VALIDATION OF 28 AND 15 ITEM VERSIONS OF THE SCORE FAMILY ASSESSMENT QUESTIONNAIRE WITH ADULT MENTAL HEALTH SERVICE USERS

Kevin O'Hanrahan[1,2], Megan Daly White[1,2], Alan Carr[1,3], Paul Cahill[1,2], Mairi Keenleyside[2], Mark Fitzhenry[1,2], Elizabeth Harte[1,2], Jennifer Hayes[2], Hester Noonan[2], Helen O'Shea[2], Avril McCullagh[2], Shaun McGuinness[2], Catherine Rodgers[2], Neal Whelan[2], Noel Sheppard[2], Stephen Browne[2]

[1] School of Psychology, University College Dublin.

[2] Health Service Executive, Ireland

[3] Clanwilliam institute, Dublin

Citation. O'Hanrahan, K., Daly White, M., Carr, A., Cahill, P., Keenleyside, M., Fitzhenry, M., Harte, E., Hayes, J., McCullagh, A., McGuinness, S., Noonan, H., O'Shea, H., Rodgers, C., Whelan, N., Sheppard, N., Browne, S. (2016). Validation of 28 and 15 item versions of the score family assessment questionnaire with adult mental health service users. *Journal of Family Therapy,* doi: 10.1111/1467-6427.12107

Emails: kevin.p.ohanrahan@gmail.com, megandalywhite@gmail.com, Alan.Carr@ucd.ie, cahillpaul@gmail.com, Mairi.keenleyside@hse.ie, mark.fitzhenry@hotmail.com, Elizabeth.harte@gmail.com, Jennifer.Hayes@hse.ie, avril.mccullagh1@hse.ie, shaunmcguinness90@gmail.com, hesternoonan@yahoo.ie, helenhayesoshea@yahoo.co.uk, nealwhelan1@gmail.com, Noel.Sheppard@hse.ie, Stephen.Browne@hse.ie

Acknowledgement. This research project was supported by HSE-South clinical psychology training sponsorships to KO'H, MDW, MF and EH.

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. Twenty-eight and 15-item short versions of the SCORE have previously been validated with samples containing some families attending adult mental health services and some attending child and adolescent services; or samples containing families exclusively from child and adolescent services. This study aimed to investigate psychometric properties of two short versions of the SCORE with a sample drawn exclusively from adult mental health services. Data were collected from 199 service users attending inpatient and outpatient adult mental health services in the south east of Ireland. Both instruments had good internal consistency reliability. They also showed construct and criterion validity, correlating with measures of global functioning, service need, and childhood trauma. The 28 and 15 item versions of the SCORE are brief psychometrically robust family assessment instruments which may be used to evaluate systemic therapy with adult mental health service users.

PRACTITIONER POINTS

- Brief 15 and 28-item versions of the SCORE may be used to assess family adjustment, monitor progress and assess outcome in adult mental health cases.
- 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.

KEYWORDS

SCORE, family assessment, family research

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 therapy (Stratton et al., 2010). The SCORE was developed for routine use in clinical practice to periodically assess therapeutic progress. SCORE data may be used to inform therapists and clients about progress of individual cases and to evaluate service effectiveness (e.g., Cassells et al., 2014; Hartnett et al., In press).

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 a large sample of cases, a brief 15-item version (SCORE-15) was developed in the UK (Stratton et al., 2010) and a 28-item version (SCORE-28) was developed in Ireland (Cahill et al., 2010). Items from these short versions of the SCORE are given in Table 1. In these independent research programs, both short versions of the SCORE were found to have a three-factor structure. The factors assess family strengths, difficulties and communication. Both short forms of the SCORE showed acceptable levels of internal consistency and test-retest reliability. The factor structure, internal consistency and test-retest reliability of the 28 and 15-item versions of the SCORE have been replicated in further studies (Fay et al., 2013; Hamilton et al., 2015). Jewell et al. (2013) developed 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 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 problem are given on 10-point scales.

