of changes in protocol items were considered inappropriate. Most of these reflected the interviewer becoming over-hesitant, confused or unclear. Of the very large number of non-protocol comments, less than 1% were considered inappropriate and these were generally due to the interviewer intervening in a therapeutic manner, or pursuing a family-initiated digression too far. Detailed examination of the interviews which were most deviant quantitatively revealed that the interviewer behaviour was called for by the nature of the family and was therefore a manifestation of interviewer empathy and flexibility. In all families, about one-third of the non-protocol utterances had a conversation-exchange function, one-fifth had an interview-maintenance function and two-fifths had a supportive function. This suggested that the interviewer maintained a consistent stance towards the families.

Evans concluded that the interview was conducted in a reasonably standardized fashion in that the interviewer, in the main, followed the

TABLE 1. *Use and alteration of protocol items (modified from Evans, 1982)*

Items	Mean	Range	Total number of items	% judged inappropriate
Items included unaltered	50.8	38 to 59	627	0
Items altered:				
By omission	16.2	8 to 29	211	18
By changing significantly	15.6	10 to 27	199	12
By re-ordering	1.9	0 to 10	23	0
Total altered			433	14

Number of items per SCFI = 67; number of interviews = 13; total number of items analysed = 871.

TABLE 2. *Non-protocol utterances and their functions (modified from Evans, 1982)*

	Mean (\pm S.D.)	Range	Grand total	Judged appropriate	Judged inappropriate
Non-protocol utterances	238.7* (\pm 60.9)	143-315	2964	2940	24 (0.8%)
Functions of non-protocol utterances (as proportion of total no.)					
a) Conversation exchange	0.34 (\pm 0.05)	0.27-0.43			
b) Interview maintenance	0.20 (\pm 0.05)	0.09-0.08			
c) Supportive	0.42 (\pm 0.07)	0.29-0.57			
d) Other	0.04 (\pm 0.02)	0.01-0.08			

No. of interviews (N) = 13.

*Excluding one incomplete tape.

protocol and abided by the rules of administration. However, he pointed out that there were no criteria for deciding whether the level of standardization achieved was satisfactory.

Study no. 2: reliability and sensitivity to context

We wished to check whether our interview was reliable. There are two major issues here. Does the SCFI elicit similar family interaction if administered (a) on two occasions (its reliability over time), and (b) by two interviewers (its sensitivity to interviewer style or personality)? To obtain evidence on these two points, we gave the SCFI twice to the same families at an interval of about three months and used two interviewers, one male and one female. The interviews were allocated alternately so that half the family had the male interviewer for their initial interview and the female interviewer for the second interview, and *vice versa* for the other half. In each case, the other interviewer observed via video cameras and closed-circuit television. Both interviewer and observer rated the family at the end of the interview using the Current Family State Assessment, fourth edition (Kinston *et al.*, 1979).

Details of the sample and the procedure have been provided in Loader

et al. (1980) and will not be repeated here as they are not directly relevant to the present purpose. As the initial interview of both interviewers was conducted in a state of anxiety, it was treated as a pilot and omitted from the analysis which was performed on the data from the remaining ten families.

Preliminary checks. (a) Inter-rater agreement on categories was calculated for each item in the CFSA using every interview. Only the items reliable by weighted kappa (K_w) were to be used in the main analysis. This revealed results similar to those reported and analysed in more detail in Kinston *et al.* (1979). Using $P < 0.01$, twenty of the thirty-three categories were rated reliably, with K_w ranging from 0.28 to 0.67. (b) Inter-rater agreement on families was also checked, because if raters saw particular family interviews very differently then unreliability might be due to such global disagreements rather than to problems agreeing on particular categories of interactions. We estimated similarity of family profiles as described in Kinston *et al.* (1979) using a Family Interview Discrepancy Score (FID). The FID is a measure of the difference between two profiles of category scores e.g. profiles by two raters or from the two interviews. This score is derived as follows: family discrepancy score = $\Sigma D^2/n$, where D is the number of points difference between the mean rating on each occasion, and n is the number of items. All categories were included to maximize the discovery of any global discrepancy, therefore $n = 21$. For the CFSA, the score can range from zero to sixteen. A cut-off of one was chosen as in the previous study, i.e. scores of greater than one indicate an excessive difference in profiles.

