7-30-2012

Laboratory Technical Supplement for the Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT)

John D. Mayer
University of New Hampshire, Durham, jack.mayer@unh.edu

Laboratory Technical Supplement for the Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT)

John D. Mayer

30 July 2012

Author Disclosure: The authors of the MSCEIT receive a royalty on a share of the sales of the MSCEIT, which is owned and copyrighted by MultiHealth Systems (MHS) of Toronto, Canada.

Introduction

The Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT) User's Manual (Mayer, Salovey & Caruso, 2002), published by MHS, provides a great deal of information about using the MSCEIT test. Since the publication of the MHS manual, additional issues have arisen within the test-user and research communities associated with the test. This brief technical document addresses two of those concerns.

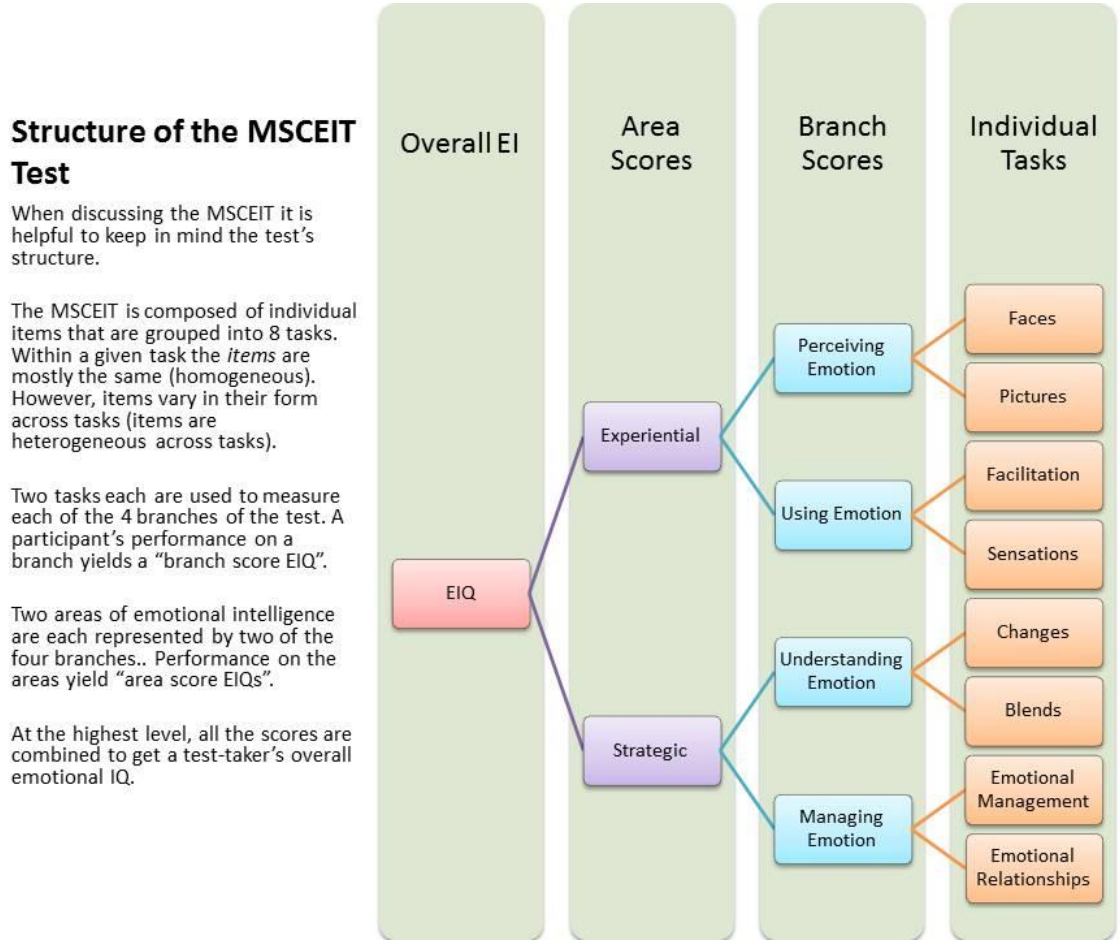
This laboratory technical supplement covers two topics that arise regarding the Mayer-Salovey-Caruso Emotional Intelligence Test: how to calculate a split half reliability for the test, and how to develop and score the test using local norms (i.e., national or regional norms, or norms for special populations). The use of the split-half reliability estimate is important with the MSCEIT because our research indicates that coefficient alpha reliabilities underestimate the reliability of the test. An in-depth discussion of this issue will appear in a forthcoming set of papers in the journal *Emotion.* The development of national and specialized norms are important to interpreting test scores outside of general populations in North America, i.e., for translated versions of the scale and for English-language versions of the scale administered outside of North America, England, Australia, and South Africa (areas that are part of the original standardization sample.

