CAMBRIDGE
UNIVERSITY PRESS

The 15-item Systemic Clinical Outcome and Routine Evaluation (SCORE-15) Scale: Portuguese Validation Studies

Margarida Vilaça[1], Bruno de Sousa[1], Peter Stratton[2] and Ana Paula Relvas[1]

[1] *Universidade de Coimbra (Portugal)*
[2] *University of Leeds (UK)*

Abstract. This study reports on the validity of the 15-item Portuguese version of the Systemic Clinical Outcome Routine Evaluation (SCORE-15; Vilaça, Silva, & Relvas, 2014), a brief and comprehensive measure of family functioning. Previous studies with SCORE-15 show that this version replicates the three-factor solution found for the original English version: Family strengths, Family communication and Family difficulties. In addition to reviewing previous studies, this article analyses the discriminant, convergent and predictive validity of the Portuguese SCORE-15. To do so, the SCORE-15 was administered to family members attending systemic family or couple's therapy at the start of the first and fourth sessions and also to a group of non-clinical individuals. Overall, data are reported from 618 participants, including 136 from families attending systemic therapy and 482 community family members. Comparisons of community and clinical samples (discriminant validity) showed statistically significant differences for the total scale and subscales ($p < .001$), with the community participants presenting healthier family functioning than the clinical ones. Analyses using SCORE-15 and the Quality of Life – adult version, another family measure applied simultaneously (convergent validity), indicate that both scales are significantly ($p < .01$) and moderately ($r = -.47$) correlated. Mean score analysis of SCORE-15's therapeutic sensitivity to change (predictive validity) showed that only the Family communication subscale was sensitive to statistically significant improvement ($p < .05$) from session 1 to session 4, whereas the SCORE-15's reliability change index points to its ability to detect clinical improvements (RCI = 14%).

The Systemic Clinical Outcome and Routine Evaluation (SCORE) is a self-report questionnaire for completion by family members aged 12 years and over designed to assess indicators of family functioning that are sensitive to therapeutic change (Stratton, McGovern, Wetherell, & Farrington, 2006). Within a context of increasing pressure to prove family therapy's effectiveness, and given the scarcity of objective measures to evaluate family therapy interventions from the family's perspective (Deacon & Piercy, 2001), the authors aimed to develop an instrument able to measure family functioning and therapeutic change (Janes, 2005; Stratton et al., 2006). Specifically, these authors aimed to create a questionnaire particularly sensitive to changes in family functioning across therapy, to identify quotidian aspects that family members would recognize as most challenging, and to evaluate family functioning markers that change across therapy and as families begin to address their problems (Stratton, Bland, Janes, & Lask, 2010).

From this process, several SCORE versions emerged, designed to be used in both clinical settings and as a research tool, whose development and validation have been described in several articles (Cahill, O'Reilly, Carr, Dooley, & Stratton, 2010; Fay et al., 2013; Hamilton, Carr, Cahill, Cassels, & Hartnett, 2015; Jewell, Carr, Stratton, Lask, & Eisler, 2013; Stratton et al., 2010; Stratton et al., 2014).

Although all the existent versions show good psychometric properties, the 15-item version (SCORE-15; Stratton et al., 2010) has been shown to be "the most practicable version for clinical use" (Stratton et al., 2014, p. 5), indicating key aspects that are relevant to the need for therapy and for therapeutic change. As a briefer version for a group of self-report measures of family processes derived from the original SCORE-40, the SCORE-15 represents the original dimensions and items, explaining 95 per cent of the variance in the means of the 40 items (Stratton et al., 2010) and 92.3 per cent of the results obtained by the SCORE-28 (Vilaça, Stratton, & Relvas, 2015). The SCORE-15 responses are given on a five-point Likert scale, which ranges from 1 = "describes us very well" to 5 = "describes us not at all". The questionnaire also includes open-ended questions about the family

Correspondence concerning this article should be addressed to Margarida Vilaça. University of Coimbra. Faculty of Psychology and Educational Sciences. Rua do Colégio Novo. 3000–115. Coimbra (Portugal). Phone: +35–1916070964.

E-mail: gui.vilaca@hotmail.com

itself, the therapeutic process, clients' expectations and the problem description. The SCORE-15 is composed of a three-factor structure with five items for each family dimension: Family strengths, Family communication and Family difficulties. It yields scores for total SCORE-15 ratings and for each dimension, where higher scores indicate more problematic family functioning.

To sum up, validation of the SCORE-15 could help close gaps between clinical practice and research in systemic family therapy interventions. Previous studies with the Portuguese SCORE-15 indicate that it is a measure of family functioning with very good psychometric properties (internal consistency alpha reliability coefficient for the SCORE-15's total score is .84, for Family strengths is .85, for Family communication is .83 and for Family difficulties is .82) (Vilaça, Silva, & Relvas, 2014). The results also indicate that the Portuguese version of the SCORE-15 represents its original factorial structure: $\chi^2 = 215.082$, $df = 86$, $p < .001$; CFI = .97 (> .9); GFI = .95 (> .9) and RMSEA = .06 (< .07) (Vilaça et al., 2014). Finally, the SCORE-15 takes less than 10 minutes to complete and is free for use.

