



Theses and Dissertations

---

2018-12-01

## The Development of a Reliable Change Index and Cutoff for the SCORE-15

Cara Ann Nebeker Adams  
*Brigham Young University*

Follow this and additional works at: <https://scholarsarchive.byu.edu/etd>



Part of the [Social and Behavioral Sciences Commons](#)

---

### BYU ScholarsArchive Citation

Nebeker Adams, Cara Ann, "The Development of a Reliable Change Index and Cutoff for the SCORE-15" (2018). *Theses and Dissertations*. 7694.

<https://scholarsarchive.byu.edu/etd/7694>

This Thesis is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact [ellen\\_amatangelo@byu.edu](mailto:ellen_amatangelo@byu.edu).

The Development of a Reliable Change Index  
and Cutoff for the SCORE-15

Cara Ann Nebeker Adams

A thesis submitted to the faculty of  
Brigham Young University  
in partial fulfillment of the requirements for the degree of  
Master of Science

Richard B. Miller, Chair  
Shayne Anderson  
Angela Bradford

School of Family Life  
Brigham Young University

Copyright 2018 Cara Ann Nebeker Adams

All Rights Reserved

## ABSTRACT

### The Development of a Reliable Change Index and Cutoff for the SCORE-15

Cara Ann Nebeker Adams  
School of Family Life, BYU  
Master of Science

The Systemic Clinical Outcome and Routine Evaluation version 15 (SCORE-15) is an assessment used to assess for clinical change in family functioning. The SCORE-15 has been demonstrated in the past to be a reliable and valid measure for assessing for clinical change and is largely used throughout the UK. However, the SCORE-15 lacks the ability to determine whether an individual's change in family functioning is clinically significant. This study aims to establish a reliable change index and clinical cutoff score based on a US sample so that researchers and clinicians can determine clinically significant change. A sample of 63 clinical participants and 244 community participants completed the SCORE-15, including 165 community participants who completed the SCORE-15 a second time. Results established a cutoff of 51.92 and a reliable change index of 17.51 for the SCORE-15. This indicates that therapy clients who improve their SCORE-15 score by at least 17.5 points and who cross the threshold of 52 during the course of therapy are considered to have experienced clinical significant improvement.

Keywords: systemic therapy, family functioning, SCORE-15, RCI, clinical cut-off

## ACKNOWLEDGMENTS

I would like to thank my chair, Rick Miller, for all his work and faith in me as a researcher and clinician. I would also like to thank my mother and husband for all their late-night help with editing and writing, as well as for their never ending support for me and my pursuits.

## TABLE OF CONTENTS

ABSTRACT.....	ii
ACKNOWLEDGMENTS .....	iii
The Development of a Reliable Change Index and Cutoff for the SCORE-15 .....	1
The SCORE-15 .....	4
SCORE-15 Clinical Cut-off and RCI Scores.....	8
Method .....	9
Procedures.....	9
Clinical Sample.....	9
Community Sample.....	10
Measures .....	11
Analysis.....	12
Results.....	13
Discussion.....	14
Limitations .....	15
Clinical Implications.....	16
References.....	18

## The Development of a Reliable Change Index and Cutoff for the SCORE-15

Assessing change in family therapy is important for two reasons. First, outcome studies using randomized clinical trials of family therapy need to be able to measure change in order to demonstrate the efficacy and effectiveness of family therapy (Carr, 2016; Carr & Stratton, 2017; Sexton, Datchi, Evan, LaFollette, & Wright, 2013; Sprenkle, 2012). That is, families who receive treatment need to experience more change from the beginning of therapy to the end of therapy compared to families who do not seek treatment during that same time period.

Second, there is a robust body of psychotherapy literature that provides evidence that regularly assessing clients' progress during the course of therapy and providing feedback to the therapist significantly improves treatment outcomes (Lambert & Shimokawa, 2011; Reese, Slone, & Miserocchi, 2013). This practice of systematically monitoring clients' progress first began in individual therapy, and random clinical trials have demonstrated that clients whose progress is regularly monitored experience significantly lower rates of premature termination and higher rates of clinically significant improvement than clients who do not (Lambert & Shimokawa, 2011; Reese, Slone, & Miserocchi, 2013).