This technical report is a slightly edited and rewritten version of two web pages that were developed originally at www.unh.edu/emotional_intelligence: http://www.unh.edu/emotional_intelligence/ei%20About%20the%20MSCEIT/eiMSCEIT%20loc

alnorms.htm, and
http://www.unh.edu/emotional_intelligence/ei%20About%20the%20MSCEIT/eiMSCEIT%20Calculating%20Reliability.htm

Part 1: The Plan of the MSCEIT

For the following discussions of reliability and norms, it is helpful to have an overview of the structure of the MSCEIT test, because the text refers to such aspects of the test as its "Individual tasks," "Branches," and Overall EIQ score. The figure below provides an overview of the organization of the MSCEIT task from the level of its individual tasks (each composed of multiple test items), to the overall MSCEIT EIQ score.



Part 2: How to Calculate the Reliabilities for the MSCEIT

The MSCEIT Reliability is Best Estimated with Split-Half Reliability Estimates

Many researchers in psychology have gotten into the habit of using coefficient alpha reliabilities for estimating the reliability of the tests they employ, and indeed, coefficient alpha often provides a wonderful estimate of reliability -- when items are homogenous (i.e., all the same in their nature). Coefficient alpha, however, provides an inappropriate reliability to report for the

MSCEIT because the branches and full-scale scores of the MSCEIT are based on items which vary task-by-task. In statistical terms, this means that the items are heterogeneous. The standard coefficient alpha reliability estimate was not designed for use with such heterogeneous items. For that reason, our research group always uses split-half reliabilities when reporting internal-consistency reliabilities for the MSCEIT (with the exception of *within* a single, individual task, where items are homogeneous). An fuller discussion of this issue can be found in Mayer, Salovey & Caruso (in press).

For that reason, the reliability of the MSCEIT is best estimated by using a split-half reliability estimate. "Best estimated" in this case, refers to employing a relatively basic technique that produces and fairly accurate estimate. There may be alternative specialized reliability estimation techniques that are equally good or represent slight improvements in such estimates.

Which Item Responses are Relevant

At the basic level, when people respond to the MSCEIT, they enter various responses for each item (e.g., 1, 4, 3, etc.). These raw responses are not scored responses, and so reliability should not be estimated for them. Any reliability they possess would reflect individual differences in the use of continuous response scales on Branches 1, 2, and 4, and would not reflect the reliability of measured emotional intelligence. (Branch 3 is multiple choice and the reliability estimate for such items would be even more difficult to interpret). Rather, the reliability calculation should be applied to the scored data. Once the data are scored, individual item data is typically reported as fractional values (e.g., .32, .45, .56, etc.). Those are the values that are employed with the split half estimation method (or any other method of estimating reliability).

Assigning Items to "Half Tests"

The key to calculating a split-half reliability estimate for a test with heterogeneous items is to assign equivalent items to the two test halves. Put another way, the aim is to create two versions of the test that can be thought of as, "Half-test A" and "Half-test B," where the two halves are as parallel to one another as possible – that is, as similar to one another as possible in their item compositions.

To create these two halves for the MSCEIT, the best and easiest plan is to divide (i.e., assign) items from a given task equally to "Half-test A" and "Half-test B" – the test halves that will ultimately form the bases of the split half estimate.

So, let's say a researcher wanted to know the reliability of Branch 3 on the MSCEIT, which is made up of subtests C (20 items) and G (12 items). The researcher would first create a summed score of all the odd items across *both* tasks (i.e., C1 + C3 + C5... + ...C19 + G1 + G3...G11), and then a second summed score of all the even items across *both* tasks (i.e., C2 + C4 + C6... + ...C20 + G2 + G4... G12). These would form the "Half-test A" and "Half-test B" for Branch 3.

The correlation between those two sums equals (by classical true-score theory) the reliability of *half* of Branch 3. To obtain a reliability estimate for all of Branch 3, it is necessary to apply the Spearman-Brown Prophecy formula. The specific correction, in this instance, is that the reliability of the whole test (all of Branch 3) is equal to twice the reliability of the half-test, divided by 1 plus the reliability of the half test. (This formula is available in most psychometric texts; see, for example, Nunnally, or Allen & Yen.)

