Emotion
2001, Vol. 1, No. 3, 196–231

1528-3542/01/$5.00 DOI: 10.1037//1528-3542.1.3.196

Does Emotional Intelligence Meet Traditional Standards for an Intelligence? Some New Data and Conclusions

Richard D. Roberts
University of Sydney

Moshe Zeidner
University of Haifa

Gerald Matthews
University of Cincinnati

Performance-based measures of emotional intelligence (EI) are more likely than measures based on self-report to assess EI as a construct distinct from personality. A multivariate investigation was conducted with the performance-based, Multi-Factor Emotional Intelligence Scale (MEIS; J. D. Mayer, D. Caruso, & P. Salovey, 1999). Participants ($N = 704$) also completed the Trait Self-Description Inventory (TSDI, a measure of the Big Five personality factors; Christal, 1994; R. D. Roberts et al.), and the Armed Services Vocational Aptitude Battery (ASVAB, a measure of intelligence). Results were equivocal. Although the MEIS showed convergent validity (correlating moderately with the ASVAB) and divergent validity (correlating minimally with the TSDI), different scoring protocols (i.e., expert and consensus) yielded contradictory findings. Analyses of factor structure and subscale reliability identified further measurement problems. Overall, it is questionable whether the MEIS operationalizes EI as a reliable and valid construct.

Emotional intelligence (EI) is a relatively new domain of psychological investigation, having recently gathered considerable momentum with widespread, international media attention. Daniel Goleman's (1995) book on the topic appeared on *The New York Times* best-seller list, which led to a *Time* magazine article devoted to detailed exposition of the topic (Gibbs, 1995). More recently, the influential electronic magazine *Salon* devoted a lengthy article to the discussion of its application in the workforce (Paul, 1999). Clearly, this attention was inspired by a veritable plethora of trade texts (and Web sites) dealing with self-help and management practices, assessment, and organization-based applications implicit to the concept of EI (see, e.g., Abraham, 2000; Bar-On, 1997, 2000; Bar-On, Brown, Kirkcaldy, & Thome, 2000; Cooper & Sawaf, 1997; Epstein, 1998; Ryback, 1998; Saarni, 1999; Weisinger, 1998).

EI as a concept has prospered, in part, because of the increasing personal importance of emotion management for individuals in modern society. Indeed, researchers have commonly claimed that EI predicts important educational and occupational criteria beyond that predicted by general intellectual ability (e.g., Elias & Weissberg, 2000; Fisher & Ashkanasy, 2000; Fox & Spector, 2000; Goleman, 1995; Mehrabian, 2000 Saarni, 1999; Scherer, 1997). Furthermore, EI's chief proponents appear to have made strides toward understanding its nature, components, determinants, effects, developmental track, and modes of modification (see Matthews, Zeidner, & Roberts, in press, for a critical review).

EI first appeared in the scientific literature in the early 1990s (Mayer, DiPaulo, & Salovey, 1990; Sa-

Richard D. Roberts, Department of Psychology, University of Sydney, Sydney, Australia; Moshe Zeidner, Department of Psychology, University of Haifa, Mount Carmel, Israel; Gerald Matthews, Department of Psychology, University of Cincinnati.

The data gathered in this investigation were collected by personnel of the (now defunct) Human Effectiveness Directorate of the United States Air Force Research Laboratory, Brooks Air Force Base, Texas. We acknowledge the support of this institution. We thank Janice Heresford, Ralph Sayler, Rich Walker, and Mary Lawson for their efforts in data collection and the scoring of test protocols. We also thank Lazar Stankov and Gerry Pallier.

Correspondence concerning this article should be addressed to Moshe Zeidner, Center for Interdisciplinary Research of Emotions, University of Haifa, Mount Carmel 31905 Israel. Electronic mail may be sent to zeidner@research.haifa.ac.il.

lovey & Mayer, 1990), where the term was used to denote a type of intelligence that involved the ability to process emotional information. Subsequently, researchers have proposed that EI incorporates a set of conceptually related psychological processes involving the processing of affective information. These processes include: (a) the verbal and nonverbal appraisal and expression of emotion in the self and others, (b) the regulation of emotion in the self and others, and (c) the utilization of emotion to facilitate thought (see Mayer & Geher, 1996; Mayer & Salovey, 1997; Salovey & Mayer, 1990). Although various authors have proposed that EI is a type of intelligence, in the traditional sense, contemporary research and theory lacks any clear conceptual model of intelligence within which to place the construct. For example, Spearman's (1927) model of *g* (general ability) affords no special role for EI. Neither is emotional (or social, for that matter) intelligence included in Thurstone's (1938) list of primary mental abilities or Guttman's (1965a, 1965b) radex model of intelligence.

Although EI has captured the public's imagination during the past 5 years, the concept's origins trace back to a number of constructs emanating from traditional psychometric models of intelligence. We now briefly examine the construct's historical roots and conceptual linkages, pointing to similarities and differences between EI and a variety of cognate constructs (i.e., social intelligence, crystallized ability, behavioral cognition, personal intelligence) identified within influential theories of intelligence.

EI Within the Context of Intelligence Theory and Assessment

Intelligence: Overview

General intelligence refers to a person's overall capacity for adaptation through effective cognition and information processing. It may be seen as a general competence of the mind (mental ability) or of higher order faculties such as understanding, reasoning, problem solving, and learning, especially of complex, structured material (cognitive ability; Brody, 1992). However, the concept of general intelligence says little about the more specific competencies that comprise it. Thus, psychologists have sought to partition the domain of intelligence into more manageable chunks, including less narrow (but still broad) categories of abilities (e.g., crystallized intelligence [Gc]) or more specific abilities (e.g., verbal comprehension). These various levels of conceptualization have led to taxonomic models, which have recently been synthesized inside Carroll's (1993) three-stratum model. Carroll found, after reanalysis of virtually all data sets collected in the 20th century, a hierarchy of structures. At the lowest level of the hierarchy there are around 70 fairly narrow primary abilities. Correlations between primary abilities allow them to be clustered together to define eight broad abilities at the second stratum, and these broad abilities cluster to define general intelligence at the third stratum.

The identification of general intelligence through psychometric studies has stimulated over a century of debate on the nature of intelligence and the proper way to assess this construct. As the American Psychological Association's (APA's) Task Force on Intelligence (APA Public Affairs Office, 1997) has stated, it is generally (though not universally) agreed that the conventional psychometric approach has successfully identified a reliable quality of the individual that predicts important real-world criteria. The conventional psychometric approach is seen as the most influential and the most systematically researched, although other conceptions of intelligence also have much to offer. The intelligence literature also contains criticisms of the notion that there is a consensual definition of intelligence shared by most psychologists (see various chapters in Sternberg, 2000b), especially in view of cultural differences in conceptions of intelligence (see Sternberg, 2000a). The importance of conventional, cognitive intelligence has been challenged by recent suggestions that there are many different kinds of intelligence (e.g., Gardner, 1983). As these include abilities such as musical intelligence, it is difficult to assume that the same criterion for inclusion holds true for all intelligence constructs. Thus, many psychometricians would concur that the defining attribute of a cognitive ability test is that there is one correct answer based on logical, empirical, semantic, or even normative criteria (Guttman, 1965a, 1965b; Nunnally, 1978; Zeidner & Feitelson, 1989). Equally important, however, the psychometric criteria developed in studies of cognitive ability may not be applicable to other domains of intelligence, such as managing emotion.

Having discussed some of these issues, we now turn to examine conceptual linkages between EI and a variety of related constructs identified within disparate models of intelligence.

EI and social intelligence. Many commentators suppose that EI derives from the broader construct of social intelligence (e.g., Bar-On, 2000; Gardner, 1983; Goleman, 1995). Contemporary individual-differences perspectives on the construct of social intelligence have their origins in Thorndike's (1920) influential, tripartite division of intelligence into the following broad classes: (a) *abstract–scholastic intelligence*—the ability to understand and manage ideas, (b) *mechanical–visuospatial intelligence*—the ability to understand and manipulate concrete objects; and (c) *social (practical) intelligence*—the ability to understand and manage people and act wisely in social contexts. Thorndike's abstract definition of social intelligence as wisdom in social contexts was translated quickly into standardized instruments for measuring individual differences in this construct.

In the 1930s, the study of social intelligence was largely a study of how people make judgments regarding others and the accuracy of such social judgments. By the 1950s, however, this body of work had polarized to form two distinct traditions: (a) an intelligence tradition, which was interested in the abilities of person perception, and (b) a social–psychological tradition, which focused on the social determinants of person perception. In recent times, there has been growing convergence between these distinctive domains. Thus, researchers from the domain of individual differences have become more interested in social facets of ability, and social psychologists have become more interested in cognitive determinants of perception (Mayer & Geher, 1996).

Despite considerable interest and numerous attempts to define and measure social intelligence over the past eight decades, these attempts have proved problematic (see Kihlstrom & Cantor, 2000, for a review of attempts to conceptualize and measure this construct). Although defining social intelligence seemed easy enough, the measurement of the construct proved to be an almost insurmountable task. Thus, social intelligence has been studied less than other forms of intelligence because it seems the hardest of the three broad classes of intelligence to distinguish from the others, both theoretically and empirically (Mayer & Geher, 1996). The inability to discriminate between general and social intelligence, coupled with difficulties in selecting external criteria against which to validate experimental scales, led to a decline in research focusing on social intelligence as a distinct intellectual entity, until the recent upsurge of interest in EI.

EI and the behavioral facet of the structure of intellect model. EI involves the processing of both information that refers directly to emotion (e.g., one's own mood) and information on behaviors that have emotional connotations (e.g., violent behaviors). Intelligence in understanding behaviors and their significance already appears in Guilford's (1959) structure of intellect model. Guilford postulated a facet model of intelligence, on the basis of all possible combinations of three major facets: (a) operations (i.e., cognition, memory, divergent production, convergent production, and evaluation); (b) content (i.e., figural, semantic, symbolic, and behavioral); and (c) products (i.e., units, classes, relations, systems, transformation, and implications).

Figural, semantic, and symbolic content largely corresponds to the abstract material contained in standard intelligence tests. However, the behavioral domain is neglected in conventional tests and seems to correspond to social–emotional intelligence. In particular, EI overlaps with the cognition of behavioral content (e.g., ability to identify internal status of individuals, interpretation of consequences of social behavior, etc). In fact, the test items designed to gauge behavioral cognition, constructed by Guilford's team (e.g., O'Sullivan, Guilford, & de Mille, 1965), are reminiscent of current behavioral measures of EI. Furthermore, references to *empathic ability* (whose definition parallels that of major facets of EI remarkably closely) can be found in some of Guilford's (1959) earliest writings on the structure of intellect model.[1]

EI and crystallized intelligence. The theory of fluid (Gf) and crystallized (Gc) intelligence proposed by Cattell (1971), Horn (1988), and their associates (see Horn & Noll, 1994; Horn & Stankov, 1982) is arguably the most efficacious empirically based psychometric model of intelligence (see Stankov, Boyle, & Cattell, 1995). Researchers have speculated that, within this theory, EI will constitute an additional aspect of (possibly one or more primary underlying mental abilities) Gc. This assertion is based on the assumption that the appraisal, expression, regulation, and utilization of emotion develops through experience and social interaction in much the same way as

[1] As testament to the preceding assertion consider the following question: "[E]fforts are being made to bring within the sphere of psychological measurement what is often called 'empathic ability.' This is an assumed ability to know psychological dispositions of other persons—their perceptions, thoughts, feelings, and attitudes, as well as their traits. . . . The understanding of such abilities would be of utmost importance to all those who deal directly with people in any professional way—politicians and teachers as well as psychiatrists, psychologists, and social workers" (Guilford, 1959, p. 395).

other psychological processes comprising Gc (see Davies, Stankov, & Roberts, 1998).

EI and personal intelligence. The concept of EI strongly overlaps with Gardner's (1983) notion of social intelligence, which he referred to as a type of *personal intelligence.* Indeed, part of Gardner's definition focuses specifically on the processing of affective information. Current conceptualization of EI (e.g., Mayer, Salovey, & Caruso, 2000a) focuses on one's ability to accurately identify, appraise, and discriminate among emotions in oneself and others, understand emotions, assimilate emotions in thought, and regulate both positive and negative emotions in self and others. This conceptualization encompasses the following subtypes of personal intelligence described by Gardner (1983) within his theory of multiple intelligence: (a) *intrapersonal intelligence*—the ability to access one's own feeling life; to identify, label, and discriminate among one's feelings; and to represent them symbolically; and (b) *interpersonal intelligence*—the ability to discern the moods, intentions, and desires of others. Thus, whereas intrapersonal intelligence refers to the person's ability to gain access to his or her own internal emotional life, interpersonal intelligence represents the individual's ability to understand other people, to know what they feel, and to notice and make distinctions among other individuals. In sum, the current definition and conceptualization of EI, as a cognitive ability, overlaps considerably with Gardner's notion of personal intelligence, subsuming both intrapersonal and interpersonal forms of intelligence.

In attempting to locate these intelligences within the traditional psychometric domain, Carroll (1993) suggested that interpersonal intelligence is a specialized type of acquired knowledge (i.e., Gc). However, Gardner's intrapersonal intelligence—access to one's own feelings—finds no counterpart in Carroll's taxonomic model. However, it may be argued that this situation has arisen because adequate assessment of this type of intelligence has never appeared in the extant factor-analytic literature.

Conceptualizing and Assessing EI

Models of EI

One of the difficulties currently encountered in research on EI would appear to be the multitude of qualities covered by the concept (see Roberts, in press). Indeed, many qualities appear to overlap with well-established personality constructs, such as the Big Five personality factor model (see Davies et al., 1998; McCrae, 2000). Mayer, Caruso, and Salovey (1999, 2000) warned that careful analysis is required to distinguish what is (and what is not) part of EI (see also Mayer, Salovey, & Caruso, 2000a, 2000b). Throughout, Mayer and colleagues distinguished between (a) *mental ability models,* focusing on aptitude for processing affective information, and (b) *mixed models* that conceptualize EI as a diverse construct, including aspects of personality as well as the ability to perceive, assimilate, understand, and manage emotions. These mixed models include motivational factors and affective dispositions (e.g., self-concept, assertiveness, empathy; see Bar-On, 1997; Goleman, 1995).