In a national telephone survey, Fay et al. (2013) developed norms for the SCORE-28 and SCORE-15 in an Irish context. Clinical cut-off points for families of children with

clinically significant behavioural and emotional problems were identified. The criterion validity of 15 and 28 item short versions or the SCORE was also established by showing that they discriminated between cases with, and without clinically significant problems. This finding was replicated by Hamilton et al. (2015).

Cahill et al. (2010), Fay et al. (2013) and Hamilton et al. (2015) 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.

Two studies have shown that the SCORE is responsive to change. In a large UK 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. In a large Irish multisite study, Hamilton et al. (2015) found that both the 28 and 15-item versions of the SCORE were responsive to change arising from a variety of systemic interventions spanning 3-5 months.

To date 28 and 15-item short versions of the SCORE have been validated with samples containing some families attending adult mental health services and some attending child and adolescent services; or samples containing families exclusively from child and adolescent services. The validity of the SCORE when used exclusively with adult mental health service users is unclear. The present study aimed to address this issue by investigating the factor structure, reliability, criterion and construct validity of the 28 and 15-item versions of the SCORE in a sample of adult mental health service users.

METHOD

The study was conducted with ethical approval of the Irish Health Service Executive (HSE) and University College Dublin, and informed consent of participants. Data collection occurred between July 2011 and June 2014 in the public mental health service in the south east of Ireland.

Participants

Participants were 199 adult mental health service users. There were 100 inpatients and 99 outpatients. Consecutive referrals for inpatient and outpatient care at the HSE Waterford mental health service were accepted into the study unless they were under 18 years; had an intellectual disability or acquired brain injury; were inappropriately referred to the service with problems such as homelessness or neurological illness; or were unable or unwilling to provide informed consent or to complete the assessment protocol.

With regard to demographic characteristics 52.8% were males; 47.2% were female, and the mean age was 40.2 years (SD = 14.0, Range = 18 - 75 years). With regard to family status, 37.2% were married, cohabiting or in a relationship; 54.3% had children; and the average number of children was 1.31 (SD = 1.61, Range = 1 - 7 children). The unemployment rate was 46.2% and employed participants came from a range of socio-economic groups.

Participants had attended mental health services for an average of 7.3 years (SD = 10.05 years), and the average duration of past inpatient treatment was 2.7 months (SD = 6.32). Participants had received multimodal treatment involving medication and individually-oriented (rather than systemic) psychotherapy. Of 199 cases, 196 (98.4%) met the diagnostic criteria for a current or lifetime DSM-IV axis I disorder, and 77 (38.7%) met the criteria for both an axis I psychiatric disorder and an axis II personality disorder. The axis I disorder rates were 64.8% for anxiety disorders, 55.8% for depressive disorders, 46.2% for alcohol and substance use disorders, 22.1% for psychotic disorders. 9.5% for

bipolar disorders, 4% for eating disorders and 2% for adjustment disorders. The personality disorder rates were 13.6% for avoidant, 9.5% for obsessive compulsive, 9.5% for borderline, 8.5% for paranoid, 8% for antisocial, 3.5% for narcissistic, 3% for dependent and for schizoid, 1% for schizotypal, and 0.5% for histrionic and for personality disorder not otherwise specified. Three cases did not meet criteria for a current or lifetime DSM axis I diagnosis. All three had subclinical mood disorders. One had made a recent suicide attempt and another had schizoid personality disorder.