There is evidence that systematically assessing the progress of couples and families in therapy also improves clinical outcomes. For example, a group of researchers (Anker, Duncan, & Sparks, 2009) randomly assigned 205 couples into either a condition where a systematic feedback system was used each week or one that did not use a feedback system. Results indicated that couples whose progress was systematically monitored experienced, on average, twice as much improvement and were four times more likely than couples in the treatment-as-usual group to reach clinically significant levels of change. These positive results were

subsequently replicated in a similar study (Reese, Toland, Slone, & Norsworthy, 2010). Thus, there is emerging evidence that utilizing a systematic client-feedback driven approach to CFT clients will enhance clinical outcomes (Halford et al., 2012).

Within the process of monitoring improvement from session to session, as well as from the beginning to the conclusion of therapy, it is important to appropriately assess for change for each client. At an individual level, it is not possible to conduct traditional statistical tests to determine significant change. In addition, when assessing change for groups of clients receiving therapy, the traditional method of determining change in therapy involves using statistical tests to determine whether outcome measures from one point in therapy to another are different at a predetermined level of probability. However, knowing that two scores are statistically significantly different provides no information about the magnitude of the difference or whether the differences are clinically meaningful (Jacobson & Truax, 1991). For example, it is possible that change in family functioning during the course of therapy could reach the level of statistical significance, but in a study with a large sample size (Cohen, 1994), the average change being measured could actually be miniscule, with nearly all of the families still being clinically distressed. Thus, statistical significance gives little practical information about clinical improvement that occurs during the course of therapy.

The method of assessing clinical significance has been developed in order to determine the clinical meaningfulness of change in therapy (Ogles, 2013). Jacobson and colleagues (Jacobson, Roberts, Berns, & McGlinchey, 1999; Jacobson & Truax, 1991) define clinical significance as returning to normal functioning. They further stipulate that recovery consists of two criteria; first, clients' level of functioning must exceed the clinical cut-off point of the measure being used. In other words, the clients during the course of therapy move from the

distressed to the non-distressed range of the outcome measure. Second, the magnitude of change needs to be statistically reliable, that is, “beyond the scope of what could reasonably be attributed to chance or measurement error” (Jacobson et al., 1999, p. 300). Classical test theory assumes that an observed score is a mixture of the measurement error (i.e., the portion of the score that is due to factors not associated with family functioning) and the true score (i.e. actual family functioning). A reliable change index (RCI) must be established to increase the confidence that changes in scores across administrations signify real changes in level of functioning. The RCI represents the level of change required to be confident that the difference in scores across administrations is not due to chance or error, but, rather, represents real change (Jacobson & Truax, 1991).

Reflecting the importance of clinical significance in understanding change in therapy, the reporting of the clinical significance, in addition to the statistical significance of change in therapy, is becoming required by a number of clinical mental health journals, including the *Journal of Consulting and Clinical Psychology* (Ogles, 2013). In order to measure the clinical significance of change in family therapy, it is necessary for researchers to use a measure that has a clinical cut-off score and RCI that have been previously calculated. Unfortunately, there has been a lack of research on the clinical significance of family-level measures (Sprenkle, 2012).

In recent years, the SCORE (Systemic Clinical Outcome and Routine Evaluation) has been developed by researchers in the United Kingdom to assess family functioning (Fay et al., 2013; Hamilton & Carr, 2016; Stratton et al., 2014). Further development has led to a 15-item version that is substantially shorter than the original 40-item scale. Psychometric analyses have found the SCORE-15 to be a reliable and valid measure of family functioning (Carr & Stratton, 2017; Fay et al., 2013). In addition, there is evidence that the SCORE-15 is clinically useful



(Carr & Stratton, 2017). Researchers have found it to be a sensitive measure of clinical change (Fay et al., 2013; Hamilton, Carr, Cahill, Cassells, & Hartnett, 2015; Stratton, Bland, Janes, & Lask, 2010), which is important as therapists use it to monitor client change.

Since its development in the 2000's, the use of SCORE has spread dramatically. It is now used throughout Europe, and it has been translated into 22 languages (Józefik, Matusiak, Wolska, & Ułasińska, 2015; Moran, 2017; Vilaça, de Sousa, Stratton, & Relvas, 2015). It has been adopted by the UK Association for Family Therapy (AFT) and the European Family Therapy Association (EFTA) as the main instrument for assessing outcomes in systemic family and couple therapy (O'Hanrahan et al., 2017).