In contrast, Mayer and colleagues have proposed a four-branch mental ability model of EI, which encompasses the following psychological processes (see e.g., Mayer, Caruso, & Salovey, 1999, 2000; Mayer & Salovey, 1997; Mayer, Salovey, & Caruso, 2000a, 2000b; Salovey & Mayer, 1990):

1. *The verbal and nonverbal appraisal and expression of emotion in the self and others.* EI has been defined as "the ability to perceive emotions, to access and generate emotions so as to assist thought, to understand emotions and emotional knowledge, and to reflectively regulate emotions so as to promote emotional and intellectual growth" (Mayer & Salovey, 1997, p.5). Inside this definitional framework, the most fundamental level of EI includes the perception, appraisal, and expression of emotions (Mayer, Caruso, & Salovey, 1999). In other words, implicit in this aspect of EI is the individual's awareness of both their emotions and their thoughts concerning their emotions, the ability to monitor and differentiate among emotions, and the ability to adequately express emotions.
2. *The utilization of emotion to facilitate thought and action.* This component of EI involves assimilating basic emotional experiences into mental life (Mayer, Caruso, & Salovey, 1999, 2000). This includes weighing emotions against one another and against other sensations and thoughts and allowing emotions to direct attention (e.g., holding an emotional state in consciousness long enough to compare its correspondence with similar sensations in sound, color, and taste). Marshaling emotions in the service of a goal is essential for selective attention, self-monitoring, self-motivation, and so forth.
3. *Understanding and reasoning about emotions.*

This aspect of EI involves perceiving the lawfulness underlying specific emotions (e.g., to understand that anger arises when justice is denied or when an injustice is performed against oneself or one's loved ones). This process also involves the understanding of emotional problems, such as knowing what emotions are similar and what relation they convey.

4. *The regulation of emotion in the self and others.* According to Mayer, Caruso, & Salovey (1999), the *highest level* in the hierarchy of EI skills is the management and regulation of emotions. This facet of EI involves knowing how to calm down after feeling stressed out or alleviating the stress and emotion of others. This facet facilitates social adaptation and problem solving.

The Assessment of EI: Self-Report and Performance Approaches

Although several measures have been (or are currently being) designed for the assessment of EI, it remains uncertain whether there is anything about EI that psychologists working within the fields of personality, intelligence, and applied psychological research do not already know. Moreover, the increased media attention and the vast number of trade texts devoted to the topic of EI often subsume findings from these fields in a faddish way rather than deal directly with the topic as defined by its chief exponents. In short, like many psychological constructs, EI is often loosely defined in the literature, causing considerable confusion among researchers in the field.

Nevertheless, since the term first appeared, there has been a rapid propagation of measures of EI (for a review, see Ciarrochi, Chan, Caputi, & Roberts, 2001). Popular measures of EI include the Bar-On Emotional Quotient Inventory (Bar-On, 1997, 2000), the EQ Map Test (Cooper & Sawaf, 1997), the Schutte Self-Report Inventory (Schutte et al., 1998), the Trait Meta-Mood Scale (Salovey, Mayer, Goldman, Turvey, & Palfai, 1995), and the Multi-Factor Emotional Intelligence Scale (Mayer, Caruso, & Salovey, 1999).[2] The content of these EI measures varies as a function of the different theoretical conceptualizations and interpretations of EI appearing in the literature (Mayer, Salovey, & Caruso, 2000a, 2000b). However, many commentators classify indicators of EI according to whether they derive from self-reports of typical behaviors in everyday life, as opposed to objective performance in controlled experimental settings. A brief overview and critique of these two distinctive approaches to the assessment of EI follows.

Self-Report Measures of EI

Self-report measures have been designed to assess beliefs and perceptions about an individual's competencies in specific domains of EI (Salovey, Woolery, & Mayer, 2000). These indexes generally ask a person to endorse a series of descriptive statements, usually on some form of rating scale. For example, in the Schutte Self-Report Inventory (Schutte et al., 1998) individuals rate themselves on a scale from 1 *strongly disagree* to 5 *strongly agree* on 33 statements (e.g., "I know why my emotions change," "I expect good things to happen"). Self-report measures typically sample a diversity of constructs, and hence assume a mixed model of EI (i.e., as both ability and personality trait), in Mayer, Caruso, & Salovey's (e.g., 1999, 2000) terminology.

A number of problems and serious omissions currently plague the research on EI that uses self-report methodologies (cf. Petrides & Furnham, 2000). These self-report scales rely on a person's self-understanding; if the self-reports are inaccurate, these measures yield information concerning only the person's self-perception (rather than his or her actual level) of EI. Self-perceptions may not be particularly accurate or even available to conscious interpretation, being vulnerable to the entire gamut of response sets and social desirability factors afflicting self-report measures, as well as deception and impression management. These problems are, of course, common to all scales based on self-report, including personality assessment. To counteract this criticism in other fields where self-reports are used, researchers have devised a number of procedures, including comparing self-assessed responses with reports provided by a respondent's peers (see, e.g., Costa & McCrae, 1992; Pavot,

[2] A number of measures of psychological constructs developed before EI gained widespread notoriety have also been used as proxies in the assessment of EI. These include instruments designed to assess alexithymia (e.g., Toronto Alexithymia Scale; Bagby, Parker, & Taylor, 1994), empathy (e.g., Questionnaire Measure of Emotional Empathy; Mehrabian & Epstein, 1970), emotional control (e.g., Emotional Control Questionnaire; Roger & Najarian, 1989), and defensive avoidance (e.g., repression–sensitization scale; Weinberger, Schwarz, & Davidson, 1979).

Diener, & Suh, 1998; Stoeber, 1998).[3] However, validation studies of this type appear not to have been conducted with respect to self-report measures of EI. Hence, whether extant scales are free from response biases and social desirability effects remains an open, empirical question in urgent need of detailed investigation.[4]

This issue notwithstanding, it is questionable whether items asking participants to self-appraise intellectual ability (e.g., "I am an extremely intelligent student") would make for a valid measure of general intelligence. Under the assumption that EI constitutes a traditional form of intelligence, the usefulness of analogous items about one's EI seems doubtful (Salovey et al., 2000). Note that past research has reported rather modest associations between self-rated and actual ability measures, with self-report accounting for less than 10% of intelligence-score variance. Thus, a meta-analytic review of 55 studies by Mabe and West (1982) yielded a mean correlation (validity coefficient of self-rating) of .34 between self-evaluations of intelligence and objective intelligence test scores. More recent studies (see, e.g., Paulhus, Lysy, & Yik, 1998) have concurred that the correlations between self-reports of intelligence and mental test performance tend to be rather modest (about $r = .30$).

Finally, tests of EI that assess noncognitive traits (e.g., assertiveness, optimism, impulse control) seem to be tapping dimensions of individual differences that are entirely different from contemporary notions of what constitutes intelligence (Davies et al., 1998). Indeed, the information derived from these instruments appears more pertinent to constructs comprising existing personality models (see McCrae, 2000). Empirical data pointing to the substantial relationship between EI and existing personality measures have, curiously, actually been used in support of the discriminant validity and conceptual soundness of EI (see, e.g., Bar-On, 2000). For example, a recent study by Dawda and Hart (2000) revealed *average* correlations approaching .50 between measures of the Big Five personality factors (i.e., Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness) and general EI derived from the Bar-On (1997) Emotional Quotient Inventory (see Table 7, p. 807). Noting the relative independence of each of the Big Five personality factors (e.g., Costa & McCrae, 1992), we believe these data suggest that the Bar-On Emotional Quotient Inventory is nothing but a proxy measure of a composite of Big Five personality constructs, weighted most strongly toward low neuroticism.

Performance-Based EI Measures

In view of the foregoing problems associated with the use of self-report measures, several authors have advocated the development of more objective, ability-based indicators of EI (e.g., Mayer, Caruso, & Salovey, 1999, 2000; Mayer & Salovey, 1997; Mayer, Salovey, & Caruso, 2000a, 2000b). According to these authors, ability testing is the "gold standard" in intelligence research, because intelligence refers to the actual capacity to perform well at mental problems—not just one's beliefs about such capacities (see also Carroll, 1993). Under this framework, a psychological instrument directly measures ability by having a person solve a problem (e.g., identify the emotion in a person's face, story, or painting). In addition, the examinee's answer should be available for evaluation against accuracy criteria (Mayer & Geher, 1996). Consequently, task-based measures engage participants in exercises designed to assess competencies supporting EI skills. The ability-based mode of assessment proposed by Mayer and Salovey (1997) and its underlying four-branch conceptual model of EI, has gained currency, largely because it appears ability-oriented and empirically based. Their four-branch model, described above, is currently operationalized through the Multi-Factor Emotional Intelligence Scale (MEIS; Mayer, Caruso, & Salovey, 1999), and the recently developed Mayer–Salovey–Caruso Emotional Intelligence Test (MSCEIT; Mayer, Caruso, & Salovey, 2000).

As further discussed below, there is considerable difficulty in determining objectively correct responses to stimuli involving emotional content and in applying

[3] It is worth noting that this is considered by researchers working within the field of personality as a measure of consensus. Given the consensus scoring procedure alluded to previously (and described shortly in some depth for performance-based assessment of EI), an intriguing conceptual question may be posed: To what extent are these two forms of consensus empirically equivalent?

[4] It is noted that peer assessment may not be a foolproof remedy to self-report biases. It may be subject to its own particular biases, such as the nature of the situations in which the peer encounters the person rated. One of the issues frequently encountered is the problem of peer agreement, which for many traits is rather low. The additional problem for emotional reaction is that it often may not be visible, even to peers.

truly veridical criteria in scoring tasks of emotional ability. Proponents of EI ability measures have thus promoted three alternative scoring procedures to discriminate right from wrong answers on ability tests of EI (Mayer, Caruso, & Salovey, 1999), which are described as follows:

Consensus scoring. An examinee receives credit for endorsing responses that the group endorses. Thus, if the group agrees that a face (or design, passage of music, etc.) conveys a *happy* or *sad* emotion, then that becomes the correct response. This approach assumes that observations for a large number of people can be pooled and can serve as reliable measures.

Expert scoring. Experts in the field of emotions (e.g., psychologists, psychiatrists, philosophers, and so forth) examine certain stimuli (e.g., a face, passage of music, or design) and then use their best judgment to determine the emotion expressed in that stimulus. Presumably, the expert brings professional know-how (along with a history of behavioral knowledge) to bear on judgments about emotional meanings. However, researchers have argued that an expert's assessment may be no more than a reliable indicator of the group consensus, albeit a particularly sensitive one (Legree, 1995). The test taker receives credit for ratings that correspond to those used by the experts.

Target scoring. A judge (i.e., the test taker) assesses what a target (artist, photographer, musician, and so forth) is portraying at the time the target individual is engaged in some emotional activity (e.g., writing a poem, playing a musical score, painting, sculpting, photographing a picture, etc.). A series of emotion-rating scales is then used to match the emotions conveyed by the stimuli to those reported by the target. It is commonly held that the target has more information than is available to the outside observer (Bar-On, 1997; Mayer, Caruso, & Salovey, 1999, 2000; Mayer & Geher, 1996) and is used as the criterion for scoring judges' responses. Target scoring has received rather little attention in previous research, ostensibly because it is suitable only for emotion-identification tasks and not for other, higher level aspects of EI. Hence, we will not discuss target scoring at length in the current article, although it seems promising for measuring some aspects of EI and might be explored further.

Issues pertaining to the scoring of EI tests. The use of multiple scoring methods in objective assessment of EI contrasts with the scoring of conventional intelligence tests. The logic of facet-analytic thinking (see, e.g., Guttman & Levy, 1991; Most & Zeidner, 1995; Zeidner & Feitelson, 1989) is that the main criterion for an intelligence task is the application of a veridical criterion against which one judges a response as correct or incorrect. Often, intelligence test items are based on some formal, rule-bound system that indicated unequivocally whether an answer is correct. Various formal systems are used depending on item content, such as mathematics (numerical tests), logic (reasoning tests), geometry (spatial tests), and the semantics of language (verbal tests). It is also relatively straightforward to determine which individuals are expert in these areas and thus are professionally qualified to act as arbiters. In contrast, items used in early IQ tests that depended on subjective judgment, such as deciding which of several faces was most attractive, have been largely removed from tests, due, in part, to the risk of cultural bias. This is not to say that conventional intelligence testing is entirely free from scoring problems. An anonymous reviewer of this article pointed out that series completion problems such as "2, 4, 6, . . . ??" could be completed in any way;[5] use of the simplest rule (add 2) is arbitrary (but consensual). In addition, individual testing, especially of children, may require a judgment on the part of the tester as to whether a question has been correctly answered. Concerns also linger over the extent to which intelligence testing is truly culture-fair, despite efforts to remove obvious sources of cultural bias. Nevertheless, there is generally a clear rationale for justifying the correctness of an answer, and it is rare for well-informed people to dispute the correct answer to an item.

The assessment of EI as a mental ability depends on the presumption that answers to stimuli assessing various facets of feelings can be categorized as correct or incorrect (Mayer & Salovey, 1997). If this presumption is incorrect, no scoring method can meet the basic psychometric criterion for ability tests, namely, the existence of a true and unequivocal veridical standard against which to judge responses. In fact, the likelihood of there being a veridical standard depends on the nature of the EI test item. As with cognitive intelligence, items may refer to psychological processes at different levels of abstraction from raw-sense data. EI may, in principle, be assessed through lower order processes linked to sensation and perception, such as detecting the presence of an emotion in a face stimulus presented tachistoscopically or deciding that two words have similar valence. Alterna-

[5] We are grateful to two anonymous reviewers for comments on earlier versions of this article.

tively, EI test items may refer to higher order reasoning processes, such as choosing how to cope with a stressful enounter.

Mayer, Salovey, and Caruso (2000a) arranged the four branches in a hierarchy beginning with lower level or basic skills of perception and appraisal of emotion, and finishing, at the highest level, with synthetic skills for emotion management that integrate lower level skills. Basic skills appear to be those most open to objective assessment (although it is likely that perception and appraisal of emotion also involve high-level inference). For example, facial expression of emotion is sufficiently well understood (e.g., Ekman, 1999) in that objectively scored tests of identification of facial emotion may be feasible. In such a case, expert scoring seems appropriate and there is no place for consensus scoring. Conversely, items for tests of the managing emotions branch are more problematic. Certain emotional reactions may be assessed according to logically consistent criteria only by reference to personal and societal standards (Matthews & Zeidner, 2000; Matthews, Zeidner, & Roberts, in press). For example, what is the best or right response to being insulted or mocked by a coworker? Clearly, the best response would depend on the situation, the person's experience with insults, cultural norms, the individual's position in the status hierarchy, and so forth. Even within a single specified situation, it is often difficult to specify the best response—there are multiple criteria for adaptation that may conflict (e.g., preserving self-esteem, maintaining good relationships with others, advancing in one's career).

