Emotional Intelligence: Examining Construct Validity Using the Emotional Stroop

Scott L. Martin*
Justin Thomas
College of Business Sciences, Zayed University
PO Box 478, Abu Dhabi, UAE
Email: scott.martin@zu.ac.ae*, Fax: 9712 4434847*

Abstract

Measures of emotional intelligence (EI) have been linked to a variety of important outcome variables, but there has been little empirical research examining the relationship between EI measures and specific cognitive-emotional processes. This study examined the relationship between the Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT) and the Emotional Stroop. Results indicated that the MSCEIT was related to the Emotional Stroop and added incremental validity beyond general intelligence.

Keywords: Emotion, emotional intelligence, measurement, emotional stroop, cognitive processes, Middle East

1. Need for Construct Validity

Emotional intelligence (EI) has been defined as the ability to perceive, understand and regulate emotions in oneself and others (Salovey & Mayer, 1990; Mayer & Salovey, 1997; Mayer, Salovey, & Caruso, 2008). A significant body of research has accumulated demonstrating that various measures of EI are correlated with a range of important criteria such as academic success (Schutte, Malouff, Hall, Haggerty, Cooper, Golden, & Dornheim, 1998), interpersonal relationships (Brackett, Rivers, Shiffman, Lerner, & Salovey, 2006), individual and team task performance (Carmeli & Josman, 2006; Law, Wong, & Song, 2004; Offermann, Bailey, Vasilopoulos, Seal, & Sass, 2004) and leadership (Kerr, Garvin, & Heaton, 2006; Rosete & Ciarrochi, 2005). However, the topic of emotional intelligence remains somewhat controversial (e.g., Ashkanasy & Daus, 2005; Davies, Stankov, & Roberts, 1998; Joseph & Newman, 2010; Landy, 2005; Locke, 2005). A central question is whether EI is simply another label for or aspect of general intelligence. That is, there is debate regarding the “construct” associated with EI. Research establishing relationships between measures of EI and outcome variables such as academic success or job performance is certainly valuable, but such research does not shed direct light on the construct of EI. It is important to conduct research demonstrating that EI measures are linked to specific emotional-cognitive processes, and that such measures provide incremental validity beyond traditional measures of intelligence. Aside from a few exceptions (e.g., Ciarrochi, Chan, & Caputi, 2000; Mayer, DiPaolo, & Salovey, 1990), there has been relatively little empirical research examining the specific cognitive processes associated with EI. Accordingly, this study examined the relationship between a measure of EI and the cognitive processing of emotion-related stimuli. We also examined whether the EI measure explained incremental variance beyond that accounted for by a traditional measure of intelligence.

2. Cognitive Processing and Emotional Stimuli

Emotions can influence physiological, affective, cognitive and behavioral systems within half a second (Lord & Harvey, 2002). An individual’s initial response to emotion-related experiences generally plays a significant role in an overall evaluation of emotional maturity or capability. The speed of initial emotion-related processes precludes the use of controlled or purposeful information processing. As a result, the effectiveness of automated processing regarding emotions is critical to the successful management of emotions. A measure of automated processing that has a long history in psychology is the Stroop task (1935).

The original Stroop task presents subjects with names of colors and requires participants to name the color of the ink in which the word is written. In one condition, referred to as the “congruent” condition, the names of the colors are the same as the color of the ink. For example, the word blue is written in blue ink. In a second condition, referred to as the “incongruent” condition, the names of the colors do not match the color of the ink. For instance, the word blue is written in red ink. Participants are tasked with naming the ink color of each word as quickly and accurately as possible. This original Stroop paradigm consistently reports slower response times in the color incongruent condition (MacLeod, 1991).
This is typically explained in terms of an interference effect caused by the need to suppress an automatic reading response. The Stroop task has been extended to many other contexts including the assessment of emotions and emotional disorders. The Emotional Stroop requires participants to name the ink colors of both neutral and affective words. The affective words are typically associated with anxiety or dysphoria. The findings from such studies have consistently indicated that individuals with lower levels of emotional functioning tend to have slower response times to such affective stimuli (e.g., Gotlib & McCann, 1984; Mittershiffthaler, Williams, Walsh, Cleare, Donaldson, Scott et al., 2008). Williams, Mathews, and MacLeod (1996) explain these delayed color naming responses for affective stimuli in terms of an attentional bias. That is, dysphoria and anxiety related words are typically more salient for individuals with lower levels of emotional functioning, and this diverts cognitive resources from the task to name the color of the ink. Additionally, the disproportionate attention given to such stimuli is thought to lead to further emotional disturbance. Thus, the Emotional Stroop would appear to serve as a useful assessment of automated processing related to emotions. Emotional intelligence involves the effective understanding and management of emotions. Over time, it is logical to expect those high in EI to have a “healthier” perspective regarding emotions so that emotional stimuli are less likely to interfere with one’s ability to focus on the task at hand. Thus, we would expect more emotionally intelligent individuals to perform better on the Emotional Stroop. In this study, emotional intelligence was measured using the Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT) Version 2.0, which is the primary task-based measure of EI and was developed by researchers who are generally recognized as the main proponents of emotional intelligence (Mayer et al., 2008). The MSCEIT is described in detail in the Method section.

