Developing an indicator of family function and a practicable outcome measure for systemic family and couple therapy: The SCORE.

Peter Stratton[1], Julia Bland[2], Emma Janes[2] and Judith Lask[3]

1. Professor of Family Therapy, Institute of Health Sciences, University of Leeds

2. Consultant Psychiatrist in Psychotherapy, Maudsley Psychotherapy Service, South London and Maudsley NHS Foundation Trust

3. Programme Leader, Family Therapy Training, Institute of Psychiatry, London

Final draft, now published at:

Developing an indicator of family function and a practicable outcome measure for systemic family and couple therapy: The SCORE.

Abstract

There is a need for a measure of outcomes in systemic family and couples therapy (SFCT) that reflects current theory and practice. To meet the needs of SFCT practice the measure needs to use self report by family members, take a short time to complete and be easy to understand. The development of such a measure, called the SCORE, is reported. Substantial piloting, consultation and review in terms of clinical judgement led to the construction of the SCORE 40 which has 40 items about how the family functions, rated by family members over 11 years of age on a Likert scale, in addition to independent ratings of the family and its difficulties. The SCORE 40 is shown to be a viable instrument but is too substantial for everyday clinical use. In a research project to reduce and refine the measure and determine its psychometric properties the SCORE 40 was administered to510 members of 228 families at the start of their first appointment for family therapy at clinics throughout the UK. The scale has good psychometric properties and could operate with either three or four dimensions. The analyses of these data, combined with data from a convenience sample of 126 non-clinic families allowed a reduction to 15 items while retaining most of the information provided by the SCORE 40. A version is offered with three dimensions of: 1. Strengths and adaptability; 2. Overwhelmed by difficulties; and 3. Disrupted communication. It is hoped that the ready availability of the SCORE 15 will encourage routine evaluation of outcomes in clinics as well as the SCORE being used flexibly for both therapy and research.

Introduction

The importance of measuring routine outcomes has been enshrined in the government document 'Organising and Delivering Psychological Therapies' (Department of Health, 2004).   The collection and analysis of such data from everyday practice is commensurate with clinical governance goals to develop and maintain standards of practice (Department of Health 1997).   Measures designed to

provide information on family relationships and functioning are not in routine clinical use within the context of family therapy clinics in the UK. This is perhaps due to various problems with applicability in service settings (e.g. excessively lengthy or expensive questionnaires), or possibly due to a lack of knowledge or general consensus on the most appropriate tools to use. Such difficulties with available instruments can compound concerns that quantitative measurement is incompatible with some of the epistemologies to which the field currently subscribes. The authors of this paper believed that it should be possible to create an indicator of the effects of family therapy that is compatible with social constructionist thinking and with current practice. We also felt that the field of family therapy was seriously disadvantaged by the fact that therapists do not routinely evaluate their own outcomes, and that research into the efficacy of SFCT is currently conducted with inadequate measures of therapeutic effects.

Approaches to measuring outcome

Therapist estimates of their own effectiveness are notoriously unreliable (Hatfield et al, 2009) and feedback obtained by a therapist asking the client is clearly subject to bias. The more objective existing measures can be grouped into three types:

1.Change in diagnosed problem.

Indicators of change in diagnostic symptoms are suitable in research where the DSM or ICD are clear and specific and define the sample. But FT clinical practice is not generally concentrated on a single diagnosis and measures of symptoms may not be capable of showing the kinds of improvements in family function that SFCT can achieve.

2. General indicator of clinical change

.A widely used example is the CORE (Clinical Outcomes in Routine Evaluation) instrument which was pioneered and developed by Barkham, Evans and colleagues (Barkham et al 1998, Evans et al 2000, Evans et al, 2002). Because the CORE is based on a concept of individual dysfunction and

designed to measure changes in individual clients, it is not likely to reflect all of the kinds of change for which SFCT is aiming.

3. <u>Measures of family functioning</u>

There are several measures which define good family functioning. Self report measures of family functioning were comprehensively reviewed early in this project (Janes, 2005). This review points to the infrequent usage of any of these measures in family therapy practice in the UK and explores a variety of the problems that they present. Subsequently, Sanderson et al (2009) reviewed 274 outcome studies between 1990 and 2005. They report that 480 outcome measures were used and only 26 of these were family system measures. Although most of the studies used more than one measure, no instrument was used in 15% of the studies. As they say "uniformity was not the norm" (p.253). There are two recent developments in family measurement. The Systemic Therapy Inventory of Change (STIC, Pinsof et al, 2009) covers adult and child individual functioning, family of origin, couple and family functioning but completing the forms takes an hour for each client before their first session. Schiepek's (2009) Synergetic Navigation System has clients providing daily ratings which are transmitted to the researchers and the therapist. These examples show how exciting the measurement of therapeutic change has become, but from the feedback we had received neither would be seen as practicable for regular use by most therapists.

We decided there was a need for a measure that would be usable in all applications of couples and family therapy and based on self report by family members. The measure should be designed to reflect the kinds of changes that SFCT is trying to achieve and convincing to external scrutiny. But it should also have a wider applicability in measuring changes in relationships following other interventions.

What is Family Therapy? The answer in the current leaflet of AFT, under the heading of **building family strengths**, is " ..Family and Systemic Psychotherapy – often called Family Therapy - helps people in close relationship help each other. It enables family members to express and explore difficult thoughts and emotions safely, to understand each other's experiences and views, appreciate

each other's needs, build on family strengths and make useful changes in their relationships and their lives."

