emotion review

*Emotion Review*
Vol. 4, No. 4 (October 2012) 403–408

ISSN 1754-0739
DOI: 10.1177/1754073912445815
er.sagepub.com

# The Validity of the MSCEIT: Additional Analyses and Evidence

John D. Mayer
*Department of Psychology, University of New Hampshire, USA*

Peter Salovey
David R. Caruso
*Yale University, USA*

## Abstract

We address concerns raised by Maul (2012) regarding the validity of the Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT). We respond to requests for clarifications of our model, and explain why the MSCEIT's scoring methods stand up to scrutiny and why many reported reliabilities of the MSCEIT may be underestimates, using reanalyses of the test's standardization sample of $N$ = 5,000 to illustrate our point. We also organize findings from four recent articles that provide evidence for the MSCEIT's validity based on its relations with other tests.

### Keywords

criterion evidence, emotional intelligence, MSCEIT, reliability, validity

Maul (2012) reviews the validity of an ability measure of emotional intelligence: the Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT; Mayer, Salovey, Caruso, & Sitarenios, 2003). He explores what the MSCEIT most likely measures and whether its technical attributes are adequate for the job. Maul accepts the premise that emotional intelligence (EI) is worth discussing and that tasks can be constructed to measure mental abilities in emotion-related reasoning. The MSCEIT is based on a model of emotional intelligence that specifies four specific "branches" (of a hierarchy) of problem-solving: perceiving emotion accurately, using emotions to facilitate thought, understanding emotions, and managing emotion (Mayer & Salovey, 1997). The MSCEIT produces an index of EI for each of the four branches (e.g., understanding emotions), as well as an overall emotional intelligence quotient (EIQ) score. Each of the branches is represented by two different tasks, and each task is represented by multiple items. According to Maul, our theoretical statements regarding emotional intelligence "set reasonable boundaries on the target domains of each of the four proposed branches of EI" (Maul, 2012, p. 398). In our response, we will provide some requested clarifications and new evidence addressing some of the questions Maul has raised concerning the MSCEIT's validity.

In his discussion of the MSCEIT, Maul expressed concerns about: (a) a need for clarification in portions of our theory; (b) the adequacy of the test's scoring system; (c) the content representation, reliability, and generalizability of the test scores; (d) the correlations between the test's scores and other measures that assess abilities in the emotional intelligence domain; and (e) correlations between the scores and criteria that are predicted theoretically. We will organize our comments according to these areas, and include findings so recent that neither Maul's article nor our own recent reviews included them (Mayer, Roberts, & Barsade, 2008; Mayer, Salovey, & Caruso, 2008).

## Issues of Theoretical Clarity

Although Maul notes that our theory of emotional intelligence "set reasonable bounds" on EI, he also raised issues that he would like to have clarified (Maul, 2012, p. 398). For example, regarding the perceiving emotion area, the MSCEIT asks test-takers to identify emotional content in faces (one task) and landscapes and designs (a second task). Regarding this branch, Maul asked us to clarify why we include landscapes and designs. To respond, many forms of visual stimuli connote certain emotions: Barren landscapes often connote sadness, sharp angles are associated with anger, and squiggly lines potentially indicate humor and joy (Buck, 1984; Hevner, 1935; Jansson-Boyd, 2011; Kastl & Child, 1968; Rosenhan & Messick, 1966; Windhager et al., 2008). Collectively, the ability to perceive these seemed (to us) to represent expertise in perceiving emotion and factor analyses indicated that the ability to perceive such emotional connotations was related to the overall MSCEIT test scores.

*Author note*: The authors receive a royalty on a share of the sales of the MSCEIT, which is copyrighted by MultiHealth Systems (MHS) of Toronto, Canada.
*Corresponding author*: John D. Mayer, Department of Psychology, 10 Library Way, University of New Hampshire, Durham, NH 03824, USA. *Email:* jack.mayer@unh.edu

Regarding the understanding emotions area, Maul asks what we meant in describing understanding as "appreciat[ing] [. . .] emotional meanings" (Maul, 2012, p. 398). By appreciating meanings we meant that a person who possessed emotional knowledge could understand emotional word meanings and concepts, understand the situations and other events that bring about emotions, and evaluate emotions according to their hedonic tone, and their moral value in a given context.