None of the scoring methods appear to be very satisfactory for higher level aspects of EI (which may be those most relevant to real-world functioning). Experts may be able to use psychological research to provide answers (as did Mayer, Caruso, & Salovey, 1999), but there are two fundamental limitations to expert knowledge in this area. First, research typically reveals only statistical rather than directly contingent relationships, for example, being mocked by a coworker typically (but not invariably) leads to anger. Second, there are multiple domains of expertise leading to conflicting viewpoints. If we present the question of how a child's emotional problems can best be managed to a cognitive therapist, an evolutionary psychologist, a psychoanalyst, a social worker, a high school teacher, and a gender studies professor, what is the probability that these experts will agree on a solution? (We might feel fortunate to find agreement between any two of the above.) The adequacy of consensus judgments is based on evolutionary and cultural foundations, where the consistency of emotionally signaled information appears paramount (Bar-On, 1997; Mayer, Caruso, & Salovey, 1999). Researchers have argued that the pooled responses of large normative samples is accurate (Legree, 1995), although more evidence is needed. Even if that is the case, there are serious concerns about bias in consensus judgment. Consensus may be influenced by nonveridical cultural beliefs, such as the traditional British belief that a "stiff upper lip" is always the best response to emotional problems. There are also concerns about the validity of consensus judgments that cross gender and cultural boundaries. The popular "Venus and Mars" view of gender relations is that men are good at understanding the emotions of other men but are inept at understanding women's feelings, and vice versa. In the worst case, consensus scoring may simply indicate the extent of agreement with cultural or gender-based prejudices.

If we are prepared to set such difficulties of scoring principles aside, perhaps we can proceed pragmatically, as Binet did in developing intelligence tests that would discriminate children of high academic potential. Testing EI may well be worthwhile if there is evidence that EI tests are reliable, in measuring some underlying quality accurately, and valid, in predicting relevant criteria better than other tests. Given that the MEIS is a new measure, it may be inappropriate to stifle research prematurely by applying overly-stringent criteria. However, it is essential that there is convergence between different scoring methods, or the construct may be judged as unreliable. Mayer, Caruso, and Salovey (1999) pointed out that as the different criteria represent different perspectives, it is unlikely that they would be in complete agreement. These authors go on to state that there should be a general "rough" convergence, which would substantiate the view that EI is, in fact, an intelligence. Unfortunately, it is unclear how high correlations should be to attain rough convergence or whether it is satisfactory for correlations to be substantial but considerably less than unity (e.g., in the range of 0.50–0.70). The pragmatic approach raises the issue of empirical findings that are based on the MEIS, which will be considered next.

EI: Empirical Findings

These theoretical issues notwithstanding, recent research by Mayer, Caruso, and Salovey (1999) suggests that state-of-the-art objective measures of EI meet the standards of validity and reliability expected of traditional cognitive-ability measures. Indeed, although the scientific study of EI has only recently begun, the scant empirical evidence available is contradictory. A brief examination of these conflicting results follows.

EI Measures: Positive Results

Mayer, Caruso, and Salovey (1999) have argued that standard criteria need to be met before any (new) form of intelligence can be considered to constitute a legitimate scientific domain. These authors have focused on the following three standards, which have been replicated many times in psychometric studies of intelligence (and its taxonomic structure) over the past century (see, e.g., Carroll, 1993; Cattell, 1971; Guttman & Levy, 1991; Horn & Hofer, 1992; Jensen, 1998):

1. An intelligence should be capable of reflecting "mental performance rather than preferred ways of behaving, or a person's self-esteem, or non-intellectual attainments" (Mayer, Caruso, & Salovey, 1999, pp. 269–270). In short, this so-called *conceptual criterion* asserts that the concept in question be operationalized as a set of abilities (in this case, emotion-related abilities) that have clearly defined performance components.
2. A (new) intelligence should meet prescribed correlational criteria. For example, tests for different aspects of such an intelligence should be positively intercorrelated. Measures of a new ability should be related to existing psychometric intelligence tests (specifically demonstrating the *positive manifold* phenomenon represented by a nonnegative matrix of correlation coefficients, as prescribed by Guttman's first law of intelligence (Guttman & Levy, 1991).[6]
3. Measures of intelligence should vary with experience and age.[7]

Researchers have claimed that available evidence supports the notion that EI meets all three criteria and so is a legitimate form of intelligence (Mayer & Cobb, 2000; Mayer & Salovey, 1993, 1997; Mayer, Salovey, & Caruso, 2000a, 2000b; Salovey et al., 2000). With respect to operationalization criteria, EI has been measured by a series of ability tasks on state-of-the-art instruments, such as the MEIS, and has been objectively scored by using consensus, expert, and (for some scales) target criteria. These criteria are claimed to converge (i.e., were positively correlated) to a satisfactory degree (Mayer, Salovey, & Caruso, 2000b). In the Mayer, Caruso, and Salovey (1999) data, correlations between consensus and expert test scores ranged from −.16 to .95, with half of the 12 correlations exceeding an *r* of .52. A median of .52 suggests the desired rough convergence, though it is questionable whether correlations of this magnitude are sufficient to establish a reliable common element to the two forms of scoring. Moreover, Mayer, Caruso, and Salovey (1999, 2000) asserted that the four-branch model has (more or less) been vindicated by a series of factor analyses, such that the component tests adhere to the stated performance model. Finally, subtests comprising the MEIS are generally claimed to exhibit satisfactory levels of internal consistency reliability (see also Ciarrochi, Chan, & Caputi, 2000).

In fulfilling the second criterion, which essentially captures major features of construct validation, measures of EI have been shown to have concurrent validity with cognate measures of EI, such as empathy, parental warmth, and emotional openness (Mayer, Caruso, & Salovey, 1999; Mayer & Geher, 1996), which serve as criteria for validity assessment. More important, consensus and target scores appear to correlate to a similar degree with selected outside criteria (e.g., empathy, self-reported Scholastic Assessment Test [SAT] scores, decreased emotional defensiveness) in student populations (Mayer & Geher, 1996), although comparability of consensus and expert scores as predictors has been neglected. Other evidence comes from studies using questionnaire measures of EI. For

[6] In interests of economy of expression, we use the term *positive manifold* throughout this article to refer to a nonnegative matrix of correlation coefficients. Strictly speaking, however, as one reviewer pointed out, positive manifold more correctly refers to an untested, mathematical hypothesis first put forward by Thurstone (1931).

[7] As one reviewer noted, Criteria 2 and 3, put forth by Mayer, Caruso, and Salovey (1999), may be problematic as criteria because these conditions may be construed as empirical findings rather than definitional features. This would appear a distinct possibility, especially with respect to Criterion 3, perhaps less so with respect to Criterion 2, as many intelligence researchers take positive manifold to be a lawful phenomenon of ability measures. Thus, for example, Guttman and Levy (1991) proposed, as the first law of intelligence, that positive intercorrelations of test items is an empirical law for a certain class of items. This issue aside, we do not wish to suggest that we wholeheartedly endorse any of these criteria (as evidenced later in our exposition). Rather we present them here as the three standards espoused by Mayer, Caruso, and Salovey (1999) for establishing that a new domain (such as EI) constitutes a form of intelligence.

example, this form of EI predicts first-year college students' success (Schutte et al., 1998). Self-reported EI is also negatively related to alexithymia (i.e., difficulties in identifying, evaluating, describing, and expressing feelings), as measured by the Toronto Alexithymia Scale (e.g., Schutte et al., 1998; Taylor, 2000).

However, arguably, the most important construct validation criterion is the extent to which EI overlaps with other intelligence(s). In their pioneering study, Mayer, Caruso, and Salovey (1999) claimed that MEIS measures were sufficiently differentiated from verbal intelligence to provide unique variance but also sufficiently correlated to indicate that concepts underlying the MEIS form an intelligence. Somewhat curiously, the verbal intelligence measure, used in the Mayer, Caruso, and Salovey (1999) study (i.e., the Army Alpha), is seldom used in contemporary investigations of cognitive ability. Moreover, another study, which included an oft-used measure of cognitive abilities, came up with a notably different finding that might be construed as questioning the claim that EI meets the standards expected of an intelligence. In particular, Ciarrochi et al. (2000) found near zero correlations between general EI, measured by total MEIS scores, and the Australian version of the Ravens Standard Progressive Matrices test (RSPM; Australian Council of Educational Research [ACER], 1989), and negative correlations between an understanding and managing emotions factor and RSPM score!

With respect to their criterion, Mayer, Caruso, and Salovey (1999) reported that differences in mean EI scores observed for adolescents and adults serve as evidence supporting the developmental criterion. Note, however, that the above study was based on a cross-sectional design and thus allows interpretation only in terms of age group differences—not developmental differences. There is another interesting issue implicit to the issue of developmental differences raised by consensus scoring. In particular, if one takes the consensus of the younger group, as the measure by which one should score these scales, it remains plausible that these age trends will reverse. In their study, Mayer, Caruso, and Salovey (1999) actually used an independent adult sample to obtain the consensus scores, meaning that this rival hypothesis certainly cannot be ruled out. In any event, the developmental criterion espoused by Mayer, Caruso, and Salovey (1999) is imprecise. In the intelligence literature, a particularly important finding is that certain classes of cognitive ability (e.g., Gf) actually decline with age (see, e.g., Carroll, 1993; Cattell, 1971; Horn & Hofer, 1992). It is difficult to envisage what developmental trend, other than complete insensitivity to age, would call into question the validity of any given measure.

EI: Negative Results

Mayer and Salovey (1993) had originally described EI as a type of social intelligence. However, despite much research, the independence of social intelligence from other types of intelligence (i.e., verbal) has not been successfully demonstrated (Carroll, 1993; Cronbach, 1960). Indeed, there is some evidence relating EI to Gc, through its mutual relationships with putative measures of social intelligence (Davies et al., 1998). Davies et al. (1998) found a range of measures purportedly assessing EI to have poor psychometric properties. These authors found low correlations among three factors defining the EI construct in their study—appraisal of emotions in the external world (perception) and appraisal of emotions in the self (awareness and clarity). A positive outcome evidenced in the Davies et al. investigation was that the perception of consensus-judged emotion in external objects represents a clearly defined unifactorial construct. However, two problems exist with emotion perception as a facet of EI. First, the scales have evidenced relatively low reliability. Second, consensus scoring may define the factor rather than emotional content per se. This methods-factor issue is an important one, certainly worthy of more careful consideration than it has been given to date.

One of the main criticisms subsequently leveled at the Davies et al. (1998) investigation was the EI measures were still in their infancy such that their conclusions appeared premature (e.g., Mayer, Caruso, & Salovey, 1999; Mayer & Cobb, 2000; Mayer, Salovey, and Caruso, 2000a, 2000b). Thus, explicitly citing this reference, Mayer and Cobb (2000) noted that "the Davies et al. study preceded publication of the highly reliable MEIS" (p. 173). The question that should then be posed is, To what extent do available data support the efficacy of the MEIS, which debatably would now appear as the premier vehicle for the assessment of EI (see, e.g., Ciarrochi et al., 2001)?

In their recent psychometric analysis of scores obtained from the MEIS, Mayer, Caruso, and Salovey, (1999) demonstrated that for consensus scores, reliabilities ranged from .49 to .94.[8] Indeed, Ciarrochi et

[8] The reader should note that, for comparative purposes, we present each of the reliabilities for respective tests and scoring protocols obtained by Mayer, Caruso, and Salovey (1999) as a companion to similar analysis we conduct as part of the present study.

al. (2000) found remarkably similar reliabilities in what essentially amounted to a replication of this study. Some of these reliabilities are not in the acceptable range, certainly for applied use of the measure for selection, intervention, and treatment purposes (see Anastasi & Urbina, 1997). For expert scores, the reliabilities obtained by Mayer, Caruso, and Salovey (1999) were even lower, ranging from .35 to .86. Consequently, it would appear that the accuracy of measurement is a function of the scoring procedure. It may well be that the expert scoring protocol is a misguided one, particularly because the test constructors made these judgments alone. Given the variation of reliabilities with disparate scoring procedures, it would appear that more detailed attention needs to be given to investigating the reliability of performance-based measures of EI.

There also seems to be some inconsistency between the theory underlying the MEIS and factor-analytic research attempting to uncover its factor structure. Although Mayer and Salovey (1997) claimed that a four-factor model underlies the MEIS, exploratory and confirmatory analyses of various data sets suggest only a three-factor solution (e.g., Ciarrochi et al., 2000; Mayer, Caruso, & Salovey, 1999). Of interest Mayer, Caruso, and Salovey (1999) found that in using confirmatory factor analysis, the model estimates of Branch-2 and -3 facets correlated .87 and thus correctly asserted that these two facets are virtually indistinguishable. The facets of understanding and assimilation of emotion coalesced into a single factor. This noteworthy finding would seem important to replicate in a comparably large and relatively homogeneous sample.

Finally, the overlap between self-report scales of EI and existing personality scales represents a serious challenge to the conceptualization of EI as a cognitive ability rather than a personality trait and may also extend to performance-based measures. In short, it is unclear whether EI tests possess discriminant validity with respect to existing measures. Because no study has yet examined the relation between performance-based EI measures and the Big Five personality factors, this would appear as an issue worthy of detailed investigation in its own right (see also, Mayer, Caruso, & Salovey, 1999). In addition, the dependence of validity coefficients on the various scoring methods has been neglected. Mayer, Caruso, and Salovey (1999) only reported correlations between consensus-scored MEIS scales and criteria such as empathy, with no indication of whether similar correlations were obtained with expert scoring.