Only one FID score exceeded one (family no. 13, first interview: FID = 1.44) and the remainder ranged from 0.31 to 1.00. Removal of the unreliable items and repeating the FID calculations improved scores slightly. Item unreliability was therefore not due to one or a few poorly rated interviews.

Main findings. (a) Profile similarity between the first and second interviews was examined using an FID score, calculated using mean rater scores on reliable categories, and the results are displayed in Table 3. Only one family score was of concern (family no. 5: FID = 1.39); and in that case, two items were 0.5 points apart, four items 1 point apart, one item 1.5 points apart, and one item 2.5 points apart. Investigation of category discrepancy scores using a similar formula revealed that no items were particularly responsible for profile changes. The highest scoring item was *intrusiveness* (category discrepancy score = 1.30).

TABLE 3. *Similarity of family profiles of interaction at the two interviews*

Family no.	Discrepancy score*
1	0.36
2	0.65
3	0.88
4	0.52
5	1.39
6	0.40
7	0.73
8	0.60
9	0.27
10	0.20

*See text for an explanation of the discrepancy score and how it is calculated.

(b) Systematic differences in interaction due to the order of the interviews or due to the interviewer were examined using the Wilcoxon Matched Pairs Signed Ranks Test on reliably rated CFSA categories. This test establishes whether the scores for any category of interaction consistently shift up or down. If the SCFI elicited interaction reliably and independently of the interviewer or order, we would not expect to find significant differences. However, even if there were not such changes, it could be that this was because the scores of families moved considerably but cancelled each other out (e.g. one family scored much higher on *flexibility* with the male interviewer, while another scored much lower). To check for this, we calculated Spearman's rank correlation to see whether families moved relative to one another with respect either to order of administration or the particular interviewer. Moderately positive correlations are to be expected. Many neutral and negative correlations would be highly disturbing and suggest that the SCFI was an unreliable or excessively sensitive tool.

Results of the two comparisons (the ten interviews given first *vs* the ten given second; and the ten interviews given by the male interviewer *vs* the ten given by the female interviewer) are provided in Table 4. Order effects appeared significant ($P < 0.05$) for only two items, *intrusiveness* (3.2 at first interview to 2.4 at second interview) and *resonance* (3.9 to 3.4). Because these categories are conceptually linked, it is likely that the finding is a genuine one. The correlations are mostly moderately positive, as expected. The items showing minimal correlations are

TABLE 4. *Reliability and sensitivity. Examination for an effect of interview order or interviewer on interaction in the SCFI using the CFSA categories*

Categories	Interview order (First vs second)		Interviewer (male vs female)	
	WMPSTR	r_s	WMPSTR	r_s
Continuity	N.S.	0.78	N.S.	0.73
Information exchange	N.S.	0.15	N.S.	0.17
Interruptions	N.S.	0.58	N.S.	0.71
Laughter	N.S.	0.63	N.S.	0.86
Self-affirmation	N.S.	-0.05	N.S.	0.34
Agreement	N.S.	-0.11	<0.05	0.29
Disagreement	N.S.	0.45	N.S.	0.58
Intrusiveness	<0.05	0.59	N.S.	0.50
Range of affects	N.S.	0.57	N.S.	0.56
Intensity of affects	N.S.	0.63	N.S.	0.80
Comfort	N.S.	0.78	N.S.	0.67
Humour	N.S.	0.49	N.S.	0.78
Effective parental coalition	N.S.	0.78	N.S.	0.82
Alliances	N.S.	0.47	N.S.	0.53
Resonance	< 0.05	0.20	N.S.	0.14
Conflict acknowledgement	N.S.	0.50	N.S.	0.65
Relation to environment	N.S.	0.48	N.S.	0.80
Conflict resolution	N.S.	0.06	<0.01	0.65
Overprotection	N.S.	0.36	N.S.	0.67
Neglect	N.S.	0.43	N.S.	0.36

WMPSTR, Wilcoxon Matched Pairs Signed Ranks Test *P* value; r_s , Spearman's non-parametric correlation coefficient; *N*, ten (ten families interviewed twice; two ratings each time); N.S., not significant.

resonance (0.20) *information exchange* (0.15), *self-affirmation* (-0.05), *agreement* (-0.11) and *conflict resolution* (0.06).