Regarding managing emotion, Maul would like us to clarify what we meant by "personal understanding and growth" (Maul, 2012, p. 398). We believe, in this regard, that people apply emotional reasoning so as to understand themselves and to develop more mature relationships with others. Because such positive development can occur in multiple ways depending upon the individual's personal goals and context, we did not specify exactly how a person would carry this out (see, e.g., Helgeson, Reynolds, & Tomich, 2006; Ryan, Huta, & Deci, 2008; Ryff & Singer, 2008). Our aim for the MSCEIT was limited to assessing whether a person knows how to manage their own and others' emotions, assuming that they often will do so to promote their own and others' positive development.

Of the four areas, Maul expresses his greatest reservations about the using emotion measures, raising concerns that we may not have distinguished the branch well enough, and indicating that the tasks we employ to assess "using" may not reflect our theory (cf., Joseph & Newman, 2010, p. 55). We believe it is worth distinguishing between the often naturally-occurring emotions that people use to guide their thinking (using emotion) versus managing emotional states themselves (managing emotion). Using one's emotions occurs when, for example, a person who feels sad decides it is a good time to undertake some detailed proof-reading, or when a person employs different perspectives on a problem, that were brought about by transitions from one mood to another (which prompts alternative viewpoints; e.g., Blanchette & Richards, 2010; Bower, 1981; Mayer, Gaschke, Braverman, & Evans, 1992). By contrast, managing emotions involves changing the feeling states themselves, as when a sad person manages her attention so as to keep her sadness within manageable boundaries, or decides to cheer herself up. The "managing" branch measures people's abilities to change their own and others' emotions through active coping mechanisms—either intrapsychically or through interpersonal acts.

## Issues of Scoring the MSCEIT

### *Item Generation was Guided by Theory*

Maul (2012) argues that good theories of emotion reasoning are necessary for constructing good items in the EI domain, in part so as to ensure that there are correct and incorrect answers to choose from for each question. In developing the MSCEIT we were guided by many such theories. For example, we used detailed theories of the affective lexicon to construct items for the understanding area of the test (Clore, Ortony, & Foss, 1987; Plutchik, 2000), and also employed theories for the management, perception, and using areas (e.g., Isen, Daubman, & Nowicki, 1987; Rosenhan & Messick, 1966). Our theoretically informed approach surely added to the quality of the overall test, and it is one reason (as we will show) that there is considerable evidence for the MSCEIT's validity.

### *The Scoring System is Adequate*

Regarding the MSCEIT, Maul suggests that "support for the adequacy of the scoring system [. . .] does not seem sufficient" (Maul, 2012, p. 397). The MSCEIT gives the user a choice to employ one of two scoring keys for the test, which are extremely similar: one based on a consensus among 21 experts and the other based on a consensus among 5,000 test-takers (Mayer, Salovey, & Caruso, 2002). Consensus scoring, common to both keys, works as follows: Consider a multiple choice item where 70% of those 21 emotions experts identified "b" as a correct answer. The score of a participant who chose "b" would be incremented by .70 (and if the respondent chose "d" and 10% of the experts had done so, the score would be incremented .10). The MSCEIT also can be scored using the consensus of the general standardization sample. These two methods produced highly similar scoring keys: The correlation between the weights of alternatives calculated based on the experts versus those calculated using the test-taker sample varied from $r$ = .88 to .91 depending upon the subsample studied (all $N$s > 2,000). Moreover, the two scoring keys generated branch, area, and total scores with $r$s = .96 to .98—so high as to be nearly indistinguishable from one another. These high correlations result whether using the first approximately 2,000 participants studied or the full normative sample of 5,000 individuals (Mayer et al., 2002, p. 34, Table 5.12; Mayer et al., 2003).

