Hamilton, E., Carr, A., Cahill, P., Cassells, C. & Hartnett, D. (2015). Psychometric properties and responsiveness to change of 15 and 28 item versions of the SCORE: A family assessment questionnaire. *Family Process*, 54, 454–463.

PSYCHOMETRIC PROPERTIES AND RESPONSIVNESS TO CHANGE OF 15 AND 28 ITEM VERSIONS OF THE SCORE: A FAMILY ASSESSMENT QUESTIONNAIRE.

Elena Hamilton[1]

Alan Carr[1,2]

Paul Cahill[1,3]

Ciara Cassells[1,3]

Dan Hartnett[1]

[1] **School of Psychology, University College Dublin.**

[2] **Clanwilliam institute, Dublin**

[3] **Health Service Executive, Ireland**

Submitted in June 2014, revised and resubmitted in October 2014 to: Jay Lebow, Editor of Family Process,

Correspondence address: Alan Carr, Professor of Clinical Psychology, School of Psychology, Newman Building, University College Dublin, Belfield, Dublin 4, Ireland. E. alan.carr@ucd.ie. P. +353-1-716-8740

Running head: SCORE

Citation. Hamilton, E., Carr, A., Cahill, P., Cassells, C. & Hartnett, D. (2015). Psychometric properties and responsiveness to change of 15 and 28 item versions of the SCORE: A family assessment questionnaire. *Family Process*, 54, 454–463.

Emails: Elena Hamilton <Elena.hamilton@ucdconnect.ie>, Paul Cahill <cahillpaul@gmail.com>, Ciara Cassells <ciaracassells@gmail.com>, Dan Hartnett <daniel.hartnett@ucdconnect.ie>

Acknowledgement. This research project was supported by grants from the following agencies: The Daughters of Charity Child and Family Service, Crosscare Teen Counseling and Archways Families First.

Abstract

The SCORE (Systemic Clinical Outcome and Routine Evaluation) is a 40-item questionnaire for completion by family members 12 years and older to assess outcome in systemic therapy. This study aimed to investigate psychometric properties of two short versions of the SCORE and their responsiveness to therapeutic change. Data were collected at 19 centers from 701 families at baseline and from 433 of these three to five months later. Results confirmed the three-factor structure (strengths, difficulties, and communication) of the 15 and 28-item versions of the SCORE. Both instruments had good internal consistency and test-re-test reliability. They also showed construct and criterion validity, correlating with measures of parent, child and family adjustment, and discriminating between clinical and non-clinical cases. Total and factor scales of the SCORE-15 and 28 were responsive to change over three to five months of therapy. The SCORE-15 and SCORE-28 are brief psychometrically robust family assessment instruments which may be used to evaluate systemic therapy.

Introduction

The SCORE (Systemic Clinical Outcome and Routine Evaluation) is a questionnaire for completion by family members 12 years and older to assess outcome in systemic family therapy and other family-orientated services (Stratton et al., 2010). The SCORE was developed to address the need for a self-report family assessment measure for routine use in clinical practice in the UK and Ireland to periodically assess families engaged in therapy. It was intended that SCORE data could be used to inform therapists about the progress of individual cases and to inform managers and funders about service effectiveness.

A 40-item version of the SCORE was developed through a process of expert practitioner consultation (Stratton et al., 2006, 2010). Using multisite data from large samples of cases and multivariate statistical methods a brief 15-item version (SCORE-15) was developed by in the UK (Stratton et al., 2010) and a 28-item version (SCORE-28) was developed in Ireland (Cahill et al., 2010). In these independent research programs, both short versions of the SCORE were found to have a three-factor structure. The factors assessed family strengths, difficulties and communication. Both short forms of the SCORE showed acceptable levels of internal consistency and test-retest reliability. Jewell et al. (2013) development a child version of the SCORE-15 using focus groups and practitioner collaboration. In a non-clinical sample of school children aged 7 – 11 years the Child SCORE showed an acceptable levels of internal consistency and test-re-test reliability. On all versions of the SCORE responses are given on five or six-point Likert scales to a series of statements about family life. There is also an open-ended question about the main problem. Severity and impact ratings of this problems are given on 10-point scales.

In a national telephone survey Fay et al. (2013) established norms for the SCORE-28 and SCORE-15 in an Irish context. Clinical cut-off points for families of children with clinically significant behavioral and emotional problems were identified and 90th percentile

points were also established. The criterion validity of 15 and 28 item short versions or the SCORE were also established by showing that they discriminated between cases with, and without clinically significant problems.

