A review and critique of emotional intelligence measures

JEFFREY M. CONTE*

Department of Psychology, San Diego State University, San Diego, California, U.S.A.

Summary

Emotional intelligence measures vary widely in both their content and in their method of assessment. In particular, emotional intelligence measures tend to use either a self-report personality-based approach, an informant approach, or an ability-based assessment procedure. In this paper, the measurement and psychometric properties of four of the major emotional intelligence measures (Emotional Competence Inventory, Emotional Quotient Inventory, Multifactor Emotional Intelligence Scale, Mayer–Salovey–Caruso Emotional Intelligence Test) are reviewed, the comparability of these measures is examined, and some conclusions and suggestions for future research on emotional intelligence measures are provided.

Copyright © 2005 John Wiley & Sons, Ltd.

Introduction

Interest in emotional intelligence (EI) has increased greatly over the last decade. Although some researchers and practitioners have been quite optimistic about the importance of EI in organizations, critical questions remain about the concept, theory, and measurement of EI (Landy & Conte, 2004; Matthews, Zeidner, & Roberts, 2002). In separate papers in this issue, Landy and Locke consider historical, scientific, and conceptual concerns about EI. The present paper reviews and critiques EI measures, which vary widely in both their content and in their method of assessment. In this paper, the measurement and psychometric properties of four of the major EI measures (Emotional Competence Inventory, Emotional Quotient Inventory, Multifactor Emotional Intelligence Scale, Mayer–Salovey–Caruso Emotional Intelligence Test V.2) will be considered, the comparability of these measures will be examined, and some conclusions and suggestions for future research on EI measures will be provided. Additional summaries and reviews of EI measures can be found in Gowing (2001) and Matthews et al. (2002).

* Correspondence to: Jeffrey M. Conte, Department of Psychology, San Diego State University, 5500 Campanile Drive, San Diego, CA 92182-4611, U.S.A. E-mail: jconte@sunstroke.sdsu.edu

Accepted 5 January 2005

Copyright © 2005 John Wiley & Sons, Ltd.
Emotional Competence Inventory (ECI)

Developed by Boyatzis, Goleman, and colleagues, the ECI is designed to assess emotional competencies and positive social behaviors (Boyatzis, Goleman, & Rhee, 2000; Goleman, 1995; Sala, 2002). The ECI has 110 items and assesses 20 competencies that are organized into four clusters: (1) Self-Awareness, (2) Social Awareness, (3) Self-Management, and (4) Social Skills. The ECI includes 360-degree assessment techniques that can include self-ratings, peer ratings, and supervisor ratings.

The internal consistency reliability of the self-assessment ECI scales ranges from 0.61 to 0.85. For the peer and supervisor rating scales, internal consistency reliability ranges from 0.80 to 0.95 (Gowing, 2001; Sala, 2002). The developers of the ECI suggest that it is supported by validity evidence from the Self-Assessment Questionnaire (SAQ), which is a predecessor of the ECI. However, for proprietary reasons, the developers of the ECI have allowed very few items to be evaluated by other researchers. Thus, few independent, peer-reviewed assessments of the reliability and validity of the ECI have been undertaken and published. Without independent replication, these reported findings on the ECI are tentative at best.

Researchers who have examined the content of the ECI competencies have concluded that they overlap with four of the Big Five personality dimensions (Conscientiousness, Emotional Stability, Extraversion, and Openness) and other psychological concepts in the motivation and leadership literatures (Matthews et al., 2002; Van Rooy & Viswesvaran, 2004). Overall, discriminant and predictive validity evidence for the ECI has not been provided, and the scale does not deserve serious consideration until peer-reviewed empirical studies using this measure are conducted.

Bar-On Emotional Quotient Inventory (EQ-i)

The EQ-i is a 133-item self-report measure that takes approximately 30 minutes to complete (Bar-On, 2000). The measure yields an overall EQ score as well as scores for five composite scales: (1) intrapersonal, (2) interpersonal, (3) adaptability, (4) general mood, and (5) stress management. However, it is not clear how each of these composites is related conceptually to EI. Matthews et al. (2002) noted that the theory behind this measure is vague, boiling Bar-On’s theoretical approach down to ‘EI is what emotional quotient tests test’ (p. 206).