Maul suggested that perhaps the experts were not very expert. The group of 21 individuals consisted of 10 men and 9 women (2 did not identify their gender); all were members of the International Society for Research on Emotion, whose expertise involved research and scholarship on emotions. The group included 16 professors and lecturers, 2 "researchers," and 2 doctoral students (and one non-identified) with a median age of 38. We consider this adequate evidence of their expertise. Despite their diversity (from North America, Europe, and the Middle East), their level of agreement as to answers was also higher than that in the general sample (Mayer et al., 2003).

Given that the expert group was highly credentialed, it is worth briefly exploring why they agreed so highly with the general sample as to the correct answers. First, much emotion knowledge reflects common use of an emotion language. Almost everyone knows, for example, that fear arises in response to a threat. In such instances, experts can help identify the common consensus, but so can a scoring key that simply employs information from how a general sample responds. Second, emotional information is a domain best modeled by fuzzy logic (or probabilistic computation), in which multiple conditions may apply and more than one correct answer is possible. A statement that a person "is angry" can indicate a range of possible levels of anger depending upon the context, as well as many different outcomes of such anger—to let the anger pass, to express it, to reframe it. Consequently answers to

emotional problems often involve a lack of certainty, and are dependent on emphasis ("how much anger?"). In such instances, consensus across a group of people who have everyday language skills and who all experience emotions may approximate expert opinion quite closely.

We have argued that consensus is a legitimate means of determining correct answers and it has been used in other contexts, such as the measurement of traditional intelligence (e.g., the comprehension subscale of the Wechsler Adult Intelligence Scale [WAIS]-III). The *WAIS-III Technical Manual* (Psychological Corporation, 1997) suggests that a form of expert consensus was employed for subtests such as vocabulary, similarities, and comprehension: Two team members collected potential responses (e.g., definitions of a vocabulary word), placed them into groups, and examined the discrepancies among their groupings. The manual explains how, at a later stage, "team members had to agree on the grouping of responses and [. . .] evaluated the quality of the responses and assigned a score value (0, 1, or 2) to each [. . .] on the basis of the accuracy of the response" (Psychological Corporation, 1997, p. 37). We would have been delighted to reference all the response alternatives on the MSCEIT to an authoritative reference book. In 2000, however, there was no widely agreed upon dictionary of emotional meanings. As that changes, veridical scoring becomes more possible.

## Issues of Construct Representation and Reliability

### *The MSCEIT Has Good Construct Representation*

Maul (2012) argued that, because the MSCEIT does not test each and every emotional intelligence skill specified in our theory, the MSCEIT suffers from concept underrepresentation. Because human skills are diverse in most intellectual domains, it is neither advisable nor possible to measure all possible skills therein. An intelligence test that measures verbal-comprehension intelligence, for example, does not do so by representing *all* such skills—that would require including the abilities to sound out words, to know their roots, to find rhymes, to know how to write a sentence, to be able to deliver a persuasive speech, to manipulate mathematical symbols in equations, and so forth. Prudent sampling within the verbal-comprehension domain is generally regarded as sufficient. Similarly, the MSCEIT is designed to sample important skills within each of the theory's four branches rather than to measure them exhaustively. That said, the MSCEIT was designed as a relatively brief assessment so as to encourage its use and thereby accumulate evidence regarding its validity. The fact that the test employs two tasks per branch (rather than more) means that it is less useful for some research purposes than others (i.e., discovering the covariance structure of EI).