Cahill et al. (2010) and Fay et al. (2013) found that both the 28 and 15-item versions of the SCORE showed construct validity insofar as their total and three factor subscales correlated significantly with other indices of family adjustment and measures of parent and child adjustment.

In a large multisite study, where family members completed the SCORE-15 before therapy, and again after 4 sessions, Stratton et al. (2014) found that the SCORE-15 was sensitive to therapeutic change in the early stage of family therapy.

The study described in this paper aimed to (1) confirm the three-factor structure, reliability, construct and criterion validity of the SCORE-15 and 28, and (2) further investigate the responsiveness of the SCORE-15 and 28 to treatment-related improvement following 4 months of systemic therapy.

Method

Ethics

The study received ethical approval from University College Dublin Human Research Ethics Committee and participants gave informed consent.

Multisite context

This multisite study was conducted at 19 centers associated with 3 clinical agencies offering family therapy and other systemic interventions to socially disadvantaged families in Dublin, Ireland. Twelve centers, which included 8 family counseling units and 4 early childhood development preschool services (ECDS), were managed by the first agency.

The family counseling units provide services to families of children aged 4-18 years with behavioral and emotional problems and in which there are a range of concerns including parenting difficulties, family conflict, parental separation and bereavement. ECDS centers offer preschool placements for disadvantaged children at risk for educational problems or with developmental difficulties, such as speech and language delay. Data were collected by EH from 347 families attending 8 family counseling centers and from 145 families attending 4 ECDS centers. Six centers were managed by the second agency which offers the Positive Systemic Practice (PSP, Carr et al., 2013; Cassells et al., 2014) model of family therapy to families of adolescents referred for help with behavioral and emotional problems. Data were collected by CC from 120 families as part of a waiting-list-control trial of PSP. One center was managed by the third agency which provides Functional Family Therapy (FFT, Graham et al., 2014; Sexton, 2011) to families of adolescents presenting with behavioral and emotional problems. Data were collected by DH from 89 families as part of a waiting-list-control trial of FFT.

Design

A cross-sectional design was used in the first part of the study to investigate the factor structure, internal consistency reliability, construct validity and criterion validity of the SCORE-15 and 28. For this, there were data from 701 cases from the three participating agencies. In the second part of the study a longitudinal design was used to determine the responsiveness of the SCORE-15 and 28 to therapeutic change and the test-retest reliability of these instruments. For this, there were data from 433 cases collected at baseline (Time 1) and 3-5 months later (Time 2). Cases who attended the ECDS centers in the first agency and cases on the waiting lists of the other two agencies who did not receive family therapy during the course of the study were classified as untreated. Of the

433 cases with Time 2 data, 175 were untreated, and provided data to evaluate the test-retest reliability of the SCORE. Data from the remaining 258 treated cases were used to evaluate the responsiveness of the SCORE to systemic therapy.

Participants

701 families participated in the study and mothers from these families were the principal informants in 88% of cases. 32% were lone parents. 44% were unemployed. 24% had one child. 40% had 2 children, and 36% had 3 or more children. 53% of children for whom help was sought were female and 47% were male. The mean age of these children was 10.06 years (SD = 5.00).

Instruments and Procedures

Respondents completed a 29-item version of the SCORE which incorporated items from the SCORE-15 and SCORE-28 (see Table 1). They also completed the Strengths and Difficulties Questionnaire (SDQ, Goodman, 2001) to assess adjustment of the child for whom they had sought clinical services and the Mental Health Inventory (MHI-5, Berwick et al., 1991) to assess parental adjustment. Therapists rated family adjustment with the Global Assessment of Relational Functioning scale (GARF, Yingling et al., 1998), child adjustment with the Children's Global Assessment Scale (CGAS, Shaffer et al., 1983), and parental adjustment with the Global Assessment of Functioning scale (GAF, Luborosky, 1962).

Data analysis

Data analysis was carried out using Version 20 of the Statistical Package for the Social Sciences and Version 7 of Mplus (Muthén & Muthén, 1998 – 2012). The false discovery

rate to control for type 1 error associated with conducting multiple statistical tests was used in analyses where multiple t-test or correlations were conducted (Benjamini & Hochberg, 1995).