Objectives of the Present Study

In light of the preceding review, which highlights several contradictory findings, the present study attempts to provide further information that is pertinent to a balanced evaluation of the empirical and conceptual status of EI. To this end, we examined the most comprehensive and contemporary performance-based measure of EI, the MEIS (e.g., Mayer, Caruso, & Salovey, 1999). Although it is possible to focus on any number of research questions bearing on the MEIS, it seemed expedient (because the measure is relatively new) to focus on the following objectives of relatively major significance:

1. *Is the construct of EI, as assessed by the MEIS, psychometrically sound?* In particular, this study sets out to examine the factorial validity of the MEIS, using both exploratory and confirmatory methods. Thus far, the one confirmatory factor analysis conducted with performance-based measures of EI (Mayer, Caruso, & Salovey, 1999) yielded rather equivocal results, including marginal fit statistics and evidence that two branches (e.g., understanding and assimilation) could not be differentiated. In addition to exploratory factor analyses, the current study used structural-equation modeling procedures to test the goodness of fit between the four-branch model of EI and the data. In addition, we examined subtest reliabilities and the patterns of intertest correlations.
2. *Do the two different scoring criteria used in the MEIS (i.e., consensus and expert scoring) demonstrate convergent validity? Do they yield similar reliability coefficients?* The Mayer, Caruso, and Salovey (1999) model predicted a positive manifold, or a nonnegative correlation matrix among the subtests, supporting three converging factors associated with emotional identification, assimilating emotions, understanding emotions, and managing emotions. The same factors should be found by using both scoring methods, as they are construed as alternative (yet analogous) scoring protocols. Following Mayer, Caruso, and Salovey's (1999) decision to focus almost all of their reported analyses on consensus scoring, Ciarrochi et al. (2000) conducted an investigation where no consideration was given to expert scores. Arguably, both studies highlighted the need to examine alternative scoring procedures in close detail. In the present investigation, all responses were scored with both consensus and expert criteria, allowing us to determine the convergent validity of these

measures. Thus, one of the major goals of this study is to examine in greater depth the relationship between consensus and expert scoring and to ascertain any problems inherent in these two ways of scoring behavioral measures of EI. Mayer and his colleagues are not clear as to whether they believe these two forms of scoring are directly equivalent or more loosely related. Indeed, they generally encourage consensus coding because of its facility. In addition, we examine the personality and individual differences correlates of the two scoring procedures.

3. *What are the relationships between EI, personality, and ability factors?* Put somewhat differently, to what extent does EI vary as a function of personality and intelligence constructs? Is the pattern of relations between EI and personality variables invariant across the types of scoring criteria used? According to prior research by Mayer, Caruso, and Salovey (1999), and the notion of divergent validity, the principal prediction is that EI should relate modestly to general cognitive ability. Mayer and his colleagues have not specified the likely personality correlates of MEIS scores. On the basis of past empirical research (e.g., Dawda & Hart, 2000), we might expect EI to relate to higher Agreeableness, Conscientiousness, Extraversion, and Openness, and to lower Neuroticism. Associations should generalize across scoring criteria.
4. *What is the nature (and magnitude) of gender, ethnic, and age differences in performance-based assessment of EI?* The strongest prediction from previous research (e.g., Mayer, Caruso, & Salovey, 1999) is that EI should be higher in women, irrespective of the scoring method used. In addition, we assess to what degree individual and group differences vary with the type of scoring criteria used.

Advances of Current Research

This study replicates and expands on available research by using the MEIS in a number of ways. With one or two exceptions, previous investigations have relied on rather modest student samples, sometimes tested under conditions that are far from optimal. This may have served to decrease the power and generalizability of the results. In contrast, the present study examines a sizable sample of young adults in a military setting, all of whom participated under nearly optimal and standardized conditions. Indeed, much is also known of this sample regarding gender and ethnic differences. Thus, Carretta and Ree (1995, 1997) have demonstrated comparable cognitive factor structures for gender and ethnic groups (specifically with the test we use in this study) and found negligible gender differences in psychomotor and cognitive abilities.

Furthermore, in contrast to prior studies that have used single tests (e.g., Ravens Progressive Matrices; Raven, Court, & Raven, 1979) as measures of intelligence and other external indicators, our study used a complete battery of EI, general intelligence, and personality measures. In particular, the measure of cognitive abilities used—the Armed Services Vocational Aptitude Battery (ASVAB)—is not only widely used for military selection purposes but it has proved to be a valid and useful instrument in a variety of educational, organizational, and research settings (see Roberts et al., 2000). It also has demonstrably better psychometric properties than many extant intelligence batteries (e.g., Foley & Rucker, 1989; Jensen, 1985; Murphy, 1985). This assertion notwithstanding, numerous psychological studies have been devoted to (or use) the ASVAB somewhere within their experimental design, generating a multimillion dollar research industry. For example, the recent controversies engendered by the publication of *The Bell Curve* (Herrnstein & Murray, 1994), with all its accompanying literature, have direct links to the ASVAB. It was, after all, from the 1980 standardization sample of the ASVAB (U.S. Department of Defense, 1982) that much of *The Bell Curve's* data on general intelligence were computed. The ASVAB also shares an important place in the history of the mental testing movement (see, e.g., Carroll, 1997; Gregory, 1996; Hunt, 1995). Indeed, within contemporary psychology, it has boldly been proclaimed that "the ASVAB is representative of the state of the art in multiple aptitude batteries" (Ree & Carretta, 1995, p. 269).

Similarly, beyond rather narrow (e.g., facet) measures, no past study of performance-based EI has attempted to examine a measure of personality that yields the higher order constructs underlying the Big Five personality factor model. This gap in the literature is curious given that this has almost become standard practice when assessing the validity of self-report measures of EI (e.g., Davies et al., 1998; Dawda & Hart, 2000; Schutte et al., 1998). The instrument we chose for this purpose was the Trait Self-Description Inventory (TSDI), because of its operational convenience (it was available gratis from the United States Air Force [USAF]), robust factor structure, and development by one of the pioneers of the

Big Five personality factor model (Christal, 1994). Indeed, in a recent study, Krause, Lee, Stankov, Pallier, and Roberts (in press) demonstrate that each of the five factors assessed by this instrument shares high correlation (i.e., aproaching an *r* of .90), with respective factors extracted from the Neo Personality Inventory–Revised (NEO-PI-R), and remarkably similar correlations with other external measures. Also, the TSDI has also been used as a core measure of personality in other studies examining the relationship between EI and the Big Five personality factor model (see Davies et al., 1998, Study 2).

Finally, we note that in the area of cognitive abilities, consistent replication of factor structure (across disparate populations) is a hallmark of this field of science. For example, Carroll's (1993) meta-analysis of some 477 data sets consistently supported the derivation of three strata of primary mental abilities, secondary-order factors, and a general factor (psychometric *g*) across widely disparate populations. Because the full scale MEIS is presently examined on a notably different sample from those investigated previously, the extent of factorial similarity (and invariance) may be determined. In addition, comparisons and contrasts between the present sample and the sample reported by the creators of the MEIS (i.e., Mayer, Caruso, & Salovey, 1999) on a variety of psychometric properties is required to determine the replicability of their findings.

Method

Participants

Participants were 704 USAF trainees, the majority of whom were male (89%). Participants ranged from 17 to 23 years of age ($M = 19.74$, $SD = 2.21$). About 30% of the sample had some college education, with the remaining 70% having completed full or partial high school education. Participants were ethnically diverse, distributed as follows by ethnicity: Caucasian, 69%; African Americans, 14%; Latino–Hispanic, 9%; Asian American, 4%; Native Americans, 2%; unidentified, 2%. The majority of the recruits (over 61%) were engaged in technical occupations in the USAF, with the remainder serving as security police (14%), medical–dental staff (5%), general staff (7%), and support staff (11%). The remaining recruits were unidentified or had not yet been placed in occupational categories.

Psychological Tests

The main instruments used in this study were the MEIS, the ASVAB, and the TSDI, a measure of the Big Five factor model of personality. Descriptions of each of these instruments are provided in the passages that follow.

A Multi-Factor Emotional Intelligence Scale (MEIS)

The MEIS (Mayer, Caruso, & Salovey, 1999) is a multi-factor ability battery divided into four branches: (a) *emotional identification (perception)* (4 tests: Faces, Music, Designs, and Stories); (b) *assimilating emotions* (2 tests: Synesthesia and Feeling Biases); (c) *understanding emotions* (4 tests: Blends, Progressions, Transitions, and Relativity); and (d) *managing emotions* (2 tests: Others and the Self). Capsule descriptions of the various subtests composing each branch are given in Table 1, with the reader encouraged to consult the primary source (i.e., Mayer, Caruso, & Salovey, 1999) for full details on each respective test. Across all 12 subtests, responses were scored according to both consensus and expert scoring criteria, the procedures for which are described below.

Consensus scoring. Mayer, Caruso, and Salovey's (1999) normative sample provided the weights for each item used in the consensus scoring procedure. This approach appears judicious because it allows ready comparisons with that sample. Moreover, in a previous study, Garcia and Roberts (2000) reported high correlations between consensus scores provided by the standardization group and consensus scores calculated from their specific sample. Thus, the response of each participant was scored according to its agreement with participants from the normative sample endorsing each alternative. For example, a participant who chose 5 in the present investigation would receive a score of .52 for that item if in the Mayer, Caruso, and Salovey (1999) study 52% of the participants answered that anger was definitely present. If the participant reported that anger was definitely not present (1), and this matched only 5% of the normative sample, then the person would receive a score of .05 for that item and so forth.

Expert scoring. Criteria for expert scoring were based on Mayer and Caruso (from Mayer, Caruso, & Salovey, 1999), who served as the "experts." Thus, each response was scored according to its agreement with the alternative identified by these authors as the "best" answer for each item. In each case, endorsing the selected best answer value (or the integer on either side of it) was scored as 1; otherwise the answer was scored as 0.

Branch scores. MEIS branch scores, calculated separately for consensus and expert scoring criteria,

Table 1
Capsule Descriptions of the 12 Subtests Composing the MEIS

Test	Task and stimuli	Response
Branch 1: Emotion identification-perception		
Faces	Eight photos of faces, each rated for degree of anger, sadness, happiness, disgust, fear, and surprise present	Five-point scale ranging from 1 (*definitely not present*) to 5 (*definitely present*)
Music	Eight original musical scores rated as for Test 1	Five-point scale ranging from 1 (*definitely not present*) to 5 (*definitely present*)
Designs	Eight computer-generated graphic designs rated as for Test 1	Five-point scale ranging from 1 (*definitely not present*) to 5 (*definitely present*)
Stories	Six narratives; participant determines characters' feelings on seven emotion scales that vary among stories	Five-point scale ranging from 1 (*definitely not present*) to 5 (*definitely present*)
Branch 2: Assimilation of emotions		
Synesthesia	Six scenarios; participants are asked to imagine a feeling until they experience corresponding emotion(s)	Five-point semantic differential scale for warm–cold, yellow–purple, sharp–dull, fast–slow, dark–light, low–high, orange–blue, pleasant–unpleasant, good–bad, and sweet–sour
Feeling Biases	Four scenarios where participants are required to assimilate their current mood state with judgments as to how they feel about a fictional person described in the scenario	Five-point scale ranging from 1 (*definitely does not describe*) to 5 (*definitely does describe*) for seven trait scales (e.g., sad, trusting, etc.) that varied across scenarios
Branch 3: Understanding emotions		
Blends	Eight items; participants analyze how blended emotions comprise two or more simple emotions	Multiple-choice; e.g., optimism combines which two emotions? (a) pleasure and anticipation, etc.
Progressions	Eight items assessing participants' understanding of how feelings and emotions progress and intensify over time	Multiple-choice; e.g., if you feel increasingly guilty and begin to question your self-worth, you feel (a) depression, etc.
Transitions	Four scenarios designed to gauge understanding of how emotions follow one another (e.g., a person is afraid and later calm: *In between,* what are the likely ways the person might feel?)	Five-point scale ranging from 1 (*extremely unlikely*) to 5 (*extremely likely*) for six emotion scales (e.g., fear, anger) that varied across scenarios
Relativity	Four scenarios depicting social encounters between two fictional persons (often in conflict); participant is asked to judge how characters are feeling	Five-point scale ranging from 1 (*extremely unlikely*) to 5 (*extremely likely*) for ten emotion scales that varied across scenarios
Branch 4: Managing emotions		
Managing Others	Six vignettes of which participants evaluate plans of action in response to fictional persons requiring assistance	Five-point scale ranging from 1 (*extremely ineffective*) to 5 (*extremely effective*) for four alternative courses of action varying across vignettes
Managing Self	Six vignettes focusing on the shelf rather than on others	Five-point scale ranging from 1 (*extremely ineffective*) to 5 (*extremely effective*) for four alternative courses of action varying across vignettes

were formed by converting the respective branch subtest scores to standardized z scores and linearly summing these standardized scores to yield a composite branch score. For example, emotional identification branch scores, scored by consensus criteria, were formed by converting consensus-based Faces, Music, Design, and Stories scores from raw to z scores, and then linearly summing these four standardized scores to form the branch measure. A similar procedure was used for constructing expert emotional identification branch scores, and so on for the other branches.

General EI score. A score representing what might tentatively be called *general EI* was formed separately for consensus-based and expert-based scores, by linear summation of respective standardized MEIS subtest scores (across all 12 subtests).

Armed Services Vocational Aptitude Battery (ASVAB)

The ASVAB (U.S. Department of Defense, 1984) is a particularly prominent multiple aptitude battery, widely implemented in a variety of educational, selection, and research settings (Murphy & Davidshofer, 1998). For example, in the selection context, performance on the ASVAB appears to be a major determinant in the career choices of over 1.3 million young Americans per annum (Kaplan & Saccuzzo, 1997, p. 365). In addition, researchers have consistently established that the ASVAB possesses levels of high reliability and validity (e.g., Carroll, 1993; Herrnstein & Murray, 1994; Ree & Carretta, 1995; Roberts et al., 2000). Collectively, these features suggest it to be a particularly efficacious measure of cognitive abilities for inclusion in the present design.

A description of the 10 subtests comprising the ASVAB, along with sample items follows. In numbering these tests (and the ones following them), we retain 1–12 for the MEIS tests previously described.

13. *General Science.* This test consisted of 25 science-fact items. For example: "Which of the following foods contain the most iron? (a) eggs, (b) liver, (c) candy, or (d) cucumbers."

14. *Arithmetic Reasoning.* This test consisted of 30 arithmetic word problems. For example: "Pat put in a total of 16.5 hours on a job during 5 days of the past week. How long is Pat's average workday? (a) 3 hours, (b) 3 hours 15 minutes, (c) 3 hours 18 minutes, or (d) 3 hours 25 minutes."