Despite the rapidly increasing utilization of the SCORE-15 in Europe, it has only recently been introduced in the United States. Although an RCI has been developed for families being seen in therapy in the UK, an RCI hasn't been calculated for U.S. samples. Moreover, no valid clinical cut-off score for the SCORE-15 has been developed. With the emerging use of the SCORE-15 in the United States (U.S.), it is important to develop an RCI and clinical cut-off score for the SCORE-15 based on U.S. norms. Consequently, the goal of this paper will be to report on a psychometric study that assessed the clinical-cutoff score and RCI of the SCORE-15 in a US sample.

### **The SCORE-15**

Over the years, several measures have been developed to assess the functioning of family systems, including Family Assessment Device (FAD), Circumplex Model of Family Adaptability and Cohesion Evaluation Scales (FACES), the Beavers Systems Model Self-Report Family Inventory (SFI), and the Family Environment Scale (FES; Hamilton & Carr, 2016.) The FAD has been the most widely used by clinical and nonclinical researchers (Hamilton & Carr, 2016),

with at least 148 studies using it to measure family functioning (Staccini, Tomba, Grandi, & Keitner, 2015). Substantial research has demonstrated its reliability and validity (Mansfield, Keitner, & Dealy, 2015; Ryan, Epstein, Keitner, Miller, & Bishop, 2012). However, the sensitivity of the measure has been called into question because it was designed to screen problem areas of family functioning (Hamilton & Carr, 2016), rather than assess change in family functioning over time. Ogles (2013), in his review of measuring change in psychotherapy, has argued that “Instruments used in the evaluation of change should meet an additional standard beyond reliability and validity. A measure of change must be capable of identifying or detecting differences as they occur over time—they must be sensitive to change” (p. 139).

In response to the need for a measure of family functioning that is sensitive to clinical change, especially in the current environment of measuring progress in family functioning throughout the course of therapy, researchers in the UK developed the SCORE. The development of the SCORE was inspired by the Clinical Outcomes in Routine (CORE) measure, which is widely used in the UK to assess clinical change in individual symptoms. Researchers in the UK recognized the need for a clinical evaluation measure that was appropriate for systemic therapy, where the focus is on relationship functioning, rather than individual symptoms (Stratton et al., 2010; Carr & Stratton, 2017).

The SCORE was developed to measure various characteristics of family functioning, changes in family functioning, and the outcomes of systemic therapy (Stratton et al., 2010; Moran, 2017). The SCORE was also developed to be used regularly in assessing therapeutic progress in clinical practice. The results of a client’s SCORE assessments are used to better educate therapists and clients concerning progress towards therapy goals, as well as to evaluate the effectiveness of therapy (e.g. Cassells et al, 2014).

The development of the SCORE began with a comprehensive review of self-report measures of family functioning, as well as research that examined antecedents of good family functioning (Janes, 2005). The goal in examining other measures and the research on family functioning was to determine which aspects of family functioning should be included in a new measure.

Based on this literature review, the developers of the SCORE developed a questionnaire that contained nine items. They distributed the nine-item version of questions through an email discussion list of the Association of Family Therapy (AFT)—an association for family therapists in the U.K.—for feedback and suggestions. Based on this feedback, as well as the earlier literature review (Janes, 2005), five dimensions of family functioning were derived: danger/hostility, communication, atmosphere/mood, flexibility/adaptability, rules/roles/individuation (Stratton et al., 2010). The researchers then developed three questions for each of the five dimensions, plus an introductory question. This 16-item version was presented for feedback to an assortment of professional audiences, including families, therapists attending workshops, trainees in family therapy courses, and students attending a class on psychological research methods. Respondents either completed this 16-item version of the SCORE for their own family or their family of origin when they were age 16. Participants came from a range of cultures and ethnicities (Stratton et al., 2010).

Based on the responses to this 16-item measure, the measure was refined to create an assessment that could accurately be completed by family members in therapy ages 12 years and older (Stratton et al., 2010). To establish the viability of the measure, this version of the SCORE was then extensively piloted.

Based on advice from statisticians, researchers then increased the variety and language of the original 16 questions to 55 items, 11 for each of the five dimensions of family functioning. The new 55-item SCORE was given to a non-clinical convenience sample. Based on psychometric analysis of the data from this sample, as well as discussions with and feedback from professionals, 15 items were judged to be the least informative. These items were then removed to produce a 40 item version of the SCORE called the SCORE-40. Subsequent factor analysis indicated that the SCORE-40 was best characterized as consisting of three factors, strengths, difficulties, and communication (Stratton et al., 2010).

