

The Validity of the Mayer–Salovey–Caruso Emotional Intelligence Test (MSCEIT) as a Measure of Emotional Intelligence

Andrew Maul
University of Oslo, Norway

Abstract

The concept of emotional intelligence (EI) has drawn a great amount of scholarly interest in recent years; however, attempts to measure individual differences in this ability remain controversial. Although the Mayer–Salovey–Caruso Emotional Intelligence Test (MSCEIT) remains the flagship test of EI, no study has comprehensively examined the full interpretive argument tying variation in observed test performance to variation in the underlying ability. Employing a modern perspective on validation, this article reviews and synthesizes available evidence and discusses sources of concern at every level of the interpretive argument. It is argued that a focus on causal explanation of observed variation in test performance would significantly improve the validity of the MSCEIT as a measure of EI.

Keywords

emotional intelligence, MSCEIT, validity

The process of creating, refining, and evaluating a measure of a psychological attribute has long been a mainstay of research in the human sciences. Such work can lay the groundwork for further study, both by clarifying the nature and structure of an attribute and by providing a tool for its quantification, which in turn facilitates empirical study of the attribute and its relationship to other variables. Thus valid measurement of a cognitive attribute such as *emotional intelligence* (EI) is in many ways a prerequisite for deep exploration of the nature and structure of this ability and the ways in which it connects with other cognitive and behavioral phenomena.

Through the lens of a modern, argument-based approach to validation, this review examines the accumulated evidence relevant to the argument for the validity of the Mayer–Salovey–Caruso Emotional Intelligence Test (MSCEIT Version 2.0; Mayer, Salovey, & Caruso, 2002) as a measure of the ability of emotional intelligence as articulated by Mayer and Salovey (1997). Although this review finds many aspects of the MSCEIT’s validity argument to be wanting, there is also much that has been learned and can be applied to future research on emotional intelligence and other psychological constructs.

A Brief Overview of Emotional Intelligence

Although the term *emotional intelligence* has seen a variety of uses by educators, businesspeople, and the popular press, the scientific literature on EI has focused on definitions that seek consistency with existing psychological conceptions of both *emotion* and *intelligence*. Typifying this approach, Mayer and Salovey (1997) proposed a model of emotional intelligence composed of four more specific abilities: (1) the ability to perceive accurately, appraise, and express emotion (“perceiving emotions”); (2) the ability to access and/or generate feelings when they facilitate thought (“using emotions,” also called “emotional facilitation of thought”); (3) the ability to understand emotion and emotional knowledge (“understanding emotions”); and (4) the ability to regulate emotions to promote emotional and intellectual growth (“managing emotions”). This model of emotional intelligence guided the construction of the Multifactor Emotional Intelligence Scale (MEIS; Mayer, Caruso, & Salovey, 1999) and later, the MSCEIT.

Author note: I am grateful to Maureen O’Sullivan, Mark Wilson, Dacher Keltner, Kimberly Barchard, and anonymous reviewers for comments on earlier drafts of this article.
Corresponding author: Andrew Maul, University of Oslo, Unit for Quantitative Analysis in Education (EKVA), Postboks 1099 Blindern, Oslo, Norway, 0317.
Email: andyemaul@gmail.com

Table 1. Layout of the MSCEIT

Branch name	Brief description of skills involved	Task name	Brief description of tasks
Perceiving Emotions	The ability to perceive emotions in oneself and others, as well as in objects, art, stories, music, and other stimuli	Faces	Identify the emotions expressed in pictures of faces
		Pictures	Identify the emotions expressed in pictures of artwork and landscapes
Using Emotions	The ability to generate, use, and to feel emotion as necessary to communicate feelings, or employ them in other cognitive processes	Facilitation	Rate the helpfulness of moods to activities
		Sensations	Generate an emotion on the basis of sensation words (cold, dark) and compare the feeling to emotion words
Understanding Emotions	The ability to understand emotional information, how emotions combine and progress through relationship transitions, and to appreciate such emotional meanings	Changes	Identify emotions that result from intensifications of other emotions
		Blends	Identify emotions that result from blends of other emotions
Managing Emotions	The ability to be open to feelings, and to modulate them in oneself and others so as to promote personal understanding and growth	Emotional management	Rate the effectiveness of actions to situations involving one's own emotions
		Emotional relations	Rate the effectiveness of actions to situations involving others' emotions

Note. Adapted from Mayer et al. (2002).

The MSCEIT

Interpretation of test results for the MSCEIT are proposed on the total test level, said to represent general emotional intelligence (or EI_g), and four branch levels, said to represent the abilities to perceive, use, understand, and manage emotions. In addition, the first two branches are organized into an experiential area score, defined as a person's "ability to perceive, respond, and manipulate emotional information without necessarily understanding it" (Mayer et al., 2002, p. 18), and the second two branches are organized into a strategic area score, defined as a person's "ability to understand and manage emotions without necessarily perceiving feelings well or fully experiencing them" (p. 18).

Each of the four branches is measured by two subscales, or "tasks." A task here is defined as a group of items of the same type, such as a group of items on which respondents make ratings on a 1–5 scale regarding the degree of presence of specified emotions in pictures of abstract art and landscapes (the pictures task). There are 141 questions on the MSCEIT altogether, between 10 and 30 on each task. In some cases, the tasks are comprised of several small groups of items with common prompts, such as when a respondent rates a single photograph in terms of several emotions. The general layout of the MSCEIT is shown in Table 1 (adapted from Mayer et al., 2002).