Continuing recent studies with the Portuguese SCORE-15, this article aims to determine whether the SCORE-15 discriminates adequately between a community and a clinical population (discriminant validity), to analyse how well the SCORE-15 correlates with results on another family adjustment measure, the Quality of Life – adult form (convergent validity), and to evaluate the SCORE-15's sensitivity to therapeutic change (as a possible way of assessing predictive validity). In sum, we address the following three questions: (a) can the SCORE-15 differentiate a community population from a clinical population? (b) does the SCORE-15 correlate with the Quality of Life – adult form? and (c) is the SCORE-15 a valid indicator of therapeutic change?

Method

Participants

The overall sample ($N = 618$) was collected from both contexts: clinical (systemic family therapy clients; $n = 136$) and non-clinical (community sample; $n = 482$), and we followed the same procedures as some of the previous SCORE validation studies (Cahill et al., 2010; Stratton et al., 2014). All participants that completed the protocol between November 2010 and December 2013 were Portuguese.

Sample 1

Discriminant validity study

This sample ($n = 276$) was comprised of both clinical (first session only) ($n = 136$, corresponding to 62 families) and community ($n = 140$) individuals, separated into experimental groups. In all, 156 women (56.5%) and 120 men (43.5%) participated in this study. Ranging in age from 12 to 79 years ($M = 36.79$, $SD = 14.08$), the predominant age groups were 40–49 (27.2%) and 12–24 (25.0%). They came from households in which an average of 3.55 people were living ($SD = 1.09$), and they corresponded mostly to mothers (33.0%), children (30.1%) and fathers (18.8%). With regards to family developmental stages, families with adult children (29.7%) and families with adolescents (29.3%) were the most represented. Participants' highest levels of education were predominantly college graduation (25.0%) or high-school exit certificate (24.3%). With regards to the sample's socioeconomic status, the most representative groups were from average (35.5%) or high (13.0%) socioeconomic levels, with 42.0% professionally inactive participants (students, retired or unemployed). The majority of participants lived in urban areas (48.2%) in the centre of Portugal (73.4%). The two groups were not significantly different on sociodemographic variables: age $t(273) = .812$, $p = .418$, sex $\chi^2 (1) = .001$, $p = .975$ or socioeconomic status $\chi^2 (2) = 2.38$, $p = .304$), but they differed statistically in terms of educational level $\chi^2 (9) = 25.42$, $p = .003$), with higher education in the community group.

Sample 2

Convergent validity study

This study was based on a sample of 430 subjects that completed the SCORE-15 together with a second family adjustment measure, the Quality of Life – adult form (Almeida, 2013), in a clinical (first session only) ($n = 88$, corresponding to 53 families) and community ($n = 342$) context. The participants, 257 women (59.8%) and 173 men (40.2%), ranged in age from 18 to 71 years old ($M = 34.84$, $SD = 11.30$). Respondents' marital statuses were primarily single (45.6%) or married (44.7%), and participants were mostly mothers (56.8%) and fathers (34.1%). On average, families were composed of 3.84 elements, mainly families with adolescents (34.1%) and families with school-age children (33.0%). They lived predominantly in urban areas (82.1%) in the centre of Portugal (54.2%). In terms of the highest levels of education achieved, respondents with licenses (33.2%) or master's (23.7%) degrees predominated. The most represented professionally active participants (71.7%) were from the average (47.0%) and high (17.0%) socioeconomic strata, while the non-active participants (unemployed, retired or students) constituted 28.3% of the sample.

Sample 3

Predictive validity study

The clinical sample used in this study was composed of family-member participants in family or couple's

therapy, who completed the SCORE-15 at the beginning of their first and fourth sessions. The sample consisted of 66 participants, including 39 women (59.1%) and 27 men (40.9%), from 30 families. Age ranged from 12 to 57 years old (*M* = 35.05; *SD* = 13.92). Respondents were mostly mothers (42.4%) and fathers (24.2%) who came from households in which an average of 3.91 family members were living (*SD* = 0.99). Families with adolescents (51.5%) and families with adult children (22.7%) were the most common in this study. In terms of the highest level of education achieved, participants reported from primary school (9.1%) to PhD (4.5%), with the junior cert being the most frequent level (31.8%). The most common represented professionally active participants (57.6%) were from the average (40.9%) and higher (16.7%) socio-economic groups, while the remaining participants (42.4%) were unemployed, retired or students. With regards to residence, most participants lived in predominantly (57.6%) or moderately (36.4%) urban regions, from the centre of Portugal (84.8%) and the Madeira and Azores islands (15.1%). The majority of families sought help to cope with problems of relationship difficulties with adolescents (33.3%), followed by marital problems (23.3%), significant family relationships (13.3%), individual psychological difficulties (10.0%), parenting questions (10.0%), reconstituted family difficulties (6.7%) and adaptation to school problems (3.3%).

Measures

For all three studies, in addition to the Portuguese version of the SCORE-15, a demographic questionnaire was used to identify and characterize participants and their families.