Bar-On (2000) reported that the internal consistency reliability of the overall EQ-i was 0.76. The EQ-i has shown adequate test–retest reliability of 0.85 after 1 month and 0.75 after 4 months (Bar-On, 1997). In terms of convergent validity, Gowing (2001) reported that the average correlation among EQ-i subscales was 0.50, and she noted that this average correlation is similar to correlations among various components of traditional intelligence tests. The correlation between the EQ-i and the MEIS was 0.36 in a study reported by Mayer, Caruso, and Salovey (2000). With respect to discriminant validity, the EQ-i correlated 0.12 with the Wechsler Adult Intelligence Scale (Bar-On, 2000), and the average correlation between the EQ-i and the Big Five personality measures was approximately 0.50 (Dawda & Hart, 2000). The EQ-i correlated −0.77 with the anxiety scale from Cattell’s 16PF test (Newsome et al., 2000), indicating that this EI measure overlaps strongly with a well-established measure of trait anxiety.

In terms of criterion validity, the EQ-i was significantly correlated with morale (0.55), stress (−0.41), general health (−0.50), and supervisor ratings of performance (0.22) in a study of retail
managers by Slaski and Cartwright (2002). Another important criterion that might be predicted by EI is academic success, which is commonly assessed via student grade point average (GPA). Although it might be argued that GPA is mainly based on cognitive (non-emotional) tasks and thus should not necessarily be related to EI, Bar-On (1997) proposed that the EQ-i measures non-cognitive aspects of personal functioning, such as a student’s ability to cope with environmental pressures and demands. Based on unpublished studies cited in the EQ-i Technical Manual, Bar-On (1997) concluded that EI is an important predictor of academic success. In addition, Goleman (1995) proposed that EI could predict success both at work and in school as well as or better than traditional intelligence measures. However, in a sample of 160 Canadian college students, the EQ-i total score had a correlation of 0.01 with grade point average (Newsome, Day, & Catano, 2000). Similarly, none of the five composite EQ-i scores was significantly associated with GPA. In contrast, cognitive ability (i.e., the Wonderlic Personnel Test) and some personality dimensions (e.g., self-control) were significant predictors of GPA. Based on their results, Newsome et al. (2000) concluded that there is inadequate data at this time to justify use of the EQ-i as a selection device. In sum, although the EQ-i demonstrates adequate reliability and some validity evidence, it is lacking in discriminant validity evidence, and few studies have examined whether it provides incremental predictive validity above the contribution of established predictors such as cognitive ability and Big Five personality dimensions.

MEIS and MSCEIT V.2

According to Mayer et al. (2000), emotional intelligence involves the capacity or ability to reason with and about emotions. They have developed two different EI tests, both of which were developed in an intelligence testing tradition. First, they developed the Multifactor Emotional Intelligence Scale (MEIS), which had some subscales with low reliability and some problems with scoring procedures. Second, they developed the Mayer–Salovey–Caruso Emotional Intelligence Test (MSCEIT), which is an update of the MEIS. The most recent version of the MSCEIT is Version 2 (V.2). Both EI tests are discussed below because, although the MSCEIT V.2 appears to have improved some of the problems of the earlier test, the MSCEIT V.2 is new, and few studies have been published using it.

The MEIS includes 402 items and produces four subscales: Perception, Assimilation, Understanding, and Managing Emotions (Mayer, Caruso, & Salovey, 2000). The MEIS is an ability test and, as such, the test developers have tried different approaches to identify the correct answers, including target scoring, consensus scoring, and expert scoring. Target scoring involves determining the correct answer by asking the person (i.e., the target) whose facial expressions are depicted in an item how he or she actually felt or what he or she was portraying when engaged in some emotional activity. Consensus scoring involves determining the correct answer by pooling the judgments of hundreds of people. This scoring technique assesses the extent to which the test taker’s choice matches majority opinion. Thus, consensus scoring techniques are ‘in direct contrast to traditional measures of intelligence where an objective measure of truth is considered’ (Matthews et al., 2002, p. 186). Expert scoring involves determining the correct answer by pooling the judgments of experts in emotions. This type of scoring technique is most similar to that used in cognitive ability tests. Overall, determining the correct type of scoring to use for ability-based EI tests is critical and, as is discussed below, controversial.

