A Further Consideration of the Issues of Emotional Intelligence

John D. Mayer
Department of Psychology
University of New Hampshire

Peter Salovey
Department of Psychology
Yale University

David R. Caruso
Work-Life Strategies

An American Psychological Association (APA) task force was assigned the difficult task of creating a consensus document regarding intelligence and related mental abilities. Its members acknowledged much ongoing debate in the field—even regarding what intelligence is. Such debate, however, can still reflect progress. As the task force noted, “Scientific research rarely begins with fully agreed definitions, though it may eventually lead to them” (Neisser et al., 1996, p. 77). Writing more generally of such controversies, the Neisser task force counseled:

In a field where so many issues are unresolved and so many questions unanswered, the confident tone that has characterized most of the debate on these topics is clearly out of place. The study of intelligence does not need politicized assertions and recriminations; it needs self-restraint, reflection, and a great deal more research. (Neisser et al., 1996, p. 97)

The task force document also concluded that there exist other forms of intelligence about which less has been researched and understood (Neisser et al., 1996).

It has been our goal to learn more about one of those less understood intelligences: an emotional intelligence. Herein, we continue the discussion of emotional intelligence we began in our target article, but now addressing the points raised by the commentators. We have organized the concerns they raised into four areas: (a) theoretical issues, (b) methodological issues, (c) empirical issues, and (d) looking toward the future. We wish to state our appreciation at the outset for the valuable insights offered by the commentators on this issue, as well as by the authors of our co-target article.

Theoretical Issues

The State of the Field

From our vantage point, there exist extremes in the debate over emotional intelligence. On the one hand, there are popularizers who overclaim the breadth and promise of emotional intelligence, defining it as many different things and inflating its predictive power. We have not been shy to criticize them (Mayer, 1999; Mayer & Cobb, 2000; Mayer, Salovey, & Caruso, 2000). We agree with Gohm (this issue) that academics need to be clearer in drawing a distinction between popular claims and real scientific research; not to do so potentially means not just emotional intelligence research, but other relevant fields more generally.

At the other end of this continuum are those who question and criticize the concept of emotional intelligence itself. In our reading, some of those critics have progressed from “there is little evidence it exists,” to “it cannot be measured reliably,” to “it does not predict important outcomes”—a substantial change of position (e.g., Davies, Stankov, & Roberts, 1998; Epstein, 1998; Roberts, Zeidner, & Matthews, 2001). And yet, as in the intelligence field more generally, the idea of an emotional intelligence continues to generate a great deal of controversy. Brody (this issue), for example, notes that we “have presented convincing evidence that the MSCEIT [Mayer–Salovey–Caruso Emotional Intelligence Test] may be scored using consensus measures and that the test has adequate reliability.” He also concludes, however, that “there is no foundation for the use of the test in applied settings nor is there evidence indicating that the test measures an important dimension of individual differences.”

We are not as pessimistic as Brody (this issue) in our view of the state of the field of emotional intelligence or its measure with the MSCEIT. Nor are we as pessimistic in our assessment of the state of the field of general intelligence, about which Brody has noted that “We know how to measure something called intelligence, but we do not know what has been measured” (Brody, 2000, p. 30). Rather, our own reading and research indicates to us that the studies of both emotional and general intelligence can be a far more informative and rewarding enterprise than suggested by such a view. Our perspective
is, perhaps, more consistent with conclusions of the APA task force that progress is being made (of which Brody was a member).

How Do We Conceive of the "Emotion" of Emotional Intelligence?

Averill (this issue) begins his commentary with a very productive reading of what we mean by the term emotions. Averill interprets our position on emotion as being that:

(a) each kind of emotion (anger, fear, etc.) shares certain essential features that are biologically based, (b) simpler emotions may combine to form more complex emotions, and (c) emotions may be regulated but not fundamentally altered by display rules. (Averill, this issue)

To Averill’s (this issue) useful summary, we would add that (d) emotions have the functional purpose of signaling relationships and changes in relationships, real or imagined, principally between people and their environments (including other people). Averill further suggests that emotion and cognition may be produced by the same physiological functional area. He compares such processes to a radio that employs the same components to broadcast music and news. We see the emergence of emotions and cognition somewhat differently. This leads us to add (e) that emotions and cognitions represent different functions of the mind, if not the brain, recognizing that the two often interact and are expressed in an integrated form (Mayer, 2001). We would draw an analogy to how, often, one division of a broadcast organization produces news, while another produces musical programming, and the use of the two types of communication can be integrated to varying degrees. Our view stems in part from a brilliant theoretical analysis of how emotional behavior emerges, also by Averill (1992).