The latter inquires about respondents' personal data (e.g., gender, age, marital status, profession, nationality), but it also inquires about the family's data for use as information about the household. Specifically, to analyse the convergent validity, we used the short version of the Quality of Life – adult form (QOL; Almeida, 2013), an adaptation of the original Quality of Life developed by Olson and Barnes (1982). This 20-item questionnaire evaluates the family's quality of life as a whole, including four subscales with five items each: Family, friends and health, Time, Media and community and Financial well-being. Acceptable psychometric properties have been reported for the use of the QOL (Almeida, 2013). Cronbach's Alpha coefficients for the current combined sample were .85 for the global scale, .72 for Family, friends and health, .79 for Time, .71 for Media and community and .87 for Financial well-being.

The SCORE-15 translation into Portuguese was developed in the context of the European Family Therapy Association (EFTA)'s SCORE project, responsible for the instrument's development and for psychometric studies across Europe. Thus, the translation process, undertaken by Relvas, Vilaça, Sotero, Cunha and Portugal (2010), followed the procedures proposed in the EFTA translation protocol: development of two independent translations, one by a fluent translator in Portuguese and an English translator and the other by a systemic clinical psychologist; four translators with similar linguistic skills developed new translations and contributed to item discussions, resulting in a consensus version. Finally, two independent translators developed two back translations that were finalized by the Portuguese SCORE team. To secure a culturally sensitive adaptation, the agreed-upon Portuguese version was tested in a community sample (*N* = 21) who were asked to record any errors or ambiguities in the item formulations or any other difficulties in fulfilling the questionnaire. As a result of this procedure, some minor adjustments were made to the administered version, namely modifications in the potentially ambiguous item formulations (items 5, 12 and 13) and clarifications in the instructions for completion.

The internal consistency of the SCORE-15 is presented in Table 1. In the discriminant validity study (study 1), Cronbach's alpha reliability coefficients for both samples (community and clinical) ranged from .68 to .87. With the exception of the Family communication subscale of the community and clinical samples, all scales and subscales yielded alphas greater than 0.70. In the study testing convergent validity (study 2), Cronbach's alpha reliability coefficients ranged from .61 to .87, with only the Family communication subscale presenting an alpha under .70, both in the community and the clinical samples. In the predictive validity study (study 3), Cronbach's alpha reliability coefficients at the first and fourth sessions ranged from .62 to .84. In all instances, again with the exception of the Family communication subscale in both sessions, total scales and subscales yielded alphas greater than .70. For total scores and separate domains, the consistency and reliability were generally good, with general higher internal consistencies for the SCORE-15 total scale and Family strengths subscale.

Procedures

As mentioned above, we followed the EFTA suggested procedures, specifically the EFTA research-SCORE guide (Association for Family Therapy and Systemic Practice, UK n.d.), which contains all of the necessary materials to set up SCORE research in a standardized fashion across different countries. The study was conducted with the ethical approval of each involved institution and the informed consent and assent of all participants. Participants were administered the

Table 1. *Internal Consistency, Cronbach's Alpha (95% CI), for Study 1, 2 and 3*

| Variables | Study 1 | | Study 2 | | Study 3 | |
| | Community | Clinical | Community | Clinical | Session 1 | Session 4 |
	(n = 140)	(n = 136)	(n = 342)	(n = 88)	(n = 66)	(n = 66)
S-15 Total	0.84 (0.82–0.84)	0.85 (0.83–0.86)	0.85 (0.83–0.87)	0.87 (0.86–0.88)	0.78 (0.75–0.80)	0.84 (0.82–0.85)
S-15 FS	0.82 (0.76–0.81)	0.87 (0.82–0.86)	0.70 (0.61–0.84)	0.87 (0.84–0.85)	0.82 (0.75–0.83)	0.83 (0.77–0.82)
S-15 FC	0.69 (0.61–0.66)	0.68 (0.58–0.72)	0.61 (0.50–0.77)	0.67 (0.56–0.72)	0.62 (0.46–0.68)	0.64 (0.52–0.66)
S-15 FD	0.70 (0.60–0.72)	0.76 (0.68–0.78)	0.78 (0.70–0.77)	0.80 (0.73–0.80)	0.72 (0.61–0.74)	0.76 (0.70–0.77)

Note: S-15: SCORE-15; FS: Family strengths; FC: Family communication; FD: Family difficulties.

informed consent document containing the study goals, the guarantee of response anonymity and the statement of the voluntary nature of participation. They were assured that their data would be used only for statistical purposes.

The initial original community sample ($n = 406$) was collected through two different procedures: a paper and pencil version and an online version, considering the following inclusion criteria: (i) family members aged 12 years and over, (ii) Portuguese nationality, (iii) not having psychiatric support and (iv) not close friend or family member of the researcher. Given the initial difficulties collecting the protocols through SCORE's paper and pencil version, we decided to develop an online website. Specifically, 244 participants were recruited through the paper and pencil version (60.1%) and 162 were recruited through the online version (39.9%). Subsequently, we stratified this sample according to the clinical comparison group in terms of gender, age and socioeconomic status. For each stratum, a random selection of participants from the initial sample was achieved, resulting in a total of 140 non-clinical cases (see Table 1). To constitute the convergent validity study's sample (study 2), we subsequently initiated new data collection from community subjects that completed simultaneously both the SCORE-15 and the QOL, reaching a total of 342 non-clinical cases, 87.1% through the online version and 12.9% through paper and pencil version.