Mayer et al. (2000) reported that the internal consistency reliability of the overall MEIS was 0.95. For consensus scored scales the average internal consistency reliability was 0.77 across the four branch scores, and for expert scored scales the average internal consistency reliability was 0.62 (Caruso,
Mayer, & Salovey, 2002; Matthews et al., 2002). The test–retest reliability of the overall MEIS over a 2-week period was 0.75. The test–retest reliability of the MEIS branch scores ranged from 0.60 to 0.68. In contrast, reliability coefficients for cognitive ability tests typically range from 0.85 to 0.95 (Kaplan & Saccuzzo, 2001; Murphy & Davidshofer, 2001).

In terms of convergent validity evidence, Mayer et al. (2000) reported that the MEIS had a correlation of 0.36 with the EQ-i, indicating that the tests share 13 percent of their variance. For discriminant validity, correlations between the MEIS (consensus scores) and the Big Five personality dimensions ranged from 0.13 for Openness and Extraversion to 0.24 for Agreeableness (Roberts et al., 2001). Data from several studies indicate that the MEIS correlates between 0.30 and 0.40 with traditional measures of cognitive ability (Roberts et al., 2001; Van Rooy & Viswesvaran, 2004). Research conducted over 75 years ago found correlations of this magnitude between constructs such as social intelligence and verbal intelligence (Hunt, 1928). Landy’s paper in this issue covers the early history of social intelligence research in more detail.

The Mayer–Salovey–Caruso Emotional Intelligence Test (MSCEIT) V.2 is also designed to measure the four branches of Mayer and Salovey’s (1993, 1997) emotional intelligence ability model. The MSCEIT V.2 provides a total EI score and four Branch scores: (1) perception of emotion, (2) integration and assimilation of emotion, (3) knowledge about emotions, and (4) management of emotions. With 141 items, the MSCEIT V.2 is shorter and quicker to administer than the MEIS, and it provides both consensus and expert scores for all Branch scores. Whereas the MEIS has 12 subtests to assess the four Branches, the MSCEIT V.2 includes two subtests for each Branch (Salovey, Mayer, Caruso, & Lopes, 2003). In a recent study by Mayer, Salovey, Caruso, and Sitarenios (2003), reliabilities at the total scale and Branch levels were all above 0.75. For all scales in the MSCEIT V.2, the average internal consistency reliability was 0.68 for consensus scoring and 0.71 for expert scoring. Given that this is an ability measure, the reliabilities of the subscales seem far from optimal (Matthews et al., 2002). In terms of validity, the authors rely primarily on evidence from the MEIS to support the MSCEIT V.2. Nevertheless, given how different the MEIS and MSCEIT V.2 are, researchers and practitioners should be cautious about making inferences about the MSCEIT V.2 based on data from the MEIS.

Several researchers have expressed concerns about the absence of scientific standards for determining the accuracy of consensus and expert scores for the MEIS and the MSCEIT V.2. In addition, given that consensus scoring uses the most common response in determining correct answers to test items, these ability-based tests may not provide meaningful scores at the high end of the EI continuum when consensus scoring is used (Matthews et al., 2002). In the expert scoring approach, Matthews et al. (2002) also raised questions about how ‘experts’ were chosen when determining the correct answers for emotional intelligence questions and tasks.

