Emotion
2008, Vol. 8, No. 4, 540–551

1528-3542/08/$12.00 DOI: 10.1037/a0012746

New Paradigms for Assessing Emotional Intelligence: Theory and Data

Carolyn MacCann and Richard D. Roberts
Educational Testing Service, Princeton, New Jersey

This article describes the development and validation of 2 measures of emotional intelligence (EI): the Situational Test of Emotional Understanding (STEU) and the Situational Test of Emotion Management (STEM). Study 1 (N = 207 psychology students) examines multiple sources of validity evidence: relationships with EI, vocabulary, personality, and emotion-related criteria. Study 2 (N = 149 white-collar volunteers) relates STEU and STEM scores to clinical symptoms, finding relationships to anxiety and stress for both tests, and to depression for the STEM. It is concluded that new performance-based approaches to test development, such as the present ones, might be useful in distinguishing between test and construct effects. Implications for expanding theory and for developing EI interventions are discussed.

Keywords: emotional intelligence, response format, test development, appraisal theory, situational judgment test (SJT)

Supplemental materials: http://dx.doi.org/10.1037/a0012746.supp

Emotional intelligence (EI) has traditionally been measured both as a set of personality traits and behavioral tendencies and as a set of cognitive abilities, creating some confusion, because the two measurement traditions seem to assess different constructs (Roberts, Schulze, & MacCann, 2008). This article considers EI only as a set of abilities, in line with concerns that the label *emotional intelligence* should be restricted to constructs within the intelligence domain (e.g., Mayer, Roberts, & Barsade, 2008; Mayer, Salovey, Caruso, & Sitarenios, 2001). As a set of abilities, EI is usually defined in terms of four hierarchically ordered branches (Mayer, Caruso, & Salovey, 2000). Presently, assessment of these four branches is restricted largely to one instrument: The Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT; Mayer, Salovey, Caruso, & Sitarenios, 2003).

Carolyn MacCann and Richard D. Roberts, Educational Testing Service, Princeton, New Jersey.

This research was based on Carolyn MacCann's doctoral dissertation at the University of Sydney, under the supervision of Beryl Hesketh and written with the support of an Australian Postgraduate Award to Carolyn MacCann and U.S. Army Research Institute Contract W91WAW-07-C-0025 to the Educational Testing Service (ETS).

We thank Patrick Kyllonen, Lazar Stankov, Nathan Kogan, and Dan Eignor for acting as internal reviewers on behalf of ETS, and the three *Emotion* reviewers for their helpful comments and suggestions on an earlier version of this article. We also thank Kevin Bennett, David Grayson, Barbara Griffin, Ann MacCann, Adam Pringle, Frank Rijmen, and Cris Valykria for supporting the preparation of this article. The views expressed here are those of the authors and do not reflect on the ETS, the U.S. Army Research Institute, or the University of Sydney.

Appendixes are provided online as supplementary material, available from the journal *Emotion* Web page and linked to the published version of this article.

Correspondence concerning this article should be addressed to Carolyn MacCann, Center for New Constructs, R&D, MS 16-R, Educational Testing Service, Rosedale Road, Princeton, NJ 08541. E-mail: CMacCann@ets.org

Having a good deal of the research on EI emanate from just one instrument is a suboptimal state of affairs because (a) test effects cannot be distinguished from construct effects and (b) the MSCEIT is empirically rather than theoretically keyed, such that EI scores do not have a strong theoretical background. The following two studies address these issues in three ways. First, two alternative tests to assess EI—the Situational Test of Emotional Understanding (STEU) and the Situational Test of Emotion Management (STEM)—are developed, and validity evidence is collected. Second, test characteristics of the STEM are experimentally manipulated to disentangle test effects from construct effects. Third, the possibility of using a theoretical rather than expert-based scoring system is examined by using appraisal theory to score the STEU, thereby providing a theoretical basis for emotional understanding.

Measurement Characteristics of EI Tests: The Need for More Diversity

The MSCEIT and its precursor (the Multifactor Emotional Intelligence Scale [MEIS]; Mayer et al., 2000) assess four hierarchically ordered branches of emotion-related abilities: (a) the *perception* and expression of emotions; (b) the *integration* of emotions into thought processes; (c) *understanding* the relations between, and transitions among, emotions and between emotions and circumstances; and (d) the *management* of emotions to moderate negative, and enhance positive, emotions. Perception and Integrationbranches collectively form the *Experiential EI* area, with Understandingand Management forming the *Strategic EI* area (Mayer et al., 2001). MSCEIT test items have two unusual characteristics for cognitive tests: (a) the use of "rate-the-extent" scales for six of the eight subtests (i.e., test takers rate the appropriateness, strength, or extent of each alternative, rather than selecting the correct alternative); and (b) a scoring rubric based on the consensus of expert or population-representative samples rather than emotions theory. With unique measurement methods and a unique construct,

it is difficult to know whether empirical results are attributable to the constructs examined or the measurement methods used.

As a case in point, MSCEIT Understanding scores are more strongly related to cognitive ability than are Perception, or Management scores and are also the only tests given in a multiple-choice, rather than rate-the-extent, format (see Roberts et al., 2008). Currently, it is impossible to tell whether this stronger relationship is due to the multiple-choice response format of Understanding, or whether the Understanding construct is genuinely more closely related to intelligence. To address this issue, the STEM was developed to be administered in both multiple-choice and rate-the-extent formats. This design allows an empirical test of whether a multiple-choice response format (rather than the construct of emotional understanding per se) results in stronger relationships with cognitive abilities. To address the issue of consensus scoring, the STEU items were developed according to Roseman's (2001) appraisal theory of emotions, such that answers are correct or incorrect according to this theory.

Sources of Validity Evidence for the STEU and STEM: The Nomological Net for EI

The label *emotional intelligence* reasonably implies that EI is one of multiple group factors of intelligence related to the processing of emotional stimuli (rather than the visual, verbal, or auditory stimuli of other group factors; see Mayer et al., 2008). As such, there are four logical relationships that demonstrate evidence for the validity of an EI test (see Orchard et al., in press). First, EI tests should relate positively to intelligence tests, demonstrating the positive manifold that exists among tests of intelligence. Second, EI tests should relate more strongly to other EI tests than to tests of other types of intelligence, demonstrating that EI is distinct from existing group factors of intelligence (e.g., text-based EI tests should relate more strongly to other EI tests than to tests of verbal abilities if they really do assess more than text comprehension). Mayer et al. (2000) referred to these two criteria jointly as the "correlational criteria." Third, EI tests should relate to variables or outcomes reasonably indicative of facility with emotions (e.g., coping with stress and lack of emotion-related disorders), demonstrating the appropriateness of the adjective *emotional* in emotional intelligence. Lastly, EI test scores should correlate with personality only in the range that other tests of intelligence tend to (i.e., at .30 or less; Ackerman & Heggestad, 1997), demonstrating that EI is in the intelligence domain rather than in the personality domain. These four criteria are used as guidelines to evaluate the validity evidence for the STEU and STEM in the two studies that follow.

Study 1: Test Development, Validity Evidence, and Comparison of Response Formats

This study collects validity evidence for the STEU and two versions of the STEM (rate-the-extent and multiple choice) in a sample of undergraduate psychology students. In addition, this study examines the effects of different response formats on the validity of the STEM to inform prior research results using the MEIS or MSCEIT (for which all Understanding tests are multiple choice and all Management tests are in rate-the-extent format).