MSCEIT test items are scored via a technique known as consensus-based scoring, in which the score assigned to each response depends on the proportion of a group of sample respondents who selected that answer. In the *general consensus*

version, scores are assigned to each response based on the proportion of respondents from a large ($N > 5,000$), diverse standardization sample from English-speaking countries endorsing that response (Mayer, Roberts, & Barsade, 2008). For instance, when respondents are asked to rate the intensity of a particular emotion in a face, if the five options relating to the emotion are endorsed by 10%, 20%, 20%, 40%, and 10% of the sample respectively, an endorsement of the third option would receive a score of .20, while an endorsement of the fourth option would receive a score of .40. Respondents' scores on each task are the average of their weighted scores for each of the task's items.

The *expert consensus* version is identical, although here the sample comprised 21 volunteer members of the International Society for Research on Emotion (ISRE) at their conference in 2000. Scores derived thus exhibit a very high correlation with scores derived from general consensus scoring (e.g., $r = .96$; Mayer, Salovey, Caruso, & Sitarenios, 2003).

The Validity of the MSCEIT

Recent literature on validation in educational and psychological measurement has been dominated by an argumentation-and-evidence-based approach, as reflected in the writings of theoreticians such as Samuel Messick (e.g., 1989) and Michael Kane (e.g., 2006). Under this framework, validity is seen as an "integrated and evaluative judgment of the *adequacy* and *appropriateness of inferences and actions* based on test scores" (Messick, 1989, p. 13, emphasis in original).

In order to adequately evaluate the validity of a measurement enterprise, it is necessary to first adequately specify what Kane (2006) refers to as the “network of inferences and assumptions leading from observed performances to the conclusions and decisions based on those performances” (p. 23), also known as the *interpretive argument*. Kane further notes that “a failure to state the proposed interpretations and uses clearly and in some detail makes a fully adequate validation essentially impossible, because implicit inferences and assumptions cannot be critically evaluated” (p. 57). Once the interpretive argument has been clearly specified, the process of validation proceeds by examining the evidence relevant to each of its inferences.

Here MSCEIT is taken to be intended mainly as a research instrument, designed to provide information about the Mayer–Salovey model of emotional intelligence. The proposed interpretation of MSCEIT test scores, in its simplest form, is therefore that *variation in observed performance on the MSCEIT reflects true variation in emotional intelligence*. This statement is clearly rooted in the particular definition of emotional intelligence being used.

It should be noted that other interpretations of MSCEIT scores are certainly possible; in particular, in applied settings it may not be the case that measurement per se is of primary concern, and instead emphasis may be placed on practical utility. An example of such an interpretation would be that *MSCEIT scores usefully predict variation in workplace performance*. Defending this interpretation would require a different argument than the one evaluated here, with different supporting evidence.

Under Kane’s (2006) framework for argument-based validation, there are four main inferences connecting observed test performances to their proposed interpretation as reflecting emotional intelligence; briefly, these are referred to as the *scoring*, *generalization*, *extrapolation*, and *interpretation* inferences. In the context of the MSCEIT, an abbreviated statement of the interpretive argument follows:

1. The *scoring* system is adequate to make the inference from observed performances to the observed score. This includes the proposition that the consensus-based scoring system is appropriate for the measurement of emotional intelligence.
2. The observed score can be *generalized* to the universe score. This includes the propositions that the sample of observations on the MSCEIT is both representative enough and large enough to control sampling error.
3. The universe score can be *extrapolated* to the target score. This inference can be supported both by logical, theory-based support for the adequacy of the sampling of test content from the domain of possible observables associated with emotional intelligence, and on empirical support for the relationship between test scores and other observations contained within the target domain.
4. Scores on the target domain can be *interpreted* as reflective of “emotional intelligence” (and, more specifically, of the four branches of emotional intelligence). This inference can be supported by evidence

demonstrating that the component abilities of EI demonstrate expected relationships with one another and with other constructs and outcomes as predicted by theory, and that EI is both conceptually and empirically distinguishable from known constructs.

The purpose of this article is to examine the full interpretive argument of the MSCEIT, in order to provide an integrated and evaluative judgment of its validity.

Scoring of the MSCEIT

Support for the adequacy of the scoring system comes from the documentation of the procedures used to assign scores to item responses. The central idea of measurement is to have a procedure sensitive to differences in the thing being measured, such that (in this case) different responses to items are reflective of different levels of emotional intelligence. The procedure of assigning scores to performances must therefore be logically and defensibly connected to the theory of emotional intelligence.

The consensus-based scoring method employed by the MSCEIT has drawn considerable controversy (e.g., Barchard & Russell, 2004; Brody, 2004; Keele & Bell, 2009; O’Sullivan, 2007). A brief review of the logic that has been presented in the defense of this scoring method will be helpful.

Legree, Psotka, Tremble, and Bourne (2005) argue that consensus-based scoring can be used as a substitute for theory-based scoring when constructs “lack certified experts and well-specified, objective knowledge” (p. 155), such as, they argue, is the case for emotional intelligence. They develop this case first by arguing that even commonly measured domains of knowledge are “lodged in opinion and [may] have no objective standard for verification other than societal views, opinions, and interpretations” (p. 159). They give the example of developing an English language vocabulary test for American colonists prior to the efforts of Noah Webster in the late 18th century. Well-qualified subject matter experts might be hard to identify, they point out: If royal English university professors were used, opinions might be skewed in an academic direction. They propose instead that a representative sample of American colonists could be selected and surveyed to identify acceptable responses to vocabulary definition items.

This example is interesting, as it pertains to a domain in which consensual belief itself directly determines correctness. In American English, the meaning of words is determined simply by how people use them; thus statements about the correctness of early American English word definitions do not have excess meaning beyond being statements about the consensual understanding of their definitions.

