A Portuguese translation of the Systemic Clinical Outcome and Routine Evaluation (SCORE): the psychometric properties of the 15- and 28-item versions

Margarida Vilaça [iD],[a] Ana Paula Relvas[a] and Peter Stratton[b]

The Systemic Clinical Outcome and Routine Evaluation (SCORE) is one of the most effective evaluation instruments to assess family functioning and to test family therapy results. To determine its applicability in the Portuguese context, we analysed the psychometric characteristics of the short (SCORE-15) and long (SCORE-28) versions using a combined sample of 538 Portuguese participants. Results indicate that both versions demonstrate good internal consistency, replicate the original trifactorial solution and significantly differentiate between clinical and community samples. The SCORE-15 predicted 93.3 per cent of the SCORE-28 results. Both instruments' reference values and cut-off points were established. These findings allow us to claim that the current Portuguese translation provides data that are comparable with, and of the same quality as, those obtained using the English language versions. The Portuguese versions can therefore be recommended to measure family functioning in routine clinical practice and research.

Practitioner points
- SCORE-15 and 28 are effective in the measurement of the family relationships in general and the family strengths, communication and difficulties in particular
- They represent sound and useful tools that may be used for research and intervention purposes, both in clinical and community contexts, when a Portuguese version is required
- The short version is most appropriate for the clinical context due to its conciseness and stronger capacity to predict family functioning

---

---

[a] University of Coimbra. Address for correspondence: Margarida Vilaça, Faculty of Psychology and Educational Sciences, University of Coimbra, Rua do Colégio Novo, 3000-115, Coimbra, Portugal. Email: gui.vilaca@hotmail.com.
[b] University of Leeds.

The Systemic Clinical Outcome and Routine Evaluation (SCORE) is a self-report questionnaire developed by Stratton, Bland, Janes and Lask (2010). The aim of the SCORE, an instrument that is sensitive to therapeutic change, is to provide an evaluation measure of family functioning. Being inspired by social constructionist thinking and by the clinical practice of systemic family and couples therapy, the initial SCORE versions were created from a combined process of a literature review about family functioning evaluation and the considerations of an expert team of family therapists (Stratton *et al.*, 2010; Stratton, McGovern, Wetherell and Farrington, 2006). The SCORE was developed with the main goal of overcoming the limitations of available family measures able to assess the results of systemic therapy and, as such, of providing an easy-to-use outcome measure applicable to current family therapy interventions. It may be used in clinical practice as well as in the investigation of brief systemic family therapy (Carr and Stratton, 2017; Stratton *et al.*, 2010) and in the context of the general population as an indicator of how family members relate to each other (Vilaça, Silva and Relvas, 2014).

Thus far, several SCORE validation studies have been conducted and have led to the reporting of different versions, namely, the SCORE-40, 15, 28 and 29, mainly in English and Irish contexts (Cahill, O'Reilly, Carr, Dooley and Stratton, 2010; Carr and Stratton, 2017; Fay *et al.*, 2011; Hamilton, Carr, Cahill, Cassells and Hartnett, 2015; Stratton *et al.*, 2006, 2010; Stratton *et al.*, 2014). It is also available in a child version of the SCORE-15 for children between 8 and 11 years of age (Jewell, Carr, Stratton, Lask and Eisler, 2013). Among the existing versions, we studied the 15-item short version (SCORE-15) and the more detailed 28-item version (SCORE-28) for application in clinical and research contexts, as these two versions correspond to the most manageable and the most frequently mentioned versions by the authors (Cahill *et al.*, 2010; Hamilton *et al.*, 2015; Stratton et al., 2014). For this purpose, we used the SCORE-29 version (Fay *et al.*, 2013), which incorporates all of the items from the SCORE-15 and SCORE-28.

In the United Kingdom, Stratton *et al.* (2010) conducted the first studies using the SCORE, presenting the SCORE-15 as a refinement of the initial 40-item version. Based on an analysis of a combined sample of 482 participants, three dimensions emerged that were replicated in subsequent studies: family strengths (FS), family communication (FC) and family difficulties (FD). On the whole, the fifteen items, equally distributed among the three dimensions, presented good internal

consistency (Cronbach alpha of .89). Later, Stratton *et al.* (2014) conducted complementary studies testing the validity of the SCORE-15 as a measure of therapeutic change. In this study, conducted with 584 systemic family and couple therapy clients, the authors legitimized the SCORE-15 as a measure that is reliable, valid and sensitive to change during brief therapeutic interventions. Because there are Portuguese speakers throughout the world we decided it was important to have a definitive translation that was fully tested. The European Family Therapy Association (EFTA) provides a detailed protocol for translation, which we followed carefully (see 'Measures', below). It is designed to create a culturally sensitive translation that captures the therapeutic significance of each item. We would not, however, claim that a version created for mainland Portugal will necessarily be culturally appropriate in all Portuguese-speaking communities, but testing a translation and researching which version is most appropriate for family therapy is an essential first step.