Validity Evidence for the STEU and STEM

Validity Evidence for Criteria 1 and 2: Relationships Between EI and Intelligence

In this study, we examine the relationships of STEU and STEM scores to vocabulary test scores, as well as to a third measure of EI: the MEIS Stories test. (The Stories test was chosen in the interest of design efficiency; it shares the highest loading on a general EI factor. See Matthews, Zeidner, & Roberts, 2002, and Roberts, Zeidner, & Matthews, 2001). Both the STEU and STEM should relate moderately to vocabulary, demonstrating that EI is a cognitive ability. However, this vocabulary–EI relationship should not be as strong as that between the three EI tests (STEU, STEM, and Stories) if EI is a separate construct from verbal ability or crystallized intelligence. In addition, correlations between the EI tests should remain significant after controlling for vocabulary if the shared variation is due to EI rather than text comprehension.

Validity Evidence for Criterion 3: Relationships Between EI and Emotion-Related Criteria

As the label *emotional intelligence* implies, EI measures should relate to emotion-related criteria. Three emotion-related criteria are considered in this study: (a) alexithymia, (b) life satisfaction, and (c) academic achievement. The STEU's and STEM's incremental prediction of these three criteria after controlling for personality and vocabulary scores is also examined.

Alexithymia. Alexithymia was originally conceptualized as a set of symptoms defining a clinical condition (e.g., difficulty identifying and describing one's emotions and a concrete, externally oriented thinking style; Sifneos, 1973) and is theoretically linked with low levels of EI. Understanding and Management MSCEIT scores relate to (less) difficulty identifying and describing emotions, although magnitudes are small (correlations of .20 to .30; Barchard & Hakstian, 2004; Warwick & Nettelbeck, 2004). The STEU and STEM are expected to relate to alexithymia to a similar magnitude.

Life satisfaction. Understanding emotional situations and managing them effectively might reasonably relate to greater satisfaction with one's life. Existing research has found small but nontrivial relationships of life satisfaction with Management and Understanding, in the range of .10 to .25 (Bastian, Burns, & Nettelbeck, 2005; Ciarrochi, Chan, & Caputi, 2000; Mayer et al., 2000). McDonald (1999) suggested that the Satisfaction With Life Scale (SWLS) measures two separable constructs: (a) a retrospective perspective (i.e., satisfaction with the past events of one's life) and (b) a current perspective (i.e., satisfaction with the current conditions of one's life). Because of the role of positive reappraisal and (lack of) rumination about past events in Management, STEM scores might be expected to relate more strongly to a retrospective perspective of life satisfaction.

Academic achievement. Some researchers hypothesize that academic achievement is partly determined by emotion-related variables and thus might be predicted by EI (e.g., Barchard, 2003; Parker, Summerfeldt, Hogan, & Majeski, 2004). That is, the social and emotional demands of academic life contribute to success as much as the cognitive demands do. Research using the MSCEIT has found a relationship between Understanding and academic

achievement, although this relationship decreased to a trivial level after accounting for both intelligence and personality (Barchard, 2003; Woitaszewski & Aalsma, 2004). In this study, two measures of academic achievement are examined: (a) weighted average mark (WAM) over all university participants at the end of the first year of study and (b) psychology grade for the first semester of study. It is expected that the STEU and the STEM will predict university grades, with this relationship stronger for STEU than for STEM. We might further expect the correlation with psychology grades to be higher than the overall WAM, given that the subject matter of psychology requires knowledge and processing of material relating to human emotions and social phenomena. That is, in addition to the role that EI might play in coping with the social and emotional demands of academic life, emotional knowledge might actually overlap with the content of a first-year psychology program.

Validity Evidence for Criterion 4: Relationships Between EI and Personality

EI tests should be distinct from personality traits to be considered part of the intelligence, rather than personality, domain. Using a relatively brief measure of the 5 superfactors of personality, we hypothesize that STEU and STEM scores should not correlate strongly (in excess of $r = .50$; see Cohen, 1988) with any of the 5 factors. However, both tests might be expected to relate to Agreeableness and to Openness at about the $r = .30$ level, with relationships stronger for the STEM than for the STEU (in line with findings for MEIS/MSCEIT research; see Roberts et al., 2008).

Methodological Issues

Administering the STEM in both multiple-choice and rate-the-extent response formats allows two methodological issues to be examined: (a) the relationship of different branches of EI to intelligence and (b) the relationship between Management and Understanding. If response format issues affect EI's relationship to intelligence, then the multiple-choice STEM should relate more strongly to vocabulary than does the ratings-based STEM. Additionally, response format issues may affect interbranch correlations: Management and Understanding might be expected to correlate more strongly when both are in multiple-choice format (i.e., the STEU will relate more strongly to the multiple-choice STEM than to the ratings-based STEM).

Summary of Hypotheses

Hypothesis 1

Relationships between EI and vocabulary will be affected by response format: There will be stronger correlations when the STEM is multiple choice than when it is in rate-the-extent format, and the STEU will relate more strongly with vocabulary than does the rate-the-extent STEM but not the multiple-choice STEM.

Hypothesis 2

Relationships between Understanding and Management will be affected by response format: There will be a stronger relationship between STEU and STEM when both are in multiple-choice formats than when the STEM is in a rate-the-extent format.

Hypothesis 3

The STEU and STEM will meet Mayer et al.'s (2000) correlational criteria for an intelligence: STEU and STEM scores will correlate with each other and with the Stories test more highly than they correlate with vocabulary scores. In addition, correlations between STEU, STEM, and Stories will remain significant when controlling for vocabulary.

Hypothesis 4

STEU and STEM scores will not correlate strongly (in excess of $r = .50$) with any dimension of the 5-factor model of personality but may show a small degree of relationship to Agreeableness and Openness to Experience.

Hypothesis 5

STEU and STEM scores will relate meaningfully to criterion variables. Small negative correlations with alexithymia facets and small positive correlations with life satisfaction and academic achievement are expected. These relationships are further expected to hold after personality and vocabulary are controlled.

Method

Test Development: STEU

Roseman's (2001) appraisal theory (see Figure 1) is used as the basis for item construction and scoring of the STEU. This theory defines how 17 discrete emotions are generated according to specific combinations of seven appraisal dimensions. For example, the diagram in Figure 1 shows that *relief* results from circumstance–cause, certainty, motive consistency, and aversive stimuli. In lay language, relief occurs when "an unwanted situation becomes less likely or stops altogether." Thus, the STEU item generated was

> an unwanted situation becomes less likely or stops altogether. The person involved is most likely to feel: (a) regret, (b) hope, (c) joy, (d) sadness, (e) relief

with (e) as the correct answer.

A similar process was conducted for 13 other emotions in the model. In addition, analogues representing workplace and personal-life contexts were created. For example, the personal-life analogue for relief was

> an irritating neighbor of Eve's moves to another state. Eve is most likely to feel? (a) regret, (b) hope, (c) relief, (d) sadness, (e) joy.

The workplace analogue was

> a supervisor who is unpleasant to work for leaves Alfonso's work. Alfonso is most likely to feel? (a) joy, (b) hope, (c) regret, (d) relief, (e) sadness.

Thus, the STEU contains 42 items: 14 context-reduced, 14 with a personal-life context, and 14 with a workplace context. These items are reproduced in Appendix A (which is available online as supplementary material).