In the case of interviewer effects, again only two items show significant changes: *agreement* (3.4 to 2.8; $P < 0.05$) and *conflict resolution* (1.9 to 2.7; $P < 0.01$). Again, a conceptual link exists in that pursuing the issue of conflicts (as the interviewer concerned acknowledged was his style) increases the likelihood of both discovering conflict and its resolution as well as tending to decrease the amount of agreement expressed. The correlations are again mostly moderately positive. The lowest correlations are for *information exchange* (0.17) and *resonance* (0.14).

Study no. 3: validation against another method of family interviewing

Other workers in the field (Lewis *et al.*, 1976; Minuchin *et al.*, 1978), as well as ourselves (Kinston and Loader, 1986), have used an interview consisting of a series of tasks to elicit family interaction. Unlike the SCFI, such interviews are quite dissimilar to conventional clinical interviews. Our Family Task Interview (FTI) leaves the family alone in the room and all instructions are delivered from a tape recorder. Full details of the procedure are provided in the above reference. It would not be expected that two different procedures would produce similar behaviour as categorized by the CFSA, nor indeed, similar patterns of interaction as judged clinically (Stratford *et al.*, 1982). Nevertheless, ratings of overall family health on our FHS, which tap general dimensions such as communication, atmosphere, parental functioning and boundary integrity would be expected to be similar.

Consecutive families attending the coeliac clinic were screened, and demographic and household data were obtained. All families were requested to attend for two whole family interviews until seventeen had consented. Six families refused to attend and two were excluded because English was not the language spoken at home. Full demographic details are provided in Evans (1982). All families had two parents and from one to four children ranging in age from six months to sixteen years. All social classes were represented. Half the families received the SCFI initially and half received the FTI. At the second interview two to five weeks later, each family received the other research interview. For practical reasons, the FHS ratings of the two interviews were performed at different times. The SCFI was rated at the time of the interviews by the interviewer and by an observer who watched on closed-circuit television. The FTI was rated by two observers watching the twelve available videotapes three years later. The FHS ratings in the latter case were based on three of the six tasks (i.e. approximately half the interview). We justified this using Lewis *et al.*'s (1976) demonstration that family assessment based on viewing part of a family task interview is highly correlated with assessment based on viewing the whole interview.

We examined the FHS ratings for reliability using Spearman's rank correlation coefficient. For the SCFI, $r_s = 0.73$ ($N = 17$); and for the FTI, $r_s = 0.73$ ($N = 12$). As these results suggest adequate agreement, we used the mean FHS score of each family interview in our statistical analyses. We compared the two interviews, first to see whether either interview appeared overall to give an impression of better or worse family functioning (Wilcoxon); and second, to see whether the families

maintained a similar position with respect to each other (r_s). We also examined the data to check for an order effect.

Results. Correlation between mean FHS scores on the SCFI and mean FHS scores on the FTI is moderately high ($r_s = 0.49$, $P < 0.05$) as expected. The Wilcoxon Test showed no significant difference between the two sets of scores, so indicating that there is no tendency for either of the interviews to reveal the families systematically as more or less well-functioning.

Analysis to check whether the order in which the interviews were administered to the family affected these conclusions revealed no significant difference between the set of scores obtained at the first interview (i.e. half being the SCFI and half the FTI) and the set of scores obtained at the second interview. Correlation between these scores was 0.48, indicating that the families tended to keep to their relative ranked positions.