Current validation evidence

Fay *et al.* (2011) conducted validation studies of the SCORE-15 and SCORE-28 in an Irish context. In this study, based on a random sample of 403 participants, the authors reported that both the SCORE-15 ($\chi2 = 1179.19$, $df = 17$, $p < .001$, CFI = .92, TLI = .97, RMSEA = .07 and SRMR = .05) and the SCORE-28 ($\chi2 = 1877.04$, $df = 31$, $p < .001$, CFI = .91, TLI = .98, RMSEA = .06 and SRMR = .06) reproduce the original factorial structure. In relation to internal consistency, appropriate reliability values are observed for the SCORE-15 global values and dimensions (total = .83; FS = .76; FC = .58; FD = .71) as well as SCORE-28 (total = .89; FS = .86; FC = .75; FD = .76). More recently, Hamilton *et al.* (2015), in a study of the psychometric properties of the SCORE-15 and 28 that used a combined sample of 701 subjects, concluded that both versions present trifactorial structures similar to the original factor solution (SCORE-15: $\chi2 = 330.81$, $df = 87$, $p < .001$, CFI = .98, TLI = .98, RMSEA = .06 and SRMR = .04; SCORE-28: $\chi2 = 979.024$, $df = 347$, $p < .001$, CFI = .97, TLI = .97, RMSEA = .05 and SRMR = .05). Both scales also show good internal consistency with Cronbach alpha coefficients above .70 for the SCORE-15 (total = .90; FS = .83; FC = .78; FD = .85) and the SCORE-28 (total = .93; FS = .90; FC = .86; FD = .86). The results of a different study of the SCORE-28 (Cahill *et al.*, 2010), which was based

on a combined sample of 791 individuals, also support a trifactorial structure ($\chi 2 = 337.402$, $df = 91$, $p < .001$, CFI $= .92$, TLI $= .98$, RMSEA $= .08$ and SRMR $= .05$) and high internal consistency for the total scale (Cronbach alpha $= .93$) and the respective dimensions (FS $= .91$; FC $= .86$; FD $= .83$). The authors further suggest the distinct use of each version according to its goal. Thus, the SCORE-15 is used when repeated and alternate applications are required, as its administration is less demanding and it provides a reliable, valid and global vision of overall family functioning. In contrast, the use of the SCORE-28 is suggested when a more detailed perspective of family functioning is desired, as this instrument provides a reliable reading of each of the three dimensions that it represents, even though it may be less acceptable to the respondents because it is more time-consuming.

In Portugal, preliminary studies that administered both versions (Pereira, 2011) corroborated the good psychometric properties previously found for both the SCORE-15 and the SCORE-28. This study, based on a community sample of 300 individuals, showed that both versions represent a trifactorial structure similar to that found by the scales' authors (Cahill *et al.*, 2010; Fay *et al.*, 2011; Hamilton *et al.*, 2015; Stratton *et al.*, 2010). With respect to reliability, both the SCORE-15 and 28, achieved acceptable to excellent Cronbach alpha scores for the total scale (SCORE-15 $= .88$; SCORE-28 $= .92$) and the subscales (SCORE-15: FS $= .86$; FC $= .73$; FD $= .80$; SCORE-28: FS $= .92$; FC $= .81$; FD $= .83$). More recently, new studies have been conducted aiming to adapt and validate the SCORE-15 for the Portuguese population on the basis of a combined sample of 513 individuals from clinical and community contexts (Vilaça *et al.*, 2014). The results of the factor analysis, which are consonant with previous studies, demonstrate an adequate adjustment for the original three-factor model ($\chi 2 = 215.082$, $df = 86$, $p < .001$, CFI $= .97$, GFI $= .95$ and RMSEA $= .06$). With respect to scale accuracy, alpha scores indicate a good internal consistency for the total scale (.84) and all three dimensions (FS $= .85$; FC $= .83$; FD $= .82$).

The studies conducted within the English and Irish contexts, with the aim of developing and validating the SCORE-15 and 28, certify both instruments as effective assessment measures of family functioning (Cahill *et al.*, 2010; Fay *et al.*, 2011; Stratton *et al.*, 2010) and therapeutic change (Hamilton *et al.*, 2015; Stratton *et al.*, 2014). Accordingly, we followed the SCORE authors' steps and continued the investigation conducted in Portugal, thereby contributing to the

validation of the Portuguese version of the SCORE-15 and 28 and making new evidence available regarding their use in a culturally distinct context (southern Europe). Specifically, to compare the performances of the SCORE-15 and the SCORE-28 and hence to identify which of the two versions best forecasts family functioning, we analyse the psychometric properties of the two measures, i.e., ability to differentiate populations, factor validity, reliability, construct validity, predictive validity, reference values and cut-off points.

Method

Participants

The combined participant sample involved 538 Portuguese nationals from a community context (75.5%, $n = 406$) and a clinical context (24.5%, $n = 132$), all of whom were aged 12 years or over. To be included in the community sample, the participant could not be receiving psychiatric support or be closely related to any of the investigators, while the clinical sample participants were required to have had family members who had completed the SCORE-29, which corresponded to family and couple therapeutic processes, at the beginning of the first therapy session. The sample was composed of 311 woman (57.8%) and 227 men (42.2%) with an age average of 36.65 ($SD = 16.27$). The most dominant age groups were 12 to 24 (21.7%) and 40 to 49 (21.2%) years. The majority of the participants were married or in a common-law union (50.0%), and 34.2% held a bachelor's degree. The majority of the sample was from the central zone of the country (66.6%), with major emphasis on predominately urban (49.1%) and moderately urban (29.2%) areas. Regarding the socioeconomic status (SES) measurement of professionally active individuals (58.4%), we considered the intersection of data for profession and level of education (Simões, 2000). Thus, the most representative category was that of average SES (35.3%), followed by high SES (21.4%) and then low SES (1.7%). The professionally inactive participant group (unemployed, students and retired individuals) represented 35.1% of the sample. The majority of households were composed of three or four members (53.0%), and the average number of family members per household was 3.34 ($SD = 1.19$). The participants were mainly mothers (29%), fathers (15.1%), and children (33.3%), pertaining mostly to the family life cycle stage of families with adult children (31.2%) and families with adolescents (21.2%). In the specific case of