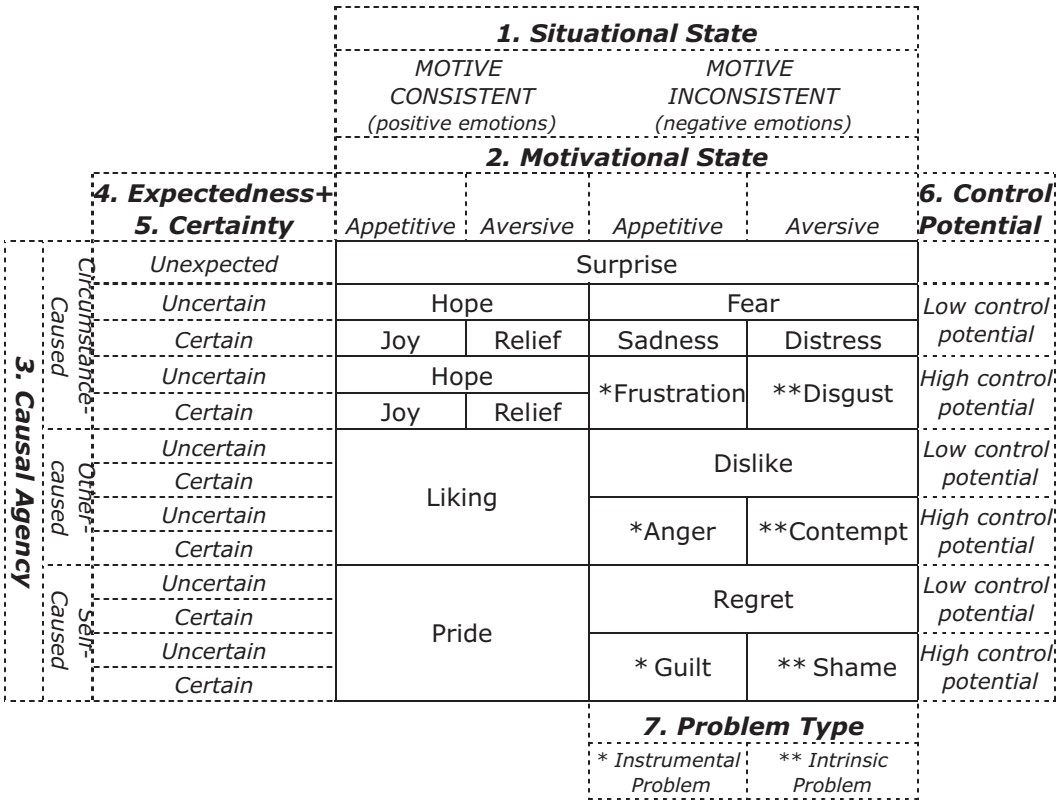

Figure 1. Roseman's structure of the emotions (adapted from Roseman, 2001, pp. 70–71). The seven appraisal dimensions may be briefly defined as follows: (1) *Situational state* = motive consistent (an event is desired) versus motive inconsistent (an event is unwanted); (2) *Motivational state* = aversive (want to minimize punishment) versus appetitive (want to maximize reward); (3) *Causal agency* = caused by self, caused by another person, or caused by circumstance (i.e., no person has caused this event); (4) *Expectedness* = expected versus unexpected; (5) *Certainty* = uncertain (event may or may not happen) versus certain (event has happened or definitely will happen); (6) *Control potential* = low (there is nothing or very little one can do to change the motive-relevant aspects of the event) versus high (there is something one can do to change the motive-relevant aspects of the event); (7) *Problem type* = instrumental (unwanted because it blocks attainment of goal) versus intrinsic (unwanted because of inherent nature of the event or stimuli).

Test Development: STEM

The STEM was developed according to the Situational Judgment Test (SJT) paradigm, with two alternative response formats: multiple choice and rate-the-extent. The following paragraphs summarize the three SJT development stages.

Step 1: Item generation. Semistructured interviews were held with 50 individuals (31 women; 30 psychology students and 20 community volunteers), who described between 3 and 11 emotional situations they had experienced in the past 2 weeks, for a total of 290 situations. These interviews were transcribed and content-analyzed to provide both a content structure for the test and the material for the test items themselves. Content coding was reasonably reliable (raw agreement ranged from .85 to .98 between two coders). Test content was structured such that items represented one of four emotions (sadness, anger, fear, or disgust) and one of several types of situations within each emotion. For example, anger situations included the situation types "fight/argument," "goal striving impeded," and "unfairness/injustice." For each situation type, at least 6 items represented "workplace" content, and 6 represented "personal-life" content. One hundred thirty-eight items were created in this way, with each item consisting of a 1- to 2-sentence description of emotionally salient aspects of a situation.

Step 2: Response-option generation. A second sample of 99 undergraduate students (56 women) was split into three groups. Each group was given approximately one third of the 138 items and asked to write both (a) the *best thing to do* in that situation and (b) what *they would do* if they were in that situation. These responses were summarized into different types, each represented by a brief phrase. For many items, there was little variability among responses (i.e., there was one obvious course of action). In such instances, these items could not be used. At the end of Step 2, there were 44 items, 18 with anger content, 14 with sadness content, and 12 with fear content.

Step 3: Expert scoring. Matthews et al. (2002) proposed multiple domains of expertise for emotion management. Relevant experts might include people with (a) academic knowledge of emotions; (b) experience in professions geared toward emotional healing (e.g., counseling, psychiatry, psychotherapy, and possibly

some forms of religious leadership); or (c) professions related to understanding and managing people's relationships and goals (e.g., human resource–related careers and life coaches). In line with these three domains of expertise, experts were members of an EI research consortium, professionally trained psychologists holding a master's degree or equivalent, or life coaches with experience in counseling or psychology. Two expert groups responded to the 44 items, one group responding to multiple-choice items ($N = 13$, 9 women, ages 24 to 64 years), and the other group rating each response option on a 6-point scale ($N = 6$, all women, ages 25 to 51 years). All experts were living and working in Australia, such that it can be assumed most were from an Australian, "Western" culture. Agreement between experts was calculated for both groups, and one expert was excluded from each group because of low agreement with the others. After exclusion, raw agreement averaged .59 for the multiple-choice group (chance = .25; mean $\kappa = .41$), and the mean Pearson correlation between ratings was .63 for the rate group. There are thus two sets of weights for each option: (a) the mean expert rating of each option and (b) the proportion of experts selecting each option.

Participants

Psychology undergraduate students ($N = 207$, 140 women) took part in the experiment for course credit. Their median age was 19.0 years ($M = 21.1$, $SD = 5.6$). Participants were drawn from both a rural campus ($N = 55$) and an urban campus ($N = 149$) of Sydney University. In terms of primary cultural identification, 110 participants reported an Australian or Anglo-Celtic identification (e.g., British or American), 77 an Asian identification, and the remaining as "other" (2 participants did not report their gender; 3 participants did not report their age; and 2 participants did not report their ethnicity).

Design

This study used a quasi-experimental design, in which two groups of participants completed two different test batteries: A or B. In Battery A, the STEM was administered as a multiple-choice test. In Battery B, the STEM was administered with rate-the-extent options. For logistical reasons, all participants completing Battery A were from the urban campus ($N = 113$), while all participants from the rural campus ($N = 55$), and a subset from the urban campus ($N = 39$), completed Battery B.

Test Battery

STEU. Participants completed the 42 multiple-choice items of the STEU, as has been described in the preceding section (see also Appendix A). Total scores are calculated by taking the mean of all item scores as is done for the MEIS and MSCEIT.