### *Reports of Reliability at or Close to Those Found in the Test Manual*

Maul reports ranges of Cronbach's coefficient alpha estimates of the MSCEIT's total and branch reliabilities that are a bit lower than those described in the test manual. The MSCEIT is a composite of diverse individual tasks (two tasks represent a branch). Although items are homogeneous (i.e., of parallel form) within a task, the items are heterogeneous across tasks by design: The diversity of items and item response methods was intentionally built into the test so as to maximize the test's overall validity. For example, the items address emotion expressed in a face in one task (faces) and the meaning of emotion blends in another task (blends). The heterogeneous items dictate, in important ways, how internal-consistency reliability estimates ought to be calculated: Such estimates are best based on a division of the test into two or more equivalent forms (i.e., roughly parallel). To construct equivalent forms of the MSCEIT, one can take half of the items *of each task* and place them on one form, and place the remaining half of the items on the other form. This way, items representing each task can be found on each split of the test—making them equivalent. When we reanalyzed the standardization data for the MSCEIT ($N = 5{,}000$) and calculated the reliability in this way, the reliability was $r = .93$, as shown in the first column of Table 1 (using the general-consensus reliability estimates that Maul discusses).

Split-halves of the same data set can be created that are non-equivalent. In such cases the estimated reliability of the MSCEIT will be far lower. SPSS, for example, uses a default split-half approach that takes the first half of all the test items (based on the order of variables entered) and compares them with the second half of the test items. If a pair of MSCEIT tasks on a given branch had equal numbers of items (in fact, the tasks vary somewhat in length), the SPSS program would estimate the

**Table 1.** Reliability of the MSCEIT and its subscales: Estimates that meet or violate assumptions regarding heterogeneous items ($N = 5{,}000$)

| | Split half estimates | | | | Coefficient alpha estimates | | | |
|---|---|---|---|---|---|---|---|---|
| | Equivalent forms | | First vs. second half[a] | | Task level | | Item level | |
| | *Meets assumptions for a test with heterogeneous items?* | | | | | | | |
| | *Yes* | | *No* | | *No* | | *No* | |
| | *K* | *r*[b] | *K* | *r* | *K* | α | *K*[c] | α |
| MSCEIT Total | 2 | .93 | 2 | .80 | 8 | .75 | 122 | .91 |
| Perception | 2 | .92 | 2 | .67 | 2 | .46 | 42 | .89 |
| Using | 2 | .81 | 2 | .61 | 2 | .56 | 24 | .74 |
| Understanding | 2 | .80 | 2 | .75 | 2 | .65 | 32 | .78 |
| Managing | 2 | .84 | 2 | .76 | 2 | .70 | 24 | .80 |

[a]This analysis employs the SPSS default split half, which chooses items based on their order of specification, treating the first half of the items as making up one form, and the second half of the items as making up the second.
[b]These estimates vary by $r = .01$ or .02 for three of the branches from the values reported in the manual for general-consensus data (the data file we employed here) owing either to different ways of creating the equivalent forms, missing data management, or a difference in statistical programs.
[c]Of the 141 items on the MSCEIT, 19 are not scored, yielding 122 functional items.

branch reliability by comparing items on one task to the non-equivalent items on the second task, yielding non-equivalent forms. For the overall MSCEIT, SPSS would compare the first four tasks with the last four, again yielding non-equivalent forms. We also analyzed the standardization sample in this way. The estimated reliability for the overall test was $r = .80$; branch values also were lower (Table 1, column 2).