Mayer, Salovey, Caruso, and Sitarenios (2001) imply that emotional information may be similar in that “it helps to know how an individual’s reactions compare with how most people would emotionally respond to a situation [. . .] such knowledge helps *define* the general meaning of emotions in regard to relationships” (pp. 236–237, emphasis added). However, available research in emotions supports the idea that emotions are not defined solely by their consensual use and interpretation.

In particular, facial displays of emotion appear to be largely biological in origin and culturally universal (e.g., Ekman & Friesen, 1978), as do many other aspects of the experience and functioning of emotions (for an overview, see Oatley, Keltner, & Jenkins, 2006). Thus the principles of emotions are not like the principles of American English in being solely determined by their common use. Moreover, the concept of *intelligence* (and Mayer and Salovey's [1997] definition of EI, as discussed previously) involves more than knowledge, and most matters of differences in intelligence are not matters of consensus.

If consensus does not itself define correctness, the connection between the consensual answer and the correct answer is not direct and needs to be defended on other grounds. Specifically, it needs to be argued that, rather than consensus *determining* a correct answer, consensus could be used to *discover* an independently existing correct answer. Nothing inherent to the idea of consensus indicates that the consensus will always be correct; in fact, examples abound in many fields of cases in which the majority of people exhibit a particular misconception. The literature on deception (e.g., Ekman, O'Sullivan, & Frank, 1999) suggests that a great deal of emotional information is routinely missed by all but a very astute minority, and that the consensus interpretations of many displays of emotion are incorrect. One well-known example of this is the non-Duchenne smile (sometimes called a "camera smile," which is almost identical to a genuine smile save for the absence of activation of small muscles around the eyes; see Ekman, Davidson, & Friesen, 1990), which might give the appearance of happiness to all but the most emotionally astute. The consensus judgment of the emotional state of a person displaying such as smile would be incorrect. Thus nothing inherent to the logic of consensus-based scoring of emotional stimuli indicates either that consensus *determines* the correctness of answers, or that consensus will reliably *discover* correct answers.

The use of a smaller sample of experts, in addition to a sample from the general population, is argued to provide a cross-validation of the consensus scoring method. Thus it seems relevant to review the expert scoring method in greater detail.

No documentation has been provided concerning the qualifications of the 21 individuals recruited to serve as experts except that they were all present at the 2000 conference of the International Society for Research on Emotion. Although this group is likely to have had more formal knowledge of emotions theories and research than the general public, the connection between formal knowledge of emotions and emotional intelligence is not necessarily clear, as was argued by Zeidner, Matthews, and Roberts (2001) and Brody (2004), who noted that even "a person who has expert knowledge of emotions may or may not be expert in the ability that is allegedly assessed by the test" (p. 234). With specific regard to facial displays of emotion, O'Sullivan and Ekman (2008) note that "there is no guarantee that emotions experts, some of whom might be philosophers or historians, are expert at identifying facial expressions accurately" (p. 35).

It may be worth pausing to reflect on a statistical point. If, in fact, the 21 members of the expert group were not experts in EI

at all—if they were instead just a random subsample of the general population—it would still be expected that scores based on their consensus would exhibit a very high degree of convergence with the general consensus. Thus the high convergence between scores obtained using expert and general consensus methods cannot by itself be taken to provide a cross-validation of the scoring method.

Legree et al. (2005) show that, often, the judgments of experts deviate from the judgments of laypeople in that the most commonly selected answer among laypeople is also the most commonly selected answer among experts, but the experts agree to a larger degree on that answer. Again this is not true by logical necessity (a point which Legree and colleagues acknowledge), but if we accept it as generally true, then statistical evidence of expertise may be found in an examination of item response patterns of the expert group compared to the general group. Unfortunately, Mayer and colleagues have not reported the proportion of experts that selected each response on the MSCEIT, so it is not possible to check the actual degree of consistency among the expert sample. They have reported that the experts exhibit central tendencies similar to the general consensus across all branches of the MSCEIT, and higher interrater agreement in their answers for the Faces task and the two tasks on the Understanding Emotions branch (thus on three out of eight MSCEIT tasks), which "may reflect the greater institutionalization of emotion knowledge among experts in these areas" (Mayer et al., 2003, p. 101). However, their results show that interrater agreement was actually higher for the general group than it was for the expert group for the two tasks on the Facilitating Emotions branch, and was not significantly different for the Managing Emotions branch's tasks or the Pictures task, leaving five out of eight of the MSCEIT tasks without observably greater interrater agreement among experts. Accepting the observations of Legree et al. (2005) discussed previously, this suggests that the expert sample was, in fact, likely no more expert than the general consensus concerning the content covered by these sections.

If there exists a well-developed body of knowledge concerning an area of human functioning, as the quote above acknowledges there is for the areas of facial expression of emotion and understanding of some emotional principles, it is not clear why deriving a scoring guide through group consensus should be considered preferable to writing questions with answers designed in advance to be correct or incorrect based on theory and prior empirical research. On the other hand, no compelling evidence has been presented that general or expert consensus scoring procedures can be used to identify correct answers to test items for the remaining abilities targeted by the MSCEIT. Support for the adequacy of the scoring system, and therefore for the first inference of the validity argument, does not seem sufficient.

Generalization of MSCEIT Scores

Evaluation of the generalization inference requires an examination of the adequacy of the sampling of the tasks on the

MSCEIT from their universe of generalization. The concept of reliability is directly relevant to this inference.