For the clinical sample collection, seven family therapy centres (six public and one private) collaborated with this investigation. Families and couples who were about to engage in systemic therapy were recruited. Participant institutions were from various regions within Portugal, specifically from the north (1), centre (3), Lisbon (1), Madeira (1) and the Azores (1). Clinicians were asked to approach all families and couples coming to their service by explaining the investigation aim and inviting them to participate to ensure that the sample was representative of their usual clinical

population and that the cases were consecutive rather than selected by therapists. The conducted systemic therapy interventions were based on a second-order epistemological perspective (Golann, 1988) combined with brief-therapy models (6 to 10 sessions, spaced from 3 to 4 weeks). Overall, interventions followed principles of a focus on relationships, an emphasis on strengths and resilience and the establishment of collaborative therapeutic alliances, adopting techniques such as circular questions, reframing, positive connotation and hypothesizing. Sessions occurred in specific therapeutic settings, with two co-therapists in the therapy room and an observing team behind the one-way mirror room. Generally, after approximately 60 minutes, therapists took a 10 to 15 minute break to plan a concluding intervention with the observing team that was delivered to the family at the end of the session. This final intervention to families varied from a simple reflection or reframing to a structured homework prescription. The significant reduction ($n = 70$; 51.5%) in participants from the first ($N = 136$) to the fourth ($N = 66$) session was mainly due to dropout ($n = 61$; 44.9%), while some family members that participated in the first session did not attend the fourth session ($n = 9$; 6.6%).

Overall, the clinical group consisted of family members that completed the SCORE-15 only before the first session (study 1), together with the QOL (study 2), and before the first and fourth sessions (study 3). For the discriminant validity study, the clinical group was composed of participants who completed the SCORE-15 before the first family or couple session because they asked for family therapy support and had not yet received therapy. For the convergent validity study, we included respondents that completed both the SCORE-15 and the QOL before the first therapy session. Finally, for the predictive validity study, we planned to examine therapeutic change by analysing results in the first session (no treatment) and fourth session (after three sessions of therapy) because this latter session has been

indicated as a therapeutic process moment that is supposed to predict short-term therapy outcomes (Lambert & Vermeersch, 2008; Stratton et al., 2014).

Data Analysis

We introduced the data in an anonymized format into an SPSS file and verified the values. There were missing data for 7.4% (46/618) of responses, specifically 18/136 in the clinical and 28/482 in the community sample. Overall, there where were no more than three missing data points (20%) per subject and those presented, on average, as 1.26% missing values in the clinical sample and 1.06% in the community one. In such instances, mean substitution was used. Statistical analyses were performed using version 20 of the Statistical Package for the Social Sciences (SPSS; IBM Corp., 2011).

Results

Can SCORE-15 Discriminate a Community from a Clinical Population?

To address this first research question, we compared the distribution of the average scores across the 140 participants in the community sample and the 136 in the clinical on the five-point scale. An independent-samples Student's *t-test* was conducted to compare the functioning of the scale for clinical and non-clinical datasets. To control for a family-wise error rate, a more stringent cut off value of $p = .01$ was used. Effect sizes comparing the means of both subsamples using Cohen's (1988) formula (d = community mean – clinical mean/pooled SD) were computed.

Results show statistically significant differences between the two groups' total scores (all items) [$t(274) = -9.084$, $p < .001$ (two-tailed)], with the community sample presenting better family functioning (lower scores) than the clinical one (see Table 2). Regarding the Family strengths [$t(274) = -6.487$, $p < .001$ (two-tailed)], communication [$t(274) = -7.719$, $p < .001$ (two-tailed)] and difficulties [$t(274) = -7.343$, $p < .001$ (two-tailed)] subscales, the clinical group participants also scored significantly higher (worse functioning) than the community group, meaning that the clinical group showed more Family difficulties than the non-clinical condition. From Table 2, it may be seen that effect sizes ranged from –0.78 to -1.09. With the exception of the Family strengths subscale, whose effect size was medium ($d < 0.8$), the magnitude of the differences in the total and subscales means was large ($d > 0.8$).

Results indicate that SCORE-15 total and subscales discriminate between the samples, with the therapy clients presenting more difficulties than the non-clinical participants.

We also examined whether the differences between clinical and community samples were influenced by differences in gender, in age or in socioeconomic status distribution in the two samples by performing stepwise linear regression modelling (see Table 3). The results show that there were no significant differences in the total and the subscales scores, between samples as a result of gender, age or social-economic level.

Does SCORE-15 Correlate with QOL?

To answer the second research question, we evaluated the convergent validity of the SCORE-15, total and subscales, by performing Spearman's correlations between the SCORE-15 and the QOL. As indicated in Table 4, total scores on the SCORE-15 correlated moderately and significantly with total QOL, with both scales sharing 22% of their variance. As expected, because lower SCORE results indicated better family functioning, and higher QOL ratings corresponded to higher quality of family life, the correlation was negative, meaning that as family functioning improved, family quality of life also increased. SCORE-15 subscales were also correlated moderately with the QOL Family, friends and health subscale.

Is the SCORE-15 a Valid Indicator of Therapeutic Change?

