Emotion
2003, Vol. 3, No. 1, 97–105

1528-3542/03/$12.00 DOI: 10.1037/1528-3542.3.1.97

BRIEF REPORT

Measuring Emotional Intelligence With the MSCEIT V2.0

John D. Mayer
University of New Hampshire

Peter Salovey
Yale University

David R. Caruso
Work-Life Strategies

Gill Sitarenios
Multi-Health Systems, Inc.

Does a recently introduced ability scale adequately measure emotional intelligence (EI) skills? Using the Mayer–Salovey–Caruso Emotional Intelligence Test (MSCEIT; J. D. Mayer, P. Salovey, & D. R. Caruso, 2002b), the authors examined (a) whether members of a general standardization sample and emotions experts identified the same test answers as correct, (b) the test's reliability, and (c) the possible factor structures of EI. Twenty-one emotions experts endorsed many of the same answers, as did 2,112 members of the standardization sample, and exhibited superior agreement, particularly when research provides clearer answers to test questions (e.g., emotional perception in faces). The MSCEIT achieved reasonable reliability, and confirmatory factor analysis supported theoretical models of EI. These findings help clarify issues raised in earlier articles published in *Emotion*.

The past 12 years have seen a growing interest in *emotional intelligence* (EI), defined as a set of skills concerned with the processing of emotion-relevant information and measured with ability-based scales. A new ability test of EI, the Mayer–Salovey–Caruso Emotional Intelligence Test, Version 2.0 (MSCEIT, V2.0; Mayer, Salovey, & Caruso, 2002a), potentially improves upon earlier measures and can inform the debate over the scoring, reliability, and factor structure of such scales (Mayer, Salovey, & Caruso, 2002b). The MSCEIT is intended to measure four branches, or skill groups, of EI: (a) perceiving emotion accurately, (b) using emotion to facilitate thought, (c) understanding emotion, and (d) managing emotion (Mayer & Salovey, 1997).

The MSCEIT is the most recent of a series of ability scales of EI. Its immediate predecessor was the MSCEIT Research Version 1.1 (MSCEIT RV1.1), and before that, the Multi-Factor Emotional Intelligence Scale (MEIS; Mayer, Caruso, & Salovey, 1999; Mayer et al., 2002b). Those tests, in turn, evolved out of earlier scales measuring related constructs such as emotional creativity, social intelligence, and nonverbal perception (e.g., Averill & Nunley, 1992; Kaufman & Kaufman, 2001; O'Sullivan & Guilford, 1976; Rosenthal, Hall, DiMatteo, Rogers, & Archer, 1979). The MSCEIT and its predecessors are based on the idea that EI involves problem solving with and about emotions. Such ability tests measure something rela-

John D. Mayer, Department of Psychology, University of New Hampshire; Peter Salovey, Department of Psychology, Yale University; David R. Caruso, Work-Life Strategies, New Canaan, Connecticut; Gill Sitarenios, Research and Development, Multi-Health Systems, Inc., Toronto, Ontario, Canada.

The Mayer–Salovey–Caruso Emotional Intelligence Test, Version 2.0 (MSCEIT V2.0) is available from Multi-Health Systems (MHS) of Toronto, Ontario, Canada in booklet and Web-based formats. MHS scores the test on the basis of the standardization sample and expert criteria; researchers have the further option of developing their own independent norms. Researchers can obtain the MSCEIT through special arrangements with MHS, which has various programs to accommodate their needs.

We gratefully acknowledge the assistance of Rebecca Warner and James D. A. Parker, who served as expert consultants concerning the structural equation models reported in this article. In addition, Terry Shepard was instrumental in preparing early versions of the tables.

Correspondence concerning this article should be addressed to John D. Mayer, Department of Psychology, Conant Hall, 10 Library Way, University of New Hampshire, Durham, New Hampshire 03824. E-mail: jack.mayer@unh.edu

tively different from, say, self-report scales of EI, with which correlations are rather low (Mayer et al., 2002b). The ability to solve emotional problems is a useful ingredient to behaving in an emotionally adaptive way, although not sufficient by itself.

The MSCEIT V2.0 standardization process required collecting data sets relevant to the several issues concerning EI. For example, in a recent exchange of articles in *Emotion,* Roberts, Zeidner, and Matthews (2001) raised concerns about the earlier developed MEIS test of EI. These concerns included whether there is one set of correct answers for an EI test, or whether expert and general (e.g., lay) opinions about answers diverge too much, whether such tests could be reliable, and whether the factor structure of such tests was fully understood and consistent with theory. In that same issue of *Emotion,* we, and others, responded (Izard, 2001; Kaufman & Kaufman, 2001; Mayer, Salovey, Caruso, & Sitarenios, 2001; Schaie, 2001; Zeidner, Matthews, & Roberts, 2001). Findings from the MSCEIT standardization data reported here have the promise of more directly informing the debate through empirical findings. The analyses we present address the following three questions: (a) Do general and expert criteria for correct answers to EI test items converge? (b) What is the reliability of such tests? and (c) Is the factor structure of such tests consistent with theoretical models of EI?