Results

Factor structure

Confirmatory factor analyses (CFA) supported the three-factor structure of both the 15 and 28-item versions of the SCORE. Figure 1 presents the results of CFAs for the SCORE-15 and SCORE-28, and Chi square, Root-Mean-Square Error of Approximation (RMSEA), Comparative Fit Index (CFI), Tucker-Lewis Fit Index (TLI) and the Standardised Root-Mean-Square Residual (SRMR) which were used to evaluate whether data fit the SCORE-15 and SCORE-28 three-factor models (Bollen & Long, 1993; Kline, 2010). Both the SCORE-15 and SCORE-28 three-factor models fit the data from the present study very well. In each case the CFI was above the required threshold of 0.90; the TLI was above the required threshold of 0.95; the RMSEA was below the threshold of 0.08; and the SRMR was below the required threshold of 0.07. In both CFAs the Chi-square statistic was discounted due to its unreliability as a fit index with large samples (Kline, 2008). From Figure 1 it may be seen that for the SCORE-15 the factor loadings ranged from 0.536 to 0.878. Error variances range from 0.405 to 0.747. The correlations between factors ranged from 0.50 to 0.97. From Figure 1 it may also be seen that for the SCORE-28 the factor loadings ranged from 0.530 to 0.892. Error variances ranged from 0.281 to 0.795. The correlations between factors ranged from 0.50 to 0.92.

Internal consistency reliability

Internal consistency reliability analyses confirmed that both brief versions of the SCORE had good internal reliability with alpha coefficients for overall scales and subscales above .7. Cronbach alphas for the SCORE-15 and SCORE-28 totals were .90 and .93 respectively. Reliabilities of the strengths, difficulties and communication subscales of the SCORE-15 were .83, .85 and .78 respectively. Reliabilities of the strengths, difficulties and communication subscales of the SCORE-28 were .90, .86 and .86 respectively.

Test-retest reliability

Test-retest reliability analyses confirmed that both brief versions of the SCORE had good test-retest reliability. Intraclass correlations between data collected at Time 1 and 3-5 months later at Time 2 for overall scales and subscales were above .7. To determine the test-retest reliability of the SCORE-15 and SCORE-28, intraclass correlations were computed between Time 1 and 2 data from 175 untreated cases. Intraclass correlations for the SCORE-15 and SCORE-28 totals were .91 and .93 respectively. Intraclass correlations for the strengths, difficulties and communication subscales of the SCORE-15 were .82, .91 and .85 respectively. Intraclass correlations for the strengths, difficulties and communication subscales of the SCORE-28 were .85, .90 and .88 respectively.

Correlations with other measures

The pattern of correlations between the SCORE-15 and SCORE-28 and other measures of family and individual adjustment provide support for the construct validity of both versions of the SCORE. Moderate rather than high correlations were expected between SCORE totals and the GARF, since the fact that ratings were made by different informants on these two measures of family adjustment were expected to affect the magnitude of

correlations. The SCORE is a self-report measure and the GARF is a clinician-rated measure. Moderate correlations were expected between SCORE totals and totals on measures of child (SDQ and CGAS) and parent (MHI-5 and GAF) adjustment, since family systems theory would predict an association, but not a perfect correlation between the adjustment of family systems and family subsystems (Carr, 2012). It was also expected that correlations between totals on measures of family and individual adjustment with SCORE totals, would be greater than those with SCORE subscales, since subscales assessed aspects of family functioning rather than overall family functioning. The subscales and totals of the SCORE-15 and SCORE-28 had significant ($p<.01$) Pearson correlations with other measures of family adjustment (GARF), child adjustment (SDQ and CGAS), and parental adjustment (MHI-5 and GAF) as shown in Table 2. The absolute magnitude of correlations between SCORE-28 and SCORE-15 subscales and totals on the one hand, and measures of family and individual adjustment on the other ranged from 0.23 to 0.56. For each measure of family and individual adjustment, correlations with SCORE totals were greater than correlations with SCORE subscales. As expected, correlations with SCORE totals were predominantly moderate.