To address the third research question and to analyse the SCORE-15's sensitivity to therapeutic change, we started by observing SCORE-15 change scores. An inspection of the mean scores indicated that participants in session 1 reported slightly higher results (poorer levels of family functioning) than in session 4, with the exception of the Family strengths dimension (Figure 1).

Thereafter, we assessed improvements in family functioning between session 1 and session 4 by conducting paired *t*-tests. The results (Table 5) showed that only Family communication ($\overline{X}_{session1} - \overline{X}_{session4} = .198$, $p = .049$) presented a statistically significant decrease ($p < .05$) from session 1 to 4, while the global SCORE-15 ($\overline{X}_{session1} - \overline{X}_{session4} = .079$, $p = .202$) and Family difficulties ($\overline{X}_{session1} - \overline{X}_{session4} = .099$, $p = .267$) presented a slight decrease ($p > .05$) following three sessions, meaning that families improved in these dimensions. On the other hand, the Family strengths subscale presented worse results in session 4 than in session 1 (deterioration) ($\overline{X}_{session1} - \overline{X}_{session4} = -.059$, $p = .510$), although for Family strengths, the mean change was lower than those obtained in other variables (Figure 1 and Table 5). From Table 5, it may be seen that the effect sizes ranged from $d = -.08$ to .30. The highest effect size was found for Family communication, although it was moderate. If we compare these results with the English study (Stratton et al., 2014), we can observe that not only were the overall SCORE-15 means results (session 1 = 2.63; session 4 = 2.50) lower than the

Table 2. *Reference Values for SCORE-15 in Clinical (n = 136) and Community (n = 140) Groups*

Variables	Community group M (SD)	Clinical group M (SD)	t	d
S-15 Total	2.01 (.56)	2.67 (.66)	−9.08*	−1.09
S-15 Family strengths	1.78 (.65)	2.39 (.91)	−6.49*	−.78
S-15 Family communication	2.09 (.69)	2.77 (.76)	−7.72*	−.93
S-15 Family difficulties	2.15 (.72)	2.85 (.85)	−7.34*	−.88

Note: S-15: SCORE-15; *p < .001; *d* = effect size; *t* = value from *t*-test.

Table 3. *Standard Multiple Regression Analyses for SCORE-15, Total and Dimensions, Controlling for Gender, Age and Socio-Economic Level*

Independent variables	Dependent variable in model* S-15 total β	p	S-15 FS β	p	S-15 FC β	p	SC-15 FD β	p
Gender	−0.07	.28	−0.04	.55	−0.03	.63	−0.10	.11
Age	0.03	.67	0.03	.73	0.00	.98	0.05	.49
Socio-economic status	0.05	.50	0.02	.81	0.04	.57	0.07	.38

Note: S-15 = SCORE-15; β = standardized beta of independent variables; * Model fit statistics (ANOVA) with *df* = 3 for all models: S-15 total: $F = 0.555$, $p = .65$; S-15 FS: $F = 0.174$, $p = .91$; S-15 FC: $F = 0.211$, $p = .89$; S-15 FD: $F = 1.155$, $p = .33$.

Table 4. *Correlations among SCORE-15 and QOL Totals and Dimensions*

| Variable	1	2	3	4	5	6	7	8
1. S-15 Total	1							
2. S-15 Family strengths	0.79	1						
3. S-15 Family communication	0.91	0.61	1					
4. S-15 Family difficulties	0.89	0.56	0.75	1				
5. QOL Total	**−0.47**	**−0.56**	**−0.33**	**−0.44**	1			
6. QOL Family, friends and health	**−0.63**	**−0.60**	**−0.53**	**−0.55**	0.72	1		
7. QOL Time	**−0.25**	**−0.29**	**−0.16**	**−0.24**	0.74	0.40	1	
8. QOL Media and community	**−0.26**	**−0.23**	**−0.19**	**−0.26**	0.57	0.31	0.27	1
9. QOL Financial well-being	**−0.23**	**−0.22**	**−0.13**	**−0.29**	0.74	0.37	0.39	0.36

Note: S-15 = SCORE-15; Correlations between SCORE-15 and QOL are given in bold; All correlations are significant at *p < .01*.

Portuguese (session 1 = 2.70; session 4 = 2.61), but the Portuguese fourth session scores (2.61) were closest to the English first session results (2.63). In contrast, in a similar study carried out in Ireland (Hamilton et al., 2015) concerning SCORE-15's responsiveness to change between Time 1 (intake) and Time 2 (after 3 to 5 months of systemic therapy), the obtained mean values for the total scale in Time 1 (2.75) were higher than the Portuguese and English ones in the first session, while mean scores from Time 2 (2.53) were closer to the English study results in the fourth session.

To inspect an individual's SCORE-15 results progression, we subtracted session 4 from session 1 global SCORE-15 ratings, where positive results corresponded to improvement, negative results corresponded to deterioration and results equal to zero corresponded to no change. Accordingly, 60.6% (*n* = 40) of the participants showed improvement, 36.4% (*n* = 24) presented deterioration and 3% (*n* = 2) had similar scores at both times.