A shorter 15-item version (SCORE-15) was developed because of the excessive length of the SCORE-40 and the desire to make it more clinically useful (Stratton et al., 2010). Psychometric information from earlier studies on the SCORE-40 was used to pare the instrument to 15 items, with five items representing each of the three factors of the measure.

Subsequent research has found the SCORE-15 to be a reliable and valid measure of family functioning. In a study of 701 families being seen at 19 clinics in Ireland (Hamilton et al., 2015), researchers reported an overall Cronbach alpha of .90, and the alphas for the subscales of strengths, difficulties, and communication were .83, .85, and .78, respectively. The test-retest correlation over a 3-5 month period was .91. The SCORE-15 correlated significantly with the GARF. In addition, the SCORE-15 demonstrated the ability to discriminate between clinical and nonclinical populations, and a confirmatory factor analysis verified the factor structure of the SCORE-15 as maintaining the three dimensions of strengths, difficulties, and communication, with five items loading appropriately on each factor. Results also indicated that the measure was sensitive to change during the course of therapy. Subsequent psychometric studies have provided substantial additional evidence that the SCORE-15 is reliable, valid, and sensitive to clinical

change (Carr & Stratton, 2017; Hamilton & Carr, 2016; O'Hanrahan et al., 2017; Stratton et al., 2014).

Since the development and validation of the SCORE-15, researchers have expanded the reach of the measure. The Child SCORE-15 has been developed for suitable use for children as young as seven years old (Jewell, Carr, Stratton, Lask, & Eisler, 2013). The SCORE-15 has also been modified and validated for use by LGBTQ persons and called the R-SCORE-15 (Teh, Lask, & Stratton, 2017). In addition, the SCORE-15 has been translated into 22 languages (Carr & Stratton, 2017).

### **SCORE-15 Clinical Cut-off and RCI Scores**

Researchers in Ireland have established Irish norms for the SCORE-15 with a clinical cut-off score of 28.5 (Fay et al., 2013). However, the version of the SCORE-15 that they used for their study was based on six response options, rather than the five options that are now used (Carr & Stratton, 2017). Using six response options, the scale has a range of 15 to 90, while the range of the scale is five response options is 15 to 75. With a difference in ranges, the cutoff score derived from the six response option version of the SCORE-15 cannot translate to the five response option version. Consequently, the findings from the Irish study are not generalizable to the current use of the SCORE-15.

Stratton and colleagues (Stratton et al., 2014) reported on a study that assessed the RCI of the SCORE-15, using the more widely used five option version of the SCORE-15. Using data from 584 participants from 239 families who were being seen in family therapy clinics throughout the UK, they used scores from the Cronbach's Alpha of the SCORE-15 and the standard deviation of assessed scores to calculate an RCI of .68. If this RCI score of .68 is multiplied by 15, which represents the number of items in the scale, the RCI is 10.2. The

problem with this calculation is that it used Cronbach's Alpha, a measure of internal reliability, instead of the test-retest method to assess reliability. Jacobson and Truax (1991) argue that using the test-retest reliability score is preferable because it more nearly approximates the temporal nature of clinical assessment, where change is measured across time.

Jacobson and Truax (1991) argue that assessing the percentage of clients who move from the distressed to nondistressed range of functioning is an important factor in determining clinical significance, which requires an established clinical cut-off score of the outcome measure. To date, no clinical cut-off score of the SCORE-15 with its current five response items has been developed. An RCI has been developed among UK families, but it was developed using a method of calculation that is not recommended by Jacobson and Truax (1991). Consequently, this study will address two research questions:

1. What is the clinical cut-off score of the SCORE-15 in the U.S. population?
2. What is the RCI of the SCORE-15 in the U.S. population?

## **Method**

### **Procedures**

**Clinical sample.** Clinical data was collected at intake from three MFT university-based clinics. Twenty-six family cases came from the University of Connecticut, 26 cases came from Florida State University and 11 cases came from Saint Louis University for a complete samples size of 63. 32.7% of the studies participants were male, and 67.3% were female with an average age of 40.38 ( $SD = 5.3$ ). Almost two-fifths (39.3%) of participants were currently single, 7.1% cohabiting, 39.3% married, and 10.7% divorced. Most (78.8%) of the participants self-identified as white, with 13.5% obtaining a HS diploma or GED, 34.6% a Bachelors degree, and 32.7% a Graduate degree.