Comparison of clinical and non-clinical groups

A comparison of means of clinical and non-clinical groups indicate that the SCORE-15 and SCORE-28 demonstrated good criterion validity in terms of their ability to differentiate between clinical and non-clinical cases. Mean scores of the 701 clinical cases from the present study were compared with those of an Irish national random sample (N = 403, Fay et al., 2011). t-test results showed that SCORE-15 and SCORE-28 means of the clinical group were significantly higher than those of normal controls Effect sizes showed that these differences were large, with Cohen's d values greater than 0.8. For the SCORE-15,

the mean of the clinical group was 2.64 (SD = 1.04) and the mean of the non-clinical group was 1.87 (SD = 0.64) (t = 13.47, p<.01, d = 0.84, 95% CI = 0.79, 0.90). For the SCORE-28, the mean of the clinical group was 2.57 (SD = 0.91) and the mean of the non-clinical group was 1.91 (SD = 0.58) (t = 13.11, p<.01, d = 0.82, 95% CI = 0.77, 0.87).

Responsiveness of group means to change

Analyses of group means of 258 cases at Time 1 and 2 showed that all SCORE-15 and SCORE-28 scales were responsive to change associated with attending systemic therapy for 3 to 5 months. The statistical significance of differences between means of treated cases at Time 1 and 2 was evaluated with dependent t-tests. The magnitudes of these differences were expressed as standardized response means (SRM) which is an effect size measure used to gauge the responsiveness of a scale to therapeutic change (Husted et al., 2000). Table 3 shows that mean values for all of the SCORE-15 and SCORE-28 scales improved significantly from Time 1 to 2. SRM values for scales ranged from 0.15 – 0.28. These changes in perceived family functioning are relatively small compared with changes that occurred in ratings of main problem severity and impact, where SRM's ranged from 0.52 to 0.71.

Reliable change

The SCORE-15 and SCORE-28 detected reliable change from Time 1 to Time 2 in a proportion of cases. The reliable change index (RCI) is an index of clinical improvement from one time point to another which takes account of the psychometric properties of the scale used to assess the reliability of improvement. RCIs greater than 1.96 occur by chance in fewer than 5% of cases (Jacobson & Truax, 1991). On the SCORE-15, 106 of

258 cases (41.1%) showed reliable change with RCIs greater than 1.96. On the SCORE-28, RCIs greater than 1.96 were obtained by 93 of 258 cases (36%).

Clinical improvement

The SCORE 15 and 28 detected clinical improvement, and both clinical improvement and reliable change in a proportion of cases. Cases were classified as having shown clinically significant improvement based on the 90th percentile clinical cut-offs established by Fay et al. (2013) in an Irish national random sample. These were 2.92 for the SCORE-15 total and 2.86 for the SCORE-28 total.

Of 258 treated cases assessed at Time 1 and 2, 113 scored above the clinical cut-off point of 2.92 on the SCORE-15 at Time 1. At Time 2, 42 of these 113 cases scored below the SCORE-15 clinical cut-off point, indicating that 37% (42/113) showed clinical improvement. Thirty-six of the 113 cases (31.9%) who scored above the clinical cut-off point on the SCORE-15 at Time 1 showed both clinical improvement and reliable change at Time 2.

Of 258 treated cases assessed at Time 1 and 2, 107 scored above the clinical cut-off point of 2.86 on the SCORE-28 at Time 1. At Time 2, 34 of these 107 cases scored below the SCORE-28 clinical cut-off point, indicating that 32% (34/107) showed clinical improvement. Thirty-one of 107 (30%) cases who scored above the clinical cut-off point on the SCORE-28 at Time 1 showed both clinical improvement and reliable change at Time 2.

Correlations between SCORE change scores and change scores of other variables

The pattern of correlations between difference scores reflecting change over the course of treatment from Time 1 to 2 for the SCORE-15 and SCORE-28 with measures of parent and child adjustment, and changes in severity and impact ratings of the main problem

supported the construct validity of the SCORE. Difference scores reflecting changes from Time 1 to 2 were computed for the SCORE, GARF, SDQ, CGAS , MHI-5, GAF and main problem severity and impact ratings. Pearson's correlations were computed between the difference scores of the SCORE-15 and SCORE-28 on the one hand, and difference scores of the GARF, SDQ, CGAS , MHI-5, GAF on the other. Significant correlations between difference scores occurred for the SCORE-15 and the SDQ ($r = 0.29$, $p<.01$), MHI-5 ($r = 0.31$, $p<.01$), GAF ($r = 0.24$, $p<.05$), and main problem severity ($r = 0.17$, $p<.05$). Significant correlations between difference scores occurred for the SCORE-28 and the SDQ ($r = 0.33$, $p<.01$), MHI-5 ($r = 0.37$, $p<.01$), GAF ($r = 0.26$, $p<.01$), main problem severity ($r = 0.23$, $p<.01$) and main problem impact ($r = 0.24$, $p<.05$). Correlations between difference scores for both versions of the SCORE on the one hand and the GARF and CGAS on the other were not significant, nor was the correlation between difference scores from the SCORE-15 and ratings of main problem impact.