A basic assumption of coefficient alpha is homogeneity of items. Oversimplifying slightly, coefficient alpha provides an estimate of reliability that approximates the average of many haphazardly chosen split-half estimates. For a heterogeneous test such as the MSCEIT, this means that, in essence, reliability estimates based on equivalent split-half form will be averaged with far more numerous haphazard splits of items. Cronbach (1951) noted that, as a consequence, the alpha reliability estimate would be low relative to a split half using equivalent half tests or, in his words, alpha may estimate an inappropriately low reliability for a test even though a given half of the test "may nonetheless have a high correlation with a carefully planned equivalent form" (Cronbach, 1951, pp. 300, 307). Coefficient alpha's estimate will be still lower if it is calculated on only a few parts (Osburn, 2000), where the number of parts is represented as $K$. Regarding the MSCEIT, if the alpha coefficient for each of the four branches is calculated at the level of the two tasks that make it up (i.e., $K = 2$), and the overall MSCEIT is calculated based on the eight tasks (i.e., $K = 8$), it would lead to a considerable underestimate of the MSCEIT's overall reliability at $\alpha = .75$, as compared to the $r = .93$ of the equivalent forms (see Table 1, column 3). As $K$ grows larger, as would happen for an alpha estimated at the item level, alpha provides less of an underestimate, as shown in Table 1 in the right-most columns. Most reports of MSCEIT coefficient alphas do not specify whether they are at the task level or not; this information would be helpful to report in the future. Even so, judging by our standardization sample, the alpha based on $K = 24$ to 42 items for individual tasks appears to underestimate the branch-level reliabilities somewhat. We repeated these analyses for the expert-scored data (which yields reliability estimates about .01 to .03 lower throughout); the same pattern of underestimated reliability held.

Our point is that researchers who report alphas for the MSCEIT thereby underestimate the test's reliability as a consequence of violating the coefficient's assumptions in ways known to lead to such underestimates. A better estimate of the MSCEIT's reliability can be obtained through equivalent-forms split-half estimates (Cronbach, 1951) and stratified alphas (Osburn, 2000; cf. Tellegen & Briggs, 1967). Another appropriate reliability estimate is the test–retest reliability estimate; the MSCEIT test–retest reliability is $r = .86$ (Brackett & Mayer, 2003).

## Recent Evidence Indicates the MSCEIT Correlates Well with Other EI Ability Measures

Most importantly, if the MSCEIT correlates in a reasonable fashion with similar other ability measures intended to assess emotional intelligence, then the MSCEIT "works" in some fundamental way, regardless of other technical issues. Correlations between the MSCEIT and several similar performance-based measures of EI not included in Maul's review are shown in Table 2.

**Table 2.** Correlations between the MSCEIT and related ability scales

| Research report | MSCEIT subscales[a] and total sample reliabilities | | | | | Criterion measure[b] | | Correlations between the MSCEIT subscale[a] and total scores, and criteria | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Per | Usi | Und | Man | Tot | Name | Reliability | Per | Usi | Und | Man | Tot |
| Roberts et al. (2006)[c] | .85 | .67 | .68 | .68 | n.r. | Vocal 1 | .45 | −.05 | .01 | .23 | .15 | n.r. |
| | | | | | | JACBART | .73 | .03 | .11 | .18 | .17 | n.r. |
| Austin (2010) | .86 | .58 | .66 | .66 | .90 | STEU | .48 | .16 | .21 | .44 | .17 | .33 |
| | | | | | | STEM | .67 | .13 | .25 | .40 | .30 | .36 |
| | | | | | | Facial Blends | n.r. | .08 | .00 | .18 | .17 | .14 |
| Maul (2011) | .87 | .62 | .62 | .66 | .85 | MEIS-Tot | .94 | n.r. | n.r. | n.r. | n.r. | .72 |
| | | | | | | MEIS-Per | .94 | .69 | .49 | .34 | .15 | n.r. |
| | | | | | | MEIS-Usi | .82 | .18 | .28 | .25 | .28 | n.r. |
| | | | | | | MEIS-Und | .67 | .19 | .17 | .46 | .37 | n.r. |
| | | | | | | MEIS-Man | .66 | −.03 | .11 | .24 | .40 | n.r. |
| Mayer, Panter, & Caruso (2012)[d] | n.i. | n.i. | .77 | .81 | .86 | Eyes of the Mind | .68 | n.i. | n.i. | .57 | .41 | .56 |