Study no. 4: validation against individual member psychopathology

To be valid, the SCFI must allow the family to reveal itself as an integrated entity. The relationship between individual disturbance and family dysfunction is not well understood, and conventional thinking might suggest that the most disturbed families would contain the most disturbed individuals. However, the rationale and assumptions of family therapy would lead us to predict many examples of a lack of congruence between individual and family disturbance. A high degree of congruence at all levels of health might suggest that the SCFI simply allowed members to present themselves as individuals rather than specifically revealing family patterns. We therefore proceeded to test the hypothesis that the SCFI would reveal incongruence.

After giving the SCFI in the project described in study no. 3, we asked members to complete well-established self-report questionnaires which provide an estimate of individual psychological health. This study started with family no. 3, so here $N = 15$. The parents completed the Rutter A Scale (PR) for school children (Rutter *et al.*, 1970) and the Behaviour Check List (BCL) for pre-school children (Richman, 1977). Parents and children over sixteen completed the sixty-item General Health Questionnaire (GHQ; Goldberg, 1972). All three instruments employ a cut-off score, above which the respondent is deemed a potential psychiatric case. FHS ratings were made in ignorance of individual health scores.

Results. The mean FHS scores ranged from 3.3 to 5.8. The FHS is constructed such that a score of five indicates adequate functioning and we selected a score of 4.75 to differentiate healthy from unhealthy families. This resulted in seven healthy families (FHS range: 5.0 to 5.8) and eight unhealthy families (FHS range: 3.3 to 4.7). The families contained sixty-six individuals comprising thirty parents and one child over sixteen years assessed with the GHQ, twenty-nine children aged five to fifteen years assessed with the PR, and six children under five years assessed with the BCL. Eleven of these individuals (from five families) scored as potential cases and were designated unhealthy. The correspondence between family and individual health is displayed in Table 5. Certain features are to be noted. Congruence and incongruence between family health as assessed by FHS from the SCFI, and family health as assessed from incidence of an unhealthy member, occur with similar frequency. Six families (40%) are congruent – four healthy and two unhealthy on both criteria; while nine families (60%) are incongruent – six families are unhealthy but their members healthy, and three families are healthy but at least one of their members is unhealthy.

TABLE 5. *Family health matrix*

	Family health assessed via health of members		Total
	Healthy	Unhealthy	
Family health assessed via FHS on SCFI			
Healthy	4 (27%)	3 (20%)	7 (47%)
Unhealthy	6 (40%)	2 (13%)	8 (53%)
Total	10 (67%)	5 (33%)	15 (100%)

The relation between family health assessed by FHS at the SCFI and family health assessed according to whether the family contains at least one unhealthy member (using standard instruments – see text for details).

Study no. 5: consumer assessment

Shortly after being interviewed, each family who received the SCFI was sent a simple questionnaire asking how they found the interview. This dealt with how comfortable they felt during the SCFI, whether or not they found it an interesting experience and how much they thought they

had been like they usually were when together. The sample included families from the coeliac clinic used in piloting exercises as well as the families seen in study 3 above ($N = 17$). If the family had received the FTI, comparisons were requested. Additional comments were also invited.

Results. All families returned the questionnaire. Of those receiving both interviews, eight preferred the SCFI and six the FTI. When asked 'What do you think of the way you were together in our interview room (for the SCFI)?', two families did not respond, eleven believed they were 'very typical' or 'typical' and four thought their interaction 'untypical'. Of this last group three also responded 'untypical' to their interaction in the FTI. No family found the SCFI 'boring', 'uncomfortable' or 'harmful' (our terms) and only two (the healthiest families on the FHS) made critical comments: one family complained of things being done in a rush; and the other complained that the interviewer had expected natural behaviour and interaction in an unnatural setting.

Discussion

The SCFI in its pilot, final and developed forms has now been formally administered on some dozens of occasions, mainly by two interviewers. The psychometric testing of the SCFI as described in this paper, though severely limited when judged by conventional standards, still reflects the most determined attempt yet to put a clinically-based research interview for the whole family through methodological hoops. In favour of the SCFI, it must be noted that the administration requires no complex equipment and minimal environmental facilities; the protocol is non-offensive to families, psychiatrically labelled or non-labelled; and its subject matter is easily handled by both interviewer and family.