Three Issues About EI Addressed in the Present Study

The Criteria for Correct Answers

One must know how to score a test's items before one can settle such issues as the test's reliability and factor structure. Our model of EI hypothesizes that emotional knowledge is embedded within a general, evolved, social context of communication and interaction (Mayer et al., 2001). Consequently, correct test answers often can be identified according to the consensus response of a group of unselected test-takers. For example, if the respondents identify a face as predominantly angry, then that can be scored as a correct answer. We have further hypothesized that emotions experts will identify correct answers with greater reliability than average, particularly when research provides relatively good methods for identifying correct alternatives, as in the case of facial expressions of emotion, and the meaning of emotion terms (Mayer et al., 2001).

If the general and the expert consensus diverge too far as to the correct answers on a test, a complication arises, because the two methods yield potentially different scores for each person. Mayer et al. (1999) used an expert criterion based on two experts and a general consensus method to score the earlier MEIS measure and found those methods only partially converged. Roberts et al. (2001) replicated the finding and raised the lack of convergence as a serious problem. We believed that some aggregation of experts beyond two will provide reliable identification of answers that more closely converges to the general consensus (Legree, 1995). Twenty-one emotions experts were used in the present study.

Issues of Reliability

The MSCEIT V2.0 must exhibit adequate levels of reliability, as did the MEIS, MSCEIT RV1.0, and comparable psychological tests (see Mayer et al., 2001, pp. 239–240). As with its predecessor tests, the four MSCEIT V2.0 branch scores (e.g., Perception, Facilitating, Understanding, Management) draw on different tasks that include different item forms; that is, the items are nonhomogeneous. Under such conditions, split-half reliability coefficients are the statistic of choice (relative to coefficient alphas), as they involve the orderly allocation of different item types to the two different halves of the test (Nunnally, 1978). The test–retest reliability of the total MSCEIT score has been reported elsewhere as $r(60) = .86$ (Brackett & Mayer, 2001).

Issues of Factor Structure

The factor structure of a test indicates how many entities it plausibly measures. In this specific case, it indicates how many dimensions of EI the test is "picking up"—one unified dimension, many related dimensions, or something else. We believe that the domain of EI is well-described by one-, two-, and four-oblique (correlated) factor models, as well as other equivalent models. If the MSCEIT V2.0 shows a similar structure to the MEIS for both expert and general scoring, it would strengthen the argument that the theory of EI we use works across tests. Using the standardization sample, we performed confirmatory factor analyses of the full scale MSCEIT V2.0, testing one-, two-, and four-factor models to examine the range of permissible factor structures for representing the EI domain.

Method

Participants

General sample. The present sample consisted of 2,112 adult respondents, age 18 or older, who completed the MSCEIT V2.0 in booklet or on-line forms before May 2001. The sample was composed of individuals tested by independent investigators in 36 separate academic settings from several countries. The investigators had requested prerelease versions of the MSCEIT booklet or on-line forms depending on Internet availability and other criteria, and had submitted documentation of their research qualifications and a detailed plan of their intended research. Only basic demographic data were collected across samples due to the diverse nature of the research sites.

Of those reporting gender, 1,217 (58.6%) were women and 859 (41.4%) were men. The mean age of the sample was 26.25 years (SD = 10.51), with roughly half the sample being college-aged (52.9%) and the rest ranging upward to 69 years old. The participants were educationally diverse, with 0.6% reporting not completing high school, 10.3% having completed only high school, 39.2% having completed some college or university courses, 33.7% having completed college, and 16.1% holding master's level or higher degrees. The group was ethnically diverse as well, with 34.0% Asian, 3.4% Black, 2.0% Hispanic, 57.9% White, and 2.3% other or mixed ethnicity. Most participants were from the United States (1,240), with others from South Africa (231), India (194), the Philippines (170), the United Kingdom (115), Scotland (122), and Canada (37); all testing was conducted in English.

Expert sample. The expert sample was drawn from volunteer members of the International Society for Research on Emotions (ISRE) at its 2000 meeting. The Society was founded in 1984, with the purpose of fostering interdisciplinary scientific study of emotion. Membership is open to researchers and scholars who can demonstrate a serious commitment to the investigation of the emotions. Twenty-one experts, 10 male and 11 female, from eight Western countries, participated. The sample of experts had a mean age of 39.38 (SD = 6.44; range = 30–52); no data about their ethnicity were collected.

The MSCEIT V2.0

The MSCEIT V2.0 is a newly developed, 141-item scale designed to measure the following four branches (specific skills) of EI: (a) perceiving emotions, (b) using emotions to facilitate thought, (c) understanding emotions, and (d) managing emotions. Each of the four branches is measured with two tasks. Perceiving emotions is measured with the faces and pictures tasks; facilitating thought is measured with the sensations and facilitation tasks; understanding emotions is measured with blends and changes tasks; and managing emotions is measured with emotion management and emotional relationships tasks.