Discussion

This study aimed to confirm the three-factor structure, reliability, construct validity and criterion validity of the SCORE-15 and 28, and investigate the responsiveness of the SCORE-15 and 28 to treatment-related improvement. Results showed that both the SCORE-15 and the SCORE-28 data from our large multisite sample fit the three-factor structure found in previous studies (Cahill et al., 2010; Fay et al., 2011; Stratton et al., 2010). Both versions of the SCORE showed good internal consistency and test-retest reliability; good construct validity in that they correlated with measures of child, parent and family adjustment; and good criterion validity insofar as mean scores of clinical and non-clinical groups differed significantly from each other.

The 28-item version had better internal consistency reliability than the 15-item version of the SCORE, which is not surprising because longer versions of scales typically have larger alpha in internal consistency reliability coefficients.

Because the high correlations (r >0.9) between the between family difficulties and family communication factors in the CFAs of both SCORE-15 and 28 data suggested the possibility of over-specification of the underlying model, ancillary CFAs were conducted to check if the data fit two-factor (factor 1 = family strengths, factor 2 = family difficulties and communication problems) or single factor models better than the three-factor model. The two-factor model, but not the single factor model fit the SCORE-15 and SCORE-28 data.

The principal novel contribution of this study was its demonstration of the responsiveness of the SCORE-15 and SCORE-28 to changes in family functioning which occurred during a moderate period of systemic therapy. There were three key findings. First, group means for all SCORE-15 and SCORE-28 scales were responsive to change occurring during three to five months of therapy. However, the magnitude of changes detected by the SCORE-15 and 28 totals, and strengths, difficulties and communication subscales were relatively small compared with changes in ratings of main problem severity and impact. Second, we found that the SCORE-15 and SCORE-28 detected clinical improvement and reliable change in about a third of cases over three to five months of therapy. Third, there were significant correlations between difference scores reflecting change over three to five months of systemic therapy from the SCORE-15 and 28 on the one hand and measures of parent and child adjustment and main problem severity and impact ratings on the other. Taken together these three results support the view that the SCORE-15 and SCORE-28 are sensitive to the sorts of changes that occur during family therapy.

Unexpectedly correlations between difference scores for both versions of the SCORE on the one hand and the GARF and CGAS on the other were not significant. This reflects the fact that where changes in parents' perceptions of family adjustment as assessed by the SCORE occurred, corresponding changes did not occur in clinicians' perceptions of family adjustment on the GARF and children's adjustment on the CGAS. Clinician's and parents' differing perspectives may have accounted for these non-significant correlations.

This study has a number of limitations. First, Time 2 data were available for only a subgroup of cases for whom Time 1 data were obtained. Because this was a self-selected subgroup, responsiveness analyses conducted on this group may not be generalizable to cases who enter services and dropped out of treatment or data collection processes. Second, inter-rater reliably data were not available for therapist-rated scales (GARF, CGAS and GAF). This means that there is some uncertainty about the reliability of results from construct validity correlational analyses conducted using these scales.

Our main conclusion is that the SCORE-28 and SCORE-15 are viable measures for routine use in clinical settings to assess outcome in systemic therapy. The briefer 15-item version has demonstrated better psychometric properties in this study than the 28-item version, takes less time to administer than the SCORE-28 and is therefore preferential for routine administration.

Clinical implications

Both versions of the SCORE described in this paper may be administered to family members periodically in clinical practice to monitor change over the course of family therapy. The SCORE-28 is in the appendix to Cahill et al. (2010). The SCORE-15 and a 29-item version that contains items for both the SCORE-28 and SCORE-15 is in Fay et al.,

(2013). A web-based system for administering and scoring the SCORE-15 and SCORE-28 is available at this link: http://scorefamilyassessment.org/login.php.

References

Benjamini, Y., & Hochberg Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological*), 57, 289–300.

Berwick, D., Murphy J., Goldman et al., (1991). Performance of a five-item mental health screening test. *Medical Care*, 29, 169-176.