the clinical sample, the 132 participants corresponded to sixty-one families that sought help to cope with family problems that included relationship difficulties with children (7%) or adolescents (28%), marital problems (28%), children with behavioural problems or problems adapting to school (8%), individual psychological difficulties (8%), family relationship difficulties (8%), co-parenting questions (5%), problems adapting to chronic disease (5%) or reconstituted family difficulties (5%).

Both samples were recruited to represent the range of communities that comprise Portugal, though with an over-representation of urban residents in both groups. We have used the term 'community sample' because we did not screen for clinical condition, and therefore the term 'non-clinical' would not have been appropriate.

Measures

A 29-item version of the SCORE (SCORE-29; Fay *et al.*, 2011), translated into Portuguese, was used, as it contains all the items of the Portuguese versions of the SCORE-15 and SCORE-28. Therefore, it was possible to perform a psychometric analysis of the two versions based on the same sample. The SCORE-29 translation and cross-cultural adaptation was initiated in the framework of the EFTA SCORE project, a research network that aimed to develop, test and validate a self-report family functioning measure, the SCORE, in several European countries. As such, the translation process followed the procedures recommended by the EFTA research SCORE guide (Association for Family Therapy and Systemic Practice, n.d.). First, two Portuguese translations were developed, one by an independent bilingual professional and another by a clinical psychologist. Taking those two initial translations as the starting point, four independent, qualified translators developed new translations and promoted, along with the Portuguese SCORE team, comparison and discussion of the versions and items in a consensus meeting. The agreed Portuguese version was then translated back into English by two independent translators, with the resulting discrepancies being discussed by the Portuguese SCORE team. Finally, the consensus version was pilot tested in a community sample ($N = 21$) that was asked to note any difficulties encountered while completing the questionnaire. Some minor adjustments were performed based on the respondents' comments regarding the readability of specific items.

Both scales (SCORE-15 and 28) include three family functioning dimensions: family strengths, family communication and family difficulties. The last two dimensions are composed of negative items, whereas the family strengths dimension is represented by positive items. With respect to the distribution of the items among the dimensions, the SCORE-15 contains five items per dimension, while the SCORE-28 presents an irregular item distribution by dimension (FS = 13 items, FC = 9 items, and FD = 6 items). The SCORE-15 and SCORE-28 responses are reported on a five-point Likert scale that ranges from 1 (*describes us very well*) to 5 (*describes us not at all*). Although the SCORE-28 usually employs a six-point Likert scale, following the original usage in development of the SCORE- 40, the procedure by which SCORE-15 was generated gave a detailed rationale for use of a five-point scale (Stratton *et al.*, 2010). We have, therefore, used the five-point scale for both versions in the interests of more direct comparability. The total and subscale scores were obtained by reversing the negative items; thus, higher scores correspond to less positive levels of family functioning. Furthermore, in addition to the item answers (quantitative), the SCORE contains a group of questions about the respondent's current state with respect to family difficulties (qualitative), which are not considered in the analysis in this study.

Procedure

As previously mentioned, for this study we recruited a wide group of individuals from two distinct contexts: community (general population) and clinical (clients attending systemic family therapy). Sample collection took place all over the country, including the continent and the islands, between November 2010 and January 2014.

The community subsample (*n* = 406) data were collected in two ways: a paper and pencil version (60.1%, *n* = 244) and an online version (39.9%, *n* = 162). Community participants were recruited by the Portuguese SCORE team and by a group of psychologist trainees enrolled in the Clinical and Health Psychology Master's degree programme at the University of Coimbra, who agreed to collaborate in the sample collection process. In this initial stage, the questionnaires were administered in person by investigators and trainees. Due to the initial difficulties in collecting the necessary responses, the Portuguese SCORE team developed an online version and disseminated it through email and social networks. The clinical subsample (*n* = 132) collection process was implemented in several national therapy

centres. Although general therapy is primarily available in the central part of the country, it is also offered in the north, in Lisbon, and in the autonomous regions of Madeira and Azores.

In addition to the SCORE, all participants also answered a sociodemographic and family information questionnaire that sought to obtain personal data regarding the respondent's gender, age, marital status, profession, number of family members, etc. Individuals who agreed to participate in the study completed an informed consent form that contained information about the study's purposes, a guarantee of confidentiality, an explanation of voluntary participation and an assurance that the data would be used for statistical purposes only. In the case of child participants, if the child assented, the consent form was then signed by the parent, guardian or legal representative. The online participants did not sign a consent form, although they were informed about all the participation conditions described above (APA, 2002). The respondents spent, on average, ten to twenty minutes completing both surveys (the SCORE-29 and the sociodemographic questionnaire).