Each of the eight MSCEIT tasks is made up of a number of item parcels or individual items. A parcel structure occurs, for example, when a participant is shown a face (in the faces task) and asked about different emotions in the face in five subsequent items. The five items make up an item parcel because they are related to the same face, albeit each asks about a different emotion (Cattell & Burdsal, 1975). Other items involve one response per stimulus, and are, in that sense, free-standing. Response formats were intentionally varied across tasks so as to ensure that results generalized across response methods, and to reduce correlated measurement error. Thus, some tasks, such as the pictures task, used 5-point rating scales, whereas other tasks, such as the blends task, used a multiple-choice response format.

Briefly, in the faces task (four item parcels; five responses each), participants view a series of faces and for each, respond on a five-point scale, indicating the degree to which a specific emotion is present in a face. The pictures task (six parcels; five responses each) is the same as the faces task, except that landscapes and abstract designs form the target stimuli and the response scale consists of cartoon faces (rather than words) of specific emotions. In the sensations task (five parcels; three responses each), respondents generate an emotion and match sensations to them. For example, they may generate a feeling of envy and decide how hot or cold it is. In the facilitation task (five item parcels; three responses each), respondents judge the moods that best accompany or assist specific cognitive tasks and behaviors, for example, whether joy may assist planning a party. In the blends task (12 free-standing items), respondents identify emotions that could be combined to form other emotions. They may conclude, for example, that malice is a combination of envy and aggression. In the changes task (20 free-standing items), respondents select an emotion that results from the intensification of another feeling. For example, they may identify depression as the most likely consequence of intensified sadness and fatigue. Respondents in the emotion management task (five parcels; four responses each) judge the actions that are most effective in obtaining the specified emotional outcome for an individual in a story. They are asked to decide, for example, what a character may do to reduce his or her anger, or prolong his or her joy. Finally, in the emotional relationships task (three item parcels; three responses each), respondents judge the actions that are most effective for one person to use in the management of another person's feelings. Interested readers may review the actual test, and its manual, for more specific task information (Mayer, Salovey, & Caruso, 2002a; Mayer et al., 2002b, p. 20).

General and expert consensus scoring. The MSCEIT yields a total score, two area scores (Experiential and Strategic), four branch scores corresponding to the four-branch model, and eight task scores. Each score can be calculated according to a general consensus method. In that method, each one of a respondent's answers is scored against the proportion of the sample that endorsed the same MSCEIT answer. For example, if a respondent indicated that surprise was "definitely present" in a face, and the same alternative was chosen by 45% of the sample, the individual's score would be incremented by the proportion, .45. The respondent's total raw score is the sum of those proportions across the 141 items of the test. The other way to score the test is according to an expert scoring method. That method is the same as the former, except that each of the respondent's scores is evaluated against the criterion formed by proportional responding of an expert group (in this case, the 21 ISRE members). One of the purposes of this study was to compare the convergence of these two methods.

Procedure. The MSCEIT administration varied, depending on the research site at which the data were collected (see the *General sample* section). The MSCEIT was given to participants to complete in large or small groups, or individually. Of the 2,112 participants, 1,368 took the test in a written form and 744 took the test in an on-line form that presented the exact same questions and response scales, by accessing a Web page. Those taking the pencil-and-paper version completed scannable answer sheets that were entered into a database. Web page answers were transmitted electronically. Prior research has suggested that booklet and on-line forms of tests are often indistinguishable (Buchanan & Smith, 1999).

Results

Comparison of test-booklet versus on-line administration groups. We first compared the MSCEIT V2.0 booklet and on-line tests. For each, there were 705 responses to the test (141 items; × five responses each). The correlation between response frequencies for each alternative across the two methods was $r(705) = .987$. By comparison, a random split of the booklet sample alone, for which one would predict there would be no differences, yields almost exactly the same correlation between samples of $r(705) = .998$. In each case, a scatterplot of the data indicated that points fell close to the regression line throughout the full range of the joint distribution and that the points were spread through the entire range (with more points between .00 and .50 than above .50). Even deleting 30 zero-response alternatives from the 705 responses lowered the correlation by only .001 (in the random split case). The booklet and on-line tests were, therefore, equivalent, and the samples were combined.

Comparison of general versus expert consensus scoring criteria. We next examined the differences between answers identified by the experts and by the general consensus. We correlated the frequencies of endorsements to the 705 responses (141 items; × five responses each) separately for the general consensus group and the expert consensus group and obtained a correlation of $r(705) = .908$ that, although quite high, was significantly lower than the correlation of .998 for the random split, $z(703) = 34.2$, $p < .01$. The absolute difference in proportional responding for each of the 705 alternatives also was calculated. The mean value of the average absolute difference between the expert and general groups was $|M_D(705)| = .080$ ($SD = .086$), which also was significantly greater than the difference between the booklet and on-line samples of $|M_D(705)| = .025$ ($SD = .027$), $z(705) = 16.3$, $p < .01$.