**Community sample.** Data were collected using a Qualtrics-affiliated survey research panel. Panel members received an invitation from Qualtrics to participate in the research with a link to the study. Those who indicated consent received the survey items. After completing the initial survey, participants were directed to a "thank you" message that contained a link to a directory of couple therapists from the American Association for Marriage and Family Therapy in case responses to questions about their family relationship led to any distress. Eligibility criteria for the study was that the person on the panel be married and have a least one child under the age of 18 living in the household. Those respondents who met these criteria were included in the dataset used for the analyses of this study.

One week following the completion of the first survey, the panel members who responded to the first survey received an invitation to complete a shorter survey containing only the SCORE-15. These additional data were used to establish the test-retest reliability for the SCORE-15. Each Qualtrics panel member had a unique identifier that was used to link the surveys while ensuring their anonymity to the research team. Payment of participants occurred through Qualtrics. The amount and form (cash, frequent flyer miles, etc.) of payment differed among respondents and depended upon their individual arrangements with Qualtrics. Researchers had no knowledge of the participants' identity or contact information.

The community sample collected through Qualtrics included 244 individuals. Participants consisted of 83.6% females and 16.0% males, with a mean age of 36.18 ( $SD = 9.7$ ). All of participants were married, with a mean of 2.13 children per participant. The mean age of the participants' youngest child was 6.05 ( $SD = 4.83$ ). Of all the participants, 78.7 % self-identified as Caucasian, 3.7% as Black or African American, 7.8% as Asian, 7.4% as Hispanic, and 10.6% as other. Participant education levels included 3.3% with education ending at Junior High or less,

27.9% graduating from high school, 4.9% attending vocational/technical School, 21.3% with some college but no degree, 10.2% with an associate or 2 year college degree, 19.3% with a Bachelor degree, and 13.1% with a graduate/professional degree.

The sample for the second wave of data collection consisted of 165 of the original survey participants. Although they received the invitation to take the SCORE-15 a second time one week after they completed the first survey, the participants retook the SCORE-15 an average of 12.10 ( $SD = 2.29$ ) days after they took it the first time. T-tests and chi-square statistics indicated that there were no demographic differences between those in the sample who completed both surveys were no different from those who only took the original survey. However, a t-test comparison of the original SCORE-15 scores of the two groups indicated that those who responded to both surveys were significantly more distressed than those who only responded to the first survey ( $t = 2.45, p < .01$ ).

### Measures

The SCORE-15 is a shortened version of the SCORE-40 (Stratton et al., 2010), which measures overall family functioning. The SCORE-15 consists of fifteen items designed to measure family functioning. The measure consists of three subscales, with five items each: family strengths, difficulties, and communication. A Likert-type five-point response format, ranging from 1 to 5, is used for each item, with responses ranging from “describes us very well” to “doesn’t describe us at all”. Applicable items were reversed scored so that high scores indicate greater difficulties within the respondent’s families. The SCORE-15 correlates highly with the original SCORE 40 and has well established validity (Fay et al., 2013; Hamilton et al., 2015; O’Hanrahan et al., 2017). Previous research has found a Cronbach’s Alpha of .89 for the total scale (Stratton et al., 2014). In this study, the Cronbach’s Alpha of the SCORE-15 for the



community sample was .92 at both the first and second wave of data collection, and it was .87 for the clinical sample.

### Analysis

Based on the conceptualization by Jacobson and Truax (1991), two conditions are required to be met in order to establish the clinical significance of change. First, a participant's score from pre- to post-test needs to pass from the dysfunctional to the functional population, normally across a pre-determined cutoff score. Second, the change needs to be statistically reliable (i.e., large enough to not be due to chance), as measured by a reliable change index (RCI).

Consistent with previous research (Anderson et al., 2014), the cut-off score for the SCORE-15 was calculated using the Jacobson and Truax (1991) recommendation. They used the following equation for determining a cutoff score based on two unequal distributions where  $s$  signifies the standard deviation,  $M$  the mean, and  $0$  or  $1$  the community and clinical samples:

$$c = \frac{s_0 M_1 + s_1 M_0}{s_0 + s_1} \quad (1)$$

Jacobson and Truax (1991) developed the most commonly used method for calculating the RCI (Ogles, 2013). Subsequently, attempts have been made to improve upon the Jacobson and Truax method by using various estimate interval approaches. However, analyses have led to the conclusion that the classic approach introduced by Jacobson and Truax is preferable (Maassen, 2000; Anderson et al., 2014). Jacobson and Truax (1991) have suggested the following equations to determine a RCI:

$$RCI = \frac{x_2 - x_1}{S_{diff}} \quad (2)$$

where  $x_2 - x_1$  represents an individual's change between administrations of the measurement.