The two contentious issues which require some discussion are the degree to which methodological demands have been met, and the feasibility of using such an interview in research. We will first tackle the methodology by reviewing and discussing the findings from each study.

Study no. 1: standardization

The findings of this study are indicative rather than definitive. Only one observer checked the interviews against the protocol and decided appropriateness and function of the non-protocol items. His checks may have been carried out in an unreliable, invalid or systematically biased

way. Bias is probably unlikely, since the observer was independent of the project and his findings were mixed—part favourable and part unfavourable. In the method used, rating categories and definitions were kept extremely simple to increase the likelihood of reliability and validity. We can therefore examine the substance of the results with some degree of confidence.

It will be recalled that non-protocol items and deviation from the protocol are permitted, but only if appropriate to the aims of the interview. However, the high incidence of inappropriate alterations of the given protocol items was notable and of concern. Although only 0.8% of the non-protocol items were judged inappropriate, 14% of changes in the protocol items were so judged. In no case (0%) was adherence to the item as provided in the protocol judged to be inappropriate. This last finding might be expected because great care had been taken in devising the protocol to suit the purpose of the interview. We therefore concluded that although the protocol probably met the specified aims of the SCFI, the findings suggest that the interviewer had difficulty in regularly adhering to it.

To understand the reason for this, we reviewed the interviews and were impressed by the impact the family can have on the interviewer. The difficulty in standardization may reflect this impact. It will be recalled that the SCFI's goal is to reveal, without modifying or bypassing, the dysfunctional interaction which is the basis of the impact. Interviewers in conventional interview settings may be more protected from such impact because their goal is to bypass dysfunction and elicit specific items of information, or to modify dysfunction by intervening therapeutically.

Study no. 2: reliability and sensitivity

The interview reliability study again revealed the influence of contextual factors on family dynamics. For example, at the second interview, the family members showed less intrusiveness and less pathological involvement (resonance). This was possibly due to a reduction in their anxiety based on having already gone through an interview. Similarly, families revealed less conflict resolution and reduced agreements when interviewed by one of the interviewers. This reflected this interviewer's particular interest in pursuing family conflicts.

The large majority of CFSA items did show stability in the face of contextual change, and so it seems that the SCFI can be judged to elicit interaction in a reliable manner. However, several of the CFSA items

showed a low correlation across interviews and across interviewers. Do these invalidate the items on CFSA or do they cast doubt on our interview? We interpret these low correlations to mean that these items refer to types of interaction which are not stable—even though the items of interaction may be reliably rated on any particular occasion or with a particular interviewer.

Only two items, *resonance* and *information exchange*, are not stable in respect of both time and interviewer, and these findings do raise questions about the usefulness of these categories. For *information exchange* this is relatively trivial; it is not surprising that the amount and quality of information exchange varies greatly with circumstances. For *resonance* it seems to be more serious. The concept plays a major part in Minuchin's influential theories (Minuchin, 1974; Minuchin *et al.*, 1978) and is used as fundamental descriptor of the family in the 'Circumplex Model' of Olsen and co-workers (Olsen *et al.*, 1979).

Study no. 3: validation against the FTI

Validity is suggested by the finding that the degree of dysfunction in the family as revealed by the SCFI was similar to that revealed by a FTI. However, the correlation of 0.49 between FHS scores indicates that these two interviews reveal the families as by no means identical in terms of overall family health. These differences may once again reflect the effect of an interviewer, but, if so, it is an effect which is not systematically depressing or elevating the quality of interaction.

Our clinical assessments based on the FHS are weak in an important respect. The FHS are designed for the rater to take into account as much information as is available, in other words, whenever possible both the family's reports of problems and events and observed family interaction. However, in these studies raters were asked to base their judgement on observed interaction because the SCFI is not designed to obtain family reports in any standardized way. The raters, however, doubted that their ratings were based on observed interaction alone, as they felt influenced by reported information. A further problem the raters encountered was how to deal with the fact that all family interaction also involved the interviewer. They felt that his influence on the family system could not be separated off. The extent of this problem on assessment, a difficulty which is of course routine in the clinical situation (Stratford *et al.*, 1982), was examined by attempting to compare SCFI-generated interaction with FTI-generated interaction. Significantly, this study was abandoned following failure of the observers to agree on

what they observed due to differing judgements of the effect of the interviewer.