15. *Word Knowledge.* This test contained 35 standard vocabulary items, such as: "The wind is *variable* today. The word *variable* means: (a) mild, (b) steady, (c) shifting, or (d) chilling."

16. *Paragraph Comprehension.* Participants in this test were presented with 15 paragraphs, each 1 to 3 sentences long, followed by a multiple-choice response question about the paragraph's content.

17. *Mathematical Reasoning.* This test consisted of 25 mathematical problems (primarily algebra but also questions about area, square roots, percentages, and simple geometry). For example: "If $3x = -5$, then $x = $ (a) −2, (b) −5/3, (c) −3/5, or (d) 3/5."

18. *Numerical Operations.* This was a 10-min speeded test. The participants' task was to respond to 50 (simple) number-fact items [e.g., "$2 \times 6 = ?$ (a) 4, (b) 8, (c) 3, or (d) 12"].

19. *Coding Speed.* This is another 10-min speeded test. An item consisted of a word followed by five 4-digit number strings [e.g., "green, (a) 6456; (b) 7150; (c) 8385; (d) 8930; (e) 9645"]. The participants' task was: (a) to look up the word's number code in a key consisting of 10 word-code pairs placed at the top of the page and then (b) to place the letter associated with that number code on the provided answer sheet. Coding speed consisted of 84 such items.

20. *Auto and Shop Information.* This test consisted of 25 questions about automobiles, shop practices, and the conventional use of mechanical tools.

21. *Mechanical Comprehension.* This test was composed of 25 questions, normally accompanied by drawings. These questions related to general mechanical and physical principles.

22. *Electrical Information.* This test contained 20 questions that related to electrical, radio, and electronic information.

ASVAB composites. Of course, it is possible to provide a breakdown of the relation that each of the 10 subtests of the ASVAB share with other constructs examined in the present investigation. However, it would appear expedient to report correlations with composites, which are often cited, particularly in the applied psychological literature. In the case of the ASVAB, there are five such scales: Administration, Electronics, General, Mechanical, and Air Force Qualifying (AFQT). These scales reflect various measures of practical utility in the selection context.[9] The last of these scales, the AFQT, is thought to provide an index of the general intelligence construct (i.e.,

[9] We provide the formula for deriving these various composites at the bottom of Table 4. Note that in certain instances, individual tests appear in the formula for more than one composite.

psychometric *g*; e.g., see Stauffer, Ree, & Carretta, 1996).[10]

To avoid any confusion, it should also be noted that researchers have reported factor-analytic solutions with the ASVAB subtests (see, e.g., Ree & Carretta, 1994). Generally, three factors emerge—a construct having salient loadings from Tests 13–17, which is often interpreted as verbal–math; a factor defined by the two clerical-perceptual speed measures (Tests 18–19); and a construct defined by salient loadings from the three mechanical–technical knowledge tests (Tests 21–22). A similar solution emerges in the present data set. However, we refrain from reporting factor analysis of the ASVAB here because it would take the reader too far afield from the central topic of the present investigation, that is, the relation between EI and intelligence constructs.[11] Notably, the first principal component extracted from the battery correlates near unity with the AFQT, whereas the various composites reflect well-established relationships with these factors (e.g., the Mechanical composite correlates .90 with a mechanical–technical factor score). It is important to note that patterns of relationships between these factor scores and other variables are also mirrored (especially with respect to various scores derived from the MEIS).

Trait-Self Description Inventory (TSDI)

This Likert-type self-report inventory was designed to assess each of the Big Five personality factors. It contained two sections. In the first section, 64 trait names (e.g., *thorough*) were presented (one after the other). The participant was required to indicate how characteristic the trait was of him or herself as compared with other individuals of the same sex and age. Seven possibilities were given, ranging from 1 (*extremely not characteristic of me*) to 7 (*extremely characteristic of me*). In the second section, participants were given 99 behavioral statements and asked to indicate the extent to which they disagree (or agree) with each assertion (e.g., "I like parties, where there are lots of people"). As for trait names, responses fell along a nine-point scale, ranging from 1 (*very strongly disagree*) to 9 (*very strongly agree*).

Thereafter, trait and behavioral responses were summed to generate five composites. These composites represent the super factors of Agreeableness, Conscientiousness, Extraversion, Openness, and Neuroticism. A validation study of 2,853 USAF enlistees provides a large reference group with which isolated studies (such as the present one) might be compared (Christal, 1994; see also Davies et al., 1998).

Information pertinent to each scale of this Big Five factor measure (including sample items from both item formats and the number of trait names and behavioral statements in each scale) is given below.

23. *Neuroticism (10 trait names + 28 behavioral statements = 38 items):* "nervous"; "I worry more than most people."

24. *Agreeableness (16 trait names + 19 behavioral statements = 35 items):* "kind"; "I am always considerate of the feelings of others."

25. *Conscientiousness (17 trait names + 14 behavioral statements = 31 items):* "organized"; "I tend to set higher standards for myself than others set for me."

26. *Extraversion (12 trait names + 18 behavioral statements = 30 items):* "sociable"; "I tend to take charge in group meetings."

27. *Openness (9 trait names + 20 behavioral statements = 29 items):* "creative"; "I have a lot of intellectual curiosity."

Procedure

Testing on the MEIS and TSDI was conducted in groups of between 30 and 40 participants, with a small team of proctors available to administer the protocols and respond to queries. Each participant received item and answer booklets that contained all necessary instructions, items, and responses. Although testing time was self-paced, because on average testing took over 2 hrs, suitable rest–pauses were provided after 50 min of work. One exception to this general rule was the Music test, which we administered first because the instructions and music were played aloud (and obviously therefore experimenter-paced). Note the ASVAB data were obtained (after informed consent) from the enlistees' records rather than at this time. Not only is this standard practice with studies using military samples (e.g., Roberts & Kyllonen, 1999; Stauffer et al., 1996), it suggests that these data were collected under particularly optimal

[10] On the basis of recent analysis, somewhat controversially, researchers have argued that the AFQT is more likely to be an index of acculturated abilities (Roberts et al., 2000). Elsewhere, however, it is considered by many commentators as a particularly highly saturated *g* measure, similar to, for example, the Ravens Matrices tests (see e.g., Jensen, 1998, for a detailed discussion).

[11] Full data on the factor analysis of the ASVAB are available from Moshe Zeidner.

Table 2
Means, Standard Deviations, and Reliabilities for Consensus Scoring of the MEIS Subtests

	Present USAF Sample			Mayer et al. sample		
MEIS subscale	*M*	*SD*	α	*M*	*SD*	α
Emotional identification						
Faces	.41	.06	.81	.40	.08	.89
Music	.45	.08	.84	.44	.11	.89
Designs	.36	.07	.85	.36	.08	.90
Stories	.36	.06	.72	.38	.07	.85
Assimilating emotions						
Synesthesia	.29	.06	.84	.31	.04	.86
Feeling biases	.28	.05	.66	.30	.05	.70
Understanding emotions						
Blends	.46	.10	.38	.49	.10	.49
Progressions	.52	.12	.37	.58	.10	.51
Transitions	.29	.05	.57	.30	.04	.94
Relativity	.29	.04	.68	.30	.04	.78
Managing emotions						
Managing others	.26	.04	.66	.28	.04	.72
Managing self	.25	.04	.68	.28	.04	.70

Note. Results also include consensus-based scores for the Mayer, Caruso, and Salovey (1999) investigation. MEIS = Multi-Factor Emotional Intelligence Scale; USAF = United States Air Force.

conditions because the participants' responses were largely determining their career path.[12]

Results

Descriptive Statistics

MEIS (EI). Summary statistics derived from the present USAF sample for the MEIS consensus-based scores are presented in Table 2, and expert-based scores are presented in Table 3. In addition, Tables 2 and 3 incorporate comparable summary MEIS subscale statistics for the sample provided by Mayer, Caruso, and Salovey (1999, Table 1). Product-moment (structural) correlations (calculated between the mean vectors of the two samples) yielded correlations of .98 and .95, respectively, for consensus-based and expert-based scores. Thus, the profile of mean scores was highly comparable for both scoring criteria. There is also remarkable correspondence with the means and standard deviations reported by Ciarrochi et al. (2000) and these studies for consensus scores.

ASVAB (cognitive abilities). Mean performance of participants on each of the ASVAB subtests was calculated, along with intercorrelations. In each case, this series of results (available from Richard D. Roberts on request) was similar to normative samples (see, e.g., U.S. Department of Defense, 1982). As previously described in the *Method* section, composites (whose derivation is given in the bottom portion of Table 4) were subsequently calculated. The means and standard deviations for these five ASVAB composites are reproduced in the top half of Table 4.

Notably the mean of the ASVAB normative sample on any one of the occupational composites is 50 (*SD* = 10) (see Foley & Rucker, 1989). In other words, the present sample has performed above average on the ASVAB tests. There are two explanations for this occurrence. The first is that the ASVAB, like many other cognitive ability measures, is likely subject to item drift, encapsulated within the so-called *Flynn* effect (e.g., Flynn, 1987). The second is that the

[12] Note that one reviewer expressed concern that the time difference between administration of the ASVAB and the other instruments "is a bit worrisome." However, in a study where ASVAB tests were given a second time, in an experimental session similar to the present one, the correlation between tests given at different times was high enough (i.e., consistent with reported test–retest reliabilities) to consider them analogous (see Roberts et al., 2000). Moreover, the fit of alternative structural-equation models to the two separate sessions where ASVAB data were collected did not differ markedly. Time of testing does not appear, therefore, to result in an experimental confound (Roberts et al., 2000).

Table 3
Means, Standard Deviations, and Reliabilities for Expert-Based Scoring of the MEIS Subtests

MEIS subscale	Present USAF sample			Mayer et al. sample		
	M	*SD*	α	*M*	*SD*	α
Emotional identification						
Faces	.64	.10	.68	.64	.11	.74
Music	.72	.09	.72	.73	.10	.86
Designs	.69	.10	.68	.69	.11	.74
Stories	.71	.10	.70	.72	.11	.72
Assimilating emotions						
Synesthesia	.65	.11	.76	.69	.09	.66
Feeling biases	.67	.13	.63	.72	.12	.60
Understanding emotions						
Blends	.55	.18	.26	.60	.19	.35
Progressions	.74	.19	.45	.83	.16	.50
Transitions	.52	.12	.47	.56	.11	.85
Relativity	.51	.12	.69	.56	.11	.63
Managing emotions						
Managing others	.60	.13	.55	.60	.12	.42
Managing self	.53	.13	.52	.55	.12	.40

Note. Results also include consensus-based scores for the Mayer, Caruso, and Salovey (1999) investigation. MEIS = Multi-Factor Emotional Intelligence Scale; USAF = United States Air Force.

Table 4
Means, Standard Deviations, and Reliabilities for the Cognitive Ability and Personality Measures for the Current Investigation, and a Comparative Sample for the ASVAB

Measure	Current USAF sample			Comparative sample			
	M	*SD*	α	*M*	*SD*	α	*d*
ASVAB composites[a]							
Mechanical	58.58	21.52	—	56.67	22.04	—	0.09
Administration	64.46	19.70	—	69.19	19.94	—	0.02
General	61.84	17.48	—	62.74	17.72	—	−0.05
Electronics	64.90	16.19	—	63.56	17.37	—	0.08
Air force qualifying	218.28	16.17	—	219.21	16.67	—	0.06
TSDI							
Neuroticism	−162.04	239.57	0.94	−177.07	246.12	.94	0.06
Agreeableness	365.24	187.34	0.94	446.40	161.30	.93	−0.43
Conscientiousness	253.22	144.20	0.93	315.48	125.45	.92	−0.43
Extraversion	152.68	181.02	0.92	180.85	191.95	.94	−0.16
Openness	89.82	186.77	0.92	149.12	172.02	.96	−0.32

Note. Reliabilities of ASVAB were not available because these tests are scored by independent agencies, although the scores used here actually serve both screening and selection purposes in the USAF. N = 420 for ASVAB, N = 2,853 for TSDI. Sample sizes vary between 668 and 702 (because of missing data). ASVAB = Armed Services Vocational Aptitude Battery; USAF = United States Air Force; TSDI = trait self-description inventory.
[a] ASVAB composites were formed by combining subtests in the following manner: (a) Mechanical: Arithmetic Reasoning, Mechanical Comprehension, Auto and Shop Information, and Electronics Information; (b) Administration: Word Knowledge, Paragraph Comprehension, Mathematics Knowledge, and Coding Speed; (c) General: Word Knowledge, Paragraph Comprehension, Arithmetic Reasoning, and Mechanical Comprehension; (d) Electronics: Arithmetic Reasoning, Mathematics Knowledge, Electronics Information, and Science; and (e) AFQT: Word Knowledge, Paragraph Comprehension, Arithmetic Reasoning, and Mathematics Knowledge (see Herrnstein & Murray, 1994, pp. 579–592, for a detailed exposition of technical issues related to this measure).

USAF has in times of (relative) global peace and tech nological advances reduced the number of its personnel (Hammer, 1997), thereby resulting in more stringent selection criteria. Consistent with both propositions is the fact that means and standard deviations for ASVAB normative composites were obtained from the 1980 American Youth Population (Foley & Rucker, 1989). Moreover, the values obtained on the ASVAB are close to those reported in various validation studies conducted on this battery in recent times (e.g., Ree & Carretta, 1995; Roberts & Kyllonen, 1999; Stauffer et al., 1996). To highlight this correspondence, the results obtained by Roberts and Kyllonen are also included in Table 4. Note that for all composites the effect sizes for intersample differences are close to zero.

TSDI. Means, standard deviations, and reliabilities for the Big Five personality factors assessed by the TSDI are presented in the bottom half of Table 4. These values bear close correspondence to the results obtained by Christal (1994) in his original validation study of 2,853 USAF enlistees for measurs of Extraversion and Neuroticism, but not for measures of Agreeableness, Conscientiousness, or Openness. To allow ready comparisons, we also include these data (along with observed effect sizes) in Table 4. Cohort differences of this magnitude are quite common in military samples, where specific occupational clusters often go through basic training together. Thus, for example, the present sample is more heavily weighted toward technical occupations (i.e., 61%) than is common to the USAF, in general. It is unlikely that differences on these three personality dimensions, however, would constitute any serious challenge to interpretation of findings with personality. Consistent with this assertion, the reliability of the subscales of the TSDI remains high, and factor analysis of the measure recovers the five personality clusters.