To my knowledge, correlations between the MEIS and the MSCEIT V.2 have not been examined. Given the notable differences between the two measures, research conducted on the MEIS cannot be used to support the validity of the MSCEIT V.2. Because the MSCEIT V.2 is too new to have been included in most EI research or in the meta-analysis by Van Rooy and Viswesvaran (2004), much of that research has examined the MEIS. Now that the MEIS has been replaced by the MSCEIT V.2, research conducted using the MEIS needs to be reconsidered with the new measure. It is likely that the MSCEIT V.2 will demonstrate discriminant validity from personality measures, but not incremental validity in predicting performance outcomes. In fact, a recent study by Barchard (2003) found that none of the many EI measures she examined (including the MSCEIT) showed incremental validity for predicting academic success over and above cognitive ability and personality. Although Brackett and Mayer (2003) found that the MSCEIT and EQ-i showed some evidence of incremental validity in predicting social deviance and alcohol use, respectively, their results indicated that neither measure provided incremental validity in predicting academic performance.
Comparability of EI Measures

The developers of EI measures have used different definitions of the EI construct, which has resulted in different types and numbers of dimensions for the various measures (Gowing, 2001). Perhaps of more importance, the measures use different response formats, including self-report, ability, and informant approaches. The self-report EI measures (e.g., ECI and EQ-i) sample a broad range of individual differences, but nearly all of the self-report scales that have satisfactory reliabilities relate to or load on well-established personality dimensions (Daus & Ashkanasy, 2003; Davies, Stankov, & Roberts, 1998). Ability-based EI measures (i.e., MEIS, MSCEIT V.2), which are more distinct from the Big Five personality dimensions, have higher correlations with general mental ability (GMA) than do self-report EI measures (Van Rooy & Viswesvaran, 2004), leaving less room for ability-based EI measures to provide incremental prediction of work criteria such as job performance and leader emergence. Given ability-based EI measures’ overlap with GMA and their lack of incremental validity evidence in predicting work criteria, it is possible that ability-based EI measures will ultimately be regarded in the same way as early measures of social intelligence. Specifically, after conducting several studies using social intelligence measures, R. L. Thorndike concluded that early measures of social intelligence were simply poor tests of GMA, which he called ‘abstract intelligence’ (Thorndike, 1936; Thorndike & Stein, 1937).

Few studies have examined both trait and ability-based EI measures to examine the extent of the overlap. Mayer et al. (2000) found that the MSCEIT and Bar-On scales correlated 0.36, indicating that they share approximately 13 percent of their variance. In a more recent study, Brackett and Mayer (2003) found that the MSCEIT and Bar-On scales correlated 0.21, indicating that they share approximately 4 percent of their variance. The low relationship between different EI measures raises serious questions about whether they are all actually measuring the same construct (Matthews et al., 2002). Compared to ability-based EI measures, self-report measures are likely to receive less attention in the coming years given that they lack psychometric support (particularly discriminant validity from the Big Five personality dimensions). Alternatively, ability-based EI measures are likely to receive continued attention, and it is essential that additional assessments of the convergent validity across EI measures are conducted.

EI Measures: Conclusions and Suggestions for Future Research

In general, EI measures have demonstrated adequate internal consistency reliability. Self-report EI measures have acceptable internal consistency as do the overall scales for ability-based measures, but these data by themselves give no indication about whether EI measures are simply assessing constructs already measured by other, more established constructs (e.g., the Big Five personality dimensions). Further, some of the subscales for the ability-based EI measures have marginally acceptable internal consistency and test–retest reliability.

Validity evidence for EI measures has lagged behind reliability evidence. Content validity evidence for EI measures is lacking because of vague theoretical development for many of the measures and because the content across EI measures varies widely. Because few EI researchers are willing to be specific about what they want to measure, it is difficult to examine content validity. Similarly, construct validity evidence in the form of convergent and discriminant validity is lacking. First, EI measures have failed to converge on a common construct. Second, self-report EI measures appear to assess existing personality characteristics or perhaps emotional competencies, but they do not appear to assess intelligence (Mayer, Caruso, & Salovey, 1999).
In their meta-analysis, Van Rooy and Viswesvaran (2004) found that EI and Big Five personality dimensions had correlations (corrected for unreliability) that ranged from 0.23 to 0.34 (these analyses included both ability and self-report EI measures). Thus, EI and the Big Five personality dimensions are more highly correlated than many EI researchers have proposed, suggesting that these EI measures are lacking in discriminant validity. Emotional intelligence measures can also be examined in terms of criterion-related and incremental validity. Van Rooy and Viswesvaran (2004) found that the percentage of variance in performance explained by EI was 5 percent, which is much lower than the claims of some EI proponents (e.g., Goleman, 1995), who have argued that it is more important than general mental ability (GMA). To provide perspective, estimates of the percentage of variance in job performance that GMA accounts for range from 10 percent to 26 percent (Hunter & Hunter, 1984; Schmidt & Hunter, 1998).