$S_{diff}$ , the standard error ( $SE$ ) of the difference between the two scores, is defined in the following equations:

$$S_{diff} = \sqrt{2(S_E)^2} \quad (3)$$

$$S_E = s_1 \sqrt{1 - r_{xx}} \quad (4)$$

$S_{diff}$  accounts for the variation in reliability of the questionnaire. It represents the standard deviation of the clinical population at intake ( $s_1$ ) and the test-re-test reliability ( $r_{xx}$ ) of the assessment in a non-clinical sample. In order for a change to be seen as statistically reliable at the  $p = .05$  level, it must reach the level of 1.96—meaning the change in an individual's scores between administration of the assessment divided by the standard error of the difference between these scores must equal 1.96.

### Results

The test-retest reliability of the community sample was .71. The mean and standard deviation of the SCORE-15 in the clinical sample was 45.54 ( $SD = 11.73$ ), and the mean and standard deviation in the community sample was 58.17 ( $SD = 11.49$ ). Using information in Jacobson and Truax's (1991) formula (Equation 1) resulted in a cutoff score of the summed scale of 51.92. The cutoff score of the average scale, which divides the summed scale by 15, which is the number of items in the scale, is 3.46

Jacobson and Truax's (1991) method for determining reliable change was used to determine the amount of change in the score from pretest to posttest that would be statistically significant at the  $p = .05$  level. We used the community sample test-retest reliability estimate (.71), as well as the standard deviation of the SCORE-15 in the clinical sample as inputs for

Equations 3 and 4. This resulted in a reliable change index of the summed scale of 17.51. The RCI for the averaged scale is 1.17.

### **Discussion**

This study established a clinical cutoff and reliable change index for the SCORE-15 using a clinical sample and a community sample within the United States. An individual whose score on the SCORE-15 moves across the clinical cutoff to a score of 52 and changes by at least 17 points from the first to the most recent administration of the SCORE-15 can be classified as experiencing clinically significant change. Changes of at least 17 points that do not cross the cutoff score can be classified as having reliably improved or reliably deteriorated, based on the direction of the change. With the SCORE-15 having a range of 15-75, an RCI of 17.5 indicates that a person needs to experience a change of at least 29.2% on the RCI in order for it to be considered statistically reliable. With the established clinical cut-off and reliable change index, it would be difficult for any one person to demonstrate change. There is a need to improve the test-retest reliability of the SCORE-15 variable to decrease the RCI.

The RCI of 17.51 that was calculated in this study is substantially higher than the one calculated by Stratton and colleagues in the UK (Stratton et al., 2014), which was .68, or 10.2 if not dividing by the number of items in the scale. The primary reason for the difference is that Stratton et al. (2014) used the Cronbach's Alpha coefficient as their measure for reliability, rather than the test-retest reliability score, which Jacobson and Truax (1991) recommend. A Cronbach's Alpha score of reliability is generally higher than a test-retest score. Consequently, when it is plugged into the equation for the RCI, a higher reliability score will lead to a smaller RCI. This makes sense from a mathematical perspective because the equation calls for the reliability score to be subtracted from 1. Thus, a larger reliability score will lead to a smaller

difference. From a conceptual level, a higher reliability score leading to a smaller RCI makes sense because a higher level of reliability in a measure will require less of a range to establish reliable change. In the case of the current study, if the Cronbach's Alpha score, which was .92, was used instead of the test-retest score of .71, the RCI would drop from 17.51 to 9.48, which is comparable to the RCI of 10.2 that Stratton and colleagues (Stratton et al., 2014) calculated.