The research team set out to create a measure that would indicate how a family conducted its life and relationships in practice. We were inspired by the example of the CORE and intended to use the way that the CORE was developed and applied as a model. We have been particularly influenced by the process of practitioner collaboration and consultation which enabled the development of this measure. However, we aimed to create a measure that was specifically sensitive to changes in how relationships are manifested and which would be in a form that clinicians could easily integrate into their practice.

A core battery has been defined by Barkham et al (1998) as 'a set of measures that are applicable to *all* patients in psychotherapy regardless of the clinical settings, mode of therapy, or specific problems (clinical population) of the patients'. The authors recommend the development of a set of extensions to the core battery to address specific problems or modes of therapy. The CORE questionnaire in its present form is clearly intended for use in therapies seeking intra-individual change and as such, would have limited value in reporting family functioning. We are not aware of any attempts to refine or adapt the CORE for use in family therapy.

The project was to create an entirely new measure, not a systemic adaptation of the CORE. This systemic measure to parallel the CORE has been called the SCORE. The basic requirements were that it was to be completed by each adult and young person in the family, and give a snapshot of the quality of family life while minimising the imposition of any normative assumptions. Secondly the authors intended to develop a measure that will be sensitive to change in family functioning over time, possibly after family or couple therapy. It would .indicate aspects of family life that therapists judged to be most relevant as targets of therapeutic intervention. It should indicate aspects of family life that could make it difficult for family members to deal with challenges to the extent that members would develop psychological problems. Finally, it would tap aspects of family functioning that would be expected to change during therapy as the family became better able to handle the difficulties in their lives.

5

Developing the SCORE 40

The SCORE was developed by a substantial team. The 'core' team throughout were Dr Julia Bland (principal investigator) ; Professor Peter Stratton, Dr Emma Janes, & Mrs Judith Lask and there was important input at different stages from Ms Annie Peppiatt, from Dr Chris Evans, Dr Anne Ward, Dr Himanshu Mistry, Dr Antonia Regojo, Dr. Nikola Kern, Dr Aspa Paspali, and Dr Malik Saouid. We are grateful to Sabine Landau (Institute of Psychiatry, London) and Wendy Harrison (Institute of Health Sciences, Leeds University) for helpful consultations on the methodology and statistics of this research.

The development of the 40 item version of SCORE was a lengthy process of reviewing relevant literature about family functioning and its measurement, carefully considering the claims in terms of the current clinical experience of the team members , circulating the progressive formulations to colleagues and testing them out in a variety of ways.

A starting point was the review of existing measures (Janes, 2005) which was combined with a review of accounts of therapeutic activity and objectives, and studies in adjacent disciplines (developmental psychology and sociology) of research-based indicators of good family functioning.

We were strongly influenced by a discussion (Cartwright, 2004) of the limitations of CORE for systemic therapy, and an accompanying list of 10 items to assess family functioning. We opted to investigate the plausibility of using a self-report measure based on descriptions of family life with responses from family members scored on a Likert scale.

A nine-item version of representative questions was posted on the AFT email discussion list and circulated to colleagues to invite comments about the feasibility of bringing therapeutic consideration into a measure of this form, and to request suggestions of further items.

Five dimensions of family functioning were derived from the reviews. These, with the summary title used during statistical analysis, were:

(1) Danger/hostility (hostil)

(2) Communication (commun)

(3) Atmosphere/mood (mood)

(4) Flexibility/adaptability (adapt)

(5) Rules/roles/individuation (roles)

A preliminary set of three items for each dimension plus an introductory item designed to be answered positively by all respondents was constructed. This 16 item version was presented to a variety of professional audiences: family and other therapists attending workshops; trainees on family therapy courses; and students attending an MSc in psychological research methods. In some cases respondents completed the SCORE for their own family, and in others, for their family of origin when they were aged 16. This included individuals from a range of cultures and ethnicities.

Feedback from these various consultations was used to refine the 16 questions to create an instrument that could realistically be used by family members aged 12 years upwards. This version was then extensively piloted to establish the viability of such a measure. In February 2005 the SCORE 16 was sent out to 24 senior family therapists who were invited to complete the SCORE 16 while making notes on a parallel version of how they went about deciding on their scoring of each question. The commentaries were returned to the research team.

Responses from the various audiences suggested that therapists were often concerned about the general principle of quantifying the state of family life at a specific time, but they found the constellation of questions to be highly informative. The detail of their responses was used for subsequent development of the SCORE as described below.

In the next study 33 trainee family therapists simulated responses of a) an adult in a family that was experiencing serious difficulty, and b) repeated the process for a person from a family that had made significant gains during therapy. The mean scores showed a clear difference between the two

somewhat stereotyped family members: 4.14 for the first and 2.62 for the second (t=10.566, df=11, p<.001). The Likert scale scores positive ratings at 1 and negative at 6 with the scoring of negative items reversed, so throughout this report a low score is a positive evaluation of the family. All of the items correlated positively with the average score, with the lowest correlation being 0.54 and the highest 0.91.

In order to anchor the development firmly in therapeutic thinking, a third, qualitative, evaluation was carried out in the form of a practitioner research network and has been reported in detail (Stratton et al, 2006). Nine experienced family therapists role played the responses of a specific person in one of the families they had seen recently that was experiencing serious difficulty, and repeated the process for a person from a family that had made significant gains during therapy. They then responded to four questions about each of the 16 items.

Their responses were recorded, transcribed, and analysed through Template Analysis (King, 1999). The study concluded that "The SCORE was found to be generally acceptable for use with clients but the exercise generated information on a variety of concerns about the content of the measure. Beyond its potential as an outcome measure there was considerable interest in how it could be used therapeutically". (Stratton et al, 2006, p. 199).