MEIS Intercorrelations

Table 5 presents the matrix of correlations for the MEIS among consensus- and expert-based branch (and composite) scores, which shows a positive manifold. One possibility is that the matrix of non-negative correlation coefficients represents a common influence of general intelligence (i.e., psychometric g) on the subscales. However, partialing out ASVAB-AQFT scores, as an estimate of g, had only minor effects on the correlations, and the positive manifold was preserved.[13]

As shown in Table 5, composite consensus-based and expert-based MEIS scores were meaningfully correlated (i.e., $r = .48$). However, the magnitude of correlation for the two scoring criteria varied by branch. Thus, the correlations among consensus- and expert-based scores was relatively high for understanding emotion ($r = .78$) and assimilating emotion ($r = .66$); moderate for managing emotion ($r = .43$); and negligible for emotional identification ($r = .02$).

To gain some clearer understanding of branch correlations, we conducted separate analysis for each subtest. Table 6 presents the intercorrelations among the 12 subtests for consensus- (lower diagonal) and

[13] We are grateful to Professor Nat Brody for suggesting this analysis; the matrix of partial correlations is available from Moshe Zeidner.

Table 5
Matrix of Intercorrelations Among Consensus- and Expert-Based MEIS Branch and Composite Scores

MEIS measure	1	2	3	4	5	6	7	8	9	10
Consensus										
1. Composite	—									
2. Identification	.76*	—								
3. Assimilation	.68*	.38*	—							
4. Understanding	.80*	.38*	.41*	—						
5. Management	.57*	.12*	.32*	.44*	—					
Expert										
6. Composite	.48*	.08	.44*	.56*	.42*	—				
7. Identification	.17*	.02	.21*	.18*	.16*	.78*	—			
8. Assimilation	.39*	.09*	.66*	.31*	.24*	.66*	.42*	—		
9. Understanding	.56*	.14*	.33*	.78*	.39*	.66*	.21*	.32*	—	
10. Management	.23*	−.04	.14*	.27*	.43*	.62*	.36*	.25*	.24*	—

Note. Sample sizes ranged from 702 to 704. MEIS = Multi-Factor Emotional Intelligence Scale.
* $p < .05$.

Table 6
MEIS Subscale Intercorrelations for Consensus (Below diagonal) and Expert (Above diagonal) Scoring

MEIS test	1	2	3	4	5	6	7	8	9	10	11	12
1. Faces	**−.22**	.44*	.47*	.42*	.26*	.20*	.07	.12*	.19*	.02	.22*	.23*
2. Music	.64*	**.29***	.42*	.41*	.20*	.22*	−.02	.11*	.11*	.10*	.19*	.19*
3. Design	.60*	.66*	**.09***	.38*	.27*	.20	.01	.16*	.06	.03	.33*	.23*
4. Stories	.43*	.51*	.52*	**.12***	.26*	.33*	.05	.19*	.26*	.11*	.20*	.23*
5. Synesthesia	.23*	.19*	.21*	.31*	**.64***	.23*	.11*	.23*	.11*	.12*	.24*	.17*
6. Biases	.24*	.24*	.26*	.35*	.33*	**.61***	.05	.15*	.22*	.24*	.12*	.13*
7. Blends	.04	.06	.14*	.14*	.14*	.12*	**.85***	.24*	.06	.15*	.15*	.10*
8. Progression	.16*	.22*	.23*	.29*	.24*	.23*	.33*	**.96***	.17*	.23*	.17*	.18*
9. Transitions	.26*	.25*	.30*	.36*	.23*	.26*	.16*	.24*	**.41***	.28*	.12*	.10*
10. Relativity	.19*	.22*	.24*	.29*	.31*	.32*	.27*	.33*	.39*	**.66***	.10*	.07
11. Others	.08	.08	.10*	.21*	.28*	.23*	.20*	.17*	.28*	.45*	**.48***	.43*
12. Self	.02	.02	.05	.13*	.25*	.16*	.21*	.18*	.24*	.36*	.53*	**.35***

Note. The correlation between expert and consensus scores for each respective test is located down the main diagonal in boldface. Sample size varies from 698 to 704 because of missing values. MEIS = Multi-Factor Emotional Intelligence Scale.
* $p < .05$.

expert-based scores (upper diagonal). Of interest, the product–moment correlation between the two correlation profiles was not significant, (i.e., $r[64] = .15$, $p > .05$), suggesting rather weak correspondence among the pattern of intercorrelations based on these two different scoring criteria. The main diagonal of Table 6 is also worth noting, because it reports correlations between consensus and expert scores for individual tests. The problematic correlation between Emotional Perception subtests scored in the two differing formats is highlighted here—with even a modest negative correlation ($r = -.22$) evidenced for the Faces test.

The overall pattern of correlations among the MEIS subscales found in the present data set are highly similar to those reported by Mayer, Caruso, and Salovey (1999). Correlations reported by Mayer, Caruso, and Salovey ranged from .07 to .68 (*Mdn* = .29), whereas in the present study they ranged from .02 to .66 (*Mdn* = .26). A product–moment correlation among the strung-out correlations in the two correlation matrices (i.e., Mayer, Caruso, & Salovey's data and the present data set) yielded a sizable correlation coefficient of .87. Furthermore, in both studies, the correlations for the emotional perception branch subscales were higher than those found for the understanding branch subscales. Thus, the pattern of relations among the subtests is quite robust and shows considerable replicability across the two data sets.[14]

Exploratory Factor Analysis of the MEIS Subscales

To shed light on the factor structure underlying the subscales of the MEIS, the subscale intercorrelation matrices for both consensus- and expert-based scores were submitted, in turn, to principal-axis factor analyses, followed by direct oblimin rotation. Oblique rotation was used throughout because the intercorrelations of first-order factors should define a second-order factor of general EI. Table 7 gives the loadings on the first, unrotated factor for both scoring methods. In these analyses, each first factor resembles a general factor, with positive loadings on all subscales, explaining 27.1% (consensus) and 20.0% (expert) of the variances for the respective matrices. In both cases, the Emotional Perception subscales had the highest loadings on the general factor (i.e., loadings in excess of .50 for each test), whereas most of the Emotional Understanding and Emotion Management subscales had relatively low loadings (in some cases, less than .30). Principal-components analysis (with communalities set to 1.00) is sometimes used to estimate general factors. We extracted principal components from each subscale intercorrelation matrix, which explained 32.7% (consensus) and 26.2% (expert) of the variances. Loadings were a little higher than for the principal factors, but otherwise the patterns of loadings were highly similar.

Following rotation, the principal-axis factor analysis for consensus-based scores yielded three interpretable factors, as was the case for consensus scores. In each instance, the fourth factor extracted had an eigenvalue less than one, and failed to meet Cattell's

[14] Comparable analyses conducted with the Ciarrochi et al. (2000) data reveal similarly high levels of correspondence.

Table 7
Factor Loadings on First, Unrotated Factor From Principal-Axis Factors Analysis of Consensus- and Expert-Scored Branch Subscale Scores

	Scoring method	
MEIS branch subscale	Consensus	Expert
Emotional identification		
Faces	.59	.61
Music	.64	.54
Designs	.66	.61
Stories	.68	.61
Assimilating emotions		
Synesthesia	.46	.42
Feeling biases	.49	.39
Understanding emotions		
Blends	.29	.17
Progressions	.44	.34
Transitions	.33	.31
Relativity	.55	.23
Managing emotions		
Managing others	.41	.42
Managing self	.32	.45
Variance	27%	20%

Note. MEIS = Multi-Factor Emotional Intelligence Scale.

(1971) scree criterion. The factor-pattern matrices for the consensus- and expert-based scores are presented in Tables 8 and 9, respectively. Interpretation of each respective solution follows.

Factor interpretation: Consensus scores. The first factor to be extracted, Perception (identification) of Emotions, accounted for 28% of the total variance. The four perception subscales loading on this factor showed both convergent and divergent factorial validity. Accordingly, the scales were strongly correlated with the Perception factor (.66 to .82) while correlating much more weakly with the remaining two factors. The second factor to emerge, Emotion Management, accounted for 12% of the total variance, with managing others and managing self loading .79 and .65, respectively, on this factor. The third factor extracted, Emotion Understanding, accounted for 4% of the total variance. Mainly, the four original scales defining the Understanding Emotions factor of the MEIS (i.e., Blends, Progressions, Transitions, and Relativity) define this factor. The two Assimilation of Emotions factors were found not to constitute unifactorial scales. By loading about equally on each of the three factors, these scales fail to evidence divergent validity. Furthermore, the Relativity subscale loaded higher on the Emotion Management factor (.60) than on the Emotion Understanding factor it was designed to mark (.54). Thus, these data do not provide much support for combining the original Understanding and Assimilating Emotions scales into one super-scale (as was done by Mayer, Caruso, & Salovey, 1999). Intercorrelations among the three obtained factors were as follows: Perception and Management (r = .32);

Table 8
Oblimin Factor-Pattern Matrix

	Factor			
MEIS branch subscale	Perception	Management	Understanding	h^2
Emotional identification				
Faces	**.80**	−.02	−.11	.57
Music	**.86**	−.07	−.06	.68
Designs	**.80**	−.06	.02	.63
Stories	**.58**	.08	.14	.47
Assimilating emotions				
Synesthesia	.21	.24	.17	.22
Feeling biases	.28	.15	.20	.24
Understanding emotions				
Blends	−.07	.06	**.46**	.23
Progressions	.06	−.09	**.68**	.43
Transitions	.28	.24	.18	.29
Relativity	.12	**.43**	.27	.44
Managing emotions				
Managing others	.29	**.84**	−.11	.63
Managing self	.29	**.67**	.03	.43

Note. Loadings above 0.30 are in boldface type. Findings are based on principal axis factor analysis of consensus subscale scores, followed by oblimin rotation. MEIS = Multi-Factor Emotional Intelligence Scale.

Table 9
Oblimin Factor-Pattern Matrix

MEIS branch subscale	Factors: Perception	Understanding	Management	h^2
Emotional identification				
Faces	**.65**	−.02	.07	.44
Music	**.61**	−.04	.00	.37
Designs	**.65**	−.15	.22	.50
Stories	**.57**	.20	−.01	.42
Assimilating emotions				
Synesthesia	.22	.16	.20	.18
Feeling biases	.28	**.34**	−.07	.22
Understanding emotions				
Blends	−.11	.28	.19	.13
Progressions	.01	**.41**	.17	.24
Transitions	.14	**.42**	−.06	.21
Relativity	−.05	**.57**	−.04	.30
Managing emotions				
Managing others	.07	−.02	**.72**	.55
Managing self	.14	.03	**.48**	.29
Variance	21%	6%	5%	

Note. Loadings above 0.30 are in boldface type. Findings are based on principal axis factor analysis of expert-based subscale scores, followed by oblimin rotation. MEIS = Multi-Factor Emotional Intelligence Scale.

Perception and Understanding ($r = .51$); Management and Understanding ($r = .62$).

Factor interpretation: Expert scores. A similar principle-axis factor analysis for expert-based scores followed by oblimin rotation yielded three interpretable factors: Perception (21% of the total variance), Understanding of Emotions (6%), and Management of Emotions (5%) (see, Table 9). The correlations among the three factors in our study were as follows: Perception and Management, $r = .26$; Perception and Understanding, $r = .28$; Management and Understanding, $r = .30$. As observed for consensus scoring, the Assimilation scales did not show good convergent or divergent validity, and only three of the four original branch scores (perception, understanding, and emotions) are supported in this analysis.

Collectively, these results diverge from those reported by Mayer, Caruso, and Salovey (1999) in their factor analysis of the MEIS scales. In short, our data do not support their observation that assimilation and understanding coalesce into a single Understanding factor. Instead, for both consensus and expert scoring, assimilation appears both factorially complex and underrepresented by whatever the other tests are assessing (note therefore also the low communalities for these subtests). Thus, our data fail to support the conclusion reached by Mayer, Caruso, and Salovey (1999) that three factors capture the four branches of their recent model of EI (see also Mayer & Salovey, 1997).

Finally, we tested the correspondence of the factor scores derived from expert-based and consensus-based scale scores. We calculated regression-model factor scores for each participant, for each of the three expert-based factors and each of the three consensus-based factors. The correlations between the two sets of factors were then computed. Although the two Perception factors were defined by similar loadings, the correlation between them was .00. The two Understanding factors correlated .74, and the two Management factors correlated .46. In addition, Understanding (consensus) correlated with Management (expert) .43, and Understanding (expert) correlated with Management (consensus) .58. All other correlations were less than .25. In general, the correlations of corresponding factors fall short of the magnitudes required to establish the equivalent of the two forms of scoring. We also tested the equivalence of the first unrotated factors (i.e., general factors) from each subscale correlation matrix, using the same factor-score method. The correlation between these two general factors was .26.

Confirmatory Factor Analysis (CFA)

To assess the goodness of fit between the pattern of covariation in the MEIS subscales and the four-factor (branch) model posited to underlie the construction of these measures, we submitted consensus- and expert-score covariance matrices, in turn, to confirmatory factor analyses. Throughout, AMOS 3.6 (Arbuckle, 1997), a structural equation model-fitting program, was used to conduct the analyses (see Table 10).

The results for both consensus and expert scores, shown in Table 10, suggest that the correspondence of the four-branch model to the data is probable. Thus, whereas the chi-square index was significant for consensus scores, χ^2 (48, N = 656) = 171.19, $p < .001$, chi-square/df = 3.57, the standard goodness-of-fit (GFI) indices were all above .90 (i.e., GFI = .96, comparative fit index [CFI] = .95) with root mean square error of approximation (RMSEA) = .06, at the borderline level of acceptability. All of the standardized factor loadings (save for two scales assessing the Understanding factor) exceeded .50. Absolute values of the factor correlations were generally moderate. Similarly, whereas the chi-square fit index for expert-based scores was significant, χ^2 (48, N = 656) = 158.98, $p < .001$, chi-square/df = 3.31, the standard GFI indices were at the acceptable level (GFI = .96, CFI = .92; RMSEA = .06). The scales assessing identification and management of emotions all showed good convergent validity, loading above .60 on their target factor. The scales assessing assimilation and understanding tended to show somewhat lower loadings, ranging from .31 to .52. Of interest, whereas for expert-based scores, the correlation between identification and management of emotions was moderately high (r = .54), for consensus-based scores it was quite low (r = .14).