[a]The abbreviations refer to the four subscales of the MSCEIT, per = perceiving emotions, usi = using emotions, und = understanding emotions, and man = managing emotions.
[b]Vocal 1: Index of Vocal Emotion Recognition; JACBART: Japanese and Caucasian Brief Affect Recognition Test; STEU: Situational Test of Emotional Understanding; STEM: Situational Test of Emotion Management; MEIS: Multibranch Emotional Intelligence Scale; Eyes of the Mind: the Reading the Mind in the Eyes Test.
[c]Scores were reported by task. Branch level task reliabilities were based on an average of the two associated tasks, stepped up via the Spearman-Brown; criterion correlations are the average of the two relevant task correlations.
[d]Calculated from the study specifically for this reply; only strategic subscales of the MSCEIT were used in the study.
n.i.—not included in study; n.r.—not reported in article.

The left-hand columns of Table 2 include the reported MSCEIT reliabilities of the studies, the middle columns indicate the criterion scales and their reliabilities, and the far right columns show the correlations of the scores with those criterion scales.

The MSCEIT correlated differently with different groups of criteria. The MSCEIT was least correlated with three measures of emotional perception: vocal 1 (vocal tone; Roberts et al., 2006), the Japanese and Caucasian Brief Affect Recognition Test (JACBART; Matsumoto et al., 2000), and "facial blends" (Austin, 2010). Of these measures, vocal 1 has a reliability of $r = .45$, facial blends is of unknown reliability, and neither scale correlates highly with the MSCEIT. Because the validity of the two criterion scales is uncertain, these results are difficult to interpret. The JACBART, however, is a widely employed test with demonstrated validity. The correlations between the MSCEIT and the JACBART scale are no higher than $r = .18$ (Table 2, right-hand side). It is not clear why the correlation between these criteria scales and the perception branch should be so low (essentially zero), but as we have acknowledged elsewhere (Mayer et al., 2008), it suggests that the perception branch of the MSCEIT may be insufficient to measure this skill. In fact, the MSCEIT understanding branch exhibits higher correlations with these scales, at $r = .23$, .18, and .18.

The MSCEIT did correlate, however, with an apparently similar measure, the Reading the Mind in the Eyes test (Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001). Baron-Cohen's test involves looking at a picture of a person's eyes and trying to estimate what the person is feeling. The MSCEIT strategic area (understanding and managing areas) shown in the "total" column, correlated $r = .56$ with the Reading the Mind test. We have suggested elsewhere that the Reading the Mind test may measure emotional understanding as well as perception because of the advanced vocabulary used on it (Mayer, Panter, & Caruso, 2012).

Another pair of criterion measures was developed to measure emotional understanding and management: MacCann and Roberts's Situational Test of Emotional Understanding and Situational Test of Emotional Management (STEU and STEM; MacCann & Roberts, 2008). The STEU is scored veridically: The authors claim there is a correct and incorrect answer for each question, determined by theory (Roseman, 2001)—an approach that Maul says he prefers. The STEM uses an expert consensus. In the study cited here, the alpha reliabilities of the STEU and STEM were $r = .48$ and .67 (lower than the original reports), which were low enough, especially regarding the STEU, to meaningfully depress their correlations with the MSCEIT (discussed next).

The STEU measures appraising emotions, a form of emotional understanding not directly measured by the MSCEIT. It can reasonably be considered a new task in the understanding area and, consistent with that idea, it exhibits a subscale-like correlation with the understanding area of $r = .44$ (Austin, 2010). The STEM and MSCEIT management branches are related $r = .30$. The factor structures of the STEM and STEU are complex, which may help explain why the STEM correlated more highly with MSCEIT understanding, $r = .40$, than with the management branch (Ferguson & Austin, 2010).

Finally, Maul (2011) compared the MSCEIT to an earlier test of emotional intelligence, the Multifactor Emotional Intelligence Scale (MEIS), a 12-task measure of EI which has several similar tasks but no item overlap with the MSCEIT (Mayer, Salovey, & Caruso, 1997). The MSCEIT and MEIS correlated $r = .72$ (Maul, 2011).