Development of the SCORE 40

This extensive piloting convincingly indicated that a self-report measure of up to 16 items could be statistically viable, informative about families, and valued by clinicians. On statistical advice we increased the variety and expression of the original 16 questions to give a pool of items from which a concise measure could be empirically derived. The full range of feedback acquired during each stage of piloting was used in conjunction with the literature reviews and discussions with therapists, to build a more comprehensive set of 55 items, 11 for each of the five dimensions. New items were constructed using alternative wordings and differences in emphasis that had been suggested for the original 16. Other novel items had been suggested by therapists and participants in the piloting or had been developed by the research team. This version was administered to a convenience sample of 69 people, only one

from any given family, with the main criterion that no member of the family had been referred for psychotherapy. The parts of this research involving non-clinical samples formed part of the research training in LFTRC and have been described by Stratton & Hanks (2008).

The non-clinical sample responding to the 55 item version generated a mean score of 2.12 (range 1.29 to 3.36, sd=0.46). It proved to be a coherent set of items with Cronbach Alpha= 0.78, split half correlation= 0.69.

Data about the performance of individual items in this non-clinical sample and the comments participants had been invited to make about which items were more, or less, informative, were combined with discussions with colleagues and with members of the Maudsley Psychotherapy Service , Service Users Group, to identify the less informative items. There was then intensive discussion of the items and their clinical relevance among the research group who are in active clinical practice with families. The 15 items judged to be least informative were removed, to produce a 40 item version.

This version, the SCORE 40, was administered to a new non-clinical sample of 57 respondents who had a mean score of 2.01 (sd=0.61). For this version the basic measure of coherence, the Cronbach Alpha, is .934, and Split-half reliability is a correlation of .833 . The reduction to 40 items therefore successfully improved the coherence and reliability of the questionnaire. In this sample respondents were asked to provide a qualitative description of their family and any difficulties it was encountering. A simple coding of this information on a 5-point scale correlated quite strongly with each person's average on the SCORE 40: Pearson correlation =0.68 (df=55, p<.01).

One refinement to the SCORE 40 was proposed during an exercise in year 1 of the Leeds MSc in systemic family therapy generating and using data from the non-clinical sample of 126 respondents. Marina Webster (2008) conducted a detailed analysis of the use of the 6-point Likert scale and demonstrated that on a number of criteria, the first two scale points: 1. That describes our family: *Extremely well*; and 2. That describes our family: *Very well* were not independently informative. In particular, some respondents used point 1 and other used point 2, and the six point scale is not evenly

weighted around a midpoint. We are grateful to Dr Webster for her detailed analyses that demonstrated that negligible information was lost by reducing the scale to five points. On statistical advice we recalibrated the data by coalescing the first two points, so the subsequent analyses are based on the 5 point scale.

Ethics of inflicting a novel measure on vulnerable clients

It was judged that the performance of the 40 item SCORE that derived from this lengthy process of development and refinement was sufficiently robust to justify its application with a clinical sample. The 40 items along with the qualitative and demographic material had been found to take significant time for most family members to complete, so a major objective was to create a version of between 12 and 15 items with well established psychometric properties and in the process, to use the variety contained in the 40 items to provide information about families who are referred for family therapy.

We wanted to ensure that the evaluation of the SCORE 40 and creation of a concise version continued to be carried out in such a way that we stayed close to clinical considerations throughout. Our statistical and methodological advice was also that the conditions of generating and analysing data should closely correspond with the ways the final versions of SCORE would be used in practice. The developmental process was therefore based on data gathered in a wide variety of family therapy clinics working generically.

Project to evaluate and refine the SCORE40

Method

The SCORE 40, consisting of 40 questions rated on a 5-point Likert scale, eight items that provide direct descriptions of the family, and six aspects of demographics, was completed by family members just before the first session. The quantitative items were explored using multivariate statistical analysis and on advice we planned for a minimum of 200 families. Resources of the Association for

Family therapy (AFT) were used to contact clinics throughout the UK to invite them to participate. The intention was to sample a wide range of clinics, both adult and child, and so recruit a heterogeneous sample of families with differing referral criteria and identified problems. The SCORE items had been constructed to be readily readable by people with normal literacy in English aged at least 12 years. It was completed by each family member independently and the analyses were conducted on the full set of individually completed questionnaires. We did not want to make assumptions about the degree of consistency within families, regarding this as an empirical question that subsequent research using the SCORE will be able to answer. As stated above, our advice was to carry out the analyses on the same data set that will be used clinically. In consultation with therapists the consensus was that in clinical practice the ratings of the questions will not be averaged, but the scores of individual family members will be used. Analyses were therefore conducted on the forms provided by individuals.

Recruitment of clinics and ethics

Full consideration was given to potential ethical issues. In particular we were aware that items provided to families in this way could carry messages about the expectations of therapists, and this concern was salient throughout the development of the items. We were also concerned about the demands on therapists and families and while these were kept to a minimum, a major objective of the research was to create a shorter version in which any items that might cause distress to family members had been identified and eliminated. The project was submitted to the South London & Maudsley NHS Trust Research Project for Research and Development Approval by Julia Bland as Principle Investigator. Once approved, it was submitted for national ethical approval through Charing Cross Research ethics. Approval was given by COREC, (subsequently NRES and now 'IRAS': integrated research application system). Participating clinics were provided with the forms and procedures for obtaining local ethical approval and were extensively supported throughout this process by Emma Janes. Participating families were given advance notice of the invitation to participate and clinics were provided with letters of invitation and consent forms.