STEM. The 44 STEM items were presented to participants in either multiple-choice format (in Battery A) or rate-the-extent format (in Battery B). The multiple-choice STEM was scored according to expert weights, and the rate-the-extent STEM was scored according to the distance from the expert ratings (distance scores were reversed for ease of interpretation, so that higher scores indicated higher levels of EI rather than a greater distance from expert opinion). Test items and scoring weights are reproduced in Appendix B (which is available online as supplementary material).

Stories test. The Stories test was taken from the MEIS Scale (Mayer et al., 2000). Participants were presented with six stories of 2–3 sentences describing events happening to a fictitious person. For each story, participants rated the degree of emotion they thought the protagonist felt for seven different emotions. Ratings were given on a 5-point scale from *definitely present* to *definitely NOT present*. The Stories test was scored by the test authors' consensus weights, as this has been shown to be the most reliable (as well as most frequently used) scoring format (e.g., Mayer et al., 2000; Roberts et al., 2001). Total scores are calculated by taking the mean of all item scores as is done for the MEIS and MSCEIT.

Vocabulary. This 18-item test, representing the primary mental ability of verbal comprehension and crystallized intelligence (Gc) at the second order, was taken from Stankov (1997). Participants were given a target word and asked to select the word most similar to the target from five alternatives. For example:

Revolve: 1. A gun, 2. Uprising, 3. Turn around, 4. Grow, 5. Decide.

Openness Conscientiousness Extraversion Agreeableness Neuroticism Index Condensed 20-item version (OCEANIC-20). A 20-item short form of the OCEANIC was constructed for this study by taking 4 items from each dimension of the existing 60-item OCEANIC (Schulze & Roberts, 2006). Items were selected for their high loadings on each dimension as has been reported by Roberts (2000). Participants rated the 20 items on a 5-point scale from *never* (1) to *usually* (5): for example, *I am talkative*.

Toronto Alexithymia Scale (TAS-20; Bagby, Parker, & Taylor, 1994). Participants rated 20 self-report items on a 5-point scale from *strongly disagree* (1) to *strongly agree* (5). The TAS-20 indexes three aspects of alexithymia: (a) difficulty identifying feelings (DIF; e.g., *I have feelings that I cannot quite identify*); (b) difficulty describing feelings (DDF; e.g., *People tell me to describe my feelings more*); and (c) an externally oriented thinking style (EOT; e.g., *I find examination of my feelings useful in solving personal problems*).

SWLS (Diener, Emmons, Larsen, & Griffin, 1985). Participants rated five items on a 7-point scale from *strongly disagree* (1) to *strongly agree* (7): for example, *In most ways my life is close to ideal*. In line with previous structural analysis of the SWLS, scores are analyzed in terms of two components: current satisfaction and retrospective satisfaction (McDonald, 1999).

Procedure

Participants read through an information form and then completed a consent form indicating their willingness to participate and for the experimenter to access their grades from university records (128 psychology grades and 117 weighted average grades over all tertiary subjects were available for analysis; the majority of these were for participants completing Battery A). Participants from the city campus read paper-and-pencil information and consent forms and completed the test battery outlined below on PCs in groups of (approximately) 10, in an on-campus laboratory. Participants from the rural campus were e-mailed information, consent forms, and a computerized version of the test battery, which they

completed at computer terminals of their choosing. Although all tests were self-paced, testing took approximately 1 hr. This study was approved by the Sydney University Human Research Ethics Committee.

Results

Reliability and Descriptive Statistics

Descriptive statistics for the EI measures are shown in Table 1. Reliabilities are acceptable for group assessment for all tests (i.e., greater than .60; Wasserman & Bracken, 2003). Reliabilities for the 20-item OCEANIC were slightly lower than for the full form (Roberts, 2000), but never by more than .10, which seems reasonable for a two- thirds reduction in the number of items. Descriptive statistics and reliability estimates were in line with prior findings for all measures. There were no gender differences for the EI or vocabulary measures, and gender differences were generally consistent with prior findings for the other scales. Thus, men scored (a) significantly higher on Open to Experience and Externally Oriented Thinking and (b) significantly lower on Extraversion, Neuroticism, and Difficulty Identifying Feelings than did women (see, e.g., Feingold, 1994; Montebarocci, Codispoti, Baldaro, & Rossi, 2004).

Correlations of EI with Other Variables

Table 2 shows the Pearson correlations among the EI tests (STEU, STEM, and Stories), as well as between the EI tests and other variables in the study. Partial correlations controlling for vocabulary and all five dimensions of personality are also shown. In line with Hypothesis 1 (that relationships of EI with vocabulary would be effected by response format), the multiple-choice STEM and the STEU showed a similar relationship to vocabulary ($z = 0.79$, $p = .430$), whereas the rate-the-extent STEM showed significantly lower correlations with vocabulary than did the STEU ($z = 2.67$, $p = .008$). However, the difference between the multiple-choice and rate-the-extent STEM's respective relationships to vocabulary was not itself significant ($z = 1.76$, $p = .078$). In line with Hypothesis 2 (that relationships between Understanding and Management would be affected by response format), the multiple-choice STEM correlated significantly more strongly with the STEU than did the rate-the-extent STEM ($z = 3.22$, $p = .001$).

Hypothesis 3 (that the STEU and STEM would meet Mayer et al.'s, 2000, correlational criteria for EI) was partly supported. Thus, although the STEU and STEM correlated more highly with each other than with vocabulary (with high correlations remaining significant after controlling for vocabulary and personality), only the STEU related to the Stories test. More problematic for this hypothesis, the relationship between the STEU and Stories test was smaller than that between the STEU and vocabulary.

Hypothesis 4 was confirmed. Thus, performance-based EI tests showed independence from personality, with the highest correlation at $r = .24$. The STEU and multiple-choice STEM showed small, positive relationships with Agreeableness, although the expected relationship with Openness was not observed. Of note, the magnitude of the relationship between Agreeableness and Man-

Table 1
Descriptive Statistics, Reliabilities, Gender Differences, and Comparison of Alpha and Mean Differences to Published Sources

	Current study				Comparison group	
Source	*N*	α	*M (SD)*	Gender *d*[a]	α	*d*
1. STEU	200	.71	0.60 (.13)	−0.11	—	—
2. STEM (multiple choice)	112	.68	0.52 (.07)	−0.35	—	—
3. STEM (rate the extent)	91	.92	2.57 (.46)	−0.10	—	—
4. Stories test[b]	201	.83	0.34 (.07)	−0.11	.76	−0.19
5. Vocabulary[c]	201	.79	0.49 (0.20)	0.00	.75	0.01
6. Openness (O)[d]	199	.71	13.21 (3.05)	0.39**	.80	—
7. Conscientiousness (C)	199	.82	13.99 (3.19)	0.04	.87	—
8. Extraversion (E)	199	.85	12.42 (3.56)	−0.32*	.90	—
9. Agreeableness (A)	199	.77	16.97 (2.39)	−0.14	.88	—
10. Neuroticism (N)	199	.75	12.71 (3.15)	−0.36*	.87	—
11. Difficulty identifying feelings[e]	200	.81	16.87 (5.52)	−0.35*	.85	0.01
12. Difficulty describing feelings	200	.75	13.56 (4.32)	−0.10	.84	0.05
13. Externally oriented thinking	200	.73	17.68 (4.83)	0.41**	.63	0.01
14. Satisfaction with Life Scale (SWLS)[f]	201	.85	23.31 (6.23)	0.01	.87	−0.03
15. SWLS: retrospective	201	—	23.31 (6.23)	0.09	—	—
16. SWLS: current	201	—	14.49 (3.89)	−0.09	—	—
17. Psychology grade	128	—	68.03 (11.67)	−0.26	—	—
18. Weighted average mark	117	—	66.76 (13.40)	0.02	—	—