Assessment protocol

The assessment protocol included the Systemic Clinical Outcome and Routine Evaluation (SCORE, Cahill et al., 2010; Fay et al., 2013), the Structured Clinical Interviews for DSM-IV axis I and II disorders (SCID I, First et al., 1996; SCID II, First et al., 1997), the Global Assessment of Functioning rating scale (GAF, Luborsky, 1962), the patient version of the Camberwell Assessment of Need Short Appraisal Schedule (CANSAS, Trauer et al., 2008), and the Childhood Trauma Questionnaire (CTQ, Bernstein & Fink, 1998)

A 29-item version of the SCORE was used which included all items in the SCORE-28 and SCORE-15. The SCID I yielded DSM-IV axis I diagnoses. Axis II personality disorders were assessed with the SCID II. The GAF provided an interviewer's assessment of overall client functioning on a 100-point scale. The CANSAS provided an index of self-reported unmet service needs, based on a list of 22 needs for mental health service users. The CTQ assessed recollections of childhood physical, sexual and emotional abuse, and physical and emotional neglect.

All variables in the assessment protocol were reliably assessed. DSM diagnoses based on the SCID I and II had adequate inter-rater reliability. Kappa coefficients using data from pairs of raters for 19 cases ranged from 0.6 (95% CI [0.11, 1.00]) to 1.0 (95% CI [1.00, 1.00]), with the majority of values being above 0.7. The GAF had high inter-rater

reliability. The intraclass coefficient from pairs of raters for 19 cases was 0.98 (95% CI [0.94, 0.99]). For self-report scales (SCORE, CANSAS-P, and CTQ) there were acceptable levels of internal consistency reliability with alpha coefficients ranging from 0.76 (95% CI [0.71, 0.81]) - 0.95 (95% CI [0.94, 0.96]).

Procedure

Recruitment was conducted in collaboration with administrative and clinical staff at inpatient and outpatient centres. Research team interviewers were trained in administration and scoring of all instruments, notably the SCID I and II diagnostic interviews. All interviewers had primary degrees in psychology. Interviews were conducted at Waterford Regional Hospital or Saint Patrick's Hospital, Waterford.

Data management

Data were entered item-by-item into an SPSS file and verified by checking ranges for all items. There were missing values in 38 cases, and in these, values for fewer than 20% of items were missing No values were substituted for missing diagnoses. For other variables, in cases with missing data, for multi-item scales, scale means were substituted for missing items in all analyses except the confirmatory factor analyses (CFA). The Mplus (Muthén & Muthén, 1998 – 2012) default weighted least squares estimation procedure, which assumes data points are missing at random, was used to deal with 4 missing data points in the CFAs of SCORE data (Asparouhov and Muthen, 2010). This procedure yields results consistent with those from maximum likelihood estimation.

Items were combined into multi-item scales and all continuous variables were checked for normality. With few exceptions continuous variables in this study were normally distributed, justifying the use of parametric statistical tests. Data were analysed with Version 20 of the Statistical Package for the Social Sciences and Version 7 of Mplus.

In analyses where multiple t-tests or correlations were conducted the false discovery rate was used to control for type 1 error (Benjamini & Hochberg, 1995). FRD was used because many statistical tests were conducted and controlling the probability of type I errors using the Bonferroni correction would have led to a very conservative p value and increased the probability of type II errors. The FDR approach tolerates some Type I errors by controlling the rate rather than the presence of them.

RESULTS

Factor structure

The weighted least squares with mean and variance adjustment estimator was used for CFAs (Muthén et al., 1997). The weighted least square estimator was used (rather than the maximum likelihood estimator) because is it a good estimator of latent variable models (Muthén, & Satorra, 1995); it is a robust estimator for modelling categorical data (Brown 2006); it yields comparable results to maximum likelihood estimation in large data sets when there is evidence supporting the model from previous studies (Mislevy, 1986; Forero & Maydeu-Olivares, 2009); it is more efficient and less computationally demanding than maximum likelihood estimation (Agresti, 2002); and it yields similar results to maximum liklihood estimation under general missing data assumptions (Asparouhov & Muthen, 2010).