Procedure

Data collection at multiple sites commenced in March 2006. Each clinic undertook to follow the research protocol and in particular to offer the SCORE to every family that attended for a first appointment during the period of data collection. A principal investigator was appointed at each clinic who took responsibility for gaining local ethical approval and for the conduct of the research in accordance with the protocol. A log was to be kept of families who did not complete the SCORE (choosing not to participate, for any reason e.g. limited time, limited knowledge of English, reading difficulties or reasons of therapy or administration). This was to ensure that, as far as practicable, results were not affected by any form of selection of families. Each family had been referred for family therapy and completed the SCORE 40 before the start of their first session.

Each family member completed the SCORE privately, in the presence of a member of the therapy team or an administrator. For the purposes of the research, each form was coded so that anonymity could be retained, and posted to the research team. It was left to each clinic whether they would negotiate complete anonymity or would agree with the family that their forms could be copied and used within the therapy. However in all cases the family members were assured that their responses would remain confidential and not be shared with other members of the family.

Samples

When data collection closed and analysis started, 15 clinics had supplied data from a total of 228 families with 510 individual SCORES completed.. Numbers of cases from each clinic varied widely and a substantial proportion came from the Maudsley clinics. Because of the types of referral to the Maudsley there is a substantial number of cases of couples therapy in the sample. There are some cases of systemic therapy with individuals and families in which all of the children are under 12 years who would therefore only provide data from the adults. Of the 510 individual SCOREs, age was indicated in 497 cases. 82 (16.5%) were aged less than 20 years. 41.6% were male and 58.4% female.

Data were also available from a convenience sample of 126 non-clinical cases. These data, as described above, had not been collected under such rigorously controlled conditions. However, they offer a useful comparison to the clinical data and a preliminary indication of the ability of items to discriminate between families in the general population and those referred for family therapy. They are included in the analyses where appropriate on the rationale that the SCORE is not attempting to define an optimal family operation. Rather, that the types of response of an unselected sample of non-clinical families is a reasonable indication of the kinds of response that we would hope families would provide after successful therapy.

Data processing

The data from each form were recorded in an SPSS file with columns for the text of the qualitative responses and for the demographic information. A random sample of 80 cases was re-recorded using a different system and no discrepancies were found. The qualitative items (See Appendix 1) were each coded on a five-point scale to indicate the perception by that respondent of their family and the extent of its problems. The description of the family could be coded with good reliability (as '*family rating*' in Table 1 below) but the invitation to specify the 'biggest challenge' did not generate data that could be reliably discriminated and so data from this item were not used. Ratings of the severity of the main problem and of its effect on the family on the 10 cm analogue scales were quantified. On the basis that the problem and the effect on the family compound rather than add to each other, a single measure ('*problem rating*' in Table 1 below) was obtained by multiplying the two scores.

In 28 of the 510 questionnaires four or more of the 40 items had not been rated. These cases were excluded from the analyses since averages and other statistics could be biased if 10% or more of the items had been ignored for a particular reason. Analyses are therefore reported on 482 cases.

Statistics of individual items, their correlation with the average of the other 39 items, and their correlation with the two measures that were provided independently of the 40 questions, were computed. The correlations of each item with average SCORE for the non-clinical population was

also computed for comparison. This analysis by individual item played an important role in the subsequent process to reduce the number of questions for practicable clinical usage.

Coherence of the SCORE 40.

The standard measure of whether the items in the scale are coherent with each other is Cronbach's alpha. In the clinical sample alpha was 0.93 and in the non-clinical sample, 0.90. these indicate high levels of coherence across the 40 items. An initial test of reliability is provided by correlating the score derived from half of the items with the score s from the other half. Split-half reliability was 0.84 for clinical and 0.82 for non-clinical with Guttman's lower bounds for true reliability 0.92 for clinical and 0.90 for non-clinical. These indicate high reliability within a single completion of the SCORE.

Performance of the SCORE 40

The distribution of the average scores across the 482 participants in the clinical sample and the 126 non-clinical, on the 5-point scale, were compared. The two sets of scores were not compared statistically because the samples were not matched. But comparisons of the averages and of individual items are indicative of the functioning of the scale. The mean for the clinical sample at 2.58 (sd=0.64) is towards the mid-point of the scale while the mean for the non-clinical sample of 1.75 (sd=0.45) is closer to agreement with the positive end of the scale.

Performance of the individual questions was examined and is presented in Table 1.

Column 1 lists the 40 questions as presented in the SCORE40 (Appendix 1). The subtitle indicates which of the five *a priori* dimensions (see above) that item was constructed to indicate.

Column 2 shows the number of each question that was not scored in the full sample of 510 participants. It is clear that none of the items was particularly avoided.

Columns 3 give the means and standard deviations for the 482 people who scored more than 90% of the items. Q1 was designed to give most family members a chance to start with a positive answer and it has the most positive mean. In all, only four items averaged below 2, the second point on the scale,

and two of these were the only items with a standard deviation below 1. We concluded that almost all of the questions were capable of indicating a degree of difficulty in these families.

Column 4 gives the correlation of each item with the average of the remaining 39 items. Every item correlated significantly with the SCORE average with all but one with $p < .01$ (two-tailed). Q39 correlated with $p < .05$.

Columns 5 and 6 show the correlations of each person's SCORE 40 average with two indicators provided by that person independently of the SCORE questions. These comparisons are particularly useful because they show whether the SCORE corresponds to overall descriptions and evaluations in forms completely different to their responses on the Likert scale. No question generated a negative correlation with either item. Thirty-four items correlated strongly ($p < .01$) with both measures; three (Qs 1,21, and 39) correlated with $p < .05$, and three (Qs 16, 27, and 35) did not correlate. This is strong validation of the SCORE 40 but also provided information on which weaker items to eliminate. The final two columns provide the means and correlations of each item with the total for the non-clinical sample. These data were also used to reduce the number of items in SCORE when the clinical average was similar to the non-clinical average.