The *MSCEIT User's Manual* (Mayer et al., 2002) reports split-half reliability coefficients of .93 for general consensus scoring of the MSCEIT and .91 for expert scoring, with lower estimated reliabilities at the area, branch and task levels. Independent investigations (e.g., Lopes, Salovey, & Straus, 2003; Maul, 2011a; Palmer, Gignac, Manocha, & Stough, 2005; Roberts et al., 2006) have reported lower reliability coefficients at all strata of testing (usually in the low .80s for the full test, in the .60s and .70s for the four branches of EI, and as low as .40 for individual tasks). Two recent studies (Føllesdal & Hagtvet, 2009; Maul, 2011b) that have employed methods which statistically correct for coefficient inflation due to local item dependence from the MSCEIT's multifaceted structure (i.e., the collection of items into groups around common prompts, and further into tasks) have suggested that true reliability could be even lower than these estimates indicate (estimates from these studies ranged from the .70s down to the .40s at the branch level). Even prior to these recent studies, Matthews, Zeidner, and Roberts (2004) summarized that MSCEIT estimated reliabilities were "far from optimal" (p. 198).

There are no absolute standards on what constitutes acceptable reliability for a test designed as a research instrument. Reliability coefficients in the .60s and .70s are lower than some researchers would desire; a reliability of .70, for example, translates into a standard error of measurement of .55 standard deviations. In practical terms, on the MSCEIT's intelligence quotient (IQ)-like scale with a mean of 100 and a standard deviation of 15, a respondent would have to get a score of higher than 116 or lower than 84 to be statistically significantly ($p < .05$) above or below average. Thus a note of caution is appropriate when generalizing observed scores to universe scores.

Extrapolation of MSCEIT Scores

The third inference in the MSCEIT's interpretive argument requires extrapolation from the universe of generalization to a target domain of possible observations associated with variation in emotional intelligence. This inference can be supported by rational, theory-based evidence of the adequacy of the sampling of test content from the target domain, and by empirical demonstration that MSCEIT scores are associated with other observations (including test scores, behaviors, and other outcomes) that fall within the target domain. It should be noted that evidence of relations to other variables that fall outside the target domain of EI are not relevant to the evaluation of the extrapolation inference, but may be relevant to the theory-based interpretation inference that follows.

Evidence Based on Instrument Content

A comprehensive account of content-based evidence depends on an articulation of how responses to MSCEIT test items constitute a representative sample of observations associated with the target domain of emotional intelligence. This first requires a

definition of emotional intelligence clear enough to set the boundaries of the target domain.

The definition of emotional intelligence. The definition of emotional intelligence as presented by Mayer and Salovey (1997) provides the construct framework for the MSCEIT. Table 1, adapted from the *MSCEIT User's Manual* (Mayer et al., 2002), gives more specific quotes that generally set reasonable boundaries on the target domains of each of the four proposed branches of EI, although some specific confusions remain. In particular, the description of the Perceiving Emotions branch refers to the Perception of Emotions in "objects, art, stories, and other stimuli"; emotion is a property of conscious beings, and therefore strictly speaking cannot be present in these stimuli. In the case of art and stories this phrase may refer to perceiving the emotions communicated by a human in the act of creation, or the emotions that would be commonly elicited in observers, or (where applicable) the emotions felt by the persons depicted, or the emotions the respondent feels when exposed. It is less clear what the perception of emotions in other objects or stimuli would mean (as well as what objects and stimuli are covered, and what are not). Additionally, it is not clear what is meant by "appreciat[ing] such emotional meanings" in the description of the Understanding Emotions branch. "Appreciate" could refer to having awareness or knowledge of something, or could indicate valuing (and in particular having gratitude for) something. The phrase "emotional meanings" is also ambiguous, as "meanings" could refer to communication (e.g., the meanings of words and phrases about emotions), or the causes of something (e.g., the meaning of one's heart rate increasing in the presence of spiders), or personal significance, among other possibilities. With respect to the Managing Emotions branch, it is not clear what is meant by "personal understanding and growth"; this is a subjective phrase that could have any number of interpretations.

As emotional intelligence is articulated as a type of *intelligence*, it seems reasonable to inquire about its relationships to existing models of intelligence. Neubauer and Freudenthaler (2005) suggest that the conceptual relationship between emotional intelligence and fluid (Gf) and crystallized (Gc) intelligence is not clear and that the construction of the MSCEIT, rather than the construct definition itself, seems to have led this model of EI to resemble Gc more than Gf. Related to this, Zeidner and colleagues (2001) point out that much emotional and social knowledge can be implicit, procedural, and difficult to verbalize, and the relationship between this implicit knowledge and explicit knowledge about emotions is unclear. Whether, for example, the understanding emotions branch conceptually refers to explicit knowledge, implicit knowledge, or both, is not specified.

Any lack of conceptual clarity in the construct definition of EI makes the task of evaluating the match between test items and the target domain more difficult. Nevertheless, for the most part the construct definition of EI seems clear enough to set reasonable bounds on the target domain.

The match between MSCEIT test items and the EI construct.

An examination of the match between the content of MSCEIT items and the target domain reveals problems with what Messick (e.g., 1989) referred to as both *construct underrepresentation* and *construct-irrelevant variance*.

The Perceiving Emotions branch is described by Mayer et al. (2002) as referring to “the ability to perceive emotions in oneself and others, as well as in objects, art, stories, and other stimuli” (p. 7) and also involving “the capacity [. . .] to express feelings” (p. 19). Tasks related to expression of emotion are not present on the MSCEIT, nor are tasks related to the ability to perceive emotions in oneself. Within the ability to recognize others’ emotions, respondents make judgments only about photographs of people’s faces and pictures of abstract art and landscapes, thus omitting a wide range of other relevant modalities, such as tone of voice and posture. Further, within facial displays of emotion, the inclusion of only context-free still photographs excludes a range of potentially relevant stimuli such as micro- and brief-affect displays and other dynamic and contextual factors. Finally, it does not appear that the four pictures of faces on the MSCEIT were selected with a specific plan concerning what emotions should be represented (O’Sullivan & Ekman, 2008, p. 31).