We hypothesized that emotions experts would be more likely than others to possess an accurate shared social representation of correct test answers; their expertise, in turn, could provide an important criterion for the test. If that were the case, then experts should exhibit higher interrater agreement than the general group. To assess interrater agreement, we divided the expert group randomly into two subgroups of 10 and of 11 experts each and computed the modal response for each of the 705 responses for the two subgroups of experts. The kappa representing agreement controlling for chance across the two expert subgroups for the five responses of the 141 items was $\kappa(110) = .84$. We then repeated this process for two groups of 21 individuals, randomly drawn from the standardization (general) samples and matched to the expert group exactly on gender and age. We used two control groups, rather than one, to enhance our confidence in the results; the education level of the comparison groups was comparable to that of the rest of the general sample. When we repeated our reliability analysis for the two matched control groups, we obtained somewhat lower kappas of $\kappa(110) = .71$ and .79.

The same superiority of the expert group exists at an individual level as well, where disaggregated agreement would be lower (Legree, 1995). The average interrater kappa coefficients of agreement across the five responses of the 141 items, for every pair of raters within the expert group, was $\kappa(210) = .43$, $SD = .10$, which significantly exceeded the average kappas of the two control groups $\kappa = .31$; $SD = .082$; $\kappa = .38$, $SD = .098$; zs(418) = 4.8, 1.85, $p < .05$–.01, one-tailed test.

Because general and expert groups both chose similar response alternatives as correct, and experts have higher interrater reliability in identifying such correct alternatives, members of the standardization sample should obtain somewhat higher test scores when the experts' criterion is used (before normative scaling corrections are applied). Moreover, the expert group should obtain the largest score advantages on skill branches where the experts most agree, owing to the experts' greater convergence for those responses. For example, one may expect increased expert convergence for Branches 1 (emotional perception) and 3 (emotional understanding) because emotions experts have long focused on the principles of coding emotional expressions (e.g., Ekman & Friesen, 1975; Scherer, Banse, & Wallbott, 2001), as well as on delineating emotional understanding (e.g., Ortony, Clore, & Collins, 1988). By contrast, research on Branches 2 (emotional facilitation of thought) and 4 (emotion management) is newer and has yielded less consensus, and so experts' responses may be more similar to the general sample in such domains.

To test this idea, we conducted a 4 (branch) × 2 (consensus vs. expert scoring criterion) analysis of variance (ANOVA) on MSCEIT scores. The main effect for scoring criterion was significant, $F(1, 1984) = 3464$, $p < .001$, indicating as hypothesized, that participants obtained higher raw scores overall when scored against the expert criteria. The main effect for branch was significant as well, $F(3, 5952) = 1418$, $p < .001$, indicating, unsurprisingly, that items on some branches were harder than others. Finally, there was a Branch × Scoring Criterion interaction, $F(3, 5952) = 2611$, $p < .001$.

Orthogonal contrasts indicated that participants who were scored according to the expert criterion on Branches 1 and 3 obtained significantly higher scores than when scored against the general consensus (see Table 1), $F(1, 1984) = 1631$ and 5968, respectively; ps < .001. Branch 2 (using emotions to facilitate thought) showed a significant difference favoring general consensus, $F(1, 1984) = 711$, ps < .001; and Branch 4 showed no difference, $F(1, 1984) = 1.57$, *ns*. The advantage for expert convergence on Branches 1 and 3 may reflect the greater institutionalization of emotion knowledge among experts in these two areas.

In a final comparison of the two scoring criteria, we scored participants' tests using the general criterion, on the one hand, and the expert criterion, on the other. The correlation between the two score sets ranged from $r(2004–2028) = .96$ to .98 across the Branches, Areas, and Total EI scores.

The evidence from this study reflects that experts are more reliable judges and converge on correct answers where research has established clear criteria for answers. If further studies bear out these results, the expert criteria may prove preferable to the general consensus.

Reliability of the MSCEIT V2.0. The MSCEIT V2.0 has two sets of reliabilities, depending on whether a general or expert scoring criterion is used. That is because reliability analyses are based on participants' scored responses at the item level, and scores at the item level vary depending on whether responses are compared against the general or expert criterion. The MSCEIT full-test split-half reliability is $r(1985) = .93$ for general and .91 for expert consensus scoring. The two Experiential and Strategic Area score reliabilities are $r(1998) = .90$ and .90, and $r(2003) = .88$ and .86 for general and expert scoring, respectively. The four branch scores of Perceiving, Facilitating, Understanding, and Managing range between $r(2004–2028) = .76$–.91 for both types of reliabilities (see Table 1). The individual task reliabilities ranged from a low of $\alpha(2004–2111) = .55$ to a high of .88. However scored, reliability at the total scale and area levels was excellent. Reliability at the branch level was very good, especially given the brevity of the test. Compared with the MEIS, reliabilities were higher overall at the task level (e.g., Mayer et al., 1999; Mayer et al., 2002b) but were sometimes lower than is desirable. We therefore recommend test interpretation at the total scale, area, and branch levels, with cautious interpretations at the task level, if at all.