### **Limitations**

Within the present study we have used a convenience sample collected from a US based national study, which is not the ideal random community sample, and the clinical sample used is small; these are the main limitations for our proposed RCI and clinical cutoff. Small sample sizes may reduce the confidence level of the present study and increase the margin of error. In the future, researchers should focus on a larger clinical sample to determine if the presented RCI and Clinical cutoff are an accurate depiction of true change. It is also possible, as proposed by Anderson et al. (2014) in regards to the RDAS reliable change index and clinical cutoff development, that families presenting at university-based training clinics for therapy may vary in meaningful ways from those presenting for therapy in other clinical sites. It is also important to note that the clinical sample contained participants with various relationship statuses, and we can only assume what the family structures of the participants in the clinical sample are. When generalizing these results to other clinical settings, this limitation should be considered.

With regards to the difficulty that participants would have demonstrating change with the established clinical-cutoff and reliable change index presented in this study, we believe our community sample had some unique limitations. Participants within Qualtrics research panels could be seen as "professional survey takers". Their responses may not be reliable and or accurate in representing a normative population because the amount of survey they take

compared to other populations. In the future, we would encourage the use of attention-check survey questions to ensure the data quality or that community samples be randomized and collected through the phone, mail, or face-to-face interviews.

### **Clinical Implications**

The SCORE-15 can now be used as a more practical measurement tool in providing systemic therapy to families and couples. Previous research has shown the SCORE-15 has been shown to be sensitive to change early in systemic family therapy and a valid indicator of family functioning and clinical change (Stratton et al., 2014). This study now offers clinicians and researchers alike a way to use these established clinical cutoff and reliable change index scores for the SCORE-15 (which only takes 10 mins for clients to complete) to better measure the effectiveness of treatment and the efficacy of a variety of therapy treatment models. It also assists in describing clinical change (Hamilton et al., 2015). Consistent monitoring of family therapy using measurements like the SCORE-15 also permits clinicians to modify treatment plans when their clients' family functioning is not responding to the treatment given (Lambert & Shimokawa, 2011; Reese et al., 2013).

Using the improved reliable change index and clinical cutoff score presented in our study, based on a US population, also allows researchers who use the SCORE-15 within the US to more correctly measure whether or not the interventions they are analyzing lead to clinical meaningful changes in systemic therapy. Being able to measure true clinical change in treatment also helps to move professionals towards practicing according to results-based accountability standards (Anderson et al., 2014). Consistent with research that demonstrates the positive effects of monitoring client progress (Anker et al., 2009; Reese et al., 2010), we encourage clinicians and researchers to use the information presented in this study to better track the effectiveness of their

clinical work and to provide clients with evidence of the effectiveness of the given treatment. We also hope that as the recognition and use of the SCORE-15 increases, more diverse and larger clinical datasets will be available to further improve the precision and clinical utility of the SCORE-15.

## References

- Anderson, S. R., Tambling, R. B., Huff, S. C., Heafner, J., Johnson, L. N., & Ketring, S. A. (2014). The development of a reliable change index and cutoff for the Revised Dyadic Adjustment Scale. *Journal of Marital and Family Therapy, 40*(4), 525-534.
- Anker, M. G., Duncan, B. L., & Sparks, J. A. (2009). Using client feedback to improve couple therapy outcomes: A randomized clinical trial in a naturalistic setting. *Journal of Consulting and Clinical Psychology, 77*, 693-704.
- Carr, A. (2016). How and why do family and systemic therapies work? *Australian and New Zealand Journal of Family Therapy, 37*(1), 37-55.
- Carr, A., & Stratton, P. (2017). The SCORE family assessment questionnaire: A decade of progress. *Family Process, 56*(2), 285-301.
- Cassells, C., Carr, A., Forrest, M., Fry, J., Beirne, F., Casey, T., & Rooney, B. (2014) A controlled trial of Positive Systemic Practice for families of adolescents with emotional and behavioral problems in Ireland. *Journal of Family Therapy, 26*(1), 5-28. doi: 10.1111/1467-6427.12038
- Cohen, J. (1994). The Earth is round ( $p < .05$ ). *American Psychologist, 49*, 997-1003.
- Fay, D., Carr, A., O'reilly, K., Cahill, P., Dooley, B., Guerin, S., & Stratton, P. (2013). Irish norms for the SCORE-15 and 28 from a national telephone survey. *Journal of Family Therapy, 35*(1), 24-42.
- Halford, W. K., Hayes, S., Christensen, A., Lambert, M., Baucom, D. H., & Atkins, D. C. (2012). Toward making progress feedback an effective common factor in couple therapy. *Behavior Therapy, 43*, 49-60.