Statistical analysis

Descriptive statistics of the SCORE-15 and 28 variables were developed for the combined, community-based and clinical-based samples. Independent sample *t*-tests were conducted to determine the existence of significant differences between the subsamples. Subsequently, we calculated effect sizes (Cohen's *d*) for the verified significant differences. To confirm the factor validity of both scales, a confirmatory factor analysis (CFA) of the original structures was conducted. We then checked the factors' internal consistency scores using Cronbach alpha and Spearman correlation coefficients for the SCORE-15 and 28 scales and subscales as both samples failed the normality assumption ($p < .001$). We also computed multiple regressions (stepwise) to ascertain the predictive power of the short scale with regard to the long version. Finally, we explored the normative scores for the different samples by calculating the distribution of percentiles and the obtained results for the possible SCORE punctuations. Additionally, we conducted a receiver operating characteristic analysis (ROC) of both versions to identify the cut-off scores that optimally distinguish between community and clinical participants. To control for family-wise error and to increase the sensitivity of our analyses to identify effects, we chose to adopt a more stringent significance level

of 1% ($p \leq .01$) (Keselman, Cribbie, & Holland, 1999). Analyses were performed with SPSS software (version 20.0) and AMOS (version 4.0.1) for the operating system Windows.

Results

Differences between clinical and community samples

The SCORE-15 and 28 means and standard deviations (total and dimensions), on the basis of the different samples (combined, community and clinical), were calculated to better understand the scales' performances using the data being studied. We conducted *t*-tests to compare the two samples, community and clinical, and to evaluate the capacity of both SCORE versions to differentiate the general population from the clinical population. In relation to the observed results for the community- and clinical-based samples (Table 1), the two total scales and the corresponding subscales presented significant statistical differences between samples ($p < .001$), with the clinical participants presenting higher results (worse family functioning), as expected. The magnitude of the results varied between moderate and large effect sizes (Cohen, 1988). However, although the SCORE total scales and subscales presented similar differences for the short and long versions, with respect to the global scales, we verified that the SCORE-15 generated higher average scores than the SCORE-28 in relation to the combined, community and clinical samples.

TABLE 1 *Comparison between SCORE-15 and 28 global scales and the corresponding subscales in the combined, community-based and clinical-based sample*

	Combined ($N = 538$)		Community ($n = 406$)		Clinical ($n = 132$)			
	Mean	SD	Mean	SD	Mean	SD	t	d
SCORE-15 Total	2.24	0.67	2.10	0.61	2.69	0.65	−9.46*	.94
SCORE-15 FS	2.01	0.78	1.87	0.69	2.41	0.90	−6.23*	.67
SCORE-15 FC	2.34	0.79	2.19	0.75	2.79	0.75	−8.03*	.80
SCORE-15 FD	2.38	0.85	2.23	0.79	2.85	0.85	−7.77*	.76
SCORE-28 Total	2.15	0.60	2.02	0.54	2.55	0.59	−9.68*	.94
SCORE-28 FS	1.94	0.65	1.84	0.59	2.27	0.73	−6.20*	.65
SCORE-28 FC	2.32	0.74	2.17	0.69	2.78	0.71	−8.77*	.87
SCORE-28 FD	2.34	0.84	2.18	0.78	2.83	0.82	−8.20*	.81

Note. FS = Family strengths; FC = Family communication; FD = Family difficulties; SD = Standard deviation; t = value from t-test; *$p < .001$; d = effect size.

TABLE 2 *Statistical fit indices for the original SCORE model*

Version	Trifactorial Model	χ^2	df	CFI	GFI	RMSEA (CI 90%)
SCORE-15	Initial	274.051*	87	.940	.936	.063 (.055; .072)
	Re-specified	226.296*	85	.954	.948	.056 (.047; .064)
SCORE-28	Initial	945.166*	347	.901	.844	.057 (.052; .061)
	Re-specified	817.476*	342	.928	.901	.051 (.046; .055)

Note. *$p < .001$.

The gender differences analysis indicated that the mean scores for men and woman do not differ: the combined version [SCORE-15: $t(536) = -0.707$, $p = .480$; SCORE-28: $t(536) = -0.815$, $p = .415$], the community-based sample [SCORE-15: $t(404) = -1.243$, $p = .215$; SCORE-28: $t(404) = -1.670$, $p = .096$], and the clinical based-sample [SCORE-15: $t(130) = 0.765$, $p = .446$; SCORE-28: $t(130) = 1.210$, $p = .229$].

Factor validity

We proceeded to analyse the internal structure of both SCORE versions to examine the adequacy of the original trifactorial model for the sample data in this study. To do so, we computed a CFA on the basis of the usual procedure of maximum likelihood and using a data covariance matrix. After the model specification and estimation, its adequacy was evaluated by a set of fit indices: the Chi-square index, the Chi-square index divided by the degrees of freedom ($\chi2/df$), the Comparative Fit Index (CFI), the Goodness-of-Fit Index (GFI) and the Root Mean-Squared Error of Approximation (RMSEA) with the respective confidence intervals. Usually, values of $\chi2/df$ under or near 2 are considered reasonably good indicators of fit (Ullman, 2001) – although values under 5 are tolerable (Byrne, 2001). CFI values near to or higher than .95 (Hu and Bentler, 1999), GFI values near to or higher than .90 (Tanaka, 1993) and, finally, RMSEA values under .06 (Byrne, 2001; Maroco, 2010) are acceptable.