Insert Table 1 about here

Factor analyses were carried out to determine the structure of the SCORE 40. Principle components analysis was used with Varimax rotation which allocates the variables (items in the SCORE) in such a way as to keep the factors as distinct as possible. The basic analysis identified nine factors but the scree plot indicated that the variance explained levelled off after three factors, and only the first four factors had more than two items with coefficients greater than 0.4. The analyses were repeated forcing a 3 factor and then a 4 factor solution.

The character of each factor can be indicated by the strongest questions in each case. In both 3 and 4 factor solutions the first two factors were:

Factor 1: In my family we talk to each other about things that matter to us (30); Each of us gets listened to in our family (14); We are good at finding new ways to deal with things that are difficult (31).

Factor 2: We seem to go from one crisis to another in my family (38); Life in our family is very difficult (13); Things always seem to go wrong for my family (26).

In the three factor solution **Factor 3** included: People in our family lie to each other (24); People often don't tell each other the truth in my family (8); People hit each other a lot in my family (28); One person tends to get blamed for everything in my family (7).

In the 4 factor solution, **Factor 3** included: People often don't tell each other the truth in my family (8); When people in my family get angry they ignore each other on purpose(12); People in our family lie to each other (24);

And **Factor 4** which only included four items consisted of: People hit each other a lot in my family (28); People slam doors, throw things or make a lot of noise if they are upset (36); People in the family are nasty to each other (32) and My family is very strict (39).

Although each SCORE questionnaire was completed privately and without discussion, there is a possibility that the inclusion of members of the same family could distort the analysis as they are not strictly independent data. Therefore the factor analysis was repeated with one adult member randomly selected from each family. This is the procedure followed by Pinsof et al (2009). 211 cases were selected (101 men and 110 women). The 3 Factor solution was almost identical to that generated by the full sample of 482 SCORES. The first 12 variables, having coefficients >0.5, on Factor 1 are identical though with slightly different weightings. In Factor 2 the eight variables with coefficient > 0.4 in the full sample were present at the same level in the selected sample, with the addition of Q19 'Other people look down on my family because we are different'. For Factor 3, the eight variables

with coefficients > 0.4 were present in Factor 3 of the selected sample, which also had Q29, which was present at a lower level in the analysis of the full sample.

In the multiple regression (reported below) the full sample multiple regression coefficient after inclusion of 30 items was 0.995, and with the sample of one adult from each family was 0.996. Of the ten last items for the full sample, 8 were the same. In the sub-set of one adult from each family, Q39 'My family is very strict' came earlier at step 24, and was replaced in the last 10 by Q14 'Each of us gets listened to in our family' which had been the 11th least significant in the full sample.

As the differences between the full clinical sample and the sample with only one adult selected from each family are minimal, it is clear that there was no spurious inflating nor significant distortion arising from treating each family member's SCORE as a separate data point.

Acceptability

There were several indicators of the acceptability of the SCORE 40. Table 1 shows that no question was consistently avoided by any of the respondents. They were also invited to indicate which questions were especially informative and which were uninformative (see the SCORE 40 in Appendix 1 for the full wording). At an interim point when 388 SCORES had been completed, each question had been rated informative on average 16 times and uninformative 3 times. So in general the questions were seen as much more often informative. There was considerable variability in which questions were selected for comment and these variations were used in the final selection of items.

Some of the clinics that had initially volunteered to participate found that therapists were unwilling to have the families complete such a lengthy instrument and so did not provide data. Of those that accepted the protocol, some did not report how many families chose not to provide data, but in those that did, less than 5% declined.

We concluded that every question in the SCORE 40 was viable as an indicator of family functioning. All of them related significantly to different measures and elicited variability in responses with none being answered at an extreme by all respondents. We concluded that the SCORE 40 is a viable

instrument with clear psychometric properties. It offers interpretation with dimensions of: I strength and adaptability; II overwhelmed by difficulty; and III disrupted communication. With the possibility of separating out a fourth dimension of hostility and aggression.

Although the SCORE 40 is an effective and acceptable instrument, in this form it takes too long to be viable for everyday clinical use. We therefore turn to the process of investigating whether we can substantially reduce the length while retaining its functionality.

Refining the measure: towards the SCORE 15

Identification of specific items for removal.

The SCORE team engaged in a substantial process to select items and a structure for a short version of SCORE. The process was continually guided by considerations of therapeutic relevance. It had the following stages:

Multiple regression with stepwise inclusion was carried out with the mean on SCORE 40 as criterion variable. The first 30 items to be included gave a multiple regression coefficient of .995, accounting for 99.1% of the variance in the mean score. Items that appear after this point can be expected to have nearly all of the information that they provide already provided by previous items. The last 10 items (in descending order of significance: Q1adapt, q36commn, q31adapt, q17commn, q21roles, q39roles, q15adapt, q24commn, q13mood, Q2mood), were examined for clinical significance and it was decided to delete them. Only Q36 had been a strong component in the factor analyses but it was felt that in the culture of some families, slamming doors and making a lot of noise are seen as a legitimate indication of being upset.

Two more items were deleted on the basis of content. Q7 (one person gets blamed for everything) and Q25 (we blame each other) correlate highly and both concern blaming. Q25 came earlier in the multiple regression, had a higher correlation with the mean score and a bigger difference between its averages for the clinical and the non-clinical samples. So Q7 was deleted on the basis of duplication.

18

Q28 (People hit each other a lot) raised two concerns. First that people might not feel able to be truthful in answering it, and also that providing an extreme answer might raise expectation that action would be taken, whereas the anonymity would prevent this.