Confirmatory factor analyses (CFA) showed that data fit the 3-factor structure of the SCORE-28 quite well, but fit the 3-factor structure of the SCORE-15 less well. 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 Weighted Root-Mean-Square Residual (WRMR). The RMSEA, CFI, TLI and WRMR were used to evaluate the extent to which data fit the SCORE-28 and SCORE-15 three-factor models (Bollen & Long, 1993; Hooper et al., 2008;

Kline, 2010). The Chi-square statistic was discounted due to its unreliability as a fit index with large samples (Kline, 2008). Hu and Bentler (1995, 1999) and Yu (2002) offered the following standards for ‘good-fit’: RMSEA <0.06, CFI >0.95, TLI >0.95 and WRMR <1.0.

For 3 of 4 indices the data fit the SCORE-28 three-factor model. The RMSEA of 0.065 (95% CI [0.058, 0.073]) was outside of its required threshold of 0.06; while the CFI, TLI and WRMR were within their respective parameters indicating an adequate fit rather than a good fit (CFI of 0.961, TLI of 0.957 and WRMR of 0.948).

For the SCORE-15, two of the four indices of data-fit were satisfactory. The CFI of 0.954 and the WRMR of 0.921 were within their parameters indicating a good-fit, while the TLI of 0.945 and the RMSEA of 0.098 (95% CI [0.084, 0.112]) failed to indicate satisfactory goodness of fit.

From Figure 1 it may be seen that for the SCORE-28 the factor loadings ranged from 0.494 to 0.915. Error variances ranged from 0.163 to 0.756. Correlations between factors ranged from 0.69 to 0.88.

From Figure 1 it may also be seen that for the SCORE-15 factor loadings ranged from 0.553 to 0.858. Error variances ranged from 0.264 to 0.694. Correlations between factors ranged from 0.712 to 0.939.

For both the SCORE-28 and the SCORE-15, the highest correlations occurred between the family difficulties and communication factors and the lowest were between the family strengths and difficulties factors.

Internal consistency reliability

Internal consistency reliability analyses confirmed that both brief versions of the SCORE had good internal consistency reliability, with alpha coefficients for total scales and subscales above .7. Cronbach alphas for the SCORE-28 and SCORE-15 totals were .95 (95% CI [0.94, 0.96]) and .92 (95% CI [0.90, 0.93]) respectively. Reliabilities of the

strengths, difficulties and communication subscales of the SCORE-28 were .94 (95% CI [0.92, 0.95]), .82 (95% CI [0.78, 0.86]) and .86 (95% CI [0.83, 0.89]) respectively. Reliabilities of the strengths, difficulties and communication subscales of the SCORE-15 were .88 (95% CI [0.85, 0.90]), .83 (95% CI [0.79, 0.87]) and .77 (95% CI [0.72, 0.82]) respectively.

Comparison of clinical and non-clinical groups

A comparison of means of clinical and non-clinical groups indicated that the SCORE-28 and SCORE-15 demonstrated good criterion validity in terms of their ability to differentiate between clinical and non-clinical cases. Mean scores of 199 clinical cases from the present study were compared with those of an Irish national random sample (N = 403, Fay et al., 2013). Results of this analysis are given in Table 2. It was expected that clinical cases would score higher than non-clinical cases because the SCORE is keyed so that high scores indicate more severe adjustment problems. t-tests showed that SCORE-28 and SCORE-15 total and subscale means for the clinical group were significantly higher than those of normal controls. For both the SCORE-28 and SCORE-15 totals, effect sizes reflecting differences between means of clinical and non-clinical groups were large (d = 0.91, 95% CI [0.85, 0.98]). Effect sizes for subscales were medium to large, ranging from 0.57 (95% CI [0.52, 0.68]) to 1.1 (95% CI [1.03, 1.17]).