Effective performance on the MSCEIT is reflected in higher scores, whereas effective performance on the Emotional Stroop is reflected in lower scores (i.e., faster response times or shorter response latencies). As a result, we predicted the following:

**Hypothesis 1:** The MSCEIT will be negatively correlated with performance on the Emotional Stroop.

Given the Mayer et al. (2008) view that EI is a unique ability, we also examined whether the EI measures added incremental validity over and above that explained by a traditional measure of intelligence. We predicted:

**Hypothesis 2:** The MSCEIT will have incremental validity beyond general intelligence in predicting performance on the Emotional Stroop.

### 3. Method

#### 3.1. Subjects

The subjects were 87 undergraduate students enrolled in a mid-size university in the United Arab Emirates. The students were enrolled in either management or human resource courses and received optional research credit for participating in the study. The research was conducted on an all-female campus of the university so all subjects were women. The subjects ranged from age 19 to 28 ($M = 21.2$, $SD = 1.3$). The native language for all students was Arabic, but all students were fluent in English.

#### 3.2. Measures and Procedures

There were three measures used for this investigation and they were administered over a three-week period. The first test administered was the MSCEIT Version 2.0. The test measures four skills: perceiving emotions, using emotions to facilitate thought, understanding emotions, and managing emotions. Each of these skills is measured by two different tasks for a total of eight sections on the test. Perceiving emotions is measured by showing faces (Faces Task) and pictures of landscapes or abstract designs (Pictures Task) and asking how much of a given emotion is displayed in the face or picture. Using emotions to facilitate thought is measured by indicating the sensations that are created by different emotions (Sensations Task) and how certain moods can help with cognitive tasks or behaviors (Facilitation Task). A sample item is: “What moods might be helpful to feel when creating new, exciting decorations for a birthday party?” Various emotions are presented such as “Joy” and the respondent answers using a five-point scale ranging from “Not Useful” to “Useful.” Understanding emotions is measured by indicating how emotions may combine to form other emotions (Blends Task), and how emotions change over time (Changes Task). A sample item is: “Marjorie felt more and more ashamed, and began to feel worthless. She then felt...” and the respondent is presented with the following response options: “overwhelmed, depressed, ashamed, self-conscious, jittery.”
Managing emotions is assessed by having respondents indicate the actions that are most effective for obtaining a given emotion (Emotion Management Task) or for managing another person’s feelings (Social Management Task). A sample item is: “Mara woke up feeling pretty well. She had slept well, felt well rested, and had no particular cares or concerns. How well would each action help her preserve her mood?” The respondents are then presented with a series of actions such as “She used the positive feeling to call her mother, who had been depressed, and tried to cheer her up” and respond to each action using five-point response options ranging from “Very Ineffective” to “Very Effective.” The test includes 141 items and took approximately one hour to complete. The test can be scored using either consensus or expert responses as the criteria for identifying the correct answer. We used expert scoring since it yields slightly higher estimates of inter-rater reliability (Mayer et al., 2003). Additional details on the MSCEIT is provided by Mayer, Salovey, Caruso, and Sitarenios (2003) and in the User’s Manual (Mayer, Salovey, & Caruso, 2002). The MSCEIT produced a split-half reliability of .91 for expert scoring (Mayer et al., 2003) and a test-retest reliability over a three-week period of .86 (Brackett & Mayer, 2003). Consistent with our comments in the introductory section, the test clearly appears valid in predicting a variety of important outcomes (see Mayer et al., 2008 for a review). For instance, Brackett and Mayer (2003) found the MSCEIT to be correlated -.27 with social deviance, .21 with high school rank and .16 with college GPA. Similarly, Brackett, Mayer, and Warner (2004) found significant negative correlations between the MSCEIT and illegal drug use, alcohol abuse and deviant behavior.