Both versions achieved a good match to the values found in previous studies (Cahill *et al*., 2010; Fay *et al*., 2011; Hamilton *et al*., 2015; Stratton *et al*., 2010, Vilaça *et al*., 2014). Despite the fact that both

scales showed similar values, we may highlight a slight superiority of the SCORE-15 model ($\chi2/df = 2.662$; CFI = .954; GFI = .948; RMSEA = .056) over the SCORE-28 model ($\chi2/df = 2.390$; CFI = .928; GFI = .901; RMSEA = .051). To obtain those adjustment indices, both models contained error variances between factors, with higher correlations between family communication and family difficulties, which, in theory, is understandable because both dimensions lead to complexities or less positive aspects of the family functioning. Apart from this adjustment, we assumed the connection among certain items whenever the error variance was high and represented a credible theoretical association, for example, the connection between item 22 ('things always seem to go wrong for my family') and 28 ('we seem to go from one crisis to another in my family') from the family difficulties dimension, which points to the existence of obstacles within the family, emphasizing an external locus of control.

Reliability

The internal consistency of the SCORE-15 and 28 totals and subscales based on the combined, community and clinical samples is presented in Table 3. For both total scales and separate dimensions, consistency was good, with values close or superior to .80, except for the SCORE-15 family communication and family difficulties dimensions (combined, clinical and community sample) and the SCORE-28 family difficulties dimension (clinical sample), which represented lower yet acceptable alpha values (between .68 and .77) (DeVellis, 1991).

On the whole, we may affirm that the alpha values suggest that the items are consistent with the dimensions they belong to, indicating

TABLE 3 *SCORE-15 and 28 Cronbach alpha for combined, community and clinical samples*

	Combined ($N = 538$)	Community ($n = 406$)	Clinic ($n = 132$)
SCORE-15 Total	.88	.87	.85
SCORE-15 FS	.86	.84	.86
SCORE-15 FC	.73	.71	.68
SCORE-15 FD	.78	.76	.77
SCORE-28 Total	.92	.91	.91
SCORE-28 FS	.92	.91	.91
SCORE-28 FC	.83	.81	.80
SCORE-28 FD	.82	.81	.77

Note. FS = Family strengths; FC = Family communication; FD = Family difficulties.

TABLE 4 *Correlations between the SCORE-15 and 28 total scales and subscales based on a combined sample*

	S-15 Total	S-15 FS	S-15 FC	S-15 FD	S-28 Total	S-28 FS	S-28 FC	S-28 FD
SCORE-15								
Total	1							
FS	.729*	1						
FC	.886*	.486*	1					
FD	.872*	.443*	.722*	1				
SCORE-28								
Total	.967*	.768*	.831*	.819*	1			
FS	.712*	.935*	.494*	.453*	.802*	1		
FC	.890*	.542*	.914*	.767*	.894*	.556*	1	
FD	.832*	.391*	.722*	.945*	.798*	.401*	.735*	1

Note. FS = Family strengths; FC = Family communication; FD = Family difficulties; * = The correlation is significant at the $p < .01$ level (2 tailed).

good internal consistency regarding the three dimensions, except for the SCORE-15 family communication and family difficulties dimensions (combined, community and clinical samples) and the SCORE-28 family difficulties dimension (clinical sample) that presented an acceptable internal consistency.

Construct validity

Table 4 displays correlations among both SCORE versions. The correlation between each dimension and the respective total scales based on the combined sample was positive and statistically significant ($p < .001$), with high associations (between .70 and .89) (Pestana and Gageiro, 2008). The highest correlations corresponded to the family communication and family difficulties dimensions, whereas the family strengths dimension presented slightly lower correlations on the global scale. The cross-correlations between the SCORE-15 and 28 regarding the total and the different dimensions results were also positive and significant ($p < .001$), with correlations ranging from low (between .20 and .39) to very high (between .90 and 1) (Pestana and Gageiro, 2008). Overall, family communication and family difficulties dimensions were highly correlated. As expected, the correlation between the SCORE-15 and 28 was very high and positive, emphasizing that both scales measure the same construct: family functioning. We should, however, note that SCORE-28 contains 14 items of SCORE-15 so a high correlation is to be expected.

Predictive validity

We also performed multiple regression procedures with the aim of comprehending how SCORE-15 data represent the information obtained with the 28-items version. The results showed that the SCORE-15 explains 93.3% of the variance in the 28-item version [$R = .966$, $R^2 = .933$; $F(1, 536) = 7456.16$, $p < .001$].

Reference values and cut-off points

Following the studies that aimed to assess the SCORE-15 and 28 norms for the Irish population (Fay *et al.*, 2011), we calculated the norms of the SCORE-15 and 28 results based on the Portuguese sample. Table 5 shows the 5th, 10th, 25th, 50th, 75th, 90th, and 95th percentiles of the short and long SCORE versions, as well as the percentages of cases with average results higher than each scale point (2 = *describes us well*; 3 = *describes us partly*; 4 = *describes us not well*; 5 = *describes us not at all*). Scale point 1 (*describes us very well*) is not mentioned in Table 5, as all individuals score necessarily above that value so the cumulative total is always 100 per cent. The results were calculated based on the general or combined ($N = 538$), community ($n = 406$) and clinical ($n = 132$) samples.

Regarding the percentile examination, the results for the SCORE totals, based on the combined sample, indicated higher values for the 15-items version compared to the 28-item version. The same occurred with the subsamples. For instance, the SCORE-15 exhibited, for the most part, higher values than the SCORE-28, especially after the 25th percentile. An analysis of the results, bearing in mind that each scale point ranges from 1 to 5, indicated that the subscales presented a distinct score distribution, with higher results for family communication and family difficulties.