Comparison of groups with and without personality disorders

A comparison of means of cases with axis I psychiatric disorders only, and cases with both axis I psychiatric disorders and comorbid axis II personality disorders supported the criterion validity of the SCORE-28 and SCORE-15 in terms of their ability to differentiate between these two groups of cases. It is well established that among adult mental health service users, the presence of comorbid personality disorders is associated with poorer

psychosocial adjustment (Skodol et al., 2005). It was therefore expected that service users with personality disorders would show poorer family adjustment on the SCORE. SCORE means of 77 cases with axis I psychiatric disorders and comorbid axis II personality disorders were compared with those of 119 cases with axis I disorders only. Results of this analysis are given in Table 3. t-tests showed that SCORE-28 and SCORE-15 total and subscale means of the group with personality disorders were significantly higher than those of cases without personality disorders. For both the SCORE-28 and SCORE-15 totals, effect sizes reflecting differences between means of these two groups were large ($d > 0.8$). Effect sizes for subscales were medium to large, ranging from 0.66 (95% CI [0.48, 0.83]) to 0.88 (95% CI [0.75, 1.01]). On the SCORE main problem severity scale (but not the problem impact scale), the mean for the group with personality disorders was significantly higher than that of the group without personality disorders, and the effect size was medium ($d = 0.49$, 95% CI [0.16, 0.82]).

Correlations with measures of functioning, service need, childhood trauma and motivation for psychotherapy

The pattern of correlations between the SCORE and measures of global personal functioning, service need, and childhood trauma provided some support for the construct validity of both the 28 and 15-item versions of the SCORE. It was expected that SCORE scales would correlate significantly and moderately with the GAF, CANSAS, and CTQ, since service users reporting greater family adjustment problems, were expected to show poorer global functioning and report greater service needs, and recollections of childhood trauma. Pearson product-moment correlations between SCORE scales and the GAF, CANSAS, and CTQ are given in Table 4. From the table it may be seen that for both versions of the SCORE small to moderate, significant correlations occurred between all SCORE scales on the one hand, and the GAF, CANSAS and CTQ on the other. The

absolute magnitude of correlations between SCORE scales, and the GAF, CANSAS, and CTQ ranged from 0.17 (95% CI [0.31, 0.03]) to 0.46 (95% CI [0.33, 0.58]). For the GAF and CANSAS, largest correlations occurred with the family difficulties subscales on both versions of the SCORE. For example, there was a correlation of r = 0.46 (95% CI [0.33, 0.58]) between the CANSAS and the family difficulties scale of the SCORE-28. In contrast, correlations between the CANSAS and the SCORE-28 total, and the family strengths and family communication subscales of the SCORE-28 were 0.40 (95% CI [0.27, 0.53]), 0.28 (95% CI [0.14, 0.41]) and 0.33 (95% CI [0.20, 0.47]) respectively.

DISCUSSION

This study aimed to confirm the three-factor structure, reliability, criterion and construct validity of the 28 and 15-item versions of the SCORE in a sample of adult mental health service users. Results showed that SCORE-28 data from our sample, on 3 of 4 indices, fit the three-factor structure found in previous studies (Cahill et al., 2010; Fay et al., 2011; Hamilton et al., 2015). In contrast SCORE-15 data from our sample fit the three-factor structure on only two indices. This shows that the SCORE-28 had greater factorial validity in an adult mental health context.

Both versions of the SCORE showed good internal consistency reliability; good criterion validity insofar as mean scores of clinical and non-clinical groups, and groups with and without personality disorders differed significantly from each other; and good construct validity in that all SCORE scales correlated with measures of global personal functioning, service need, and childhood trauma. These findings further supported the construct validity of both short versions of the SCORE.

The principal novel contribution of this study was demonstrating the reliability and validity of brief versions of the SCORE in an adult mental health service context, with

service-users who had had received long-term multimodal treatment involving medication and individually-oriented (rather than systemic) psychotherapy.