A further four questions: Q16, Q19, Q27 and Q35 were examined as each had, as shown in Table 1, low correlations with the four measures of mean score, problem rating, family rating, and mean score for non-clinical, and also were among the seven items with a small difference between clinical and non-clinical means. In a factor analysis of 28 items, after the first 12 items identified above had been removed, none of these 4 correlated strongly with any of the factors. They were therefore removed from the data set and factor analysis was run on the remaining 24 items.

Factor analysis with Varimax rotation was now run to generate four factors. However the outcome was unhelpful in terms of being a basis for final selection of items, as Factor 4 had only two items with strong coefficients and the next also loaded on Factor two. The three factor solution met statistical requirements (Kaiser-Meyer-Olkin Measure of Sampling Adequacy=0.929, Bartlett's Test of Sphericity, Approx. Chi-Square=4093, df=300, p<<.001) and generated three clear dimensions of the SCORE. Factor loadings were considered as the basis for a version of SCORE in terms of the original objective of 12-15 items. The weightings in the factor analysis, and avoidance of items that appeared in more than one factor were a starting point. But we were also concerned to consider each item in terms of clinical implications, greatest likelihood of showing therapeutic change, and ensuring that the final version represented all the main features of the SCORE 40. Final factor analyses and multiple regressions were run on the data set that combined the clinical and non-clinical samples. Total number of respondents 608.

Taking all of these considerations into account, we concluded that three factors with 5 questions for each would cover the range of issues, with reasonable robustness of the dimensions. Factor analysis of the 15 final items gave a 3-factor solution that is shown in Table 2. The analysis fully meets criteria: Kaiser-Meyer-Olkin Measure of Sampling Adequacy=0.92, Bartlett's Test of Sphericity, Approx. Chi-

Square=3271, df=105, p<<.001; Cronbach's alpha 0.89, split-half correlation 0.81; Guttman split-half coefficient 0.89.

Insert Table 2 about here

Our specification of items for the SCORE 15 is:

Factor 1. Strengths and adaptability

In my family we talk to each other about things that matter to us (30)

We are good at finding new ways to deal with things that are difficult (31)

When one of us is upset they get looked after within the family (20)

Each of us gets listened to in our family (14)

We trust each other (33)

Factor2. Overwhelmed by difficulties

We seem to go from one crisis to another in my family (38)

Things always seem to go wrong for my family (26)

In my family we blame each other when things go wrong (25).

It feels miserable in our family (22)

We find it hard to deal with everyday problems (11)

Factor 3. Disrupted communication

People often don't tell each other the truth in my family (8)

It feels risky to disagree in our family (4)

People in my family interfere too much in each other's lives (5)

People in the family are nasty to each other (32)

When people in my family get angry they ignore each other on purpose (12)

Considering the five dimensions that we originally wanted to cover, the three factors of the SCORE 15 were constructed of:

Factor 1: 3 communication, 1 mood, 1 adaptability

Factor 2: 2 mood, 2 adaptability, 1 hostility

Factor 3: 3 hostility, 1 communication, 1 roles

So all of the original dimensions are represented. No attempt had been made to select on the basis of the dimensions as they had been used to generate the variety of 55 initial items and after this, we wanted to be guided by clinical judgement applied to the empirically derived responses.

As a final check on how well these 15 items represent the original SCORE 40 a Multiple Regression with SCORE40 mean as criterion, on the full sample of 608 cases gave a multiple regression coefficient of 0.975, explaining 95% of the variance in the means of the 40 items.

Conclusion and Next Stages

The SCORE 15 is the product of an extensive process of development, has excellent psychometric properties and a dimensional structure that is clinically interpretable. The authors hope that this measure will prove acceptable and user friendly, both for family members attending clinics who may be feeling distressed and exposed to scrutiny by being asked to complete even a brief measure, and to busy clinicians who hesitate to demand form filling from vulnerable families but who also recognise the need for evidence based practice and empirical self scrutiny by therapists.

The SCORE 15 will be used in Stage 2 of the project to determine its criterion validity in terms of corresponding to therapeutic change. Concurrent validity will also be assessed where other measures are applied to the same families. We will also complete a formal assessment of the SCORE 15, in a

non-clinical sample to assess the measure's capacity to differentiate adequately between a clinical and non-clinical population, i.e. to detect 'caseness' , as CORE has successfully achieved, and to check the test-retest reliability.

The brevity and clear structure of the SCORE 15 lends itself to both research and clinical applications.

Some Research possibilities

- Generating an evidence base appropriate to relational therapies

- Examining effects of therapy with:

    - Different lengths

    - Different client groups

    - Different approaches

- Collaboration across countries. The research committees of the European Family Therapy Association are supporting a project involving translation of the SCORE 15 in 12 European countries, and English language collaboration is already under way in Ireland and Australia.

- A national data-base

- Exploration of alternative subsets of items from the SCORE 40 for example the items relating to risk and hostility.

- Practitioner Networks of researchers

- Exploring cultural differences between families

Potential Clinical uses

- Pre-therapy information

- Discussing the items that are significant for clients

- Indicating major areas of change, and of no change, between sessions

- A context for discussions of usefulness

- Using the items to alert family members to disregarded aspects

- Checking for difference between therapist and client perceptions

The SCORE 15, current format available from the website of the Association for Family Therapy, is an effective indicator of family functioning, and designed to be sensitive to the kinds of change to which SFCT is directed. It is designed to be used routinely in clinical practice and we hope it will make a significant contribution towards outcomes evaluation becoming routine.