Overall, the new findings reported in Table 2 make the case that the MSCEIT correlates meaningfully with a variety of ability based criteria of EI. The MSCEIT correlates in meaningful fashions with the Reading the Mind test, the STEU, the STEM, and the MEIS. Collectively, they provide encouraging support of the adequacy of the MSCEIT's relations with similar ability measures, excepting relatively pure measures of emotional perception. When Maul (2012, p. 399) wrote "evidence [is] still scant on this topic," he had not included these most recent findings.

## Criterion Correlations More Generally Support the Validity of the MSCEIT

A test's validity should be supported by theoretically-expected correlations with a variety of criteria. The recent *Annual Review* article on EI found several trends in criterion correlations that are predictable by theory: in particular, that EI enhances social relationships both among friends and among colleagues at work (Mayer et al., 2008). Maul basically agrees that "MSCEIT scores are associated with [. . .] other psychological variables and positive outcomes in a manner fairly consistent with the idea that the MSCEIT measures emotional intelligence" (Maul, 2012, p. 400).

## General Conclusion

In this comment we did not reply to all of Maul's criticisms of the MSCEIT, but only to those we regarded as most crucial. Taking into account the new findings we have reported (as well as previous findings), the argument for the MSCEIT's overall validity is growing and arguably quite strong, notwithstanding the technical imperfections that are a part of any real-life form of measurement, and acknowledging that improvements in the MSCEIT and measurement in the area are desirable.

### References

Austin, E. J. (2010). Measurement of ability emotional intelligence: Results for two new tests. *British Journal of Psychology*, *101*, 563–578. doi:10.1348/000712609X474370

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry*, *42*, 241–251. doi:10.1111/1469-7610.00715

Blanchette, I., & Richards, A. (2010). The influence of affect on higher level cognition: A review of research on interpretation, judgement, decision making and reasoning. *Cognition & Emotion*, *24*, 561–595. doi:10.1080/02699930903132496

Bower, G. H. (1981). Mood and memory. *American Psychologist*, *36*, 129–148. doi:10.1037/0003-066X.36.2.129

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

Buck, R. (1984). *The communication of emotion*. New York, NY: Guilford Press.

Clore, G. L., Ortony, A., & Foss, M. A. (1987). The psychological foundations of the affective lexicon. *Journal of Personality and Social Psychology*, *53*, 751–766. doi:10.1037/0022-3514.53.4.751

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297–334. doi:10.1007/BF02310555

Ferguson, F. J., & Austin, E. J. (2010). Associations of trait and ability emotional intelligence with performance on theory of mind tasks in an adult sample. *Personality and Individual Differences*, *49*, 414–418. doi:10.1016/j.paid.2010.04.009

Helgeson, V. S., Reynolds, K. A., & Tomich, P. L. (2006). A meta-analytic review of benefit finding and growth. *Journal of Consulting and Clinical Psychology*, *74*, 797–816. doi:10.1037/0022-006X.74.5.797

Hevner, K. (1935). Experimental studies of the affective value of colors and lines. *Journal of Applied Psychology*, *19*, 385–398. doi:10.1037/h0055538

Isen, A. M., Daubman, K. A., & Nowicki, G. P. (1987). Positive affect facilitates creative problem solving. *Journal of Personality and Social Psychology*, *52*, 1122–1131. doi:10.1037/0022-3514.52.6.1122

Jansson-Boyd, C. (2011). Designing aesthetic concepts: Can it be done? *Psychology of Aesthetics, Creativity, and the Arts*, *5*, 279–290. doi:10.1037/a0023490

Joseph, D. L., & Newman, D. A. (2010). Emotional intelligence: An integrative meta-analysis and cascading model. *Journal of Applied Psychology*, *95*, 54–78. doi:10.1037/a0017286

Kastl, A. J., & Child, I. L. (1968). Emotional meaning of four typographical variables. *Journal of Applied Psychology*, *52*, 440–446. doi:10.1037/h0026506