Given these inconclusive results, based on the mean change significance of differences between sessions, we decided to examine clinical improvement rates based on the global SCORE-15. To do so, we followed the authors' procedures used in a recently published study to evaluate the SCORE-15's sensitivity to change during
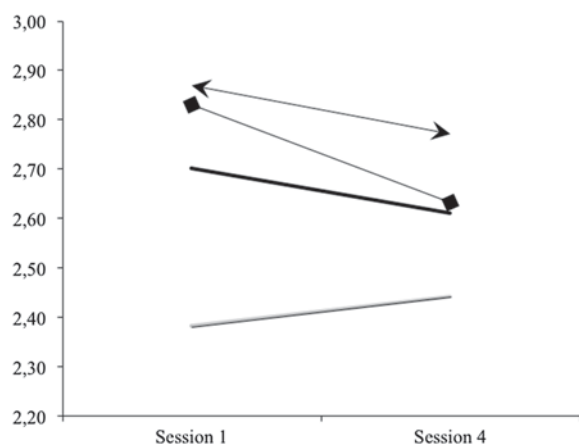
Figure 1. Evolution of SCORE-15 total (bold line), Family strengths (simple line), Family communication (rhombus line) and Family difficulties (arrows line) at Session 1 and Session 4.

Table 5. *Participants' Family Functioning Improvement Patterns at Session 1 and Session 4*

	Session 1	Session 4		
Variables	M (SD)	M (SD)	t	d
S-15 Total	2.70 (.56)	2.61 (.53)	1.29	.17
S-15 Family strengths	2.38 (.84)	2.44 (.73)	−.663	−.08
S-15 Family communication	2.83 (.72)	2.63 (.59)	2.00*	.30
S-15 Family difficulties	2.87 (.79)	2.77 (.71)	1.20	.13

Note: S-15 = SCORE-15; *p < .05; d = effect size; t = value from *t*-test.

the initial phase of therapy (Stratton et al., 2014), and we calculated the percentage of all sixty-six family cases with a reliable change index (RCI). The RCI was developed by Jacobson and Truax (1991) to estimate the clinical significance of change, i.e., in cases of change, whether the magnitude of the change is statistically reliable. This approach is, therefore, an attractive and complementary alternative to the traditional methods of evaluating psychotherapy outcomes (Evans, Margison, & Barkham, 1998). According to Evans et al. (1998), it consists in calculating the standard error of the differences (SE_{diff}) between both measurements (in this case, between the 1st and the 4th session) to subsequently multiply this result with the specific level of 1.96. If an individual's result is greater/less than 1.96 x SE_{diff} (RCI = .727), then the change, improvement/deterioration, is considered statistically reliable (Evans et al., 1998; Jacobson & Truax, 1991). The use of this test-retest reliability to calculate the standard error of measurement allows us to compare the achieved reliable change values with the 5% that would be expected to occur in the retest if there was no intervention at all. Thus, no more

than 5% of RCI values are likely to be due to only measurement unreliability (Evans et al., 1998).

The proportion of participants showing statistically reliable improvement (< -.727) was 7.6% (*n* = 5), while 6.1% (*n* = 4) showed statistically reliable deterioration (> .727). The changes in the remaining 57 participants (86%) were not statistically significant. These results are depicted in Figure 2, where it can be seen that most points from both sessions were close to each other. Nevertheless, given that the proportion of subjects that demonstrated clinically meaningful change (13.7%) was higher than the expected 5%, one may assume that SCORE-15 was able to detect therapeutic change.

Discussion

The current study provided a test of a newly developed practice-based assessment of family functioning, the SCORE-15, in both community and clinical samples. Our primary goal was to analyse the capacity of this questionnaire to differentiate family functioning patterns between community and clinical populations and from the moment after initiating therapy to the middle of the therapeutic process. We also wanted to evaluate if this measure correlated with another corresponding family adjustment scale.

Concurrent Validity

Results indicate that the SCORE-15 total and dimension scores differentiated the community from the clinical population because the non-clinical group presented significantly lower scores (healthier family functioning) than the clinical group. These are expected results if we consider that the community participants did not seek therapy support, while the clinical sample corresponded to participants that asked for family therapy support but were as yet without treatment. These results are similar to those obtained in the Hamilton and colleagues study (Hamilton et al., 2015), whose analyses of total SCORE-15 scores pointed to significantly higher means (*M* = 2.64, *SD* = 1.04) for the clinical group (*N* = 701) than the community one (*N* = 175; *M* = 1.87, *SD* = 0.64). Likewise, a previous study using SCORE-40 (Stratton et al., 2010) also pointed to lower mean scores (*M* = 1.75, *SD* = 0.45) for the community sample (*N* = 126) and higher mean scores (*M* = 2.58, *SD* = 0.64) for the clinical sample (*N* = 482).

We also note that participants' gender, age and socioeconomic status did not impact the results, as expected.