Note. STEU = Situational Test of Emotional Understanding; STEM = Situational Test of Emotion Management; *M* = Mean; *SD* = Standard Deviation.
[a] Effect sizes were calculated using $(\mu_1-\mu_2)/\sqrt{\sigma_1^2(n_1-1)+\sigma_2^2(n_2-1))/(n_1+n_2-2)}$ (using sample variances) with group 1 as men or as the current sample, such that positive values indicate higher scores for men than for women or higher scores for the current sample than for the comparison group.
[b] Stories test comparison data from Ciarrochi et al., (2000).
[c] Vocabulary comparison from MacCann et al.'s (2004) data.
[d] Openness Conscientiousness Extraversion Agreeableness Neuroticism Index Condensed (OCEANIC) comparison data from Roberts (2000); note that *d* scores were not calculated because this study used a short form of the OCEANIC.
[e] Toronto Alexithymia Scale (TAS-20) comparison from Ciarrochi et al., (2003).
[f] SWLS comparison from Diener et al., (1985).
* $p < .05$. ** $p < .01$ for a *t* test of group differences.

Table 2
Correlations of EI Scores with Other EI Scores, Vocabulary, Personality, and Criterion Variables

Source	STEU		STEM (multiple choice)		STEM (rate the extent)		Stories test	
STEM (multiple choice)	.70**	(.65**)						
STEM (rate the extent)	.37**	(.35**)						
Stories	.40**	(.33**)	.18	(.07)	.04	(−.02)		
Vocabulary	.49**		.41**		.18		.30**	
Openness (O)	.11		−.11		−.07		−.08	
Conscientiousness (C)	−.02		−.05		−.12		−.03	
Extraversion (E)	−.06		.09		.13		−.05	
Agreeableness (A)	.16*		.24*		.05		−.02	
Neuroticism (N)	−.04		.02		−.14		.02	
TAS-20 (DIF)	−.12	(−.07)	.01	(.04)	−.13	(−.08)	.13	(.17*)
TAS-20 (DDF)	−.12	(−.09)	−.07	(−.02)	−.20	(−.17)	.00	(−.04)
TAS-20 (EOT)	−.38**	(−.27**)	−.43**	(−.40**)	−.05	(−.02)	.21**	(.20**)
SWLS (current)	.12	(.13)	.10	(.11)	.07	(.02)	.01	(.02)
SWLS (retrospective)	.12	(.11)	.28**	(.29**)	.08	(.01)	−.03	(−.04)
Psychology grade	.42**	(.31**)	.34**	(.27**)			.18*	(−.09)
Weighted average mark	.37**	(.26**)	.16*	(.06)			.30**	(.23*)

Note. EI = Emotional Intelligence; STEU = Situational Test of Emotional Understanding; STEM = Situational Test of Emotion Management; TAS-20 = Toronto Alexithymia Scale; DIF = Difficulty Identifying Feelings; DDF = Difficulty Describing Feelings; EOT = Externally Oriented Thinking; SWLS = Satisfaction with Life Scale. Partial correlations controlling for vocabulary and personality are shown in parentheses. Not enough grades were available for the subsample taking the rate-the-extent STEM for these correlations to be stable.
* $p < .05$. ** $p < .01$.

agement is consistent with similar findings for the MSCEIT (see Roberts et al., 2008).

Hypothesis 5 received partial support. The STEU and multiple-choice STEM related to externally oriented thinking, psychology grade, and WAM. The multiple-choice STEM also related to retrospective life satisfaction. After controlling for vocabulary and personality, all but one of these relationships (the one between the multiple-choice STEM and WAM) remained significant. However, the rate-the-extent STEM did not relate with any of the criteria (although note that relationships with grades were not examined), and two of the aspects of alexithymia (difficulty identifying feelings and difficulty describing feelings) did not relate to any EI measures (i.e., neither the STEU and STEM, nor the Stories test).

Study 1: Discussion

Validity evidence for the STEU and STEM as measures of EI is reasonable, although some caveats should be noted. On the one hand, scores from both measures related to each other, to the externally oriented thinking component of alexithymia, and were independent of personality. The STEU also related to the Stories test and incrementally predicted academic achievement, and the multiple-choice STEM incrementally predicted retrospective life satisfaction. On the other hand, relationships with vocabulary may indicate considerable overlap between EI and crystallized ability, while the lack of relationship with two aspects of alexithymia might be considered problematic. With these provisos, further research into the correlates of the STEU and STEM is probably necessary to ascertain whether these tests can be used as valid measures of EI.

Results supported the suspected "method effect" of response format. When both the STEU and STEM were multiple choice, there was no difference in the strength of their relationship to vocabulary. However, the multiple-choice STEU related more strongly to vocabulary than did the rate-the-extent STEM. That is, the difference in correlations with vocabulary was between multiple-choice and rate-the-extent formats, rather than between Understanding and Management constructs. In addition, Understanding and Management intercorrelated more strongly when both tests were multiple choice than when Management was assessed with the rate-the-extent STEU. Both of these findings have clear implications for interpreting research on the MSCEIT, for which Understanding tests are all multiple choice and Management tests are all in rate-the-extent format.

Study 2: Further Validity Evidence from a Community Sample

The purpose of this study is to provide further validity evidence for the STEM and STEU, examining two key issues: (a) whether the STEU and STEM are reliable in an older, work (i.e., nonstudent) sample and (b) whether the STEU and STEM relate to state anxiety, depression, and stress, thereby providing further evidence for the validity of the measures. These issues are described in more detail in the next section.

The Importance of a Non-University Sample

In traditional intelligence assessments, test stimuli are designed to be equally familiar to all test takers, such that demographics and life experience should not affect test scores (e.g., Jensen, 1980). However, this practice is difficult to implement in situational tests, in which the construct of interest may well relate to familiarity with certain stimuli (i.e., when assessing knowledge of emotion-related situations, group differences may relate to familiarity with such situations, or interest in emotional phenomena). Landy (2006)

suggested that the common use of student samples (especially psychology students) is problematic for social and emotional intelligence research, as such samples are atypical in their exposure to, and interest in, social and emotional phenomena. In line with these concerns, this study examines whether the reliability of the STEU and STEM generalizes beyond psychology undergraduates to a sample of participants drawn from outside the university.

Relationship of Strategic EI to States of Distress

This study assesses state distress with the Depression Anxiety Stress Scales (DASS; Lovibond & Lovibond, 1995), which differentiates among states of depression, anxiety, and stress. Some existing research has found that EI (as assessed by the MSCEIT in particular) modestly predicts lower states of depression, anxiety, and stress, with correlations in the .10 to .30 range (e.g., Bastian et al., 2005; Brackett & Salovey, 2006; Goldenberg, Matheson, & Mantler, 2006; Matthews et al., 2006, cf., however, Gohm, Corser, & Dalsky, 2005, who found no relationship with stress). Collective findings tend to suggest that this relationship is stronger for Management than for Understanding, and stronger for anxiety and depression than for stress (Mayer et al., 2008). The STEU and STEM should provide similar relationships in order to provide meaningful (and comparable) criterion-related validity evidence.

Summary of Hypotheses

Hypothesis 1

The STEU and STEM will show the same degree of internal-consistency reliability in the current sample as in the psychology undergraduate sample from Study 1.