References

Association for Family Therapy www.aft.org.uk

Barkham, M; Evans C; Margison, M; McGrath, G; Mellor-Clarke, J; Milne, D; Connel, J. (1998) The rationale for Developing and Implementing Core Outcome Batteries for Routine use in Service Settings and Psychotherapy Outcome Research. *Journal of Mental Health,* 7 (1): 35-47

Campbell. D (2001) The other side of the story: The client's experience of therapy. In (Ed) Papadopoulos.R, K. and Byng-Hall. J. 2001 Multiple Voices. Narrative therapy in systemic family psychotherapy. London. Karnac.

Cartwright, T. (2004) CORE and the politics of evidence. Posting on aft@psyctc.org on 04-04-04.

Department of Health (1997) *The New NHS: Modern, Dependable*. London: HMSO

Department of Health (2001) *Treatment Choice in Psychological Therapies and Counselling: Evidence Based Clinical Practice Guideline.* London: HMSO

Department of Health(2004) *Organising and Delivering Psychological Therapies*. London: HMSO,

Evans, C; Mellor-Clark, J; Margison, F; Barkham, M; Audin, K; Connell, J; McGrath, G. (2000) Core: Clinical Outcomes and Routine Evaluation. *Journal of Mental Health*, 9 (3): 247-255

Evans, C; Connell, J; Barkham, M; Margison, F; McGrath, G; Mellor-Clark, J; Audin, K. (2002) 'Towards a standardised brief outcome Measure: Psychometric Properties and Utility of the CORE-OM'. *British Journal of Psychiatry*, 180: 51-60

Hatfield, D., McCullough, L., Frantz, S., & Krieger, K. (2009). Do we know when our clients get worse? An investigation of therapists' ability to detect negative client change. Clinical Psychology & Psychotherapy

Janes, E. (2005) Self Report Measures of Family Function & Change Following Family Therapy: a Review of Conceptual Issues, Existing Measures and Proposals for Improvement. Available at: www.psyc.leeds.ac.uk/staff/p.m.stratton/

King, N. (1999) Template analysis. In Symon, G & Cassel, C. (Eds), Qualitative methods and analysis in organisational research, (pp. 118-134). London: Sage.

http://www.hud.ac.uk/hhs/research/template_analysis/

Pinsof, W.M. et al. (2009) Laying the foundations for progress research in family, couple, and individual therapy: The development and psychometric features of the initial systemic inventory of change. *Psychotherapy Research*, 19, 143-156.

Sanderson, J, Kosutic, I, Garcia, M, Melendez, T, Donoghue, J, Perumbilly, S, Franzen, C, & Anderson, S. (2009) The Measurement of Outcome Variables in Couple and Family Therapy Research. *The American Journal of Family Therapy*,37:3,239 — 257.

Stratton, P., McGovern, M., Wetherell, A. & Farrington, C. (2006) Family therapy practitioners researching the reactions of practitioners to an outcome measure. *Australian and New Zealand Journal of Family Therapy*. 27 pp. 199–207

Stratton, P. & Hanks, H. (2008) From therapeutic skills to research competence: Making use of common ground. *Human Systems*. 19. 153-171

Webster, M. (2008) Research Report. LFTRC, University of Leeds.

Tables

Table 1 statistics for individual items
     Clinical N =482 non-clinical N=126

Descriptive Statistics				Corrected item-total correlation	Problem rating correlation	Family rating correlation	Non-clinical item-total correlation	Non-clinical means
	N of 510 missing	Mean	Std. Deviation					
Q1adapt	19	1.18	.517	.318	.100	.301	.386	1.063
Q2mood	2	1.73	.945	.530	.263	.413	.462	1.230
Q3adapt	7	2.52	1.198	.486	.271	.427	.473	1.993
Q4hostil	9	2.82	1.419	.497	.205	.375	.501	1.952
q5roles	13	2.77	1.318	.349	.171	.256	.458	1.976
q6mood	15	2.10	1.155	.543	.413	.556	.418	1.341
q7hostil	10	2.88	1.438	.546	.228	.314	.536	1.960
q8commun	15	2.97	1.361	.424	.231	.238	.343	2.087
q9adapt	12	2.57	1.212	.570	.266	.433	.491	1.743
q10roles	8	2.13	1.164	.585	.199	.479	.422	1.603
q11adapt	12	3.22	1.331	.504	.245	.311	.539	1.824
q12hostl	4	3.28	1.323	.469	.232	.222	.419	2.297
q13mood	8	3.21	1.356	.626	.425	.408	.538	1.635
q14commn	12	2.57	1.227	.599	.211	.465	.504	1.675
q15adapt	10	2.64	1.158	.475	.204	.365	.351	2.137
q16roles	13	2.63	1.445	.245	.073	.009	.158	2.012
q17commn	11	2.00	1.128	.573	.215	.424	.535	1.453
q18commn	10	3.36	1.291	.384	.177	.325	.450	2.429
q19mood	8	2.02	1.292	.274	.200	.135	.447	1.492