Convergent Validity

Study 2 supports convergent validity because results showed a statistically significant and modest correlation between total SCORE-15 and total QOL, as predicted
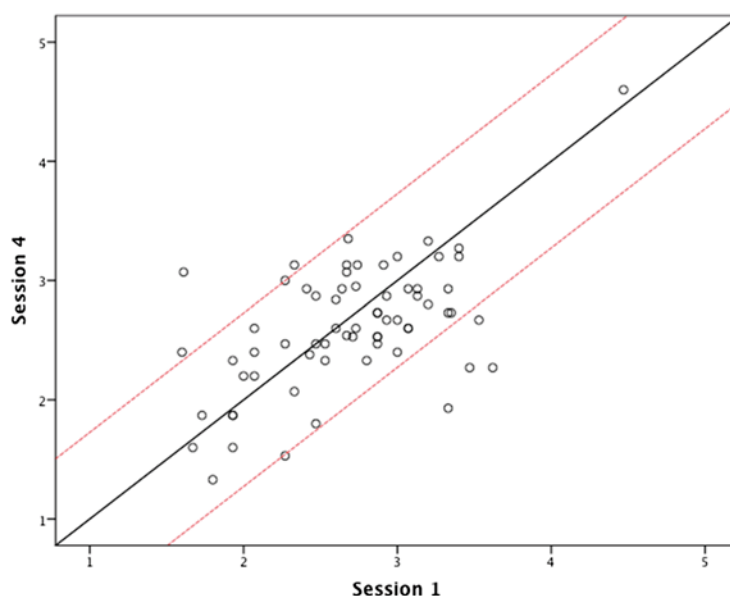
Figure 2. Scatterplot of pre- and post-therapy results on SCORE-15.

because both measures rely on general family well-being. In terms of the measured subscales, the strongest relations were between the SCORE-15 (total and subscales) and the Family, friends and health subscale. This is, in fact, the QOL dimension that most approaches the SCORE-15's item content. In other words, the Family, friends and health subscale addresses key topics covered by the SCORE-15, such as satisfaction with family (item 1) or relationships with relatives (item 4), while the remaining QOL subscales are more related to specific topics, as time, community and economics, that are not central to the SCORE-15.

Predictive Validity

The results, based on a sample of 30 families and 66 participants, indicate that there was no significant improvement from the first to the fourth therapy session if we looked at the SCORE-15 as a whole. Although there was a downward trend in the total mean scores, from session 1 to session 4, pointing to a slight family functioning improvement, a significant improvement in the mean scores was expected. If we consider the dimension mean scores, we can conclude that only the Family communication subscale was sensitive to statistically significant positive change, while the Family difficulties subscale presents such promising slight trend. On the contrary, in the Family strengths subscale the difficulties increased (worse functioning) after three sessions of therapy. In fact, this increase is questionable because it still represents considerably lower values (better functioning) than any other variable mean variable. On the other hand, we might think that this unexpectedly opposite evolution of Family strengths,

after three therapy sessions, may be related to possibly greater family awareness of the required effort to change, which is not reflected in family communication or difficulties. In sum, although there are no statistically significant differences on the total SCORE-15, their dimension mean rates differed between sessions, indicating a significant improvement in Family communication.

Though the total means analysis revealed no statistically significant group change, our analysis of the individuals' global SCORE-15 rates indicate that the majority (60.6%) had evolved positively, with the remaining 36.4% evolving in the opposite way. In terms of the individuals' clinical change, 8% showed reliable clinical improvements and 6% showed reliable clinical deterioration, representing a 14% change, a proportion superior to the 5% given by chance, which seems to confirm the previous SCORE-15 study results trending towards sensitivity to clinical change in early brief therapy, either improvement or deterioration. In the same vein, Stratton and collaborators (2014) showed that participants ($N = 247$) presented similar reliable deterioration (6%) but higher reliable improvement (13%) rates, reflecting a total of 19% reliable change. The Hamilton et al. study (2015), in turn, concluded that of the 258 participants, 41.1% presented with reliable change after 3 to 5 months of systemic therapy.

As previous studies indicate, SCORE-15 may be effectively implemented in a different context from the English-Ireland settings, specifically the Portuguese one (Vilaça et al., 2014). Given the scarcity of family measures in Portugal, this validation study represents the first attempt to investigate the Portuguese SCORE-15's functioning simultaneously in clinical and non-clinical samples. It also constitutes the first study of SCORE-15's

sensitivity to therapeutic change, in this specific context, using RCI for analyses of a clinically significant change. Thus, this research contributes to the reliable application of a measure that focuses not on the individual but highlights relationships and differences between family members in their views of the family. Furthermore, it certifies its use with families or only individuals, allowing it to yield data for clinical practice and research.

The main limitation of this study is a difficulty that is largely associated with studies that comprise clinical samples, especially if we think of systemic family therapy processes. The small clinical sample sizes constitute one important limitation of this study, particularly because results show some important trends, especially in study 3 where we can observe promising family functioning improvement in session 4. It would be interesting to observe if these trends intensify with larger clinical samples (as the ones used in the English and Irish studies) and to follow-up later in therapy. Practical constraints with collecting data from families throughout the therapeutic process have to do with the probable dropout that occurs in clinical settings and the fact that most of the time, family members that participated in one session did not coincide with family members present in the other session. The community sample was collected through the investigators' connections and online, which means that the sample is possibly not representative of the Portuguese population.