Are Emotional Intelligence and Emotional Creativity the Same?

We agree with Averill (this issue) that our scale(s) measure emotional intelligence, not, in fact, emotional creativity. Fortunately, it is now possible to move from theory to empirical findings in this regard. Icvevic, Brackett, and Mayer (2004) found nonsignificant correlations between Averill’s measures of emotional creativity and the MSCEIT (see target article) in two studies. When emotional creativity is measured as an ability, the correlations across the two studies (Ns = 107; 113) were r = .07 and -.06 with the MSCEIT; when measured via self-report, r = .02 and .12. Also of note, Averill’s measures of emotional creativity outperformed the MSCEIT in predicting creative artistic behavior. We believe that emotional creativity provides a very useful comparison variable to study along with emotional intelligence and a welcome reminder that emotional skills extend beyond the intelligences.

Methodological Issues

Can Emotional Intelligence be Studied With Methods Beyond Intelligence Testing?

Oatley (this issue) asks both sets of authors, “Have they not, perhaps, become too concerned with psychometrics?” and then continues, “My commentary is based around Salovey and Mayer’s (1990) original idea, and asks how, without denigrating important uses of testing, we can avoid the conclusion that emotional intelligence is what emotional intelligence tests measure.”

It is our strong belief that psychometrically sound tools must be created for us to examine the theory of emotional intelligence. At the same time, we recognize that any single viewpoint on the topic is incomplete in itself. Oatley’s (this issue) intriguing commentary pro-
vides an alternative emphasis to our own that focuses on examining emotional expertise: That is, he seeks to study experts at emotional skills and how they do things. Oatley notes that such experts are to be found in imaginative literature. Averill (this issue), too, notes that “Art or literary criticism may be a better model for evaluating emotions.”

We agree that the field can be enriched through such observations. For instance, we have traced back the first definite use of the term emotional intelligence (uncovered so far) to literary criticism (Van Ghen, 1961, pp. 103–107). Literary masterworks that have addressed emotional knowledge in a relatively direct fashion include such examples as Charles Dickens’ reflections over his undisciplined (and then overly disciplined) heart in his semiautobiographical novel, David Copperfield (Dickens, 1850/1987, pp. 664, 697–698, 818, 830), James Joyce’s structural emotional thesis “Love loves to love love,” in Ulysses (Joyce, 1914/1961, p. 333), and Swift’s (1727/1970; see chapters 8–9) meditations on a society without compassion, in his description of Gulliver’s visit to the Houyhnhnms. As Oatley rightly points out, there are many further examples, and his illustrations of emotional experts provide exciting additions.

Testing by itself cannot answer everything necessary about a concept. Other approaches can work together to enrich a topic. The particular value of testing, to our minds, is to provide clear limits and measures of concepts. With that in mind, we go on to consider methods specifically related to assessment in the area.

Do Experts Disagree Over the Answers to Emotional Intelligence Test Items More Than Over the Answers to Cognitive Intelligence Test Items?

One of the methods we employ for scoring our ability measures of emotional intelligence involves using the agreed on answers of a group of experts. Brody (this issue) notes that disagreements among emotion experts as to correct answers are likely to be wider than, say, disagreements among experts as to the correct answer on a scale of cognitive intelligence. We agree with this point. For Brody, such disagreements undermine any sense that there are correct answers to tests of emotional intelligence. Here, however, we offer an alternative interpretation: Emotion problems, more so than the usual cognitive IQ test problems, often involve multiple correct and multiple incorrect answers. For example, good ways to manage an unpleasant feeling could include both (a) changing one’s attributions as to what brought on the bad feeling and (b) distraction, but not (c) replaying the negative event over in one’s mind, or (d) finding someone else feeling bad and talking about negative things (see the work of Gross, 1998). So any apparent greater disagreement among experts is a product of the fact that there are multiple correct answers. Using more flexible response methods in which experts are allowed to indicate multiple patterns of responses to a test item could help clarify this issue.

Do Emotional Intelligence Tests Allow for Nonconsensual, Correct Answers?

Brody also points out that the existence of correct answers to cognitive ability items implies that it is possible for a person with unusually high cognitive ability to provide a response to an item that is nonconsensual and correct (Brody, this issue). This point, which has been raised before, may represent a true existing difference between cognitive and emotional intelligence tests. Note that this does not undermine the function of the test, so long as people vary reliably in choosing correct and incorrect answers—which they do.