The second instrument was the Emotional Stroop. The task was computerized using the Microsoft.Net development platform. The application serially presented target words on the computer screen and recorded the time when the subject pressed any key. The time between word presentation and subject key-press was recorded in milliseconds. Subjects were familiarized with the instrument by first completing the original Stroop task for color-congruent and color-incongruent words. Participants next completed the Emotional Stroop which consisted of 20 neutral words and 20 dysphoric words. The two sets of words were approximately matched for word length and frequency of usage in the language. The dysphoric words were derived from previous Emotional Stroop studies reported in Williams et al. (1996). Examples of the dysphoric words are “hopeless,” “failure,” and “useless.” Examples of the neutral words are “horse,” “passenger,” and “flight.” The order in which the word lists were presented was counterbalanced with half the participants responding to the neutral words before the dysphoric words and the other half responding to the dysphoric words before the neutral words. Scores on the Emotional Stroop task were derived by subtracting the average response time (all times were in milliseconds) for the 20 neutral words from the average response time for the 20 dysphoric words. The third instrument we administered was the Raven’s Advanced Progressive Matrices (APM) Set II (Raven, 1976). This is a nonverbal measure of cognitive ability and was used to determine if the EI measures added incremental validity beyond a traditional measure of intelligence. The APM Set II consists of 36 items.

Each item involves a set of figures that change in a logical manner and the examinee must identify the next logical figure in the sequence. The examinee is presented with eight possible figures with only one being the correct response. The items get progressively more difficult throughout the test. The subjects were given three practice items in advance to familiarize them with the test. Subjects were given 40 minutes to complete the instrument. The APM was originally developed in 1943 and has been subjected to extensive item analyses, validation research and normative studies using samples from many different countries. The test manual (Raven, Raven, & Court, 1998) reports that the APM is internally consistent (in the .80 to .90 range), significantly related to other measures of intelligence, and predictive of success in academic and employment settings. In terms of the relationship between the APM and emotional intelligence, Ciarrochi et al. (2000) found that an earlier version of the MSCEIT and the Raven’s Standard Progressive Matrices (which is similar to the APM but not as difficult) were not related ($r = .05$).

4. Results

Table 1 presents descriptive statistics and correlations for the primary variables in the study. The overall MSCEIT score had a mean of 78.8 and standard deviation of 12.2. The scores ranged from 36 to 112. The MSCEIT is normed to have a mean of 100 and standard deviation of 15. Although our standard deviation is similar to the normative data, our mean is approximately 1.5 standard deviations below the norm of 100. The MSCEIT was clearly extremely difficult for our sample of Arab students. Table 2 presents the means, standard deviations and intercorrelations for the eight scales on the MSCEIT.
Understandably, the highest mean, which was 98.2 and essentially the same as the norm of 100, was obtained on the Perceiving Emotions—Faces Task. This task involves assessing the degree of emotion displayed in a face and is largely nonverbal. The next highest mean was 92.7, which was obtained on the Perceiving Emotions—Pictures Task, which is also largely nonverbal. The means for the remaining six scales, which are all entirely verbal tasks, are below 90. These results clearly suggest that language issues depressed the MSCEIT scores. Regarding the intercorrelations among the MSCEIT scales, the results generally show significant, positive relationships, but are not so high as to suggest redundancy. Recall, the MSCEIT consists of four emotional skills with two tasks measuring each skill. The two tasks (i.e., Faces and Pictures) measuring the Perceiving Emotions skill are correlated .28 ($p < .01$). The two tasks (i.e., Changes and Blends) measuring the Understanding Emotions skill are correlated .37 ($p < .01$).