Chi-square statistics were used to test the significance of the difference in relative fits for a three-factor model (identification, understanding, and management) nested under the four-factor model. When the correlation between the Assimilation and Understanding factors was fixed at 1, the CFA tests for a three-factor model yielded χ^2 (51, N = 656) = 192.46 for consensus scores and χ^2 (51, N = 656) = 203.25 for expert-based scores. On the basis of the chi-square statistic, the fit of the three-factor model is significantly worse for both consensus, χ^2 (3, N = 656) 21.27, $p < .001$, and expert scores, χ^2 (3, N = 656) = 44.27, $p < .001$. Thus, a four-factor model is more plausible than the three-factor model suggested by Mayer, Caruso, and Salovey (1999).

Moving from this issue, we tested the extent to which a second-order EI factor accounts for the observed covariation among the four MEIS branch–factors. To that end, we conducted a higher order CFA for the consensus-based and expert-based MEIS branch scores, in turn, with the covariances among the first-order factors set to zero. (It is noted that the first-order model represents an upper bound for the target higher order CFA.)

For consensus scores, the higher order model provided an unsatisfactory fit, χ^2 (50, N = 656) = 215.38, $p < .001$, and chi-square/df = 4.31. A test for differences in chi-square statistics comparing higher and lower order models proved to be significant, χ^2 (2, N = 656) = 44.19, $p < .001$, suggesting that the higher order, more restricted model is significantly worse than the first-order model.

The goodness-of-fit analysis for expert-based scores provided a somewhat better fit than for the consensus scores, χ^2 (50, N = 656) = 176.64, $p <$

Table 10
Factor Loading Matrices for CFA of the Fit of MEIS Scales to the Four-Branch Model

MEIS scale	Identification	Assimilation	Understanding	Management
Faces	.74 (.68)			
Music	.82 (.63)			
Designs	.81 (.65)			
Stories	.64 (.64)			
Synesthesia		.57 (.48)		
Feeling biases		.58 (.47)		
Blends			.37 (.31)	
Progressions			.47 (.52)	
Transitions			.55 (.42)	
Relativity			.71 (.49)	
Managing others				.78 (.69)
Managing self				.68 (.62)

Note. Loadings for expert scores appear in parentheses. CFA = confirmatory factor analysis; MEIS = Multi-Factor Emotional Intelligence Scale.

Table 11
Means and Standard Deviations for MEIS Branch Scores by Gender

MEIS branch score	Male (n = 617) *M*	Male *SD*	Female (n = 76) *M*	Female *SD*	*d*
Perception					
Consensus	−0.13	3.27	0.75	3.22	−0.27
Expert	0.16	2.96	−1.22	3.64	0.42
Assimilation					
Consensus	−0.01	1.62	0.04	1.78	−0.03
Expert	0.02	1.54	−0.16	1.85	0.11
Understanding					
Consensus	−0.04	2.76	0.38	2.43	−0.16
Expert	−0.02	2.52	0.23	2.59	−0.09
Management					
Consensus	−0.02	1.75	0.34	1.66	−0.22
Expert	0.01	1.69	−0.15	1.73	0.09
No. of composite MEIS scores					
Consensus	−0.22	6.82	1.51	6.59	−0.26
Expert	0.19	6.00	−1.30	6.82	0.23

Note. *d* scores represent male minus female branch score means, divided by the average standard deviations of group scores. MEIS = Multi-Factor Emotional Intelligence Scale.

.001, and chi-square/df = 3.53. The goodness-of-fit indices were acceptable (GFI = .96, CFI = .90, RMSEA = .06). Again, a test for differences in chi-square values between higher and lower order models proved to be significant, χ^2 (2, N = 656) = 17.66, $p < .001$, suggesting that the higher order, or more restricted, model is significantly worse than the first-order, unrestricted model. Thus, a four-factor first-order analysis seems to be the most plausible model tested.

EI: Relation to Individual and Group Differences

Gender differences. Table 11 presents descriptive statistics for the MEIS consensus- and expert-based branch and composite scores by gender. When consensus scores were used, female participants (M = 1.51, SD = 6.59) scored higher on composite MEIS scores than their male counterparts (M = −0.22, SD = 6.82), $t(691) = -2.10$, $p < .05$, by the order of about a quarter of a deviation (d = −0.26). In contrast, when expert scoring criteria were used, male participants (M = 0.19, 1.51, SD = 6.00) scored higher than their female counterparts (M = −1.30, 1.51, SD = 6.80), $t(691) = 2.01$, $p < .05$, by the order of a quarter of a standard deviation (d = 0.23). Thus, the direction of gender group effects varies as a function of the scoring criteria used.[15]

To assess to what extent the relationship between gender and MEIS scores was mediated by cognitive ability, we calculated partial correlations between gender (1 = male, 2 = female) and composite MEIS scores, statistically controlling for ability (i.e., ASVAB-AFQT score). The partial correlations for consensus scores, $r(691) = .08$, $p < .01$, and expert scores, $r(691) = -.08$, $p < .01$, were each about the same level of magnitude as the point-biserial correlations between gender and composite scores. This outcome suggests that the relationship between gender and scores on the MEIS (no matter how derived) is not mediated by general cognitive ability.

Ethnic group differences. Table 12 presents descriptive statistics for the MEIS, consensus- and expert-based branch and composite scores, by ethnic group. When consensus-based scores were used, White (M = 0.28, SD = 6.58) and ethnic minority (M = −0.63, SD = 7.12) group participants were not reliably discriminable. However, when expert scoring criteria were used, White participants (M = 1.16, SD = 5.86) scored significantly higher than their ethnic minority counterparts (M = −2.43, SD = 5.98),

[15] One reviewer expressed concern that the findings related to gender differences might be somewhat problematic, given the imbalance in numbers between male and female participants. To ensure that this was not the case, we compared the female sample with a randomly selected male sample (i.e., n = 76) and obtained differences that were remarkably similar to those reported for the full sample.

Table 12
Means and Standard Deviations for MEIS Branch Scores by Ethnic Group

MEIS subscale	White (n = 476) *M*	*SD*	Ethnic minority (n = 214) *M*	*SD*	*d*
Perception					
Consensus	–0.05	3.16	0.06	3.46	–0.03
Expert	0.50	2.79	–1.07	3.38	0.51
Assimilation					
Consensus	0.13	1.49	–0.30	1.88	0.26
Expert	0.24	1.43	–0.54	1.73	0.49
Understanding					
Consensus	0.16	2.69	–0.30	2.74	0.17
Expert	0.22	2.49	–0.44	2.52	0.26
Management					
Consensus	0.06	1.75	–0.08	1.73	0.09
Expert	0.19	1.67	–0.40	1.66	0.35
No. of composite MEIS score					
Consensus	0.28	6.58	–0.63	7.12	0.13
Expert	1.16	5.86	–2.43	5.98	0.60

Note. *d* scores represent ethnic minority minus White mean (branch) scores, divided by the average standard deviations of group scores. MEIS = Multi-Factor Emotional Intelligence Scale.

$t(688) = 7.39$, $p < .001$, by the order of over half a standard deviation ($d = .60$). Partial correlations between ethnic group membership and composite MEIS scores, holding ability constant, indicated that the partial correlations, $r(655) = .22$, $p < .001$, were not meaningfully different from the bivariate correlations, $r(688) = .27$, $p < .001$. These results suggest that the relationships between ethnic group and MEIS composite scores are not mediated by traditional cognitive abilities.

Educational and age differences. MEIS scores, whether scored by way of consensus or expert criterion, were not meaningfully or significantly correlated with educational background or age. This may be due, in part, to restriction of range in the age and educational background of the present sample.

Ability and Personality Correlates of the MEIS

The relationship between MEIS branch scores and cognitive abilities. Table 13 presents the correlations between the MEIS branch and general scales (scored by both consensus and expert criteria) and the five ASVAB composite scores. In general, correlations are highest between MEIS scores and the general and AFQT composite scores of the ASVAB. Notably, both the consensus-based ($r = .29$) and expert-based ($r = .40$) general EI scores correlated positively with the ASVAB-AFQT composite, which is the closest measure to general intelligence (see, e.g., Herrnstein & Murray, 1994; cf. Roberts et al., 2000). Of the EI concepts, the understanding facet correlated most meaningfully with ASVAB scores, whether consensus-based or expert-based scores were used. This outcome is somewhat surprising. According to the four-branch model proposed by Mayer, Caruso, and Salovey (1999), because management of emotions taps into the highest order of cognitive processes, it should (in turn) have the highest correlation with intelligence measures of all MEIS branches. Overall, however, MEIS branch and composite scores, whether scored with expert or consensus criteria, tend to be meaningfully related to intellectual ability, although correlations with ability measures tend to be slightly higher for expert than for consensus scores.

The relationship between MEIS branch scores and the Big Five personality factors. Table 14 presents the correlations between MEIS consensus and expert scores (branch and total) and the Big Five personality factors. As shown in Table 14, total MEIS consensus-based scores correlated significantly, though modestly, with each of the Big Five personality factors—correlating negatively with Neuroticism ($r = -.18$), and positively with Conscientiousness ($r = .16$), Agreeableness ($r = .24$), Extraversion ($r = -.13$), and Openness ($r = .13$). In contrast, MEIS expert-based scores failed to correlate significantly with any of the Big Five personality factors, save for Openness ($r = .15$). However, when ability scores are statistically controlled by using partial correlation procedures, the

Table 13
Correlations of MEIS Branch- and Full-Scale Scores With ASVAB Composites as a Function of Consensus and Expert Scoring Systems

MEIS subscale	Mechanics	Administration	General	Electric	AFQT
Perception					
Consensus	.02	.07	.06	.09*	.09*
Expert	.23*	.09*	.22*	.28*	.23*
Assimilation					
Consensus	.15*	.17*	.18*	.25*	.22*
Expert	.21*	.12*	.24*	.29*	.26*
Understanding					
Consensus	.20*	.16*	.36*	.32*	.40*
Expert	.21*	.13*	.36*	.32*	.39*
Management					
Consensus	.00	.16*	.11*	.12*	.16*
Expert	.12*	.11*	.20*	.20*	.22*
No. of composite MEIS scores					
Consensus	.14*	.18*	.27*	.28*	.32*
Expert	.29*	.18*	.41*	.42*	.43*

Note. Correlations based on 667 to 669 degrees of freedom. MEIS = Multi-Factor Emotional Intelligence Scale; ASVAB = Armed Services Vocational Aptitude Battery; AFQT = Air Force Qualifying Test.
* $p < .05$.

correlations between Openness and total expert-based MEIS scores approach zero ($r = .05$). Thus, the observed correlations between MEIS scores and the Openness factor may be mediated by cognitive ability.

Further inspection of Table 14 points to considerable variation in the nexus of relations between the Big Five personality factors and the respective MEIS branch scores as a function of scoring criteria. For example, the correlation between Assimilation and Agreeableness factors was .17 for consensus scoring and only .02 for expert scoring. Furthermore, the di-

Table 14
MEIS Branch and Total Score Correlations With Big Five Personality Factors (Assessed by the TSDI)

MEIS subscale	N	A	C	E	O
Perception					
Consensus	−.13*	.17*	.80*	.10*	.01
Expert	.07	−.09*	−.09*	−.11*	.08*
Assimilation					
Consensus	−.13*	.17*	.09*	.09*	.07
Expert	−.07	.02	−.02	.01	.07
Understanding					
Consensus	−.15*	.10*	.11*	.05	.18*
Expert	−.06	.07	.10*	.02	.20*
Management					
Consensus	−.11*	.29*	.22*	.14*	.12*
Expert	−.02	−.05	−.05	.05	.05
No. of composite MEIS scores					
Consensus	−.18*	.24*	.16*	.13*	.13*
Expert	−.02	−.03	−.02	−.03	.15*

Note. Sample sizes varied from 656 to 671. MEIS = Multi-Factor Emotional Intelligence Scale; TSDI = trait self-description inventory; N = neuroticism; A = agreeableness; C = conscientiousness; E = extraversion; O = openness.
* $p < .05$.

rection of EI-personality factor correlation tended to vary with scoring methods. For example, the correlation between Management and Agreeableness scores was .29 for consensus-based scores, and it was −.05 for expert-based scores.

Personality and ability as predictors of EI. Multiple regressions were performed to further investigate the nature of the overlap between the MEIS and conventional personality and ability dimension, as well as the consistency of results across scoring methods. The first set of regressions aimed to test the extent to which knowledge of personality and ability allows the researcher or practitioner to estimate EI. Composite MEIS consensus- and expert-based scores were regressed, in turn, on a predictor set consisting of ASVAB-AFQT scores ($k = 1$) and the Big Five personality factors ($k = 5$). The standard multiple regression analysis for consensus-based scores yielded a significant model, $F(6, 581) = 4.25$, $p < .001$, RSQ = 0.16, with the following variables found to be significant predictors of composite scores: Neuroticism, $\beta = -0.12$, $t(581) = -2.76$, $p < .001$; Agreeableness, $\beta = 0.22$, $t(581) = 4.27$, $p < .001$; and ASVAB-AFQT scores, $\beta = 0.29$, $t(581) = 7.25$, $p < .001$. A similar multiple regression for MEIS expert-based composite scores regressed on the predictor stock also showed a significant model, $F(6, 581) = 19.61$, $p < .001$, RSQ = 0.17. However, in contrast to what was reported for consensus scores, the only significant predictor was ability (ASVAB-AFQT), $\beta = 0.23$, $t(581) = 5.32$, $p < .001$. Thus, the predictor stock varies for consensus-based and expert-based MEIS scores.