MacCann, C., & Roberts, R. D. (2008). New paradigms for assessing emotional intelligence: Theory and data. *Emotion*, *8*, 540–551. doi:10.1037/a0012746

Matsumoto, D., LeRoux, J., Wilson-Cohn, C., Raroque, J., Kooken, 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 (JACBERT). *Journal of Nonverbal Behavior*, *24*, 179–209. doi:10.1023/A:1006668120583

Maul, A. (2011). 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. (2012). The validity of the Mayer-Salovey-Caruso emotional intelligence test (MSCEIT) as a measure of emotional intelligence. *Emotion Review*, *4,* 394–402.

Mayer, J. D., Gaschke, Y. N., Braverman, D. L., & Evans, T. W. (1992). Mood-congruent judgment is a general effect. *Journal of Personality and Social Psychology*, *63*, 119–132. doi:10.1037/0022-3514.63.1.119

Mayer, J. D., Panter, A. T., & Caruso, D. R. (2012). Does personal intelligence exist? Evidence from a new ability based measure. *Journal of Personality Assessment*, *94*, 124–140.

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 D. J. Sluyter (Ed.), *Emotional development and emotional intelligence: Educational implications* (pp. 3–34). New York, NY: Basic Books.

Mayer, J. D., Salovey, P., & Caruso, D. R. (1997). *Emotional intelligence test*. Needham, MA: Virtual Knowledge.

Mayer, J. D., Salovey, P., & Caruso, D. R. (2002). *Mayer-Salovey-Caruso emotional intelligence test (MSCEIT) user's manual*. Toronto, Canada: Multi-Health Systems.

Mayer, J. D., Salovey, P., & Caruso, D. R. (2008). Emotional intelligence: New ability or eclectic traits? *American Psychologist*, *63*, 503–517. doi:10.1037/0003-066X.63.6.503

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

Osburn, H. G. (2000). Coefficient alpha and related internal consistency reliability coefficients. *Psychological Methods*, *5*, 343–355. doi:10.1037/1082-989X.5.3.343

Plutchik, R. (2000). A psychoevolutionary theory of emotion. In R. Plutchik (Ed.), *Emotions in the practice of psychotherapy: Clinical implications of affect theories* (pp. 59–79). Washington, DC: American Psychological Association. doi:10.1037/10366-004

Psychological Corporation. (1997). *WAIS-III, WMS-III technical manual*. San Antonio, TX: Author.

Roberts, R. D., Schulze, R., O'Brien, K., MacCann, C., Reid, J., & 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

Roseman, I. J. (2001). A model of appraisal in the emotion system: Integrating theory, research, and applications. In K. R. Scherer, A. Schorr, and T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 68–91). New York, NY: Oxford University Press.

Rosenhan, D., & Messick, S. (1966). Affect and expectation. *Journal of Personality and Social Psychology*, *3*, 38–44. doi:10.1037/h0022633

Ryan, R. M., Huta, V., & Deci, E. L. (2008). Living well: A self-determination theory perspective on eudaimonia. *Journal of Happiness Studies*, *9*, 139–170. doi:10.1007/s10902-006-9023-4

Ryff, C. D., & Singer, B. H. (2008). Know thyself and become what you are: A eudaimonic approach to psychological well-being. *Journal of Happiness Studies*, *9*, 13–39. doi:10.1007/s10902-006-9019-0

Tellegen, A., & Briggs, P. F. (1967). Old wine in new skins: Grouping Wechsler Subtests into new scales. *Journal of Consulting Psychology*, *31*, 499–506. doi:10.1037/h0024963

Windhager, S., Slice, D. E., Schaefer, K., Oberzaucher, E., Thorstensen, T., & Grammer, K. (2008). Face to face: The perception of automotive designs. *Human Nature*, *19*, 331–346. doi:10.1007/s12110-008-9047-z