Mayer et al. (2002) define the Using Emotions branch as “the ability to generate, use, and feel emotion as necessary to communicate feelings, or employ them in other cognitive processes” (p. 7), and note that it involves “being able to use one’s emotions to help a person solve problems creatively” (p. 19). None of these abilities appear to be directly assessed. The Facilitation task, which asks respondents to rate the helpfulness of moods to specified activities, would appear somewhat relevant to the use of emotions in cognitive processes and problem solving, although its items do not involve the actual use of emotions. The Sensations task requires respondents to generate emotions and match the sensations of those emotions to colors and tastes. Although this task does call for generation of emotions, the connection between performance on this task and the Using Emotions abilities described earlier is not transparent.

The Understanding Emotions branch is defined by Mayer et al. (2002) as “the ability to understand emotional information, how emotions combine and progress through relationship transitions, and to appreciate such emotional meanings” (p. 7). There are items that concern changes (in particular, intensifications) of emotions, and combinations of emotions, both of which appear relevant to understanding emotional information.

The Managing Emotions branch is defined by Mayer et al. (2002) as “the ability to be open to feelings, and to modulate them in oneself and others so as to promote personal understanding and growth” (p. 7). There does not appear to be any content relevant to being open to feelings, or any content directly relevant to modulating feelings in either oneself or others. Items on this branch ask respondents to rate how effective various responses to emotional situations would be, which would seem to require theoretical knowledge about the practical aspects of emotions; evidence for whether and how such theoretical knowledge is related to one’s actual ability to manage emotions has not been presented.

Evidence Based on Relations to Other Variables

Evidence based on the relationship between MSCEIT scores and other variables is relevant to the extrapolation inference insofar as such evidence helps clarify the relationship between performance on test content and performance in other situations covered by the target domain of emotional intelligence. There are few, if any, tests that can be clearly interpreted as measuring content within the target domain of the Mayer–Salovey model of EI. As a consequence, very little evidence exists that can be clearly placed in this category.

There are other tests that would appear to assess abilities contained in the definition of the Perceiving Emotions branch of EI, including the Japanese and Caucasian Brief Affect Recognition Test (JACBART; Matsumoto et al., 2000), which assesses the ability to detect very briefly expressed facial expressions, and the Vocal-I (Scherer, Banse, & Wallbott, 2001), which assesses the ability to correctly identify emotions in tone of voice. However, a study by Roberts et al. (2006) found zero correlation between these tests and the MSCEIT Perceiving Emotions tasks. The Levels of Emotional Awareness Scale (Lane, Quinlan, Schwartz, Walker, & Zeitlin, 1990), which may measure content related to understanding emotions, was found by Ciarrochi, Caputi, and Mayer (2003) to weakly correlate with overall MSCEIT scores ($r = .15$) and was found by Barchard and Hakstian (2004) to load onto a common factor with MSCEIT Understanding Emotions tasks.

With evidence still scant on this topic, it is simply an open question to what extent most of the target domain of EI is related to observed performance on the MSCEIT.

Theory-Based Interpretation of MSCEIT Scores

The final inference in the interpretive argument connects the target domain of emotional intelligence to the interpretation of the idea of EI in the wider context of psychological theory. Evidence based on the internal structure of the instrument is relevant to this inference, as is evidence concerning whether the relations between MSCEIT scores and external variables fall in line with theory-based predictions. Each of these sources of evidence is considered in turn.

Evidence Based on Internal Structure

A number of studies have examined the internal structure of the MSCEIT, asking in particular whether the structure of the test conforms to a four-factor model corresponding to the four-branch theory of emotional intelligence (e.g., Day & Carroll, 2004; Gignac, 2005; Keele & Bell, 2008; Maul, 2011a, 2011b; Mayer et al., 2003; Palmer et al., 2005; Roberts et al., 2006; Rode et al., 2008; Rossen, Kranzler, & Algina, 2008). Factor-analytic studies such as these examine the extent to which associations between parts of a test are consistent with the theory upon which the test was built. In the case of the MSCEIT, the authors (Mayer et al., 2003) propose that the test should be well described by a

one-factor (“EIg”) solution, a two-factor (Experiential and Strategic) solution, and a four-factor (Perceiving, Using, Understanding, and Managing Emotions) solution.

Briefly stated, conclusions from the studies cited earlier have been heterogeneous and largely equivocal. The earliest two studies of the MSCEIT factor structure (Day & Carroll, 2004; Mayer et al., 2003) found that the results of confirmatory factor analyses supported the idea that the MSCEIT measures four distinguishable EI factors. However, a report by Gignac (2005) and a follow-up article by Palmer et al. (2005), through re-analysis of the Mayer et al. (2003) dataset and through new data collection, found one-, two-, and four-factor solutions to be ill-fitting or implausible, and concluded that a model with a general factor and with Perceiving, Understanding, and Managing (but not Using) Emotions provided the best fit to the data.

Studies by Keele and Bell (2008), Maul (2011a), and Roberts et al. (2006), Rode et al. (2008) and Rossen et al. (2008) each employed somewhat different analytic approaches, but reached similar conclusions in that a Using Emotions branch could generally not be identified apart from a general factor, and that only partial support was available for the identification of the remaining EI branches. Further, in response to both technical and conceptual challenges associated with the use of averaged task scores as indicator variables in factor models, a study by Maul (2011b) modeled the MSCEIT at the item level, using multidimensional item response models, and found no empirical support for any multifactor model when local item dependence due to task format was controlled. Thus it does not seem that the accumulated evidence provides clear support for the idea that the structure of the MSCEIT conforms to theory-based expectations.