Hypothesis 2

STEU and STEM scores will relate to state anxiety, depression, and stress, with stronger relationships observed for the STEM (over the STEU) for each of these three clinically relevant constructs.

Method

Participants

Participants were volunteers recruited from the Sydney area via advertisements and circulars, who were offered feedback on their scores as an incentive to participate. As an ethical requirement from Sydney University's Human Research Ethics Committee, participants whose scores were 2 standard deviations above the population mean on anxiety or stress did not complete the STEU or STEM (as the possibility of receiving negative feedback might be upsetting for vulnerable participants; 14 participants were excluded under these conditions). After exclusion criteria applied, the sample consisted of 149 participants (107 women) ages 18 to 59 years ($M = 35.33$, $SD = 11.03$). The majority of participants ($N = 110$; 74%) identified their cultural group as primarily Australian or Anglo-Celtic, and 24 (16%) indicated that English was not their first language. The sample was more highly educated than the Australian average: 68% of the sample had postsecondary school qualifications, compared to the Australian average of 58% (Australian Bureau of Statistics).

Procedure

After reading the initial advertisement or circular, participants e-mailed Carolyn MacCann stating their willingness to participate. After indicating their interest, participants received an e-mail response giving the URL and unique password and instructions for participating in the study. Participants completed the test battery online, completing demographic questions, followed by the DASS, the STEU, and the multiple-choice version of the STEM.

Test Battery

Participants completed the following measures.

STEU. As in Study 1, participants completed the 42 multiple-choice items of the STEU described in the *Test Development* section of this article. The test was not timed.

STEM. The STEM was given in multiple-choice format and scored according to expert proportions (as in Study 1). In an attempt to improve the reliability of the STEM, the 13 least reliable items were not included, such that the total score was based on 30 items (the excluded items are asterisked in Appendix B).

DASS (Lovibond & Lovibond, 1995). The 21-item version of the DASS consists of items rated on a 4-point scale from *did not apply to me at all* (0) to *applied to me very much or most of the time* (3). There are an equal number (i.e., 7) of depression (e.g., *I felt downhearted and blue*), anxiety (e.g., *I felt scared without any good reason*), and stress (e.g., *I found it difficult to relax*) items.

Results

Reliability and Descriptive Statistics

Table 3 shows the reliabilities and descriptive statistics for the STEU, STEM, and DASS, comparing the means and reliability of the STEU and STEM to results from Study 1. There were no significant gender differences for any of these scores. Reliabilities were acceptable for the DASS scales and STEM, but quite low for the STEU. In comparison with the psychology undergraduate sample in Study 1, reliability was comparable for the STEM but significantly lower for the STEU, and this sample scored significantly more highly on both the STEU and STEM, with a greater disparity for the STEM. Standard deviations for the STEU and STEM are also smaller than in Study 1, indicating that there may be restriction of range. Descriptive statistics for the DASS are consistent with the comparison sample, although reliabilities for the Anxiety and Stress scales were significantly lower (but still well within an acceptable range).

Relationship Between Strategic EI and The DASS

The STEU correlated with Anxiety ($r = -.25$, $p < .01$) and Stress ($r = -.17$, $p < .05$), but not with Depression ($r = -.15$, *ns*). The STEM correlated with Anxiety ($r = -.27$, $p < .01$), Stress ($r = -.26$, $p < .01$), and Depression ($r = -.17$, $p < .05$).

Study 2: Discussion

For both the STEU and the STEM, there may be restriction of range on ability in this sample, possibly due to the higher average age and higher average level of education of these participants.

Table 3
Descriptive Statistics and Reliabilities for Study 2, With Comparison of Alpha and Mean Differences to Study 1 (for the STEU and STEM) or Published Sources (for the DASS)

	Current sample			Comparison sample			Statistical comparisons of reliability and means	
	α	*M*	*SD*	α	*M*	*SD*	α (χ^2)[a]	*M* (*d*)[b]
1. STEU[c]	.43	.63	.09	.71	.60	.13	17.89**	.26*
2. STEM (30 items)	.61	.61	.08	.72	.57	.09	3.39	.53**
3. DASS: Depression[d]	.91	3.63 (2.66)	4.21 (2.67)	.91	3.60	3.27	0.00	.01
4. DASS: Anxiety	.73	2.69 (2.30)	2.89 (2.25)	.81	2.62	2.42	6.04*	.03
5. DASS: Stress	.83	5.93 (5.38)	3.83 (3.29)	.89	5.27	3.47	9.10**	.19

Note. STEU = Situational Test of Emotional Understanding; STEM = Situational Test of Emotion Management; DASS = Depression Anxiety Stress Scale; *M* = Mean; *SD* = Standard Deviation. The DASS means and standard deviations in parentheses exclude the 14 participants scoring > 2 *SD*s above the mean on anxiety or depression (who did not complete the emotional intelligence tests); *d* scores are calculated using the whole sample, including these 14.
[a] Differences between alphas are reported as the chi-square difference based on Hakstian and Whalen's (1976) formula.
[b] Mean differences were calculated as an effect size using sample variances in the formula $(\mu_1-\mu_2)/\sqrt{\sigma_1^2(n_1-1)+\sigma_2^2(n_2-1))/(n_1+n_2-2)}$, with Group 1 as the current sample (such that positive values indicate higher scores for the current sample than in Study 1).
[c] Comparison samples for the STEU and STEM are from Study 1.
[d] Comparison samples for the DASS are from Lovibond and Lovibond (1995), with parameters from the 42-item DASS halved for comparison purposes, as is instructed in their article.
* $p < .05$. ** $p < .01$ (for χ^2 distribution, or *t* test of group differences).

Previous research on the MSCEIT has found that older participants score more highly on both Understanding and Management and that this difference is larger for Management (Bastian, 2005). Current results replicated this finding, because the volunteer sample obtained higher STEU and STEM scores than did the student sample in Study 1, with a larger difference for the STEM. However, it is possible that the higher scores obtained in Study 2 might also be a function of the unproctored format.

Hypothesis 1 was supported for the STEM, but not for the STEU. Thus, the STEM showed similar reliability in this sample in comparison with the psychology undergraduate sample. However, the STEU showed considerably lower reliability in the current sample, indicating that STEU items may require modification to be appropriate for use with nonstudent samples. In summary, the STEM but not the STEU appears reliable outside the student population.

Hypothesis 2 was supported. Both STEU and STEM scores showed small correlations with state distress, in the same range as previously reported in research with MSCEIT Understanding and Management subscales. The STEU predicted anxiety more strongly than depression or stress, consistent with Matthews et al.'s (2006) and Bastian et al.'s (2005) findings for MSCEIT Understanding. The STEM predicted all three distress states more strongly than did the STEU, consistent with Hypothesis 2, and with previous findings for the MSCEIT. Both measures predicted anxiety and stress, though STEU failed to show the predicted relationship with depression.

Although this study did not account for personality or intelligence in the EI–distress relationship, the previous study found no relationship between Neuroticism and EI and only minimal relationships between Agreeableness and EI, meaning that controlling for personality is unlikely to account for results. In addition, intelligence relates only minimally to stress reaction ($\rho = -.09$ for Gc and $\rho = -.08$ for Gf [fluid intelligence, or innate reasoning ability]; see Ackerman & Heggestad, 1997), indicating that controlling for intelligence is unlikely to decrease the magnitude of the relationship with stress.