The theoretical possibility of a nonconsensual correct answer in the emotions realm is unclear. Many correct answers (e.g., the emotion expressed in a face) are themselves identified by experts according to the consensus among individuals (Mayer, Salovey, Caruso, & Sitarenios, 2001). There are exceptions, however, that allow for the theoretical possibility of correct, nonconsensual answers in the emotional intelligence realm. For example, work by Ekman (2001) on the detection of lying through signals from the face has suggested that only a very few individuals are capable of detecting the necessary indicators at better than chance levels. Are correct, nonconsensual answers relatively infrequent in the emotions realm relative to the cognitive, or have they not yet been identified as well as they could be? This is an interesting question that requires further research. It is a point that, at present, at least, distinguishes emotional intelligence from cognitive intelligence.

Are Emotional Intelligence Test Items Different From Cognitive Test Items in Their Behavioral Relevance?

Goleman (this issue) raises a criticism common to IQ tests in general: “Knowing what one should say, or how one should behave . . . does not mean that one will actually act accordingly in such a situation,” a point which we have made as well (Lopes & Salovey, in press). Some test items may be more closely linked to the desired behavioral criterion than others. So, vocabulary items on cognitive IQ tests or emotional understanding items on emotional IQ tests both reflect the respondent’s immediate comprehension of material. On the other hand, even if a participant can provide a correct answer to an Information subscale item on a
cognitive IQ test (e.g., “What should you do if you found a letter on the ground with a stamp and an address on it?”) or an emotional management question (e.g., “Given a specific situation, how could you cheer someone up?”), it would not necessarily imply the individual would carry out the smart action. Brody (this issue) believes that the case is more extreme for emotional intelligence items than for cognitive intelligence items, but his case seems to rest on a very selective sample of items drawn from emotional intelligence tests. As we have just shown, both cognitive and emotional intelligence tests possess items with comprehension-behavior links of each type.

**Empirical Predictions**

**Do We Demonstrate the Validity of the MSCEIT?**

In our target article, we review studies that support the scoring methods, content coverage, factorial structure, discriminant properties, convergence with related tests, general predictive power, and even neuropsychological patterns of the MSCEIT. Collectively, these studies are highly supportive of the MSCEIT’s validity. Although some questions were raised about incremental validity and effect sizes (see following sections), these need to be placed in context. The current Standards for Educational and Psychological Testing remind readers that “validity is a unitary concept. It is the degree to which all the accumulated evidence supports the intended interpretation of test scores for the proposed purpose …” (Joint Committee on Standards, 1999, p. 11; cf. Landy, 1986).

We believe that we have demonstrated considerable and growing evidence for ability measures of emotional intelligence in our target article and elsewhere (e.g., Brackett & Salovey, in press; Mayer, Salovey, & Caruso, 2002; Salovey, Caruso, & Mayer, in press). Here, we consider several reservations or objections raised in response to the details of that presentation.

**Do Ability Tests of Emotional Intelligence Possess Incremental Validity?**

We and others have been actively engaged in providing validity data for the MSCEIT and related scales, as described in our target article. Brody (this issue) writes that the MSCEIT test should show incremental validity above preexisting tests. We agree. One of the great difficulties with self-judgment scales of emotional intelligence is their considerable overlap with preexisting scales of personality (Brackett & Mayer, 2003). The MSCEIT’s correlations with preexisting scales are far lower than self-report scales (see target article), but the issue still arises.

As with any statistical analyses, however, some sensitivity to the context and purpose of the study needs to be considered. Brody writes that the MSCEIT test should show incremental validity above “standard tests of intelligence and the Big Five personality traits.” If one took this at face value, it would mean always statistically controlling for the effects of six or seven scales. The Big Five scores are (theoretically) orthogonal to one another and, hence, are typically studied individually. Intelligence tests (he uses the plural) would add at least two more variables. Although controlling for seven scales may be appropriate in some instances when samples sizes are large enough and the research aims demand it, it may be an inappropriately strict criterion in other research contexts.

The number of tests controlled for matters for three reasons. First, statistically controlling for multiple tests is likely to partial out legitimate variance drawn from partly overlapping tests. Second, statistically controlling for multiple tests is likely to partial out legitimate variance drawn from chance linear combinations of the scales employed. Third, the more tests included, the lower one’s statistical power.