Evidence Based on Relations to Other Variables

Evidence of relationships between MSCEIT scores and external variables is relevant to validity insofar as such evidence bears on predictions that follow from the theory that MSCEIT scores measure EI. It is important that clear, theory-based expectations motivate correlational investigations of validity; studies undertaken as exploratory research cannot then be interpreted *ex post facto* as evidence for the validity of a test, as was noted by Brody (2004).

Loosely, the other variables with which the MSCEIT’s association has been investigated can be grouped into those the MSCEIT is expected to correlate highly with (“convergent” evidence), those the MSCEIT is expected not to correlate highly with (“discriminant” evidence), and outcome variables that are expected to be predicted by the MSCEIT (“predictive” evidence).

Given that emotional intelligence is meant to represent a form of intelligence, it is expected (Mayer et al., 1999) that EI scores should be moderately, but not too-highly, associated with traditional intelligence tests. A meta-analysis by Bludau and Legree (2008) reported that MSCEIT scores—and in particular scores on the Understanding Emotions branch—are associated with crystallized intelligence (correlations averaging .40 according to the second source), but only weakly or not at all with fluid intelligence. A study by Barchard and Hakstian

(2004) found that MSCEIT understanding scores loaded onto a common factor with scales from the O’Sullivan–Guilford social intelligence measure, indicating some degree of overlap between emotional and social intelligence. Empathy is another variable with which EI may be expected to be associated, as empathy is commonly held to involve the capacity to recognize and understand others’ emotions; consistently, MSCEIT scores are associated moderately with self-reported empathy (Brackett, Rivers, Shiffman, Lerner, & Salovey, 2006).

Emotional intelligence is not expected to relate strongly to personality variables. In contrast to self-report EI scales, MSCEIT scores do not appear to correlate with Big Five personality measures, with the exception of small correlations with agreeableness (average $r = .25$) and openness ($r = .17$; Mayer et al., 2008).

A variety of studies have found small positive associations (generally below $r = .30$) between MSCEIT scores and various prosocial outcomes, including life satisfaction (Mayer et al., 2002), psychological well-being (Brackett & Mayer, 2003), and the perceived quality of social relationships (Lopes et al., 2003); and negative associations with illegal drug and alcohol use (Brackett, Mayer, & Warner, 2004), social deviance (Brackett & Mayer, 2003), and anxiety (Bastian, Burns, & Nettelbeck, 2005). Thus it appears that MSCEIT scores are associated with intelligence and other psychological variables and positive outcomes in a manner fairly consistent with the idea that the MSCEIT measures emotional intelligence. However, the overall pattern of findings is consistent with alternative explanations as well. O’Sullivan (2007), for example, argues that “it may be that sharing the emotional perceptions of one’s cultural group, whether they are actually correct or not, is a predictor of social success” (p. 6), which can account for observed associations between MSCEIT scores and prosocial outcomes, as well as Big Five agreeableness.

It should be noted that this does not indicate a deficiency in any of the evidence cited in this section, or that additional correlational evidence is required. Rather, it is simply that correlational evidence is by nature only circumstantial, and definitive theory-based interpretation of patterns of associations relies on being able to clearly articulate the meaning of MSCEIT scores (and thus, relies on addressing threats to the scoring and extrapolation inferences, as discussed previously).

Discussion and Recommendations

Stated briefly, this review has noted significant problems with the MSCEIT’s interpretive argument. The consensus-based scoring method makes it difficult to interpret the scoring system as clearly resulting in observed scores that reflect variation in emotional intelligence. Recent investigations into the effects of task format and local item dependence indicate that the reliability of measurement may be lower than desired, raising concern with the idea that MSCEIT observed scores can be generalized to universe scores with confidence. Concerns regarding underrepresentation of the EI construct in the content of MSCEIT items, and unclear logical connections between test

content and the construct, make it difficult to confidently extrapolate universe scores to target scores. Results from multidimensional studies have not returned results supporting the network of internal associations predicted by theory. Results from correlational studies have generally returned results consistent with the theory that MSCEIT scores measure EI, but these findings are consistent with alternative theories as well.

It does not seem that the findings discussed here have suggested clear directions for the improvement of measurement of emotional intelligence. This equivocation in interpretation of results could relate to the lack of focus on explanation of the connections between the construct of EI and respondents' responses to test items. The use of consensus-based scoring seems a likely cause, or symptom, or both, of this lack of focus: When there is no need to determine the correctness of item responses a priori, items can be written without careful reflection on the ways in which different responses constitute evidence of higher or lower levels of emotional intelligence. The process of scoring responses to items on theoretical grounds demands articulation of how specific item responses constitute evidence of higher or lower levels of an ability, and this in turn can lead to significant revisions in the scoring schemes, the items and types of items, and the construct itself (see, e.g., Wilson, 2005). When such articulations have been made, the usefulness of evidence from investigations of individual response processes and item-level statistics from measurement models is multiplied considerably, as these sources of evidence now relate to specific hypothesized explanations of performance.

Thus the primary recommendation of this review is to *consider an explanatory approach to the measurement of emotional intelligence*. Such an approach would demand a sufficient background of literature and theory on what constitutes better or worse performance in emotional domains. Given the lack of relevant background research and consistent failure to empirically identify a distinct Using Emotions branch in particular, it may be necessary to abandon attempts to measure individual differences in the ability to use emotions effectively until such time as the processes involved in effective emotional use can be more clearly specified.