Bollen, K., & Long, S.J. (1993). *Testing structural equation models*. London: Newbury Park.

Cahill, P., O'Reilly, K., Carr, A., Dooley, B., & Stratton, P. (2010) Validation of a 28-item version of the Systemic Clinical Outcome and Routine Evaluation in an Irish context: The SCORE-28. *Journal of Family Therapy*, *32*, 210-231.

Carr, A. (2012). *Family therapy: Concepts, process and practice* (Third edition). Chichester: Wiley.

Carr, A., Fry, J. & Casey, T., Forrest, M., Beirne, F., & Cassells, C. (2013). *Positive Systemic Practice: The Crosscare Teen Counseling model of intervention for adolescents and their parents (A practice manual).* Dublin: Crosscare Teen Counseling.

Cassells' C., Carr, A., Forrest, M., Fry, J., Beirne, F., Casey, T., & Rooney, B. (2014). A controlled trial of Positive Systemic Practice for families of adolescents with emotional and behavioral problems in Ireland. *Journal of Family Therapy,* DOI: 10.1111/1467-6427.12038

Fay, D., Carr, A., O'Reilly, K., Cahill, P., Dooley, B., Guerin, S., & Stratton, P. (2013). Irish norms for the SCORE-15 and 28 from a national telephone survey. *Journal of Family Therapy*, *35,* 24-42.

Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire (SDQ). *Journal of the American Academy of Child & Adolescent Psychiatry, 40,* 1337-1345.

Graham, C., Carr, A., Rooney, B., Sexton, T., & Wilson Satterfield, L. (2014). Evaluation of functional family therapy in an Irish context. *Journal of Family Therapy*, *36,* 20–38.

Husted, J., Cook, R., Farewell, V., & Gladman, D. (2000), Methods for assessing responsiveness: a critical review and recommendations. J*ournal of Clinical Epidemiology, 53,* 459-168.

Jacobson, N., & Truax, P. (1991) Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59,* 12-19.

Jewell, T., Carr, A., Stratton, P., Lask, J., & Eisler, I. (2013). Development of a children's version of the SCORE index of family function and change. *Family Process*, *53,* 673-683.

Kline, R. (2010). *Principles and practice of structural equation modeling.* (Third Edition). New York: Guilford.

Luborsky, L. (1962). Clinicians' judgments of mental health. *Archives of General Psychiatry*, *7,* 407-417.

Muthén, L.K. and Muthén, B.O. (1998-2012). *Mplus user's guide* (Seventh Edition). Los Angeles, CA: Muthén & Muthén.

Sexton, T. (2011). *Functional Family Therapy in clinical practice.* New York: Routledge.

Shaffer, D., Gould, M., Brasic, J., Ambrosini, P., Fisher, P., Bird, H., & Aluwahlia, S. (1983). A children's global assessment scale (CGAS). *Archives of General Psychiatry*, *40,* 1228-1231.

Stratton, P., Bland, J., Janes, E., & Lask, J. (2010) Developing an indicator of family function and a practicable outcome measure for systemic family and couple therapy: The SCORE. *Journal of Family Therapy*, *32,* 232-258.

Stratton, P., Lask, J., Bland, J., Nowotny, E., Evans, C., Singh, R., Janes, E., & Peppiatt, A. (2014). Detecting therapeutic improvement with early in therapy: Validation of the SCORE-15 Index of family functioning and change. *Journal of Family Therapy*, *36,* 3-19.

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, 199–207.

Yingling, L., Miller, W., McDonald, A. L., & Galewaler, S. T. (1998). *GARF assessment sourcebook: Using the DSM-IV global assessment of relational functioning*. New York: Brunner/Mazel.