Further, GMA typically provides incremental validity in predicting work outcomes above other measures, whereas EI has shown little or no incremental validity above GMA in predicting performance outcomes. In their meta-analysis, Van Rooy and Viswesvaran (2004) found that emotional intelligence provided incremental validity (ranging from 0.06 for Conscientiousness to 0.29 for Openness to Experience) above the Big Five personality dimensions in predicting performance. However, EI measures provided minimal (0.02) incremental validity above general mental ability (GMA) in predicting performance. Alternatively, GMA provided substantial incremental validity (0.31) above EI (Van Rooy & Viswesvaran, 2004). It should be noted that ability-based and self-report EI measures were not separated in the incremental validity analyses in this meta-analysis; thus, it is unclear how the different types of EI measures would fare in separate analyses. Nevertheless, based on empirical research to date, broad claims that EI is a more important predictor than GMA (e.g., Goleman, 1995, 1998) are unfounded and unsubstantiated.

Current EI measures typically use a four- or five-factor model, but research is needed on which dimensions are most predictive of work and non-work outcomes (Van Rooy & Viswesvaran, 2004). In addition, research is needed on the potential for faking on self-report EI measures. Mayer et al. (2003) found that women often scored higher than men on the MSCEIT V.2, but that there were no significant differences related to ethnicity. Accordingly, for the ability-based EI measures, further examination of potential adverse impact against protected groups is needed. Emotional intelligence is likely to be culturally bound; thus, an investigation of cross-cultural similarities and differences in EI is also needed. Finally, because many applications of the EI concept involve attempts to develop EI or emotional competencies (i.e., making people more socially intelligent), further investigation of the stability, and alternatively, the ‘trainability’ of EI, is needed (Slaski & Cartwright, 2003). Of course, questions about whether training can increase EI can be answered only with the use of valid EI measures and rigorous research designs (Goldstein & Ford, 2002).

In sum, serious concerns remain for all of the EI measures, ranging from scoring concerns for ability-based EI measures to discriminant validity concerns for self-report EI measures. Although ability-based EI measures appear to be most promising, many unresolved issues remain even with them. Gowing (2001) notes that many EI measures have been used for development, but that the trend is toward using them for selection as well. Managers and other organizational decision-makers should be wary of making this leap unless more rigorous discriminant, predictive, and incremental validity evidence for EI measures is shown. Although Mayer et al. (2003) have developed the most promising of the EI measures, even they state that ‘the applied use of EI tests must proceed with great caution’ (p. 104). After conducting this review of EI measures, I would have to agree. Nevertheless, I look forward to additional scientific investigations of the incremental validity of EI measures in predicting job performance and other work outcomes above the contribution of established predictors such as cognitive ability and Big Five personality dimensions.

Acknowledgements

The author would like to thank Nichole Best and Doug Chartier for assistance with the literature review and Frank Landy, Michelle Dean, and Paul Spector for valuable comments and suggestions on earlier drafts of this paper.

Author biography

Jeffrey M. Conte is an Associate Professor in the Department of Psychology at San Diego State University. He received his PhD in Industrial and Organizational Psychology from Penn State University. His research interests include personality predictors of job performance, the measurement of emotional intelligence, and the factors associated with health and stress in the workplace. He is co-author with Frank Landy of the recently published book, Work in the 21st Century: An Introduction to Industrial and Organizational Psychology (McGraw-Hill, 2004). His research has appeared in journals such as the Journal of Applied Psychology, Academy of Management Review, Journal of Business and Psychology, Journal of Managerial Psychology, and Human Performance.

References