The two tasks (i.e., Emotion Management and Social Management) measuring Managing Emotions are correlated .49 ($p < .01$). However, some interrelationships did not seem consistent with the underlying theory. For instance, the two tasks (i.e., Facilitation and Sensations) measuring the Using Emotions skill were only correlated .09 ($ns$). The Raven’s APM had a mean of 17.3 ($SD = 4.7$). The test manual includes a host of normative samples and address a variety of factors such as age, nationality (but none from the Middle East), and area of study in college (Raven et al., 1998). The means for college students tend to fall in the 20 – 25 range, so our mean of 17 is slightly lower than that found in other countries. This finding appears relevant to the relatively low scores on the MSCEIT. Since the APM is a nonverbal measure, the results suggest there are general cognitive and/or motivational factors in addition to language issues that may have contributed to the depressed scores on the MSCEIT.

Consistent with our first hypothesis, the MSCEIT was negatively correlated with the Emotional Stroop ($r = -.27, p < .01$). Table 3 presents the correlations between the eight MSCEIT scales and the primary variables. Four of the eight scales are correlated with the Emotional Stroop, but there does not appear to be any meaningful pattern (i.e., the significant relationships represent three of the four emotional skill areas). Our modest sample size does not allow us to draw any general conclusions regarding the effectiveness of the scales. In addition, although scale scores are important for theoretical and test construction purposes, the overall test score would generally be used for administrative decisions such as selection and promotion. Thus, our focus here is on the overall test score. In terms of the relationship between EI and a traditional measure of intelligence, the MSCEIT was highly correlated ($r = .47, p < .01$) with the Raven’s APM (see Table 1), and six of the eight MSCEIT scales were significantly correlated with the Raven’s (see Table 3). Regarding Hypothesis 2, we used hierarchical regression to examine the incremental validity of the MSCEIT. With the Emotional Stroop as the dependent variable, we entered the Raven’s APM as the initial independent variable, and then added the MSCEIT. The results of the hierarchical regression analyses are presented in Table 4. The Raven’s APM was significantly correlated with the Emotional Stroop ($r = -.21, p < .05$ using a one-tailed test), and explained four percent of the variance. The MSCEIT added incremental validity by increasing the variance explained from 4% to 8%. Given that we were predicting a specific cognitive process, this appears to be a noteworthy contribution over and above an established measure of intelligence.

5. Discussion

The Emotional Stroop assesses one’s ability to process information in the face of potentially distracting emotional stimuli. The MSCEIT, a task-based measure of emotional intelligence, was significantly related to performance on the Emotional Stroop, and it accounted for incremental variance beyond that explained by a traditional measure of intelligence. These results support the construct validity of the MSCEIT in predicting an emotional process. From a cognitive processing standpoint, there are alternative explanations for the relationship between the MSCEIT and Emotional Stroop. For example, it may be that emotionally intelligent individuals are more effective in handling the emotional aspects of their lives and, as a result, are less influenced by the negative emotional words on the Stroop task. Alternatively, it may be that emotionally intelligent individuals are more effective in isolating the emotional disturbance and keeping it from interfering with the task at hand. Thus, there are clearly opportunities for future research in this area. Consistent with a suggestion by Landy (2006), we have extended EI research to a non-Western sample. Although we encountered some minor challenges (e.g., lower mean scores on the MSCEIT) that were likely due, at least in part, to language or cultural issues, the overall relationships between our primary variables suggest that the EI construct and task-based measures may generalize to the Middle East.
This study has clear limitations. First, our reliance on university students arguably limits the generalizability of the results. Second, our sample consisted entirely of females, which further limits the ability to generalize. However, we examined a fundamental cognitive process, and are not aware of research suggesting that such processes would be dramatically moderated by age or gender. We offer three suggestions for future research. First, there are a number of self-report or personality-oriented measures of EI that might be examined for construct validity (see, for instance, Law et al., 2004, Schutte et al, 1998 and Tett, Fox, & Wang, 2005). Second, there would appear to be practical value in developing task-oriented instruments that are less time consuming and easier to administer than the existing MSCEIT, particularly for non-Western samples. Third, a logical extension of our study is to explore the relationship between EI measures and “real” and significant emotional events that individuals face, such as being criticized in public or losing one’s job. In sum, our results indicated that a task-based measure of EI was related to a well-established measure of emotional functioning, and had incremental validity beyond a traditional measure of intelligence. This lends support to the construct validity of EI.