Our results show that both brief versions of the SCORE may be used to assess family functioning, monitor therapeutic progress, and evaluate outcome in systemic therapy with families of adults with mental health problems. The SCORE total is a reliable and valid index of family adjustment for both the 28 and 15-item versions of instrument. The strengths, difficulties and communication subscales of the longer 28-item version of the instrument have adequate factorial validity and better factorial validity than those of the shorter SCORE-15, and therefore may be used to assess these aspects of family functioning in clinical practice. In contrast, the factorial validity of the strengths, difficulties and communication subscales of the SCORE-15 is less robust. In clinical practice scores from these subscales of the SCORE-15 require cautious interpretation. Thus, the longer SCORE-28 instrument may be used when therapists or researchers wish to assess family strengths, difficulties and communication, whereas the shorter SCORE-15 may be used when a rapid evaluation of family adjustment is required.

Two limitations of this study deserve mention, since they suggest priorities for future research. First, data from family members, other than those with mental health problems, were not collected. It would have been informative to have members of participants' families complete the SCORE and measures of personal adjustment, and also to collect observational data on family functioning. Correlations between these measures and the SCORE would have permitted a more robust evaluation of construct validity. Second, the responsiveness of the SCORE to changes arising from systemic therapy with adult mental health service users was not assessed. These data would be valuable in clarifying the degree to which the SCORE can detect therapeutic improvement in adult mental health settings.

REFERENCES

Agresti, A. (2013) *Categorical Data Analysis* (Third Edition). Hoboken, NJ: Wiley.

Asparouhov, T. and Muthen, B. (2010) Weighted least squares estimation with missing data. https://www.statmodel.com/download/GstrucMissingRevision.pdf

Benjamini, Y. and 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.

Bernstein, D. and Fink, L. (1998) *Childhood Trauma Questionnaire: A Retrospective Self-report.* Manual. San Antonio, TX: The Psychological Cooperation.

Brown, T. (2006) *Confirmatory Factor Analysis for Applied Research.* New York: Guildford

Cahill, P., O'Reilly, K., Carr, A., Dooley, B. and 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.

Cassells' C., Carr, A., Forrest, M., Fry, J., Beirne, F., Casey, T. and 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. and Stratton, P. (2013) Irish norms for the SCORE-15 and 28 from a national telephone survey. *Journal of Family Therapy*, **35***:* 24-42.

First, M., Spitzer, R., Gibbon M. and Williams, J. (1997) *Structured Clinical Interview for DSM-IV Personality Disorders, (SCID-II).* Washington, DC: American Psychiatric Press.

First, M., Spitzer, R., Gibbon, M. and Williams, J. (1996) *Structured Clinical Interview for DSM-IV Axis I Disorders, Clinician Version (SCID-CV).* Washington, DC: American Psychiatric Press.

Forero, C. and Maydeu-Olivares, A. (2009) Estimation of IRT graded response models: Limited versus full information methods. *Psychological Methods,* **14:** 275-299.

Hamilton, E., Carr, A., Cahill, P., Cassells, C. and Hartnett, D. (2015) Psychometric properties and responsiveness to change of 15 and 28 item versions of the SCORE: A family assessment questionnaire. *Family Process*. doi: 10.1111/famp.12117

Hartnett, D., Carr, A. and Sexton, T. (In press) A controlled trial of Functional Family Therapy in an Irish context. *Family Process*.

Hooper, D., Coughlan, J. and Mullen, M. (2008) Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods*, **6**: 53 – 60.

Hu, L. and Bentler, P. (1995) Evaluating model fit. In: R. Hoyle (Ed.), *Structural Equation Modeling: Concepts, Issues and Applications* (pp. 76–79). Thousand Oaks, CA: Sage Publications.

Hu, L. and Bentler, P. (1999) Cutoff criteria for fit indices in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling*, **6**: 1-55.

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

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

Mislevy, R. (1986) Recent developments in the factor analysis of categorical variables. *Journal of Educational Statistics,* **11:** 3-31.

Muthén, B., du Toit, S. and Spisic, D. (1997) Robust inference using weighted least squares and quadratic estimating equations in latent variable modelling with categorical and continuous outcomes. Unpublished technical report.

Muthén, L. and Muthén, B. (1998-2012). *Mplus User's Guide* (Seventh Edition). Los Angeles, CA: Muthén & Muthén.