- Hamilton, E., & Carr, A. (2016). Systematic Review of Self-Report Family Assessment Measures. *Family Process, 55*(1), 16-30.
- Hamilton, E., Carr, A., Cahill, P., Cassells, C., & Hartnett, D. (2015). Psychometric properties and responsiveness to change of 15-and 28-item versions of the SCORE: A family assessment questionnaire. *Family Process, 54*(3), 454-463.
- Jacobson, N. S., Roberts, L. J., Berns, S. B., & McGlinchey, J. B. (1999). Methods for defining and determining the clinical significance of treatment effects: Description, application, and alternatives. *Journal of Consulting and Clinical Psychology, 67*(3), 300.
- Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*, 12–19.
- 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, University of Leeds, United Kingdom. Retrieved from [www.psyc.leeds.ac.uk/staff/pmstratton](http://www.psyc.leeds.ac.uk/staff/pmstratton).
- Jewell, T., Carr, A., Stratton, P., Lask, J., & Eisler, I. (2013). Development of a children's version of the SCORE index of family function and change. *Family Process, 53*, 673-683.
- Józefik, B., Matusiak, F., Wolska, M., & Ułasińska, R. (2015). Family therapy process—works on the Polish version of SCORE-15 tool. *Psychiatria Polska, 50*(3), 1-13.
- Lambert, M. J., & Shimokawa, K. (2011). Collecting client feedback. *Psychotherapy, 48*, 72-79.
- Maassen, G. H. (2000). Principles of defining reliable change indices. *Journal of Clinical and Experimental Neuropsychology, 22*, 622–632.



- Mansfield, A. K., Keitner, G. I., & Dealy, J. (2015). The family assessment device: An update. *Family Process, 54*(1), 82-93.
- Moran, P. (2017). Selecting self-report outcome measures for use in family and systemic therapy. *Journal of Family Therapy, 39*(1), 41-56.
- Ogles, B. M. (2013). Measuring change in psychotherapy research. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (6 ed., pp. 134-166). Somerset, NJ: Wiley.
- O'Hanrahan, K., Daly White, M., Carr, A., Cahill, P., Keenleyside, M., Fitzhenry, M., ... & McCullagh, A. (2017). Validation of 28 and 15 item versions of the SCORE family assessment questionnaire with adult mental health service users. *Journal of Family Therapy, 39*(1), 4-20.
- Reese, R. J., Slone, N. C., & Misserocchi, K. M., (2013). Using client feedback in psychotherapy from an interpersonal process perspective. *Psychotherapy, 50*, 288-291.
- Reese, R. J., Toland, M. D., Slone, N. C., & Norsworthy, L. A. (2010). Effect of client feedback on couple psychotherapy outcomes. *Psychotherapy, 47*(4), 616–630.
- Ryan, C. E., Epstein, N. B., Keitner, G. I., Miller, I. W., & Bishop, D. S. (2012). *Evaluating and treating families: The McMaster approach*. Routledge.
- Sexton, T. L., Datchi, C., Evans, L., LaFollette, J., & Wright, L. (2013). The effectiveness of couple and family-based clinical interventions. *Bergin and Garfield's handbook of psychotherapy and behavior change, 587-639*. Somerset, NJ: Wiley.
- Sprenkle, D.H. (2012). Intervention research in couple and family therapy: A methodological and substantive review and an introduction to the special issue. *Journal of Marital and Family Therapy, 38*, 3–29. doi:10.1111/j. 1752-0606.2011.00271.x.

- Staccini, L., Tomba, E., Grandi, S., & Keitner, G. I. (2015). The evaluation of family functioning by the family assessment device: A systematic review of studies in adult clinical populations. *Family Process, 54*(1), 94-115.
- 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*(3), 232-258.
- Stratton, P., Lask, J., Bland, J., Nowotny, E., Evans, C., Singh, R., ... & 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.
- Teh, Y. Y., Lask, J., & Stratton, P. (2017). From family to relational SCORE-15: An alternative adult version of a systemic self-report measure for couples and LGB People. *Journal of Family Therapy, 39*(1), 21-40.
- Vilaça, M., de Sousa, B., Stratton, P., & Relvas, A. P. (2015). The 15-item Systemic Clinical Outcome and Routine Evaluation (SCORE-15) scale: Portuguese validation studies. *The Spanish Journal of Psychology, 18*, E87. doi:10.1017/sip.2015.95