So, how many (and which) tests should be partialed out? The current Standards for Educational and Psychological Testing provide little guidance in this regard. These standards neither define, index, nor (so far as we can tell) mention incremental validity (Joint Committee on Standards, 1999). Incremental validity is, of course, discussed in the psychological literature. It concerns such matters as the incremental predictions of psychological tests over the use of, say, clinical interviews (e.g., Hunsley & Meyer, 2003). When test-to-test comparisons are examined, incremental validity is said to occur when a new test makes better predictions than those possible with simpler and less expensive, already existing tests (e.g., Hunsley & Meyer, 2003, p. 447; Sechrest, 1963). In the present case, we understand this as meaning that an emotional intelligence scale should show incremental predictions above one or two similar tests. That is, we view partialing out six or seven alternatives to represent an exceptionally stringent approach in many instances. We are not arguing against the practice when sample sizes are large, and a case can be made that the variables to be partialed would not eliminate real variance due to emotional intelligence. On the whole, however, such demands go beyond the already conservative criteria (equivalent-length/equivalent-cost) set out by Sechrest (1963). We would not want to see unrealistic criteria for statistical controls eliminate otherwise adequate studies from consideration for publication in our journals.
Is the Variance Predicted by Emotional Intelligence Trivial?

Brody (this issue) argues that some of the variance predicted by our test can, "without hyperbole, be described as trivial." Our results may indeed be viewed as modest when compared to such claims as "EI accounts for over 85% of outstanding performance in top leaders" or that "compared to IQ and expertise [in job performance], emotional competence mattered twice as much" (Goleman, 1998, p. 31). As we and our colleagues have noted, there is little or no direct evidence to support such claims (Matthews, Zeidner, & Roberts, 2002; Mayer & Cobb, 2000; Mayer et al., 2000). Moreover, Goleman (2000, p. 22) has explicitly acknowledged that future research must determine such answers and has apparently dropped such claims.

For our own part, our initial work focused on describing and measuring the phenomenon; we did not discuss effect sizes at that time. In our first response to the popularization, we acted to reanchor expectations at a reasonable level, noting that "a 10% contribution of emotional intelligence [to life outcomes] would be considered very large indeed" (Mayer & Salovey, 1997, p. 17). Elsewhere, we further stated that the "best new variables typically increase predictions, for instance, of job performance between 1 and 4%" (Mayer et al., 2000, p. 412). Even 1%—and certainly 4% and 10% changes are far from trivial, given that they may influence a person's life course over extended periods of time. That such effects are important has been detailed extensively elsewhere (e.g., Funder, 2001, pp. 61–63, 81; Meyer, Finn, et al., 2001).

Looking at Table 7 in our target article, the average predictive validity from emotional intelligence to the eight deviant/destructive/drug behavior life space scales was $r = -.21$, accounting for 4% of the variance around the mean. The incremental validity over IQ and sex can be represented with an $r = -.11$, or 1%, for the six correlations. Focusing on the three deviant behavior relations alone, the correlations are $r = -.35$ and, incrementally, $-.18$, explaining 12% and 3%, respectively, of the variance around the mean. By comparison, the correlation between MMPI scores and prison misconduct is $r = .07$, or < 1% (Meyer et al., 2001, Table 2), and adult criminal history predicts future recidivism among mentally disordered offenders, $r = .18$, or 3% (Meyer et al., 2001, Table 1). In addition, these latter values are direct, rather than incremental, relations.

To be sure, there are much stronger test-to-criterion measures in psychology, including many involving cognitive IQ. But a recent APA task force on psychological testing concluded that "psychologists studying highly complex human behavior should be rather satisfied" with correlations in the $r = .10$ to .20 range, and they could be generally pleased with correlations (such as we also have obtained) in the .25 to .35 area (Meyer et al., 2001, p. 133). That said, we should note that as emotional intelligence’s relations to specific criteria are better understood, at least a few higher level predictions are likely to emerge. We have already seen a number of stronger relations at the $r = .40$ or higher level (Formica, 1998; Lopes et al., 2004; Rubin, 1999). It is at least possible that some of these may replicate at equally high levels in the future.

Do Validity Studies of Emotional Intelligence Depend Too Much on Self-Report Criterion?