Muthén, B. and Satorra, A. (1995) Technical aspects of Muthén's LISCOMP approach to estimation of latent variable relations with a comprehensive measurement model. *Psychometrika,* **60:** 489-503

Skodol, A., Gunderson, J., Shea, M., McGlashan, T., Morey, L., Sanislow, C., Bender, D., Grilo, C., Zanarini, M., Yen, S., Pagano, M. and Stout, R. (2005) The Collaborative Longitudinal Personality Disorders Study (CLPS). *Journal of Personality Disorders*, **19**: 487–504.

Stratton, P., Bland, J., Janes, E. and 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., and 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. and 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.

Trauer, T., Tobias, G. and Slade, M. (2008) Development and evaluation of a patient-rated version of the Camberwell Assessment of Need Short Appraisal Schedule (CANSAS-P). *Community Mental Health Journal,* **44**: 113-124.

Yu, C. (2002), Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes. Doctoral dissertation. Los Angeles: University of California.

Table 1. The SCORE-28 and SCORE-15

	SCORE-28		SCORE-15
	STRENGTHS		
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		
5	Our family shares enjoyable times together		
8	If something is going wrong in our family we know we can change it		
12	Each of us gets listened to in our family	12	Each of us gets listened to in our family
13	People in my family are willing to change their views about things		
14	In our family it is OK to show how you feel		
17	When one of us is upset they get looked after within the family	17	When one of us is upset they get looked after within the family
18	Respecting elders is important in our family		
23	In my family we talk to each other about the things that matter to us	23	In my family we talk to each other about things that matter to us
24	We are good at finding new ways to deal with things that are difficult	24	We are good at finding new ways to deal with things that are difficult
26	We trust each other	26	We trust each other
	DIFFICULTIES		
9	We find it hard to deal with everyday problems	9	We find it hard to deal with everyday problems
11	Life in our family is very difficult.		
16	Other people look down on my family because we are different		
19	It feels miserable in our family	19	It feels miserable in our family
		21	In my family we blame each other when things go wrong
22	Things always seem to go wrong for my family	22	Things always seem to go wrong for my family
28	We seem to go from one crisis to another in my family	28	We seem to go from one crisis to another in my family
	COMMUNICATION		
4	People in my family interfere too much in each other's lives	4	People in my family interfere too much in each other's lives
6	One person tends to get blamed for everything in my family		
7	People often don't tell each other the truth in my family	7	People often don't tell each other the truth in my family
10	When people in my family get angry they ignore each other on purpose	10	When people in my family get angry they ignore each other on purpose
15	In my family people prefer to watch TV than to spend time with each other		
20	People in our family lie to each other		
21	In my family we blame each other when things go wrong (This is in the Difficulties scale of the SCORE 15)		
25	People in the family are nasty to each other	25	People in the family are nasty to each other
27	People slam doors, throw things or make a lot of noise if they are upset		
		29	It feels risky to disagree in our family

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. Comparison of clinical and non-clinical groups on SCORE-28 and SCORE-15 scales.

Variable		Clinical Group N = 199	Non-clinical Group N = 403	t	d 95% CI
SCORE-28					
Total family adjustment	M	2.60	1.91	10.44***	0.91
	SD	1.04	0.58		[0.85, 0.97]
Family strengths	M	2.70	1.78	12.69***	1.10
	SD	1.19	0.59		[1.03, 1.17]
Family difficulties	M	2.30	1.70	7.57***	0.66
	SD	1.14	0.78		[0.58, 0.73]
Family communication	M	2.80	2.24	6.78***	0.59
	SD	1.19	0.81		[0.51, 0.67]
SCORE-15					
Total family adjustment	M	2.62	1.87	10.49***	0.91
	SD	1.11	0.64		[0.85, 0.98]
Family strengths	M	2.77	1.74	12.44***	1.08
	SD	1.32	0.71		[1.00, 1.16]
Family difficulties	M	2.47	1.85	7.34***	0.64
	SD	1.24	0.81		[0.56, 0.72]
Family communication	M	2.62	2.03	6.94***	0.57
	SD	1.22	0.84		[0.52, 0.68]

Note: SCORE = Systemic Clinical Outcomes and Routine Evaluation. Data for non-clinical group are form Table 8.8 in Fay (2010). M = Mean. SD = Standard Deviation. t = t-test statistic. Cohen's d = effect size. ***p<.001

Table 3. Comparison of cases with and without personality disorders on SCORE-28 and SCORE-15 scales.