Study no. 4: validation against individual member psychopathology

The SCFI was designed to reveal family dysfunction. It would not be valid if all it did was reveal the psychopathology of individual family members. Validity is therefore suggested by the finding that elicitation of dysfunctional interaction was only partly affected by the presence of psychiatrically-labelled individuals within the family. In the majority of cases (60%) there was a lack of congruence between family health assessed using the FHS as a criterion and family health assessed by the presence of a psychiatrically-labelled member.

It should be noted that whereas the measures of individual health have had empirical validation of their cut-off points, the FHS has not been empirically standardized. The cut-off point chosen on the FHS was based on clinical values rather than socially-defined disturbance. In this regard, it may be noted that the proportion of individual disturbance in the sample is not unusual—33% of families had ill members. Further study is required to appreciate the significance of finding that over half (53%) of the families were judged to be unhealthy.

Study no. 5: consumer assessment

The families generally but not invariably saw their behaviour as typical of home life. However, we have no evidence to show that this is so; and in any case, we were more concerned whether the interaction validly reflected that seen by clinicians.

Feasibility. We return now to the second contentious issue—the feasibility of using the SCFI. The technical principles of interviewing as enunciated in the accompanying paper (e.g. including children, refraining from therapeutic intervention) have been put into practice, but the SCFI is not a simple instrument to use. For example, the standardization study revealed a number of interventions judged inappropriate to our specified rules, although the impact of these on interaction was not assessed. In addition, although the similarity between interaction elicited by the two trained interviewers was marked, the sensitivity study revealed, as was to be expected, that an interviewer with a particular bias could foster specific forms of interaction.

It is well known that trained interviewers, like raters, will vary their

techniques and adherence to administration procedures over time and require monitoring and retraining (Kazdin, 1977). We believe this would be essential with the SCFI because the interviewers themselves reported that interviewing was a difficult task requiring intense concentration, high sensitivity and self-scrutiny throughout.

Conclusion

Despite the limited extent and form of our methodological studies, it appears that the SCFI shows a certain robustness, which probably emanates from its clinical origins. Our principal methods of assessing the SCFI were clinical judgements as structured by our two instruments, the CFSA and FHS. This limited the testing of the SCFI and knowledge of its properties. However, the SCFI could be used with a variety of assessment schemes, for example, measures of separation anxiety, conversational participation, or attributions.

The SCFI can be administered on more than one occasion and by different interviewers without marked fluctuations in family response, insofar as we assessed it. However, interviewer training is arduous. Because our other research interview, the FTI, is simple, totally standardized and is neither contaminated by reports of family life nor complicated by the presence of an outsider interacting with the family, it has come to be preferred for most of our current research. Where an interviewer is essential (for example in gathering family reports as well as observing interaction), our limited methodological evaluation of the SCFI has been sufficiently encouraging to lead us to use it, modified as appropriate for the particular research needs.

Acknowledgements

The assistance of Jackie Stratford is much appreciated. Generous financial support has been received from the Leverhulme Trust.

W. K. is Senior Research Fellow and Director of the Family Research Program, Brunel University and an Honorary Consultant Psychotherapist, The Hospital for Sick Children. P.L. is former Research Fellow and Honorary Senior Registrar, The Hospital for Sick children and at present Consultant Child Psychiatrist at Hornsey Rise Child Guidance Clinic.