Figure 1 shows the ROC curves of both versions, calculated to determine cut-off scores that optimize the sum of sensitivity (true positive scores) and specificity (true negatives scores) (Zweig and Campbell, 1993). The results showed that, for both versions, the area under the curve was .75, indicating that the SCORE-15 and 28 should be considered fair tests capable of distinguishing between community and clinical subjects. With regards to the SCORE-15, the optimal cut-off score of 2.4 represented the best trade-off between sensitivity (.72) and specificity (.69), with results indicating that 92 (70%) out of 132 clinical participants scored equal to or above this cut-off score and

TABLE 5 *Reference values for SCORE-15 and 28: percentiles and percentage of cases with scores above each scale point*

Sample	SCORE-15				SCORE-28			
	Total	FS	FC	FD	Total	FS	FC	FD
5th percentile								
Combined	1.27	1.00	1.20	1.20	1.29	1.08	1.22	1.17
Community	1.22	1.00	1.00	1.20	1.25	1.08	1.22	1.17
Clinical	1.58	1.00	1.40	1.53	1.61	1.15	1.53	1.50
10th percentile								
Combined	1.40	1.00	1.40	1.40	1.39	1.15	1.44	1.33
Community	1.33	1.00	1.20	1.40	1.36	1.15	1.33	1.33
Clinical	1.73	1.00	1.80	1.80	1.73	1.31	1.68	1.83
25th percentile								
Combined	1.73	1.40	1.80	1.80	1.68	1.46	1.78	1.67
Community	1.60	1.39	1.60	1.60	1.61	1.38	1.67	1.50
Clinical	2.27	1.80	2.40	2.20	2.18	1.77	2.33	2.21
50th percentile								
Combined	2.20	2.00	2.20	2.20	2.07	1.85	2.22	2.17
Community	2.00	1.80	2.00	2.20	1.95	1.69	2.11	2.00
Clinical	2.73	2.40	2.80	2.80	2.61	2.19	2.78	2.67
75th percentile								
Combined	2.73	2.40	2.80	2.80	2.57	2.31	2.78	2.83
Community	2.47	2.20	2.60	2.80	2.33	2.15	2.56	2.67
Clinical	3.07	3.00	3.20	3.40	2.93	2.77	3.11	3.33
90th percentile								
Combined	3.13	3.00	3.40	3.60	2.96	2.85	3.33	3.50
Community	2.95	2.80	3.20	3.20	2.79	2.69	3.11	3.33
Clinical	3.40	3.60	3.94	4.00	3.28	3.29	3.78	4.00
95th percentile								
Combined	3.40	3.41	3.80	4.00	3.18	3.31	3.78	4.00
Community	3.20	3.13	3.60	3.80	3.02	2.97	3.44	3.67
Clinical	3.94	4.14	4.01	4.40	3.63	3.64	4.11	4.33
% results ≥ 5								
Combined	0.00	0.20	0.00	0.70	0.00	0.00	0.00	0.00
Community	0.00	0.00	0.00	0.70	0.00	0.00	0.00	0.00
Clinical	0.00	0.80	0.00	0.80	0.00	0.00	0.00	0.00
% results ≥ 4								
Combined	1.50	2.40	4.50	6.50	0.40	0.70	2.80	6.70
Community	0.50	1.00	2.70	4.40	0.00	0.00	1.50	3.90
Clinical	4.50	6.80	9.80	12.90	1.50	3.00	6.80	15.20
% results ≥ 3								
Combined	15.10	12.30	22.70	24.70	9.10	7.80	19.70	22.90
Community	9.90	7.90	16.30	18.50	6.20	4.90	13.50	16.70
Clinical	31.10	25.80	42.40	43.90	18.20	16.70	38.60	41.70

TABLE 5 *Continued*

	SCORE-15				SCORE-28			
Sample	Total	FS	FC	FD	Total	FS	FC	FD
% results ≥ 2								
Combined	61.00	50.90	67.50	67.50	55.20	41.80	64.50	65.20
Community	53.40	44.10	61.30	61.30	47.00	35.00	57.10	58.60
Clinical	84.10	72.00	86.40	86.40	80.30	62.90	87.10	85.60

that 281 (69%) out of 406 community participants scored under 2.4. With respect to the SCORE-28, the optimal cut-off point of 2.3 represented the best balance between sensitivity (.73) and specificity (.71), with 94 (71%) out of 132 participants within the clinical group scoring equal to or above 2.3 and 298 (73%) out of 406 participants within the community group scoring under this threshold.

Discussion

The main objective of this research was to examine the psychometric properties of the SCORE-15 and SCORE-28 using univariate and
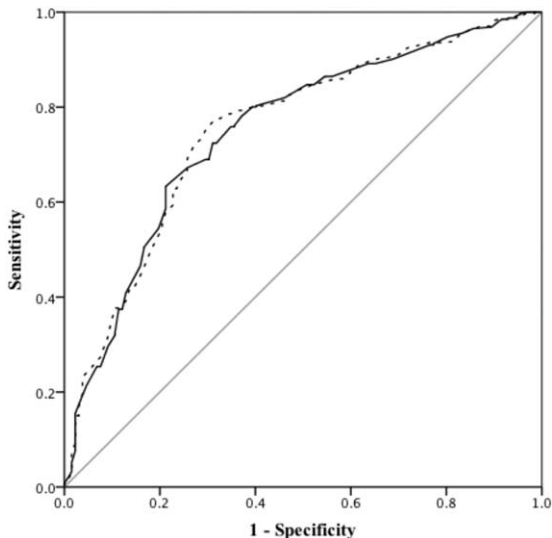

Figure 1. Receiver operating characteristic (ROC) curves for SCORE-15 (black line) and 28 (dotted line).

multivariate exploratory analysis techniques with the aim of validating both versions for the Portuguese population. For this purpose, we focused on examining the scales' ability to differentiate clinical and community contexts and the factor validity, reliability, construct validity, predictive validity and norms of both instruments, as these data allowed us to compare the scales' functioning in greater detail.