General Discussion

These studies demonstrate that standards-based and SJT approaches to EI test construction are possible and that differences in response format may affect the properties of test scores. Preliminary evidence for the validity of the STEU and STEM was reasonable. Scores showed modest and incremental prediction of some important criteria, independence from personality, meaningful relationships with other EI tests and states of distress, and a strong degree of association with an existing measure of intelligence.

Validity Evidence for the STEU and STEM

The STEU and STEM clearly assess a kind of intelligence, as was demonstrated by relationships with vocabulary, university grades, and increasing age, and independence from personality. Significant correlations of the STEU with STEM and Stories test scores after controlling for vocabulary also indicate that there is some unique EI variance not accounted for by verbal ability. However, the STEU and STEM related more strongly to vocabulary than to the Stories test, which is problematic for the claim that these tests assess emotional intelligence rather than intelligence more generally. Further research examining the relationship of the STEU and STEM to additional tests of EI (e.g., the MSCEIT, or the Intrapersonal and Interpersonal emotional abilities of Freudenthaler & Neubauer, 2005) as well as other types of intelligence is needed in order to determine whether these tests assess emotional intelligence rather than some other cognitive ability group factor.

There was a clear link between the STEU and STEM and the externally oriented thinking aspect of alexithymia, which held irrespective of the test takers' personality or vocabulary scores. It might reasonably be assumed that people with a thinking style oriented toward social and emotional phenomena rather than toward concrete external facts develop better understanding and better regulation strategies. Obviously, one has to notice emotional phenomena to understand the cause of the emotions or to develop

regulation strategies for such situations. This link to the well-researched clinical area of alexithymia may also be informative for understanding the etiology of low Strategic EI and for intervention or coaching programs aimed at the low end of the spectrum.

Similar issues arise from the links of STEU and STEM scores to anxiety and stress. If the STEU and STEM are conceptualized as knowledge (i.e., the STEU assesses whether people know what causes emotions, and the STEM assesses whether people know what to do when these emotional situations occur), then the link with anxiety and stress has implications for cognitive behavior therapy (CBT) approaches. The theoretical scoring key for the STEU makes the development of CBT approaches particularly straightforward. With a set of rules as the basis for assessment, and knowledge of these rules the operational definition of the construct, teaching these rules (and their application) ought to be a particularly useful exercise for increasing emotional understanding and (possibly) decreasing anxiety.

The prediction of academic achievement beyond the effects of intelligence and personality also links to potential applications, implicating EI as a possible component in college success. Both the STEU and STEM incrementally predicted students' psychology grades, and the STEU also incrementally predicted students' overall grades (which may constitute a less reliable measure than psychology grades, because students complete different combination of subjects with different degrees of difficulty, such that WAMs are not always comparable). This finding replicates prior research indicating that Understanding is the strongest predictor of academic success (among the EI branches). Such a result has applications for introducing social and emotional learning as a component within college-readiness courses.

New Paradigms for Test Development and Scoring

Two novel paradigms for test development in EI were used: the SJT approach and a standards-based approach grounded in Roseman's (2001) theory. Tests derived from these two different methods converged quite strongly. There are different advantages and disadvantages for each of the two methods. The SJT approach used to develop the STEM is time- and resource-consuming (three separate samples are used, and the time taken to analyze the text output is substantial). However, the test is also ecologically valid, reliable across samples, and may be used for content areas (such as Management) that do not yet have psychological theories that allow clear specifiable standards for the correct answer.

In cases in which tests can be scored according to specifiable standards for the right answer (as in the STEU), this is clearly an advantage over expertise-related scoring in terms of EI meeting J. D. Mayer et al.'s (2000) conceptual criteria for an intelligence (that intelligence reflects *mental performance* rather than dispositional qualities). However, the theory-based STEU was less reliable than was the STEM, particularly when given unproctored to a nonstudent sample. Nevertheless, scoring according to a theory-driven key unambiguously and precisely defines an a priori correct answer, allowing clear verbal specification of the construct captured by such test items. This is important both for construct validity and for the development of training programs based on the operational definition of the constructs. That is, if assessment is based on a set of lawful rules, teaching people the rules behind a scoring key forms the logical basis of the training. Schmit (2006) described the development of EI training programs in detail, beginning with the trainees learning sets of rules or general principles. Once such a set of rules is learned, trainees would practice applying these rules in workplace simulations.

Effects of Different Scoring Types and Response Formats

Some relationships were clearly affected by response format: Multiple-choice EI tests correlated more strongly with intelligence, irrespective of the construct measured (Understanding or Management). Exactly why multiple-choice EI tests correlate more strongly with intelligence is not clear, but the answer may relate to the conceptual difference between selecting the best option versus judging a variety of options. Intelligence tests are typically more similar to convergent rather than to divergent thinking, in that there is generally one correct answer rather than a number of acceptable possibilities. Multiple-choice tests would seem to involve convergent thinking (the test taker must select one best answer), whereas rate-the-extent tests seem more similar to divergent thinking (although test-takers do not generate multiple options, they must simultaneously consider several equally good options). In addition, gaining a high score on a rate-the-extent item involves knowing which options are poor as well as which option is the best or most appropriate. For example, answering a rate-the-extent Management item might involve differentiating between very effective responses, mildly effective responses, benign but ineffective responses, and highly destructive ineffective responses—possibly a different cognitive process from simply selecting the best response.

Although the mechanism is not entirely clear at this stage, differences in the patterns of correlates for the rate-the-extent versus multiple-choice STEM scores do imply that the underlying cognitive processes may be different for different formats. This has important implications for interpreting much of the prior research on EI, because the MSCEIT assesses Understanding with multiple-choice items only and assesses Management with rate-the-extent items only. Relationships between the branches of EI, and between EI branches and external correlates (such as intelligence), might conceivably show slightly different patterns depending on which item format is used.

Limitations and Future Directions

This article represents a first step within the larger goal of developing alternative measures of EI, which in turn would provide an empirical underpinning for adjustments to current theoretical models. As such, there are several future steps that would be useful for the evaluation and further development of empirical and theoretical paradigms. An obvious empirical "next step" would be to consider the relationship of the STEU and STEM with the full MSCEIT (rather than one subtest from the MEIS), informing test validity evidence as well as providing further information on the response format method effect. In addition, factor analysis to determine dimensionality of the STEU and STEM, and a comparison of the STEU and STEM to the facets of personality (in addition to broad dimensions), would inform the validity argument. A video- or audio-based (rather than text-based) presentation of items would also be useful to determine whether relationships of EI to intelligence (and especially the relationship to vocabulary

observed in this article) are due to the cognitive processing of emotional information rather than to the verbal ability required to comprehend text-based items.

Summary and Conclusions

This article demonstrates that some new ideas about test development, including standards-based scoring and the SJT approach, are possible to operationalize and can result in the valid and useful measurement of EI. New approaches to the assessment of EI are both possible and necessary, as a diversity of methods ensures that researchers can accurately interpret research findings as method-related or construct-related, illuminating the prior canon of research. In addition to such methodological issues, the current research has important implications for some practical applications of these new tools and methods in clinical psychology, in education, and in the development of intervention and training programs.

References

Ackerman, P. L., & Heggestad, E. D. (1997). Intelligence, personality and interests: Evidence for overlapping traits. *Psychology Bulletin, 121,* 219–245.