References

- COHEN, J. (1968) Weighted Kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin*, **70**: 213-220.
- CROMWELL, R. E., OLSON, D. H. and FOURNIER, D. G. (1976) Tools and techniques for diagnosis and evaluations in marital and family therapy. *Family Process*, **15**: 1-49.
- EPSTEIN, N. B. and BISHOP, D. S. (1981) Problem-centred systems therapy of the family. In: A.S.D. Gurman and D. P. Kniskern (Eds). *Handbook of Family Therapy*. New York. Brunner-Mazel.
- EVANS, D. (1982) *The Feasibility of Standardising a Clinical-style Interview Designed to Elicit Whole Family Interaction*. Paper presented to Annual Meeting, Child and Adolescent Psychiatry Section, Royal College of Psychiatrists, 1982.
- FURNISS, T., BENTOVIM, A. and KINSTON, W. (1983) Clinical process recording in focal family therapy. *Journal of Marital and Family Therapy*, **9**: 147-170.
- GOLDBERG, D. P. (1972) *The Detection of Psychiatric Illness by Questionnaire*. Maudsley Monograph No. 21. London. Oxford University Press.
- HALL, J. N. (1974) Inter-rater reliability of ward rating scales. *British Journal of Psychiatry*, **125**: 248-255.
- KAZDIN, A. E. (1977) Artifact, bias and complexity of assessment: the ABC's of reliability. *Journal of Applied Behavioural Analysis*, **10**: 141-150.
- KINSTON, W. and LOADER, P. (1986) The Family Task Interview: a tool for clinical research in family interaction. *Journal of Marital and Family Therapy*. (Submitted.)
- KINSTON, W. and LOADER, P. (1984) Eliciting whole family interaction with a standardized clinical interview. *Journal of Family Therapy*, **6**: 347-363.
- KINSTON, W., BENTOVIM, A. and LOADER, P. (1982) Constructing a focal formulation and hypothesis in family therapy. *Australian Journal of Family Therapy*, **4**: 37-50.
- KINSTON, W., LOADER, P. and MILLER, L. (1987) Quantifying the clinical assessment of family health. *Journal of Marital and Family Therapy*. (In press.)
- KINSTON, W., LOADER, P. and STRATFORD, J. (1979) Clinical assessment of family interaction: a reliability study. *Journal of Family Therapy*, **1**: 291-312.
- LEWIS, J. M., BEAVERS, W. R., GOSSETT, J. R. and PHILLIPS, V. A. (1976) *No Single Thread: Psychological Health in Family Systems*. New York. Brunner-Mazel.
- LOADER, P., KINSTON, W. and STRATFORD, J. (1980) Is there a 'psychosomatogenic' family? *Journal of Family Therapy*, **2**: 311-326.
- MINUCHIN, S. (1974) *Families and Family Therapy*. Cambridge. Harvard University Press.
- MINUCHIN, S., ROSMAN, B. L. and BAKER, L. (1978) *Psychosomatic Families: Anorexia Nervosa in Context*. Cambridge. Harvard University Press.
- OLSEN, D. H., SPRENKLE, D. H. and RUSSELL, C. S. (1979) Circumplex model of marital and family systems. 1. Cohesion and adaptability dimensions, family types and clinical applications. *Family Process*, **18**: 3-28.
- RICHMAN, N. (1977) Is a behaviour checklist for pre-school children useful? In: P. J. Graham (Ed.). *Epidemiological Approaches in Child Psychiatry*. London. Academic Press.
- RISKIN, J. (1976) 'Non-labelled' family interaction: preliminary report on a prospective study. *Family Process*, **15**: 433-439.
- RUTTER, M., TIZARD, J. and WHITMORE, K. (1970) *Education, Health and Behaviour*. London. Longmans.

- SIEGEL, S. (1956) *Nonparametric Statistics for the Behavioural Sciences*. New York. McGraw Hill.
- STRATFORD, J., BURCK, C. and KINSTON, W. (1982) Influence of context on the assessment of family interaction: a clinical study. *Journal of Family Therapy*, 4: 359-371.
- WALKER, L. G., THOMSON, N. and LINDSAY, W. R. (1984) Assessing family relationships: a multi-method, multi-situational approach. *British Journal of Psychiatry*, 144: 387-394.
- WELLS, C. F. and RABINER, E. L. (1973) The conjoint family diagnostic interview and the family index of tension. *Family Process*, 12: 127-144.