An analysis of the scales' ability to distinguish between clinical and community samples reveals that both versions (total scale and subscales) significantly differentiate between these two contexts. The average scores obtained for the total scales in the community (SCORE-15: $M = 2.10$, $SD = 0.61$; SCORE-28: $M = 2.02$, $SD = 0.54$) and the clinical (SCORE-15: $M = 2.69$, $SD = 0.65$; SCORE-28: $M = 2.55$, $SD = 0.59$) contexts are extremely close to those presented in the Hamilton *et al*. (2015) comparison study of community (SCORE-15: $M = 1.87$, $SD = 0.64$; SCORE-28: $M = 1.91$, $SD = 0.58$) and clinical (SCORE-15: $M = 2.64$, $SD = 1.04$; SCORE-28: $M = 2.57$, $SD = 0.91$) groups. Additionally, the SCORE-15 version presents higher average scores than the longer version, as found in previous research using both SCORE versions (Fay *et al*., 2011; Hamilton *et al*., 2015). With respect to the subscales, we found that, on the whole, participants present more positive family functioning patterns for family strengths compared to the other two dimensions, which also was found in the study carried out by Fay *et al*. (2011). These results may signify that both the general population and those who seek or receive therapeutic support experience greater obstacles in terms of family communication and difficulties. As expected, no substantial gender differences were found.

The structural equations analyses performed for both versions confirm that the data conform to the trifactorial model found in previous studies. The alpha values indicate good internal consistency for both scales and the respective subscales, with values similar to those found for the English and Irish samples (Cahill *et al*., 2010; Fay *et al*., 2011; Hamilton *et al*., 2015; Stratton *et al*., 2010, 2014). The slightly lower alpha values for the FC and FD subscales, found in both SCORE-15 and 28, can be expected due to the test's sensitivity to the subscales' length, as the alpha values tend to reduce with short scales, and also due to the subscales' interrelatedness, given that both scales evaluate adverse aspects of the family functioning as difficulties or obstacles (Cortina, 1993). The three subscales that form each SCORE version represent inherent aspects of family functioning. This is supported by the fact that they relate significantly to the value of the corresponding

global version. It should be noted, however, that there exists a high correlation between family communication and difficulties, while they show weaker associations with family strengths. This trend, also present in the work of Fay *et al*. (2010), indicates a possible soft distinction between the first two subscales and family strengths. With respect to the global score, we verified that the correlation between the two forms was strong, as the SCORE-15 predicted 93.3 per cent of the SCORE-28 variance.

Finally, we explored the reference values for the Portuguese population and established the SCORE-15 and 28 cut-off points for both contexts (community and clinical) to promote a richer interpretation of the results obtained from both scales. We then analysed the percentiles and percentages of given results with respect to the different possible scale points. Overall, based on the results, we emphasize the highest values for the SCORE-15 (total and dimensions) and the SCORE-15 and 28 family communication and difficulties dimensions, as these values lead us to consider the need to use higher cut-off points for the SCORE-15 and for these two dimensions. Additionally, if we consider the 90th percentile values found in the Irish research (Fay *et al*., 2011), we find that the global values generated by the Irish combined sample (SCORE-15: 2.92; SCORE-28: 2.86) were close to, albeit lower than, the values found for the Portuguese population (SCORE-15: 3.13; SCORE-28: 2.96). To distinguish healthy from clinical participants, ROC analyses led us to cut-off scores of 2.3 and 2.4, respectively, for the SCORE-28 and SCORE-15 total scales. These findings are consistent with the Fay *et al*. (2011) cut-off values for the SCORE-15 and 28, which were calculated separately for adults (1.9) and adolescents (2.9) and, as such, yield an average cut-off score of 2.4. Considering the relatively small size of our sample, it is noted that it would be inappropriate to draw definitive conclusions and that the interpretation of these values should be the focus of a future investigation or clinical attention.

To conclude, the psychometric qualities of the SCORE-15 and 28 support the results reported in previous studies (Pereira, 2011; Vilaça *et al*., 2014), thus confirming their reliability. Nonetheless, we must consider certain important limitations when examining these results. For example, one of the major limitations of this study lies in the sample characteristics. On the one hand, the community subsample is not the result of a process of a random or stratified sampling; on the other hand, the size of the clinical subsample is relatively small compared to the size of the community sample. Therefore, the results

presented herein should be interpreted with caution. In particular, with regards to the reference values found and cut-off scores, the distinct subsamples' size, as well as the use of a convenience community subsample (rather than a random non-clinical sample), limit the generalization of the Portuguese norms suggested in this study.