q20mood	7	2.06	1.153	.541	.276	.398	.447	1.484
q21roles	15	2.00	1.215	.328	.150	.314	.110	1.663
q22mood	14	2.94	1.390	.646	.432	.513	.659	1.325
q23hostl	13	2.22	1.434	.516	.374	.419	.416	1.286
q24commn	17	2.47	1.243	.486	.271	.298	.645	1.760
q25hostl	14	3.15	1.313	.591	.192	.423	.626	2.088
q26mood	17	3.03	1.344	.542	.406	.418	.442	1.704
q27hostl	11	3.06	1.313	.264	.027	.257	.267	2.444
q28hostl	13	1.73	1.150	.297	.227	.261	.312	1.143
q29adapt	18	2.85	1.331	.479	.242	.213	.445	1.967
q30commn	14	2.37	1.247	.551	.250	.467	.554	1.587
q31adapt	17	2.95	1.262	.522	.211	.477	.546	1.991
q32hostl	10	2.64	1.350	.625	.318	.478	.616	1.595
q33commn	11	2.12	1.232	.626	.307	.521	.642	1.256
q34roles	15	3.12	1.295	.422	.172	.280	.228	2.183
q35roles	11	1.87	1.125	.279	.072	.157	.254	1.222
q36commn	14	3.48	1.436	.425	.270	.449	.466	2.325
q37roles	16	2.89	1.417	.323	.187	.325	.317	2.095
q38adapt	16	3.30	1.361	.543	.377	.475	.341	1.530
q39roles	14	2.30	1.176	.109	.077	-.006	.058	2.179
q40mood	11	2.34	1.312	.550	.368	.437	.446	1.347

Table 2. Factor analysis with Varimax rotation for the final set of 3 factors. Coefficients > 0.4.

Rotated Component Matrix[a]

	Component		
	1	2	3
q30commn	.828		
q31adapt	.728		
q20mood	.714		
q14commn	.703		
q33commn	.696		
q38adapt		.824	
q26mood		.788	
q25hostl		.577	
q22mood		.560	
q11adapt		.551	
q32hostl		.476	.441
q8commun			.668
q5roles			.660
Q4hostil			.649
q12hostl			.434

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 4 iterations.

Figures

Appendix 1. The SCORE 40

Site Code   Family Number   Member code

Describing your family   (date)

We would like you to tell us about how you see your family at the moment. So we are asking for YOUR view of your family.
When people say 'your family' they often mean the people who live in your house. **But we want you to choose who you want to count as the family you are going to describe.**

All the questions are answered the same way: you put a tick ✓ in the box which best matches how you see your family. So if a statement was:
"Our family wants to stay together" and you really feel this fits you completely, you would put a tick in box 1 on that line for "extremely well".

✓					

If a statement was "We are always fighting each other" and you felt this was not especially true of your family, you would put a tick in box 5 for "not well".

				✓	

For each item, make your choice by putting ☑ in just one of the boxes numbered 1 to 6. Do not think for too long about any question, it is how they all add up that we will be interested in, rather than any specific answers. But do try to tick one of the boxes for each question.

For each line, would you say: 1. That describes our family: Extremely well 2. That describes our family: Very well 3. That describes our family: Well 4. That describes our family: A bit 5. That describes our family: Not well 6. That describes our family: Not at all	1. Extremely well	2. Very well	3. Well	4. A bit	5. Not well	6. Not at all
1) Being in this family is important to us						
2) People do things that show that they care about each other in my family						
3) We are a very organised family						
4) It feels risky to disagree in our family						
5) People in my family interfere too much in each other's lives						
6) Our family shares enjoyable times together						
7) One person tends to get blamed for everything in my family						
8) People often don't tell each other the truth in my family						
9) If something is going wrong in our family we know we can change it						
10) The rules are fair in our family						
11) We find it hard to deal with everyday problems						
12) When people in my family get angry they ignore each other on purpose						
13) Life in our family is very difficult.						
14) Each of us gets listened to in our family						

For each line, would you say: 1. That describes our family: Extremely well 2. That describes our family: Very well 3. That describes our family: Well 4. That describes our family: A bit 5. That describes our family: Not well 6. That describes our family: Not at all	1. Extremely well	2. Very well	3. Well	4. A bit	5. Not well	6. Not at all
15) People in my family are willing to change their views about things						
16) There are no rules in my family						
17) In our family it is OK to show how you feel						
18) In my family people prefer to watch TV than to spend time with each other						
19) Other people look down on my family because we are different						
20) When one of us is upset they get looked after within the family						
21) Respecting elders is important in our family						
22) It feels miserable in our family						
23) Being with some family members can be frightening						
24) People in our family lie to each other						
25) In my family we blame each other when things go wrong						
26) Things always seem to go wrong for my family						
27) We hardly ever put each other down in my family						
28) People hit each other a lot in my family						
29) In my family we ignore our problems in the hope that they will go away						
30) In my family we talk to each other about the things that matter to us						
31) We are good at finding new ways to deal with things that are difficult						
32) People in the family are nasty to each other						
33) We trust each other						
34) We get into a muddle about who should do what						
35) In my family it's OK to spend time on your own if you want to						
36) People slam doors, throw things or make a lot of noise if they are upset						
37) My family feels part of a wider community						
38) We seem to go from one crisis to another in my family						
39) My family is very strict						
40) We feel hopeful about the future						

You may feel that some questions/answers were particularly informative about your family. If so, would you please tell us the numbers of those items……

You may feel that some questions were difficult to answer informatively (e.g. confusing, badly phrased, unacceptable). If so, would you please tell us the numbers of those items…..

Now please turn over and tell us a bit more about your family.

Page 3(a) Pre-therapy version

41) What words would best describe your family?

……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………

42) What do you think is the biggest problem/challenge for the family at the moment? Please name it and mark on the line how bad it is now and how difficult it is for your family to cope with:

The main problem is………………………………………………………………………

It is now:

no problem at all	really awful
←————————→	

It doesn't affect us much	totally spoils our family life
←————————→	

43) Can you say what change you would most hope for in the near future?

I would like to see a change in...

44) Do you think that family therapy will be the right kind of approach for the problems that you have?   Please mark your answer on the line below:

totally wrong	exactly right
←————————→	

45) Can you tell us how you felt about filling in this questionnaire?

THANKYOU FOR YOUR TIME