References


| Table 1: Means, Standard Deviations, and Intercorrelations among Primary Variables |
|--------------------------------------|------|--------|-----|-----|
| Variable | M   | SD    | 1    | 2    |
| 1. MSCEIT | 78.8 | 12.2  | -   | -   |
| 2. Raven APM | 17.3 | 4.7   | .47** | -   |
| 3. Emotional Stroop | 39.7 | 137.8 | -.27** | -.21* |

Note. N = 87. All tests are one-tailed.
* p < .05. ** p < .01.

| Table 2: Means, Standard Deviations, and Intercorrelations Among MSCEIT Scales |
|-----------------------------------------------|---|--------|-----|-----|-----|-----|-----|-----|
| Scale                                   | M   | SD    | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 1. Perceiving Emotions –  Faces Task       | 98.2 | 13.3 |   |   |   |   |   |   |   |
| 2. Perceiving Emotions –  Pictures Task    | 92.7 | 12.8 | .28** |   |   |   |   |   |   |
| 3. Using Emotions – Facilitation Task      | 89.6 | 16.4 | .11 | -.03 |   |   |   |   |   |
| 4. Using Emotions – Sensations Task        | 84.7 | 12.7 | .27** | .06 | .09 |   |   |   |   |
| 5. Understanding Emotions –Changes Task    | 83.5 | 11.2 | .45** | .27** | .23* | .23* |   |   |   |
| 6. Understanding Emotions –Blends Task     | 81.3 | 11.3 | .13 | .08 | .24* | .12 | .37** |   |   |
| 7. Managing Emotions –Emotion Mgt Task     | 85.6 | 10.8 | .34** | .21* | .03 | .22* | .30** | .32** |   |
| 8. Managing Emotions –Social Mgt Task      | 85.0 | 14.3 | .41** | .24* | .25** | .26** | .40** | .37** | .49** |

Note. N = 87. All tests are one-tailed.
* p < .05. ** p < .01.
Table 3: Correlations between MSCEIT Scales and Primary Variables

<table>
<thead>
<tr>
<th>Scale</th>
<th>Raven APM</th>
<th>Emotional Stroop</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Perceiving Emotions–Faces Task</td>
<td>.22*</td>
<td>-.24*</td>
</tr>
<tr>
<td>2. Perceiving Emotions–Pictures Task</td>
<td>.14</td>
<td>-.11</td>
</tr>
<tr>
<td>3. Using Emotions–Facilitation Task</td>
<td>.05</td>
<td>-.12</td>
</tr>
<tr>
<td>4. Using Emotions–Sensations Task</td>
<td>.34**</td>
<td>-.19*</td>
</tr>
<tr>
<td>5. Understanding Emotions–Changes Task</td>
<td>.34**</td>
<td>-.01</td>
</tr>
<tr>
<td>6. Understanding Emotions–Blends Task</td>
<td>.24**</td>
<td>-.03</td>
</tr>
<tr>
<td>7. Managing Emotions–Emotion Mgmt Task</td>
<td>.33**</td>
<td>-.31**</td>
</tr>
<tr>
<td>8. Managing Emotions–Social Mgmt Task</td>
<td>.50**</td>
<td>-.21*</td>
</tr>
</tbody>
</table>

Note. N = 87. All tests are one-tailed.
*p < .05. **p < .01.

Table 4: Hierarchical Regression of Emotional Stroop on the MSCEIT

<table>
<thead>
<tr>
<th>Variable</th>
<th>β</th>
<th>R²</th>
<th>ΔR²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raven APM</td>
<td>-.21*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raven APM</td>
<td>-.11</td>
<td>.04*</td>
<td></td>
</tr>
<tr>
<td>MSCEIT</td>
<td>-.22*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. N = 87.
*p < .05.

Acknowledgments

This research was supported in part by a grant from the Research Incentive Fund at Zayed University. We thank Kaye Lushington for her assistance with data collection and management.