Another point raised in the commentaries is that the criteria for emotional intelligence scales should move beyond the use of self-report inventories (Brody, this issue). Because we used life-report, sociometric, and observer-data, and very little self-report data as criteria, we are guessing that Brody, in making this remark, must be including life-report data in the self-report category.

The terms self-report and objective tests date back at least to 1915 (Bell, 1912; Calkins, 1915). Since that time, the term self-report has been defined in so many ways that, by the beginning of the twenty-first century, it had little meaning beyond "any utterance or other sign made by the individual" (e.g., see definitions in Bordens & Abbott, 2002, p. 135; Heiman, 2002, p. 284; Shaugnessy, Zechmeister, & Zechmeister, 2003, p. 150). A little reflection makes clear that most tests, and many criteria also, would be considered self-report under these conditions, including intelligence tests, projective tests, surveys, behavioral records filled out by the individual, census data, and so on.

Contemporary classifications of data have been developed over the past few years that better accommodate the rapid proliferation of new data types (Funder, 1996, 2001; Mayer, 2004, in press). In these new classifications, life-report data is not the same as self-judgment data. Perhaps more to the point, life-report analogs to the self-judgments of, say, the Big Five personality traits yield correlations between $r = .37$ to .55—well below one’s expectations were the data types to be similar (Brackett, 2004; Mael & Hirsch, 1993). Life-report data are also less subject to demand and distortion characteristics than self-judgment data and show incremental validity over them (Mael, 1991; Mount, Witt, & Barrick, 2000; Reilly & Chao, 1982). In short, we are not entirely sure what Brody (this issue) means when he says we over-relied on self-report data in our criterion measures. We can assert, however, that we used many forms of data as criteria, including life-report, observer-report, sociometric data, institutional-record data concerning job performance, and a relatively small percentage of self-judgment data. Most recently, we examined emotional intelligence and organization-reported salary data and obtained a significant relation (Lopes et al., 2004).
Toward the Future of Emotional Intelligence: Theorizing and Testing

What is the Point of Attempting to Establish Emotional Intelligence as a Standard Intelligence?

We have published several papers intending to establish emotional intelligence as a standard intelligence, and, indeed, our belief is that we have made progress toward this goal. Gohm (this issue) suggests our ambitions and aims have been viewed as provocative by others in the intelligence community. Our attempt to establish emotional intelligence is not meant to compete with or diminish the value of general intelligence, but rather is offered in the spirit of strengthening the concept of intelligence. It can be viewed as a response to the APA task force’s conclusion that “we know much less about the forms of intelligence that tests do not easily assess: wisdom, creativity, practical knowledge, social skill, and the like.” (Neisser, 1996, p. 95). We believe work such as we are conducting strengthens the understanding of intelligence.

Where Should Research Proceed Next?

Gohm’s (this issue) commentary outlines a number of areas for validation studies that would be desirable to conduct in the future. For example, she asks for evidence that tasks included in the MSCEIT predict conceptually related but distinct tasks not included on the MSCEIT. To this we would only add that some relevant evidence already was supplied in relation to our earlier scale, the MEIS, specifically in regard to Perceiving Emotion and the ability to recognize emotions in music, videos, and related stimuli (Davies et al., 1998; Geher, Warner, & Brown, 2001; Mayer, Caruso, & Salovey, 1999).

Gohm (this issue) recommends the utility of examining the distinctive behaviors of high and low MSCEIT scorers in laboratory and other settings in which those behavioral differences can be observed and, presumably, evaluated by skilled raters. She raises the importance of studies of EI as a hierarchy of skills, and of the developmental nature of EI, and of EI as a variable in clinical studies. She suggests examining whether people high on the MSCEIT really can, relative to others, better resist criticizing a friend in the face of temptations to do so. She calls for increased cross-cultural studies. Her recommendations provide an outline of issues that can help guide future work in the field.

Continuing the Dialogue

Constructive dialogue requires all parties to engage in open and honest debate. We believe that debates concerning emotional intelligence have become much more sophisticated during the past decade. Continuing the dialogue will be fostered as researchers more clearly distinguish real scientific claims from hyperbole, whether in scientific articles or in the popular press. The number of people involved in this discussion and the number of research studies now available in the area encourage us in the belief that studies of emotional intelligence, measured as an ability, and of intelligence testing more generally, both have a great deal to offer the enterprise of understanding human performance and how to best foster it.

Note

John D. Mayer, Department of Psychology, 10 Library Way, University of New Hampshire, Durham, NH 03824. E-mail: jack.mayer@unh.edu

References