Correlational and factorial structure of the MSCEIT V2.0. As seen in Table 2, all tasks were positively intercorrelated using both general (reported below the diagonal) and expert consensus scoring (above the diagonal). The intercorrelations among

Table 1
Unscaled Score Means and Standard Deviations, Reliabilities, and Intercorrelations for the MSCEIT V2.0 for General and Expert Scoring

			Descriptive statistic[a,b]				Reliability[b,c]	
			General		Expert			
Area score	Branch score	Subtest score	*M*	*SD*	*M*	*SD*	General	Expert
Total MSCEIT V2.0			.48	.07	.50	.08	.93	.91
Experiential			.49	.08	.50	.09	.90	.90
	Perceiving		.50	.10	.54	.13	.91	.90
		Faces	.50	.12	.57	.18	.80	.82
		Pictures	.50	.13	.50	.13	.88	.87
	Facilitating		.47	.09	.45	.08	.79	.76
		Facilitation	.44	.09	.41	.07	.64	.63
		Sensations	.50	.12	.50	.12	.65	.55
Strategic			.47	.08	.51	.10	.88	.86
	Understanding		.53	.10	.60	.13	.80	.77
		Changes	.56	.10	.63	.14	.70	.68
		Blends	.50	.12	.57	.16	.66	.62
	Managing		.42	.10	.42	.09	.83	.81
		Emotion manag.	.41	.09	.40	.09	.69	.64
		Emotional rel.	.43	.12	.43	.12	.67	.64

Note. MSCEIT V2.0 = Mayer–Salovey–Caruso Emotional Intelligence Test, Version 2.0; manag. = management; rel. = relationships.
[a] Final MSCEIT test scores for both general and expert scoring are converted to a standard IQ scale, where $M = 100$ and $SD = 15$.
[b] The N for the overall scale was 2,112; ns for the branch scores were Perceiving: 2,015, with task ns between 2,018 and 2,108; Facilitating: 2,028, with individual task ns between 2,034 and 2,103; Understanding: 2,015, with individual task ns between 2,016 and 2,111; Managing: 2,088, with individual task ns from 2,004 to 2,008.
[c] Split-half reliabilities are reported at the total test, area, and branch score levels due to item heterogeneity. Coefficient alpha reliabilities are reported at the subtest level due to item homogeneity.

tasks ranged from $r(1995–2111) = .17$ to .59, ps < .01, but with many correlations in the mid .30s.

Confirmatory factor analyses. A factor analysis of the MSCEIT V2.0 can cross-validate earlier studies that support one-, two-, and four-factor solutions of the EI domain (Ciarrochi, Chan, & Caputi, 2000; Mayer et al., 1999; Roberts et al., 2001). The one-factor, general intelligence (g) model, should load all eight MSCEIT tasks. The two-factor model divides the scale into an "Experiential" area (Perceiving and

Table 2
Intercorrelations for MSCEIT V2.0 General and Expert Scoring

Scale	1	2	3	4	5	6	7	8
Branch 1: Perceiving								
1. Faces	1.000	.356	.300	.315	.191	.157	.191	.179
2. Pictures	.347	1.000	.288	.400	.286	.263	.282	.271
Branch 2: Facilitating								
3. Facilitation	.340	.328	1.000	.313	.283	.242	.262	.262
4. Sensations	.336	.402	.352	1.000	.388	.374	.384	.415
Branch 3: Understanding								
5. Changes	.225	.282	.255	.382	1.000	.575	.437	.417
6. Blends	.171	.260	.224	.375	.589	1.000	.425	.424
Branch 4: Managing								
7. Emotion manag.	.232	.300	.299	.395	.417	.416	1.000	.542
8. Emotional rel.	.191	.275	.269	.411	.395	.409	.575	1.000

Note. Correlations are based on the sample for which all data at the task level were complete ($N = 1,985$). General scoring appears in the lower left triangle; expert scoring appears in the upper right triangle. MSCEIT V2.0 = Mayer–Salovey–Caruso Emotional Intelligence Test, Version 2.0; manag. = management; rel. = relationships.

Facilitating branches) and a "Strategic" area (Understanding and Managing branches). The four-factor model loads the two designated tasks on each of the four branches (Mayer & Salovey, 1997; Mayer et al., 2001). These analyses are particularly interesting given that the MSCEIT V2.0 represents an entirely new collection of tasks and items.

We tested these models with AMOS (Arbuckle, 1999) and cross-checked them using LISREL (Joreskog & Sorbom, 2001) and STATISTICA (Statsoft, 2002). The confirmatory models shared in common that (a) error variances were uncorrelated; (b) latent variables were correlated; that is, oblique; and (c) all other paths were set to zero. In the four-factor solution only, the two within-area latent variable covariances (i.e., between Perceiving and Facilitating, and between Understanding and Managing) were additionally constrained to be equal so as to reduce a high covariance between the Perceiving and Facilitating branch scores.