Australian Bureau of Statistics. Australia Now. Retrieved 11 July 2005 from http://www.abs.gov.au/ausstats/abs@nsf/2.6.1

Bagby, R. M., Parker, J. D. A., & Taylor, G. J. (1994). The twenty-item Toronto Alexithymia Scale: I. Item selection and cross-validation of the factor structure. *Journal of Psychosomatic Research, 38,* 23–32.

Barchard, K. A. (2003). Does emotional intelligence assist in the prediction of academic success? *Educational and Psychological Measurement, 63,* 840–858.

Barchard, K. A., & Hakstian, R. A. (2004). The nature and measurement of emotional intelligence abilities: Basic dimensions and their relationships with other cognitive ability and personality variables. *Educational and Psychological Measurement, 64,* 437–462.

Bastian, V. A. (2005). *Are the claims for emotional intelligence justified? Emotional intelligence predicts life skills, but not as well as personality and cognitive abilities.* Unpublished doctoral dissertation, Adelaide University, Adelaide, South Australia, Australia.

Bastian, V. A., Burns, N. R., & Nettelbeck, T. (2005). Emotional intelligence predicts life skills, but not as well as personality and cognitive abilities. *Personality and Individual Differences, 39,* 1135–1145.

Brackett, M. A., & Salovey, P. (2006). Measuring emotional intelligence with the Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT). *Psicothema, 18*(suppl.), 34–41.

Ciarrochi, J., Scott, G., Deane, F. P., & Heaven, P. C. L. (2003). Relations between social and emotional competence and mental health: A construct validation study. *Personality and Individual Differences, 35,* 1947–1963.

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

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction With Life Scale. *Journal of Personality Assessment, 49,* 71–75.

Feingold, A. (1994). Gender differences in personality: A meta-analysis. *Psychological Bulletin, 116,* 429–456.

Freudenthaler, H. H., & Neubauer, A. C. (2005). Emotional intelligence: The convergent and discriminant validities of intra- and interpersonal emotional abilities. *Personality and Individual Differences, 39,* 569–579.

Gohm, C. L., Corser, G. C., & Dalsky, D. J. (2005). Emotional intelligence under stress: Useful, unnecessary, or irrelevant? *Personality and Individual Differences, 39,* 1017–1028.

Goldenberg, I., Matheson, K., & Mantler, J. (2006). The assessment of emotional intelligence: A comparison of performance-based and self-report methodologies. *Journal of Personality Assessment, 86,* 33–45.

Hakstian, R. A., & Whalen, T. E. (1976). A k-sample significance test for independent alpha coefficients. *Psychometrika, 41,* 219–231.

Jensen, A. R. (1980). *Bias in mental testing.* New York: Free Press.

Landy, F. J. (2006). The long, frustrating, and fruitless search for social intelligence: A cautionary tale. In K. R. Murphy (Ed.), *A critique of emotional intelligence: What are the problems and how can they be fixed?* (pp. 81–123). Mahwah, NJ: Erlbaum.

Lovibond, P. F., & Lovibond, S. H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety inventories. *Behaviour Research and Therapy, 33,* 335–343.

MacCann, C., Roberts, R. D., Matthews, G., & Zeidner, M. (2004). Consensus scoring and empirical option weighting of performance-based emotional intelligence (EI) tests. *Personality and Individual Differences, 36,* 645–662.

Matthews, G., Emo, A. K., Funke, G., Zeidner, M., Roberts, R. D., Costa, P. T., Jr., et al. (2006). Emotional intelligence, personality, and task-induced stress. *Journal of Experimental Psychology: Applied, 12,* 69–107.

Matthews, G., Zeidner, M., & Roberts, R. D. (2002). *Emotional intelligence: Science and myth.* Boston: MIT Press.

Mayer, J., Roberts, R. D., & Barsade, S. G. (2008). Human abilities: Emotional intelligence. *Annual Review of Psychology, 59,* 507–536.

Mayer, J. D., Caruso, D. R., & Salovey, P. (2000). Emotional intelligence meets traditional standards for an intelligence. *Intelligence, 27,* 267–298.

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

Mayer, J. D., Salovey, P., Caruso, D. R., & Sitarenios, G. (2003). Measuring emotional intelligence with the MSCEIT V2.0. *Emotion, 3,* 97–105.

McDonald, R. P. (1999). *Test theory: A unified treatment.* Mahwah, NJ: Erlbaum.

Montebarocci, O., Codispoti, M., Baldaro, B., & Rossi, N. (2004). Adult attachment style and alexithymia. *Personality and Individual Differences, 36,* 499–507.

Orchard, B., MacCann, C., Schulze, R. Matthews, G., Zeidner, M., & Roberts, R. D. (in press). New directions and alternative approaches to the measurement of emotional intelligence. In C. Stough, D. Saklofske, & J. D. A. Parker (Eds.), *Advances in the measurement of emotional intelligence.* New York: Springer.

Parker, J. D. A., Summerfeldt, L. J., Hogan, M. J., & Majeski, S. A. (2004). Emotional intelligence and academic success: Examining the transition from high school to university. *Personality and Individual Differences, 38,* 163–172.

Roberts, R. D. (2000). *Openness Conscientiousness Extraversion Agreeableness Neuroticism Index Condensed (OCEANIC) Test manual.* Sydney, Australia: E-ntelligence Testing Products.

Roberts, R. D., Schulze, R., & MacCann, C. (2008). The measurement of emotional intelligence: A decade of progress? In G. J. Boyle (Ed.), *Handbook of personality.* New York: Sage.

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.

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

Schmit, M., J. (2006). EI in the business world. In K. R. Murphy (Ed.), *A critique of emotional intelligence: What are the problems and how can they be fixed?* (pp. 211–234). Mahwah, NJ: Erlbaum.

Schulze, R., & Roberts, R. D. (2006). Assessing the Big Five: Development and validation of the Openness Conscientiousness Extraversion Agreeableness Neuroticism Index Condensed (OCEANIC). *Zeitschrift fur Psychologie, 214,* 133–149.

Sifneos, P. E. (1973). The prevalence of "alexithymic" characteristics in psychosomatic patients. *Psychotherapy and Psychosomatics, 22,* 255–262.

Stankov, L. (1997). *Gf–Gc quickie test battery*. Unpublished manuscript, Sydney, Australia.

Warwick, J., & Nettelbeck, T. (2004). Emotional intelligence is? *Personality and Individual Differences, 37,* 1091–1100.

Wasserman, J. D., & Bracken, B. A. (2003). Psychometric characteristics of assessment procedures. In J. R. Graham & J. A. Naglieri (Eds.), *Handbook of psychology: Vol. 10. Assessment psychology* (pp. 43–66). New York: Wiley.

Woitaszewski, S. A., & Aalsma, M. C. (2004). The contribution of emotional intelligence to the social and academic success of gifted adolescents as measured by the Multifactor Emotional Intelligence Scale: Adolescent Version. *Roeper Review, 27,* 25–30.

Received July 31, 2007
Revision received December 19, 2007
Accepted May 15, 2008 ■

Correction to Tabibnia, Lieberman, and Craske (2008)

In the article, "The Lasting Effect of Words on Feelings: Words May Facilitate Exposure Effects to Threatening Images" by Golnaz Tabibnia, Matthew D. Lieberman, and Michelle G. Craske *(Emotion*, 2008, Vol. 8, No. 3, pp. 307–317), the URL provided for the supplemental materials was incorrect. The correct URL is given below. The printer apologizes for the error.

Supplemental materials: http://dx.doi.org/10.1037/1528-3542.8.3.307.supp

DOI: 10.1037/1528-3542.8.4.551