Scale		Cases with Axis I disorders and Axis II Personality Disorders N = 77	Cases with Axis I Disorders only N = 119	t	d 95% CI
SCORE-28					
Total family adjustment	M	3.11	2.27	5.99***	0.88
	SD	1.07	0.88		[0.75, 1.01]
Family strengths	M	3.16	2.38	4.73***	0.70
	SD	1.21	1.07		[0.54, 0.85]
Family difficulties	M	2.80	2.01	5.02***	0.74
	SD	1.26	0.94		[0.59, 0.89]
Family communication	M	3.38	2.43	5.92***	0.87
	SD	1.18	1.04		[0.72, 1.02]
SCORE-15					
Total family adjustment	M	3.13	2.29	5.54***	0.81
	SD	1.16	0.95		[0.70, 0.96]
Family strengths	M	3.26	2.43	4.46***	0.66
	SD	1.35	1.22		[0.48, 0.83]
Family difficulties	M	3.04	2.13	5.39***	0.79
	SD	1.30	1.05		[0.63, 0.95]
Family communication	M	3.09	2.31	4.55***	0.70
	SD	1.30	1.08		[0.51, 0.83]
PROBLEM SEVERITY & IMPACT					
Main problem severity	M	7.40	6.26	3.33***	0.49
	SD	2.32	2.35		[0.16, 0.82]
Main problem impact	M	6.44	5.74	1.78	0.26
	SD	2.69	2.68		[-0.11, 0.64]

Note: SCORE = Systemic Clinical Outcomes and Routine Evaluation. M = Mean. SD = Standard Deviation. t = statistic. d = Cohen's d effect size. **p<.01.

Table 4. Correlations between SCORE-28 and SCORE-15 scales and the Global Assessment of Functioning Scale (GAF), the Camberwell Assessment of Need Short Appraisal Schedule (CANSAS) and the Childhood Trauma Questionnaire (CTQ).

Variable	GAF r 95% CI	CANSAS r 95% CI	CTQ r 95% CI
SCORE-28			
Total family adjustment	-0.27** [-0.41, -0.14]	0.40** [0.27, 0.53]	0.33** [0.20, 0.46]
Family strengths	-0.18** [-0.31, -0.04]	0.28** [0.14, 0.41]	0.26** [0.13, 0.40]
Family difficulties	-0.30** [-0.44, -0.17]	0.46** [0.33, 0.58]	0.33** [0.20, 0.46]
Family communication	-0.25** [-0.38, -0.11]	0.33** [0.20, 0.47]	0.29** [0.15, 0.42]
SCORE-15			
Total family adjustment	-0.24** [-0.38, -0.11]	0.36** [0.23, 0.50]	0.31** [0.17, 0.44]
Family strengths	-0.17* [-0.31, -0.03]	0.24** [0.10, 0.37]	0.21** [0.07, 0.35]
Family difficulties	-0.28** [-0.42, -0.15]	0.42** [0.30, 0.55]	0.31** [0.17, 0.44]
Family communication	-0.19** [-0.33, -0.05]	0.31** [0.17, 0.44]	0.30** [0.16, 0.43]

Note: SCORE = Systemic Clinical Outcomes and Routine Evaluation. All values are Pearson product-moment correlations. N = 199. * $p < .05$. ** $p < .01$.

Figure 1. Confirmatory Factor Analysis for SCORE-15 and SCORE-28