There was a progressively better fit of models from the one- to the four-factor model, but all fit fairly well (four vs. two factors, $\chi^2(4) = 253$, $p < .001$; two vs. one factors, $\chi^2(1) = 279$, $p < .001$; see Table 3 for further details). The chi-square values are a function of sample size, and their size reflects the approximately 2,000 individuals involved, more so than any absolute quality of fit. Fit indices independent of sample size include the normed fit index (NFI), which ranged from .99 to .98 across models, which is excellent (Bentler & Bonett, 1980), as well as the Tucker–Lewis index (TLI; Bentler & Bonett, 1980; Tucker & Lewis, 1973), which ranged from .98 to .96 and was also quite good, and Steiger's (1990) root-mean-square error of approximation (RMSEA), which ranged from .12 for the one-factor solution, which was a bit high, to an adequate .05 for the four-factor solution. A model fit using the four-factor solution with expert scoring was equivalent to that of general scoring (e.g., NFI = .97; TLI = .96; RMSEA = .04), and this correspondence between the expert and general consensus held for the one- and two-factor models as well.

MacCallum and Austin (2000) have noted that models alternative to those tested by researchers often also fit well with their data, and that was the case with a three-factor model described elsewhere that we tested on these data (Mayer et al., 1999). However, if one intentionally violates the four-factor model, by shifting the second task on each branch to the next branch up (and placing Branch 4's second task back on Branch 1), the chi square rises from 94 to 495, the fit indices become unacceptable (e.g., TLI drops from .96 to .78), and four of six correlations among branches are estimated at higher than 1.0. The four-branch model, in other words, does create a fit to the data that can be markedly superior to other models.

Table 3
MSCEIT V2.0 Parameter Estimates of the Observed Tasks on the Latent Variables and Goodness-of-Fit Statistics for the One-, Two-, and Four-Factor Models

	Model tested		
Variable[a]	One-factor	Two-factor	Four-factor
Branch 1	***I***	***I***	***I***
Perceiving			
Faces	.40	.50	.55
Pictures	.50	.59	.68
Branch 2			***II***
Facilitating			
Facilitation	.46	.54	.53
Sensations	.64	.71	.72
Branch 3		***II***	***III***
Understanding			
Changes	.65	.68	.77
Blends	.64	.67	.76
Branch 4			***IV***
Managing			
Emotion manag.	.68	.70	.76
Emotional rel.	.66	.68	.74
Goodness-of-fit index			
Model fit			
χ^2	626.56	347.32	94.28
df	20	19	15
NFI	.988	.993	.977
TLI	.979	.988	.964
RMSEA	.124	.093	.052
N	1,985	1,985	1,985

Note. The factor models were specified such that error terms were uncorrelated. In the four-branch model, the two within-area covariances (i.e., between Perceiving and Facilitating, and between Understanding and Management) were constrained to be equal to one another. Boldfaced, italicized roman numerals indicate the factors specified in each model; beneath each roman numeral are the estimated factor loadings for each of the tasks associated with that specific factor. All other loadings were fixed at zero. MSCEIT V2.0 = Mayer–Salovey–Caruso Emotional Intelligence Test, Version 2.0; manag. = management; rel. = relationships; NFI = normed fit index; TLI = Tucker–Lewis index; RMSEA = root-mean-square error of approximation.
[a] Variables are ordered according to the four-branch model of emotional intelligence (Mayer & Salovey, 1997).

Discussion

In this study, emotions experts converged on correct test answers with greater reliability than did members of a general sample. The experts' convergence was better in areas where more emotions research has been conducted. If future research confirms these findings, then an expert criterion may become the criterion of choice for such tests. Critiques of the EI concept have suggested, based on the use of one or two emotions experts, that expert and general consensus criteria might be quite different (e.g., correlate below .50; Roberts et al., 2001). Others have argued that, as more experts are used, and their answers aggregated, their performance would resemble that of the consensus of a large general group (Mayer et al., 2001). The 21 experts in this study did exhibit superior agreement levels relative to the general sample. At the same time, the expert and general consensus criteria often agreed on the same answers as correct ($r = .91$). Participants' MSCEIT scores were also similar according to the two different criteria ($r = .98$).

Reliabilities for Branch, Area, and Total test scores were reasonably high for the MSCEIT, with reliabilities at the level of the individual tasks ranging lower. Two-week test–retest reliabilities of $r(60) = .86$ are reported elsewhere (Brackett & Mayer, 2001). In addition, the findings from the factor analyses indicated that one-, two-, and four-factor models provide viable representations of the EI domain, as assessed by the MSCEIT V2.0.

No empirical findings by themselves can settle all the theoretical issues surrounding EI that were reflected in the September 2001 issue of *Emotion.* In addition, the applied use of EI tests must proceed with great caution. That said, the findings here suggest that those who use the MSCEIT can feel more confident about the quality of the measurement tool to assess EI. Ultimately, the value of the MSCEIT as a measure of EI will be settled by studies of its validity and utility in predicting important outcomes over and above conventionally measured emotion, intelligence, and related constructs. A number of such studies related to prosocial behavior, deviancy, and academic performance have begun to appear (Mayer et al., 2002b). In the meantime, we hope that the present findings inform and, by doing so, clarify issues of scoring, of reliability, and of viable factorial representations.

References

Arbuckle, J. L. (1999). *Amos 4.0.* Chicago: SmallWaters Corp.