The present study also points to a number of areas for further exploration and research, some of which is already under way. Future studies could consider focusing on comparisons of different SCORE versions, for example analysing the largest version results (SCORE-40, -28) and comparing with results from the shorter one (SCORE-15), to evaluate how a larger number of pool items works in the Portuguese context. Studying larger versions would also allow testing for whether there are different item compositions that work better compared with the 15-item version focus of the present study. Likewise, it is important to develop norms for SCORE using clinical and community samples. Furthermore, evaluation of the test-retest reliability among the non-clinical population should be a key priority in future studies with SCORE-15.

Regarding the SCORE in a clinical context, it would be interesting to investigate all family therapy process stages, from the first (start), to the fourth (middle) and last session (end). Thus, it would be possible to understand when families are expected to improve during the therapeutic process and to confirm if results from the middle point of a systemic brief therapy process are predictive of the outcome of therapy, as insinuated by some authors (Lambert & Vermeersch, 2008; Stratton et al., 2014). On the other hand, subsequent investigation should also focus on the exploration of the SCORE's criterion validity, using other outcome self-report measures and also including the therapists' perceptions.

References

Almeida S. (2013). *Escala de Qualidade de Vida Familiar (Quality of Life – QOL): Desenvolvimento de uma versão reduzida para a população portuguesa* [Quality of Life – QOL: Development of the Portuguese short version] (Unpublished master's thesis) Faculty of Psychology and Educational Sciences, University of Coimbra, Portugal. Retrieved from http://hdl.handle.net/10316/24115

Association for Family Therapy and Systemic Practice (n.d.). *SCORE-15 Index of Family Functioning and Change*. Retrieved from author Webpage http://www.aft.org.uk/view/score.html

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. http://dx.doi.org/10.1111/j.1467-6427.2010.00506.x

Cohen J. (1988). *Statistical Power Analysis for the Behavioural Sciences* (2nd Ed.). Hillsdale, NJ: Erlbaum.

Deacon S. A., & Piercy F. P. (2001). Qualitative methods in family evaluation: Creative assessment techniques. *The American Journal of Family Therapy*, *29*, 355–373. http://dx.doi.org/10.1080/01926180127627

Evans C., Margison F., & Barkham M. (1998). The contribution of reliable and clinically significant change methods to evidence-based mental health. *Evidence-Based Mental Health*, *1*, 70–72. http://dx.doi.org/10.1136/ebmh.1.3.70

Fay D., Carr A., O'Reilly K., Cahill P., Dooley B. Guerin F., & Stratton P. (2013). Irish norms for the SCORE-15 and 28 from a national telephone survey. *Journal of Family Therapy*, *35*, 24–42. http://dx.doi.org/10.1111/j.1467-6427.2011.00575.x

Golann S. (1988). On second-order family therapy. *Family Process*, *27*(1), 51–65. http://dx.doi.org/10.1111/j.1545-5300.1988.00051.x

Hamilton E., Carr A., Cahill P., Cassels C., & 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*. http://dx.doi.org/10.1111/famp.12117

IBM, Corp. (2011). *IBM SPSS Statistics for Windows, Version 20.0*. Armonk, NY: IBM Corp.

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. http://dx.doi.org/10.1037/0022-006X.59.1.12

Janes E. (2005). *Self report measures of family function and change following family therapy: A review of conceptual issues, existing measures and proposals for improvement*. Unpublished manuscript, Faculty of Medicine and Health, University of Leeds, United Kingdom. Retrieved from www.psyc.leeds.ac.uk/staff/p.m.stratton/

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*, *52*, 673–684. http://dx.doi.org/10.1111/famp.12044

Lambert J., & Vermeersch A. (2008). Measuring and improving psychotherapy outcome in routine practice. In D. Brown & W. Lent (Eds.), *Handbook of counseling psychology* (4ᵗʰ Ed., pp. 233–248). Hoboken, NJ: John Wiley & Sons Inc.

Olson D., & Barnes H. (1982). Quality of life questionnaire. In D. Olson, H. McCubbin, H. Barnes, A. Larsen, M. Muxen, & M. Wilson (Eds.), *Family inventories: Inventories used in a national survey of families across the family life cycle* (pp. 147–148). St. Paul, MN: University of Minnesota, Family Social Science.

Relvas A. P., Vilaça M., Sotero L., Cunha D., & Portugal A. (2010). Systemic clinical outcomes and routine evaluation: SCORE-15 [Portuguese translation]. (Unpublished instrument)

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. http://dx.doi.org/10.1111/j.1467-6427.2010.00507.x

Stratton P., Lask J., Bland J., Nowotny E., Evans C., Singh R., Janes E., & 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(1), 3–19. http://dx.doi.org/10.1111/1467-6427.12022

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. http://dx.doi.org/10.1002/j.1467-8438.2006.tb00722.x

Vilaça M., Silva J., & Relvas A. P. (2014). Systemic clinical outcome routine evaluation (SCORE-15). In A. P. Relvas & S. Major (Coord.), *Instrumentos de Avaliação Familiar – Funcionamento e Intervenção* (Vol. I) [Family Assessment Instruments – Functioning and Intervention (Vol. I)] (pp. 23–41). Coimbra, Portugal: Imprensa da Universidade de Coimbra.

Vilaça M., Stratton P., & Relvas A. P. (2015). *Comparison of short and long versions of the Systemic Clinical Outcome and Routine Evaluation (SCORE) in a community and clinic-based sample*. (Manuscript submitted for publication).