Table 1. The SCORE-15 and SCORE-28

SCORE 15	SCORE 28
STRENGTHS	
In my family we talk to each other about things that matter to us	In my family we talk to each other about the things that matter to us
We are good at finding new ways to deal with things that are difficult	We are good at finding new ways to deal with things that are difficult
When one of us is upset they get looked after within the family	When one of us is upset they get looked after within the family
We trust each other	We trust each other
Each of us gets listened to in our family	Each of us gets listened to in our family
	If something is going wrong in our family we know we can change it
	Our family shares enjoyable times together
	People do things that show that they care about each other in my family
	In our family it is OK to show how you feel
	Being in this family is important to us
	People in my family are willing to change their views about things
	Respecting elders is important in our family
	We are a very organised family
DIFFICULTIES	
In my family we blame each other when things go wrong	
We seem to go from one crisis to another in my family	We seem to go from one crisis to another in my family
It feels miserable in our family	It feels miserable in our family
Things always seem to go wrong for my family	Things always seem to go wrong for my family
We find it hard to deal with everyday problems	We find it hard to deal with everyday problems
	Life in our family is very difficult.
	Other people look down on my family because we are different
COMMUNICATION	
It feels risky to disagree in our family	
People often don't tell each other the truth in my family	People often don’t tell each other the truth in my family
People in the family are nasty to each other	People in the family are nasty to each other
When people in my family get angry they ignore each other on purpose	When people in my family get angry they ignore each other on purpose
People in my family interfere too much in each other's lives	People in my family interfere too much in each other’s lives
	In my family we blame each other when things go wrong (This is in the Overwhelmed by difficulties scale of the SCORE 15)
	One person tends to get blamed for everything in my family
	People in our family lie to each other
	People slam doors, throw things or make a lot of noise if they are upset
	In my family people prefer to watch TV than to spend time with each other
Clients rate each item on a 6 point scale from 1 = describes my family extremely well to 6 = does not describe my family at all	
Clients also name their main problem and rate its severity and impact on 10 point scales	

Table 2. Correlations between SCORE-15 and SCORE-28 scales and other measures of family, child and parental adjustment

	Measure of Family adjustment	Measures of Child adjustment		Measures of Parental adjustment	
Variable	**GARF**	**SDQ**	**CGAS**	**MHI-5**	**GAF**
N	391	682	432	599	368
SCORE 15					
Total family adjustment	-0.43	0.52	-0.38	-0.42	-0.35
Family strengths	-0.33	0.37	-0.23	-0.34	-0.26
Family difficulties	-0.38	0.49	-0.37	-0.37	-0.35
Family communication	-0.37	0.44	-0.33	-0.32	-0.28
SCORE-28					
Total family adjustment	-0.45	0.56	-0.38	-0.44	-0.37
Family strengths	-0.34	0.40	-0.24	-0.37	-0.28
Family difficulties	-0.38	0.48	-0.37	-0.39	-0.34
Family communication	-0.39	0.49	-0.36	-0.33	-0.33

Note: SCORE = Systemic Clinical Outcomes and Routine Evaluation. SDQ = Strengths and Difficulties Questionnaire. MHI-5 = Mental health Inventory 5. GARF = Global Assessment of Relational Functioning Scale. CGAS = Children's Global Assessment of Functioning Scale. GAF = Global Assessment of Functioning Scale. All Pearson correlations are significant at $p<.01$.

Table 3. SCORE-15 and SCORE-28 total and subscale Time 1 and Time 2 mean scores for treated cases, with dependent t-test results and standardized response means

Instrument	Variable		Time 1 N=258	Time 2 N=258	t Time 1-2	SRM Time 1-2
SCORE-15	Total family adjustment	M SD	2.75 1.05	2.53 0.98	4.51**	0.28
	Family strengths	M SD	2.51 1.15	2.26 0.98	4.21**	0.27
	Family difficulties	M SD	2.94 1.37	2.68 1.33	3.74**	0.23
	Family communication	M SD	2.78 1.26	2.64 1.23	2.37*	0.15
.						
SCORE-28	Total family adjustment	M SD	2.69 0.91	2.49 0.84	4.69**	0.29
	Family strengths	M SD	2.39 0.93	2.20 0.85	4.24**	0.26
	Family difficulties	M SD	2.88 1.35	2.66 1.34	3.22**	0.20
	Family communication	M SD	2.98 1.16	2.81 1.15	3.11**	0.19
SCORE	Rating of main problem severity	M SD	6.83 2.20	4.54 2.37	12.11**	0.52
	Rating of main problem impact	M SD	6.37 2.35	4.29 2.43	10.73**	0.71

Note: SCORE = Systemic Clinical Outcomes and Routine Evaluation. M = Mean. SD = Standard deviation. t = dependent t-test statistic. SRM = Standardized response mean. **p<.01. For SCORE rating of main problem severity N = 220 and SCORE rating of main problem impact N = 227.

Figure 1. Confirmatory factor analyses for the SCORE-15 and Score-28