To calculate the reliability for two Branches together (i.e., an area score), one simply generalizes the procedure, adding in items from additional tasks, including, for the Strategic EI area, odd and even items from Tasks D and H. One continues this process, adding in items from the whole test, to estimate reliability of the full test. (Again, after obtaining the correlation between the two test halves, it is necessary to correct upward with the Spearman-Brown formula.)

Test-retest reliability is also appropriate with the MSCEIT, and can be estimated as the simple correlation between the MSCEIT given to the same participants at two different points in time.

Part 3: Developing National, Local and/or Specialized Norms for the MSCEIT

Developing Norms for Special Groups

Introduction

The MSCEIT norms described in the test's User's Manual as well as norms provided in other published scientific articles on the test, provide comparison scores for most test-takers. Some researchers may want to develop further sets of norms, for example, for other countries, for regional areas, or for special groups such as individuals with depression or autism.

On some occasions, the research group at MHS (the test's publisher) may be interested in developing norms for a special group and researchers in psychology may first want to contact the research division of MHS to find out if they have such interests.
If the MHS researchers do not have an interest in creating special norms for a particular group, there is still a do-it-yourself option.

The Do-It-Yourself Approach

New Norms. To use the scoring of the MSCEIT as-is among your special population, you would simply administer the test to your group, have it scored in the normal fashion by MHS, and then get a new set of means and standard deviations for each area of the test. This approach will work so long as you have good reason to believe that the expert and/or consensus-identified correct answers apply to your group as much as they do to other groups.

Recalibrating the Consensus Scoring. Say, on the other hand, that members of your special group might have reason to identify answers other than the expert-identified answers as correct, or to weight good answers differently than had the experts. (Perhaps, for example, the cultural context in which your participants operate has a number of teachings about emotion that are quite different from the teachings of those teachings known to the subject-matter-experts).

In that case, you may want to develop your own consensus-scoring norms independently of MHS. In that case, you would need some of the published booklets from MHS to test people with, and some answer sheets. (The more booklets, the more people you can test in a single setting). You could then hand-enter data from the answer sheets for the 141 items of the MSCEIT into the data file of a program like SPSS. After you obtained about 300 responses or so, you would run the Frequencies program. That would tell you how many people responded to each of the alternatives for a given item.

As an example, consider the first item on the MSCEIT, named "RR01" for raw response to item 1. We also will need a "CNS01" for consensus-scored item 1. Say that:

10% of the people chose 1
15% of the people chose 2
20% of the people chose 3
50% of the people chose 4, and
5% of the people chose 5.

If you are using SPSS, you could write a set of statements that would score Let's say you named the raw response for item 1 "RR01," and "CNS01" will be the consensus-scored version. You would write a compute statement that looked something like this:

```
COMPUTE CNS01 = 0.
IF (RR01 EQ 1) CNS01 = 0.1
IF (RR01 EQ 2) CNS02 = 0.15
IF (RR01 EQ 3) CNS03 = 0.20
IF (RR01 EQ 4) CNS04 = 0.5
IF (RR01 EQ 5) CNS05 = 0.05
EXECUTE.
```

Note the proportions reflect the above percentages of those who endorsed each alternative. Next, you would repeat this process for every item on the test (incrementing the "01" to "02", etc., item-by-item).

Once you had the item-by-item consensus scores, you would add up (via more compute statements) items on each scale to get the scale scores. That would be your consensus scoring. Note that item analyses of the standardization sample indicated that 19 of the 141 items on the MSCEIT should be omitted. In your new standardization sample you may want to conduct some item analyses with an eye toward omitting any items that fail to work in your sample.

The more participants, the more accurate the scoring would become. That is, we recommend updating your consensus scoring algorithm every few hundred participants (and less frequently, after a thousand individuals).

To get EIQ scores, you would convert the obtained scores to z-scores, and then multiply them by 15 and add 100.

The process used at MHS includes a normalization procedure; this ensures a closer-to-normal distribution for scores; methods for doing this are available in psychometrics and statistical textbooks.

If you choose to develop "in-house" scoring, the above procedure should provide you with a good proxy for normed scores. That is to say, the scale you have obtained will be good for correlating with criteria.

References

Mayer, J. D., Salovey, P., & Caruso, D. R. (2002). *MSCEIT Users Manual.* MHS: Toronto.

Mayer, J.D., Salovey, P., & Caruso, D. R. (in press). The Validity of the MSCEIT: Additional Analyses and Evidence. *Emotion Review.*