Averill, J. R., & Nunley, E. P. (1992). *Voyages of the heart: Living an emotionally creative life.* New York: Free Press.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88,* 588–606.

Brackett, M., & Mayer, J. D. (2001, October). *Comparing measures of emotional intelligence.* Paper presented at the Third Positive Psychology Summit, Washington, DC.

Buchanan, T., & Smith, J. L. (1999). Using the Internet for psychological research: Personality testing on the World Wide Web. *British Journal of Psychology, 90,* 125–144.

Cattell, R. B., & Burdsal, C. A. (1975). The radial parcel double factoring design: A solution to the item-vs-parcel controversy. *Multivariate Behavioral Research, 10,* 165–179.

Ciarrochi, J. V., Chan, A. Y., & Caputi, P. (2000). A critical evaluation of the emotional intelligence concept. *Personality and Individual Differences, 28,* 539–561.

Ekman, P., & Friesen, W. V. (1975). *Unmasking the face: A guide to recognizing the emotions from facial cues.* Englewood Cliffs, NJ: Prentice Hall.

Izard, C. E. (2001). Emotional intelligence or adaptive emotions? *Emotion, 1,* 249–257.

Joreskog, K. G., & Sorbom, D. (2001). *LISREL 8.51.* Lincolnwood, IL: Scientific Software.

Kaufman, A. S., & Kaufman, J. C. (2001). Emotional intelligence as an aspect of general intelligence: What would David Wechsler say? *Emotion, 1,* 258–264.

Legree, P. I. (1995). Evidence for an oblique social intelligence factor established with a Likert-based testing procedure. *Intelligence, 21,* 247–266.

MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology, 51,* 201–226.

Mayer, J. D., Caruso, D. R., & Salovey, P. (1999). Emotional intelligence meets traditional standards for an intelligence. *Intelligence, 27,* 267–298.

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: Basic Books.

Mayer, J. D., Salovey, P., & Caruso, D. R. (2002a). *Mayer–Salovey–Caruso Emotional Intelligence Test (MSCEIT) item booklet.* Toronto, Ontario, Canada: MHS Publishers.

Mayer, J. D., Salovey, P., & Caruso, D. R. (2002b). *Mayer–Salovey–Caruso Emotional Intelligence Test (MSCEIT) user's manual.* Toronto, Ontario, Canada: MHS Publishers.

Mayer, J. D., Salovey, P., Caruso, D. R., & Sitarenios, G. (2001). Emotional intelligence as a standard intelligence. *Emotion, 1,* 232–242.

Nunnally, J. C. (1978). *Psychometric theory.* New York: McGraw-Hill.

Ortony, A., Clore, G. L., & Collins, A. M. (1988). *The cognitive structure of emotions.* Cambridge, England: Cambridge University Press.

O'Sullivan, M., & Guilford, J. P. (1976). *Four factor tests of social intelligence: Manual of instructions and interpretations.* Orange, CA: Sheridan Psychological Services.

Roberts, R. D., Zeidner, M., & Matthews, G. (2001). Does emotional intelligence meet traditional standards for an intelligence? Some new data and conclusions. *Emotion, 1,* 196–231.

Rosenthal, R., Hall, J. A., DiMatteo, M. R., Rogers, P. L., & Archer, D. (1979). *The PONS test.* Baltimore: Johns Hopkins University Press.

Schaie, K. W. (2001). Emotional intelligence: Psychometric status and developmental characteristics—Comment on Roberts, Zeidner, and Matthews (2001). *Emotion, 1,* 243–248.

Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology, 32,* 76–92.

Statsoft. (2002). Statistica 6.0 [Computer software]. Tulsa, OK: Author.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research, 25,* 173–180.

Tucker, L. R., & Lewis, C. (1973). The reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38,* 1–10.

Zeidner, M., Matthews, G., & Roberts, R. D. (2001). Slow down, you move too fast: Emotional intelligence remains an "elusive" intelligence. *Emotion, 1,* 265–275.

Received July 17, 2001
Revision received August 13, 2002
Accepted August 15, 2002 ■

United States Postal Service

Statement of Ownership, Management, and Circulation

1. Publication Title	2. Publication Number	3. Filing Date
Emotion	0 2 0 - 5 4 3	October 2002

4. Issue Frequency	5. Number of Issues Published Annually	6. Annual Subscription Price
Quarterly	4	Mbr $36 Indv $73 Inst $158

7. Complete Mailing Address of Known Office of Publication *(Not printer) (Street, city, county, state, and ZIP+4)*
750 First Street, N.E. Washington, D.C. 20002-4242
Contact Person: Barbara Spruill
Telephone: 202-336-5578

8. Complete Mailing Address of Headquarters or General Business Office of Publisher *(Not printer)*
750 First Street, N.E., Washington, D.C. 20002-4242

9. Full Names and Complete Mailing Addresses of Publisher, Editor, and Managing Editor *(Do not leave blank)*

Publisher *(Name and complete mailing address)*
American Psychological Association
750 First Street, N.E.
Washington, D.C. 20002-4242