The abilities contained in the Mayer–Salovey model of EI are quite broad, and there is a paucity of research showing that many of these abilities are associated with one another at all, let alone to the extent that they can be considered part of a general ability. Furthermore, the empirical relationships between MSCEIT branches and tasks do not themselves provide compelling evidence that the abilities targeted by these branches and tasks are related, due both to the uncertain status of the validity of the branches and tasks as measures of the abilities they purport to measure, and to the fact that observed associations among them could be explained by common features other than an underlying set of abilities, such as the consensus scoring method. It may simply not be possible at this time to speak of unitary or higher-order emotional abilities. If understanding the nature of emotional abilities and their connections is a priority, it may be necessary to start from the ground up, by first establishing defensible ways of measuring well-defined,

specific abilities and then empirically examining the relationships among them.

The process of instrument creation and validation is a crucial component of the scientific process of articulating and understanding a psychological construct. The MSCEIT, and the model of emotional intelligence that underlies it, has been the catalyst for an enormous amount of scholarly work, and this has surely contributed to the understanding of human cognition and behavior, as well as the methods used to study these things. We will do well to attend to what we have learned from these activities as interest in emotional abilities moves forward.

References

- Barchard, K. A., & Hakstian, R. A. (2004). The nature and measurement of emotional intelligence abilities: Basic dimensions and their relationships with other cognitive ability and personality variables. *Educational and Psychological Measurement, 64*, 437–462. doi: 10.1177/0013164403261762
- Barchard, K. A., & Russell, J. A. (2004). Psychometric issues in the measurement of emotional intelligence. In G. Gehr (Ed.), *Measuring emotional intelligence: Common ground and controversy* (pp. 53–72). New York, NY: Nova Science Publishers.
- Bastian, V. A., Burns, N. R., & Nettelbeck, T. (2005). Emotional intelligence predicts life skills, but not as well as personality and cognitive abilities. *Personality and Individual Differences, 39*, 1135–1145. doi: 10.1016/j.paid.2005.04.006
- Bludau, T. M., & Legree, P. (2008, April). *Breaking down emotional intelligence: A meta-analysis of EI and GMA*. Paper presented at the 23rd Annual Meeting of the Society for Industrial/Organizational Psychology, San Francisco, CA.
- Brackett, M. A., & Mayer, J. D. (2003). Convergent, discriminant, and incremental validity of competing measures of emotional intelligence. *Personality and Social Psychology Bulletin, 29*, 1147–1158. doi: 10.1177/0146167203254596
- Brackett, M. A., Mayer, J. D., & Warner, R. M. (2004). Emotional intelligence and its relation to everyday behavior. *Personality and Individual Differences, 36*, 1387–1402. doi: 10.1016/S0191-8869(03)00236-8
- Brackett, M. A., Rivers, S. E., Shiffman, S., Lerner, N., & Salovey, P. (2006). Relating emotional abilities to social functioning: A comparison of self-report and performance measures of emotional intelligence. *Journal of Personality and Social Psychology, 91*, 780–795.
- Brody, N. (2004). What cognitive intelligence is and what emotional intelligence is not. *Psychological Inquiry, 15*, 234–238. Retrieved from <http://www.jstor.org/pss/20447233>
- Ciarrochi, J., Caputi, P., & Mayer, J. D. (2003). The distinctiveness and utility of a measure of trait emotional awareness. *Personality and Individual Differences, 34*, 1477–1490. doi: 10.1016/S0191-8869(02)00129-0
- Day, A. L., & Carroll, S. A. (2004). Using an ability-based measure of emotional intelligence to predict individual performance, group performance, and group citizenship behaviors. *Personality and Individual Differences, 36*, 1443–1458. doi: 10.1016/S0191-8869(03)00240-X
- Ekman, P., Davidson, R. J., & Friesen, W. (1990). The Duchenne smile: Emotional expression and brain physiology. *Journal of Personality and Social Psychology, 58*, 342–353. doi: 10.1037//0022-3514.58.2.342
- Ekman, P., & Friesen, W. V. (1978). *Facial action coding system: A technique for the measurement of facial movement*. Palo Alto, CA: Consulting Psychologists Press.
- Ekman, P., O'Sullivan, M., & Frank, M. G. (1999). A few can catch a liar. *Psychological Science, 10*, 263–266.
- Føllesdal, H., & Hagtvet, K. A. (2009). Emotional intelligence: The MSCEIT from the perspective of generalizability theory. *Intelligence, 37*, 94–105. doi: 10.1016/j.intell.2008.08.005