Practical implications and further research

Both SCORE versions, 15 and 28, represent usable, practicable and acceptable measures that are free to use and can be used for evaluation, intervention and research purposes in a community context or, for example, in the context of brief systemic family therapy. Thus, two outcome measures of family functioning congruent with systemic thinking are accessible in the Portuguese context, thereby providing current and valid alternatives to more traditional questionnaires that focus on the individual's perspectives. We suggest the use of the shorter version, the SCORE-15, in the clinical context, as it offers good psychometric potentialities, thus making it possible, in a brief, practical, and effective way, to achieve a global view of how family members connect to each other. Accordingly, the 28-item version may be more useful for research settings, as it provides a more detailed picture of the specific constructs measured by the SCORE. As such, however, it requires more time to complete.

The findings from application of the current Portuguese translation to clinical and community samples have demonstrated that it generates data that are comparable with, and of the same standard, as those obtained with the English language versions. This translation can therefore be recommended for use in routine clinical practice and research. More generally, this research contributes to a developing consensus that whenever SCORE is carefully translated according to the established protocol, it is found to be an effective tool for measuring the quality of family functioning in that cultural context.

In keeping with the steps from previous validation studies, future research should focus on the short SCORE version, specifically when analysing the questionnaire's sensitivity to therapeutic change during systemic therapy. This study, which is presently being undertaken in a clinical context, will test the ability of the SCORE-15 to detect the evolution of family functioning in different stages of therapy. These analyses will also enrich the study of the scale norms by establishing specific cut-off points for the context of systemic family therapy. Finally, it is relevant to promote research using the SCORE to focus

on the combination of both the client's perspective (qualitative and quantitative parts) and the therapist's perspective with respect to family functioning during therapy. Consistent with this line of thought, it is essential to validate the Portuguese version of the SCORE for therapists, an aspect not yet studied. For this reason, we propose that this investigation should be performed in the near future.

References

American Psychological Association (2002) *Ethical principles of psychologists and code of conduct*. Retrieved from http://www.apa.org/ethics/code/principles.pdf.

Association for Family Therapy and Systemic Practice (n.d.) *SCORE-15 index of family functioning and change*. Retrieved 12 January 2011, from http://www.aft.org.uk/view/score.html.

Byrne, B. M. (2001) Structural equation modelling with AMOS, EQS, and LISREL: comparative approaches to testing for the factorial validity of a measuring instrument. *International Journal of Testing*, **1**: 55–86.

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.

Carr, A. and Stratton, P. (2017) The SCORE family assessment questionnaire: a decade of progress. *Family Process*, **56**: 1–15.

Cohen, J. (1988) *Statistical power analysis for the behavioural sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Cortina J. (1993) What is coefficient alpha: an examination of theory and applications. *Journal of Applied Psychology*, **78**: 98–104.

DeVellis, R. F. (1991) *Scale development: theory and applications*. Newbury Park, CA: Sage.

Fay, D., Carr, A., O'Reilly, K., Cahill, P., Dooley, B. Guerin, F. and Stratton, P. (2011) Irish norms for the SCORE-15 and 28 from a national telephone survey. *Journal of Family Therapy*, **35**: 24–42.

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

Hu, L. T. and Bentler, P. M. (1999) Cut-off 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*, **52**: 673–684.

Keselman, H. J., Cribbie, R. and Holland, B. (1999) The pairwise multiple comparison multiplicity problem: an alternative approach to familywise and comparisonwise type I error control. *Psychological Methods*, **4**: 58–69.

Maroco, J. (2010) *Análise de equações estruturais* [Structural equations analysis] Lisbon: Escolar.

Pereira, F. (2011) *Estudo de validação da versão portuguesa do SCORE-28 e 15 numa amostra não-clínica* [Validation of Portuguese SCORE-28 and 15 in a community sample] (Unpublished Master's thesis). University of Coimbra, Portugal.

Pestana, M. H. and Gageiro, J. (2008) *Análise de dados para ciências sociais – a complementaridade do SPSS* (5ª ed.) [Data analysis for social sciences – the complementarity of the SPSS (5th ed.)]. Lisbon: Sílabo.

Simões, M. R. (2000) *Investigações no âmbito da aferição nacional do teste das Matrizes Progressivas Coloridas de Raven (M.P.C.R.)* [Investigations in the context of the national standardization of the Raven Coloured Progressive Matrices test (R.C.P.M.)] Lisbon: Fundação Calouste Gulbenkian/Fundação para a Ciência e Tecnologia.

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 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.

Tanaka, J. S. (1993) Multifaceted conceptions of fit in structural equation models. In K. A. Bollen and J. S. Long (eds.) *Testing structural equation models* (pp. 10–39). Newbury Park, CA: Sage.

Ullman, J. B. (2001) Structural equation modeling. In B. G. Tabachnick and L. S. Fidell (eds.) *Using multivariate statistics* (4th ed., pp. 653–771). Needham Heights, MA: Allyn & Bacon.

Vilaça, M., Silva, J. and Relvas, A. P. (2014) Systemic Clinical Outcome Routine Evaluation (SCORE-15). In A. P. Relvas and S. Major (coord.) *Instrumentos de avaliação familiar – funcionamento e intervenção* (Vol. I) [Family assessment instruments – functioning and intervention (Vol. I)] (pp. 23–41). Coimbra: Imprensa da Universidade de Coimbra.

Zweig, M. H. and Campbell, G. (1993) Receiver-operating characteristics (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clinical Chemistry*, **39**: 561–577.