Editor *(Name and complete mailing address)* Richard J. Davidson, PhD, Department of Psychology & Psychiatry
University of Wisconsin, 1202 W. Johnson St., Madison, WI 53706-1611
Klaus R. Scherer, PhD, Department of Psychology, Universite De Geneve
40, boulevard du Pont d'Arve, CH-1205 Geneve, Switzerland

Managing Editor *(Name and complete mailing address)*
Susan Knapp
American Psychological Association
750 First Street, N.E., Washington, D.C. 20002-4242

10. Owner *(Do not leave blank. If the publication is owned by a corporation, give the name and address of the corporation immediately followed by the names and addresses of all stockholders owning or holding 1 percent or more of the total amount of stock. If not owned by a corporation, give the names and addresses of the individual owners. If owned by a partnership or other unincorporated firm, give its name and address as well as those of each individual owner. If the publication is published by a nonprofit organization, give its name and address.)*

Full Name	Complete Mailing Address
American Psychological Association	750 First Street, N.E.
	Washington, D.C. 20002-4242

11. Known Bondholders, Mortgagees, and Other Security Holders Owning or Holding 1 Percent or More of Total Amount of Bonds, Mortgages, or Other Securities. If none, check box ☐ None

Full Name	Complete Mailing Address
NONE	

12. Tax Status *(For completion by nonprofit organizations authorized to mail at nonprofit rates) (Check one)*
The purpose, function, and nonprofit status of this organization and the exempt status for federal income tax purposes:
☐ Has Not Changed During Preceding 12 Months
☐ Has Changed During Preceding 12 Months *(Publisher must submit explanation of change with this statement)*

PS Form **3526**, October 1999 *(See Instructions on Reverse)*

13. Publication Title	14. Issue Date for Circulation Data Below
Emotion	September 2002

15. Extent and Nature of Circulation			Average No. Copies Each Issue During Preceding 12 Months	No. Copies of Single Issue Published Nearest to Filing Date
a. Total Number of Copies *(Net press run)*			N/A	3150
b. Paid and/or Requested Circulation	(1)	Paid/Requested Outside-County Mail Subscriptions Stated on Form 3541. *(Include advertiser's proof and exchange copies)*		191
	(2)	Paid In-County Subscriptions Stated on Form 3541 *(Include advertiser's proof and exchange copies)*		2091
	(3)	Sales Through Dealers and Carriers, Street Vendors, Counter Sales, and Other Non-USPS Paid Distribution		
	(4)	Other Classes Mailed Through the USPS		
c. Total Paid and/or Requested Circulation *[Sum of 15b. (1), (2),(3),and (4)]*				2282
d. Free Distribution by Mail *(Samples, compliment ary, and other free)*	(1)	Outside-County as Stated on Form 3541		
	(2)	In-County as Stated on Form 3541		91
	(3)	Other Classes Mailed Through the USPS		
e. Free Distribution Outside the Mail *(Carriers or other means)*				
f. Total Free Distribution *(Sum of 15d. and 15e.)*				91
g. Total Distribution *(Sum of 15c. and 15f)*				2373
h. Copies not Distributed				777
i. Total *(Sum of 15g. and h.)*				3150
j. Percent Paid and/or Requested Circulation *(15c. divided by 15g. times 100)*				96.1

16. Publication of Statement of Ownership
☐ Publication required. Will be printed in the March 2003 issue of this publication. ☐ Publication not required.

17. Signature and Title of Editor, Publisher, Business Manager, or Owner
[illegible], Sr Director, Publishing Services
Date: [illegible]

I certify that all information furnished on this form is true and complete. I understand that anyone who furnishes false or misleading information on this form or who omits material or information requested on the form may be subject to criminal sanctions (including fines and imprisonment) and/or civil sanctions (including civil penalties).

Instructions to Publishers

1. Complete and file one copy of this form with your postmaster annually on or before October 1. Keep a copy of the completed form for your records.
2. In cases where the stockholder or security holder is a trustee, include in items 10 and 11 the name of the person or corporation for whom the trustee is acting. Also include the names and addresses of individuals who are stockholders who own or hold 1 percent or more of the total amount of bonds, mortgages, or other securities of the publishing corporation. In item 11, if none, check the box. Use blank sheets if more space is required.
3. Be sure to furnish all circulation information called for in item 15. Free circulation must be shown in items 15d, e, and f.
4. Item 15h., Copies not Distributed, must include (1) newsstand copies originally stated on Form 3541, and returned to the publisher, (2) estimated returns from news agents, and (3), copies for office use, leftovers, spoiled, and all other copies not distributed.
5. If the publication had Periodicals authorization as a general or requester publication, this Statement of Ownership, Management, and Circulation must be published; it must be printed in any issue in October or, if the publication is not published during October, the first issue printed after October.
6. In item 16, indicate the date of the issue in which this Statement of Ownership will be published.
7. Item 17 must be signed.

Failure to file or publish a statement of ownership may lead to suspension of Periodicals authorization.

PS Form **3526**, October 1999 *(Reverse)*