- Gignac, G. E. (2005). Evaluating the MSCEIT V2.0 via CFA: Corrections to Mayer et al., 2003. *Emotion*, 5, 233–235. doi: 10.1037/1528-3542.5.2.233
- Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17–64). Santa Barbara, CA: Greenwood Publishing Group.
- Keele, S. M., & Bell, R. C. (2008). The factorial validity of emotional intelligence: An unresolved issue. *Personality and Individual Differences*, 44, 487–500. doi: 10.1016/j.paid.2007.09.013
- Keele, S. M., & Bell, R. C. (2009). Consensus scoring, correct responses and reliability of the MSCEIT V2.0. *Personality and Individual Differences*, 47, 740–747. doi: 10.1016/j.paid.2009.06.013
- Lane, R. D., Quinlan, D. M., Schwartz, G. E., Walker, P. A., & Zeitlin, S. B. (1990). The levels of emotional awareness scale: A cognitive-developmental measure of emotion. *Journal of Personality Assessment*, 55, 124–134. doi: 10.1207/s15327752jpa5501&2_12
- Legree, P. J., Psotka, J., Tremble, T., & Bourne, D. R. (2005). Using consensus based measurement to assess emotional intelligence. In R. Schulze & R. D. Roberts (Eds.), *Emotional intelligence: An international handbook* (pp. 155–179). Cambridge, MA: Hogrefe & Huber.
- Lopes, P. N., Salovey, P., & Straus, R. (2003). Emotional intelligence, personality, and the perceived quality of social relationships. *Personality and Individual Differences*, 35, 641–659. doi: 10.1016/S0191-8869(02)00242-8
- Matsumoto, D., LeRoux, J., Wilson-Cohn, C., Raroque, J., Kookan, K., Ekman, P., . . . Goh, A. (2000). A new test to measure emotion recognition ability: Matsumoto and Ekman's Japanese and Caucasian brief affect recognition test (JACBART). *Journal of Nonverbal Behavior*, 24, 179–209. doi: 10.1023/A:1006668120583
- Matthews, G., Zeidner, M., & Roberts, R. (2004). *Emotional intelligence: Science and myth*. Cambridge, MA: MIT Press.
- Maul, A. (2011a). The factor structure and cross-test convergence of the Mayer–Salovey–Caruso model of emotional intelligence. *Personality and Individual Differences*, 50, 457–463. doi: 10.1016/j.paid.2010.11.007
- Maul, A. (2011b). Examining the structure of emotional intelligence at the item level: New perspectives, new conclusions. *Cognition & Emotion*, 26, 503–520. doi: 10.1080/02699931.2011.588690
- Mayer, J. D., Caruso, D. R., & Salovey, P. (1999). Emotional intelligence meets traditional standards for an intelligence. *Intelligence*, 27, 267–298. doi: 10.1016/S0160-2896(99)00016-1
- Mayer, J. D., Roberts, R. D., & Barsade, S. G. (2008). Human abilities: Emotional intelligence. *Annual Review of Psychology*, 59, 507–536. doi: 10.1146/annurev.psych.59.103006.093646
- Mayer, J. D., & Salovey, P. (1997). What is emotional intelligence? In P. Salovey & D. Sluyter (Eds.), *Emotional development and emotional intelligence: Educational implications* (pp. 3–31). New York, NY: Basic Books.
- Mayer, J. D., Salovey, P., & Caruso, D. R. (2002). *Mayer–Salovey–Caruso emotional intelligence test (MSCEIT) user's manual*. Toronto, Canada: MHS.
- Mayer, J. D., Salovey, P., Caruso, D. R., & Sitarenios, G. (2001). Emotional intelligence as a standard intelligence. *Emotion*, 1, 232–242. doi: 10.1037//1528-3542.1.3.232-242
- Mayer, J. D., Salovey, P., Caruso, D. R., & Sitarenios, G. (2003). Measuring emotional intelligence with the MSCEIT V2.0. *Emotion*, 3, 97–105. doi: 10.1037/1528-3542.3.1.97
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York, NY: American Council on Education and Macmillan.
- Neubauer, A. C., & Freudenthaler, H. H. (2005). Models of emotional intelligence. In R. Schulze & R. Roberts (Eds.), *Emotional intelligence: An international handbook* (pp. 31–50). Cambridge, UK: Hogrefe & Huber Publishers.
- Oatley, K., Keltner, D., & Jenkins, J. (2006). *Understanding emotions* (2nd ed.). Oxford, UK: Blackwell Publishing.
- O'Sullivan, M. (2007). Trolling for trout, trawling for tuna: The methodological morass in measuring emotional intelligence. In G. Matthews, M. Zeidner & R. Roberts (Eds.), *The science of emotional intelligence: Knowns and unknowns*. New York, NY: Oxford University Press.
- O'Sullivan, M., & Ekman, P. (2008). Facial expression recognition and emotional intelligence. In K. B. Leeland (Ed.), *Face recognition: New research* (pp. 25–40). New York, NY: Nova Science Publishers.
- Palmer, B., Gignac, G., Manocha, R., & Stough, C. (2005). A psychometric evaluation of the Mayer–Salovey–Caruso emotional intelligence test version 2.0. *Intelligence*, 33, 285–305. doi: 10.1016/j.intell.2004.11.003
- Roberts, R., Schulze, R., O'Brien, K., Reid, J., MacCann, C., & Maul, A. (2006). Exploring the validity of the Mayer–Salovey–Caruso emotional intelligence test (MSCEIT) with established emotions measures. *Emotion*, 6, 663–669. doi: 10.1037/1528-3542.6.4.663
- Rode, J. C., Mooney, C. H., Arthaud-Day, M. L., Near, J. P., Rubin, R. S., Baldwin, T. T., & Bommer, W. H. (2008). An examination of the structural, discriminant, nomological, and incremental predictive validity of the MSCEIT V2.0. *Intelligence*, 36, 350–366. doi: 10.1016/j.intell.2007.07.002
- Rossen, E., Kranzler, J. H., & Algina, J. (2008). Confirmatory factor analysis of the Mayer–Salovey–Caruso emotional intelligence test version 2.0 (MSCEIT). *Personality and Individual Differences*, 44, 1258–1269. doi: 10.1016/j.paid.2007.11.020
- Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion interferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology*, 32, 76–92. doi: 10.1177/0022022101032001009
- Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah, NJ: Lawrence Erlbaum.
- Zeidner, M., Matthews, G., & Roberts, R. D. (2001). Slow down, you move too fast: Emotional intelligence remains an “elusive” intelligence. *Emotion*, 1, 265–275. doi: 10.1037//1528-3542.1.3.265-275