MEIS branch scores as predictors of personality and ability criteria. We conducted a second set of regressions to test whether the different branch scores were similarly related to ability and personality, within and across scoring methods. ASVAB-AFQT score and each of the Big Five personality factors were regressed, in turn, on the four consensus-based branch scores, which in this instance served as predictors. The analysis was then repeated using the four expert-based branch scores, giving a total of 12 regressions. Table 15 summarizes the results. The predictors explained a significant part of the variance in each criterion. Overall, consensus-based branch scores appear to be more consistent predictors of personality criteria than expert-based branch scores, with the former accounting for 4% to 11% and the latter accounting for 2% to 4% of the Big Five scale variance. Expert-based scores were slightly more predictive of ASVAB-AFQT score than were the consensus-based scores. Table 15 gives beta weights for each individual predictor. In general, multiple branch factors added to the variance explained by the multiple regression. For example, with AFQT as the criterion, expert-scored perception, assimilation, understanding, and management all contributed independently to the equation. There were some exceptions to this tendency. In both consensus- and expert-scored data, Un-

Table 15
Summary Statistics for Regressions of Ability and Personality Criteria Onto MEIS Branch Scores

				Predictor			
Criterion	*R*	*F*	*df*	Perception β	Assimilation β	Understanding β	Management β
				Consensus-scored scales as predictors			
AFQT	.39	30.05**	4, 664	−.08*	.10*	.39**	−.04
N	.18	5.40**	4, 649	−.07	−.05	−.07	−.05
A	.33	19.90**	4, 654	.15**	.06	−.10*	.29**
C	.23	8.80**	4, 659	.05	.00	.01	.21**
E	.17	4.93**	4, 663	.09*	.04	−.06	.14**
O	.20	6.86**	4, 652	−.07	.00	.19**	.05
				Expert-scored scales as predictors			
AFQT	.43	37.12**	4, 664	.10*	.10*	.30**	.09*
N	.14	3.05*	4, 649	.13**	−.09*	−.05	−.03
A	.16	3.11*	4, 654	−.12**	.05	−.08*	−.04
C	.14	4.17**	4, 659	−.10*	−.01	.14**	−.04
E	.15	3.71**	4, 663	−.16**	.03	.02	.10*
O	.21	7.13**	4, 652	.05	−.01	.20**	.01

Note. MEIS = Multi-Factor Emotional Intelligence Scale; AFQT = Air Force Qualifying Test; N = neuroticism; A = agreeableness; C = conscientiousness; E = extraversion; O = openness.
* $p < .05$. ** $p < .01$.

derstanding was the only predictor of Openness. In the consensus data, there were no significant independent predictors of Neuroticism, although the overall equation was significant, suggesting that it is variance shared among two or more branches that explains the negative association between EI and Neuroticism.

Table 15 also shows (a) discrepancies between consensus- and branch-based scores as predictors and (b) discrepancies between different branch scores as predictors of the same criterion. In general, there was rather poor correspondence between the predictors identified from consensus and expert data. Table 15 shows that of the 24 predictor-criterion correlations, 4 were significant in the same direction in both consensus and expert data, 7 were significant in one data set but not the other, 9 were nonsignificant in both data sets, and 4 were actually significant in opposite directions. The Pearson correlation between the two sets of correlations is .22 (−.14 if the two substantial AQFT-Understanding correlations are excluded). Instances of relationships that generalize across both types of data include associations between Understanding and AFQT score, Extraversion and Management, and Understanding and Openness. More commonly, significant predictors fail to generalize or even show relationships of opposite sign in the two sets of regressions. Perception appears especially problematic in this regard. In the regressions for consensus-scored scales, this MEIS branch is significantly negatively related to AFQT score, and positively associated with Agreeableness and Extraversion. All these associations are reversed and significant in the expert-scored data.

The second type of discrepancy shown in Table 15 refers to cases where different branches are oppositely related to criteria. For example, although MEIS scores are generally positively related to AFQT score, there is a small but significant negative association of Perception with this criterion. Other associations that conflict with general trends seen in the bivariate data (see Table 14) include the negative association between Agreeableness and Understanding in consensus-scored data, and the tendencies for respondents high in expert-scored perception to be neurotic and low in Conscientiousness, Agreeableness, and Extraversion.

Discussion

Mayer, Caruso, and Salovey (1999) contended that if EI is to constitute a legitimate form of intelligence it should satisfy three criteria: operationalization as an ability, appropriate correlational properties, and incremental developmental trends. The data presented in this article may be considered to provide only equivocal support for the first and second of these criteria. The third criterion was not a major focus of this study, but, as we have already pointed out, is not a necessary condition for many traditional forms of intelligence. Certainly, comparison of the present data with those from studies of self-report measures of EI (e.g., Dawda & Hart, 2000) suggests that the MEIS performs better, in that it seems distinct from existing personality measures. Unfortunately, consensus- and expert-scored EI also appears to be distinct (and in some cases independent) from each other. Factor correlations are insufficient to suggest that the two forms of scoring provide alternative measures of the same underlying construct. Consensus- and expert-scored scales also differ in their relationships with sex and ethnic group. Validity is demonstrated in some respects by the correlational data. Again, however, the two forms of scoring appear to support constructs that show only partial overlap, as evidenced by the lack of consistency in the linear associations between the two sets of branch scores and the personality and ability measures.

In the passages that follow, we consider in more detail the answers provided by the present investigation to three questions posed in the introduction, which relate to the psychometric adequacy of the MEIS, group differences, and personality and ability correlates. Throughout, we also touch on the fourth issue raised therein—the extent to which consensus and expert-scoring protocols converge. We conclude by identifying the key issues that will determine whether EI comes to be seen as a real addition to our understanding of personal qualities or as a construct that is largely illusory. An important caveat is that the MEIS is a relatively new instrument still under development. Indeed, it is now being replaced by a new measure (MSCEIT) that may prove to have stronger psychometric properties. However, its theoretical underpinnings are identical to that of the MEIS, and data published so far seem similar to those obtained with the MEIS (see, e.g., Mayer, Caruso, & Salovey, 2000). It is likely that the present results should generalize (in intelligence research, older forms of a test are expected to correlate substantially with newer forms), but we remain guarded in extending present findings to the newer performance-based EI measure.

Psychometric Status of MEIS Scales

Some features of the psychometric analyses support Mayer, Caruso, and Salovey's (1999) claim that EI meets criteria for an intelligence. We replicated the finding of a positive manifold between subtests of the MEIS, and, generally, the pattern of correlations corresponded well to the Mayer, Caruso, and Salovey (1999) findings. Exploratory and confirmatory factor analyses showed broad similarities with Mayer, Caruso, and Salovey (1999) factor solutions, although there were some differences in detail, and, in the exploratory analyses, subscale communalities were often low. In fact, the confirmatory analyses tend to support Mayer, Caruso, and Salovey's (1999) initial conception of four branches of EI, rather than the three-factor model that has subsequently been derived.

However, other aspects of the data render many of the EI concepts more problematic than is acceptable for ability measures. This is no small point because, beyond research purposes, the MEIS (or its derivative, the MSCEIT) may be attractive to practitioners as a selection device. In particular, the reliability of subtests that form the highest branches of the model, and are thus probably the most important components of the MEIS for predictor of real-world social behaviors (e.g., progressions, managing others), is among the poorest in this battery. In addition, intercorrelations between subtests, although resulting in positive manifold, are notably lower than is common in research involving cognitive ability measures (compare, for example, the data presented here with various data sets presented in Carroll, 1993). Further, various factor analyses indicate a structure that is relatively unstable, certainly when compared with similar analyses that have been conducted with intelligence and personality measures.

Perhaps the most severe psychometric difficulty is the lack of convergence between expert- and consensus-scored dimensions. There are instances of agreement, especially for the Blends and Progressions tests, but in general, cross-correlations are too small to suggest convergence. The correlation between the general factors extracted from each of the two data sets was only .26. The correlations between corresponding Perception factors ($r = .00$) and Emotion Management factors ($r = .48$) seem to fall short of even the rough equivalence claimed by Mayer, Caruso, and Salovey (1999). The correlation for Understanding factors ($r = .74$) is more satisfactory, but still falls short of the correlations of .80, .90, or better that we would expect for alternative forms of a cognitive IQ test. To put these figures in perspective, consider that the cross-correlation places an upper bound on the reliability (i.e., the extent to which two test versions are measuring the same underlying construct). The standard error of measurement (SEM) of a scale score is calculated as $SD\sqrt{(1 - r)}$, where r is the reliability and SD is the standard deviation (which equals 1 for factor scores). Hence, for an r of .48, *SEM* is 0.72 *SD,* and thus the 95% confidence interval for a true score S would be $S \pm 1.41$ *SD* (i.e., $1.96 \times SEM$). For an r of .74, the interval is $S \pm 1.00$ *SD.* For an IQ test ($M = 100$, $SD = 15$) with an r of .48, the score of a person at the mean would be expected to vary between 79 and 121 on different occasions of testing—not a satisfactory state of affairs. In the case of the MEIS scales, it is likely that scores on each version would fluctuate less, in practice, because each form of scoring may identify nonshared, unique variance that will raise reliability. However, the illustrative data highlight that measurement of whatever is common to corresponding expert- and consensus-scored factors does not meet normal standards for reliability. High *SEM* adversely impacts the use of the instrument to compare individuals and to predict criteria, severely constraining practical utility. It is difficult to see how the degree of correspondence could be improved to an acceptable level by minor changes to tests or scoring procedures. In the case of the general factors, and two of the branches, it appears that expert and consensus scoring assess substantially different personal qualities.

A further difficulty derives from the patterning of loadings on the consensus- and expert-scored general factors. In cognitive-abilities research, a ubiquitous finding is that the lower order sensory processes encompassing the domain of human intelligence (e.g., auditory reception, tactile–kinesthetic perception) share lower loadings on the first principal component than do higher order cognitive processes (e.g., inductive and deductive reasoning; see, e.g., Carroll, 1993; Horn, 1988; Roberts et al., 2000). As discussed in the introduction, EI also encompasses both lower order and higher order processes, arranged hierarchically by Mayer, Salovey, and Caruso (2000a), implying that factor loadings for EI branches should vary with level in the hierarchy. In short, if EI really is a type of intelligence, then the emotion–perception branch should show the lowest loading on a general EI factor, whereas the management branch should have the highest loading (with Assimilation and Understanding–Emotion factors lying in between). As shown in Table 7, we found a result similar to those of Mayer, Caruso, and Salovey (1999) and Ciarrochi et al. (2000). That is, the highest loadings come from the Emotion Perception subtests, with Managing Emotion sharing amongst the lowest factor loading on the general EI factor. (This finding is, in any event, prefigured by the rather low correlations that management of emotion–branch tests share with traditional cognitive abilities.) The fact that with the MEIS we find exactly the opposite to results expected of traditional cognitive abilities is a significant point and one that perhaps has been ignored for too long. If EI is to be regarded truly as a type of intelligence, should it not mimic many of the replicable findings demanded of an intelligence? If not, might not the MEIS general factor be seen as some lower level attribute of emotional processing rather than as an intelligence?

Group Differences

We qualify our remarks on group differences by noting that military samples may not be representative of the general population. Nevertheless, the divergence of consensus- and expert-based scores is a cause for great concern. Within the field of cognitive abilities, what is known of gender and ethinic differences is that the patterns underlying them consistently emerge for scoring procedures (and, indeed, dependent variables) that are widely disparate. For example, Jensen (1998) has shown that salient group differences extend beyond measures of accuracy to measures of processing speed, with recent evidence indicating that this might also hold true for putative indices of meta-cognition (see, e.g., Pallier et al., 2001). In the case of performance-based EI, there is considerable overlap in the procedures used to derive objective scores, but the final results are equivocal. For example, do we take the evidence presented in this article to indicate that, in the U.S. Air Force, male enlistees are more emotionally intelligent than their female counterparts or that female enlistees are more emotionally intelligent than their male counterparts? Even though the sample of women was relatively small, and perhaps atypical of women in general, a reliable test should give an unequivocal answer to such a question.

The experts who provided the criteria for that scoring procedure were White males (with high verbal intelligence). It is troubling that gender, ethnic, and intelligence differences exist in expert scores (favoring groups to which the experts belonged) and that the obverse was often true when consensus scoring procedures were implemented. Certainly, this finding calls into question the objectivity of expert-based scoring protocols that currently comprise the MEIS. It also raises the issue of how the data used for consensus scoring should be constituted. Mayer, Caruso, and Salovey (1999) pointed out that their consensus group (that was also used here) was biased toward women (67%), although they also present analyses of one subtest suggesting that the gender difference in favor of women was not caused by differences in male–female criteria.

Personality and Ability Correlates of the MEIS

In favor of the hypothesis that the MEIS assesses an intelligence, branch scores correlated moderately and positively with general intelligence. Roberts et al. (2000) provide data indicating that the ASVAB-AFQT score is a particularly good indicator of acculturated knowledge (or Gc). Thus, the relatively substantial correlations between this measure and the assimilation-, understanding-, and management-branch scores are exactly as might be predicted of a cognitive ability that has close links to social intelligence (see, e.g., Carroll, 1993; Davies et al., 1998). Equally, lower (yet positive) correlations between the Mechanics and Administration ASVAB composites and MEIS scores are entirely consistent with the fact that these represent broad cognitive abilities that have relatively weak relationships with Gc (i.e., broad visualization and clerical–perceptual speed, respectively; see Roberts et al., 2000).

It is interesting that the association between MEIS total score and intelligence is substantially derived from the understanding branch. For consensus scores, the bivariate association between understanding and AFQT score explains 14.3% of the variance in the intelligence measure, and the remaining three branches explain only a further 1.0% of the variance. Similarly, for expert scores, understanding explained 14.0% of AFQT score, with the other three branches contributing an additional 4.3%. The subtests of this branch have a somewhat intellectual character in, for example, being able to define complex emotions in terms of simple ones (Complex Blends). Other subtests, such as those requiring the understanding of typical emotions and transitions, may have some commonality with Gc tests that aim to assess general knowledge and practical judgment, such as the Wechsler Adult Intelligence Scale—Revised (WAIS-R) information and comprehension tests. One concern about this branch is the extent to which it assesses abstract, explicit knowledge of emotions—of the type that may be acquired from an undergradate psychology course. EI may reside in the more contextually bound, and possibly implicit, knowledge that supports action in real-world situations.

As noted in the introduction, self-report assessment has consistently resulted in substantial correlations between EI and personality (e.g., Dawda & Hart, 2000). This has led several researchers to question whether EI can contribute anything to the science of psychological assessment that is not already encapsulated by personality theory (e.g., Davies et al., 1998). The correlations obtained here with Big Five personality factors were qualitatively similar to those found with self-report EI measures. However, the relatively small magnitudes of the associations between the Big Five personality factors and each of the MEIS branch-scores also offer promise for the conceptual independence of performance-based EI. Performance-based EI shares relationships with measures of personality that resemble the relationships that traditional, objective measures of intelligence share with personality. Both EI and traditional intelligence tend to correlate only modestly (i.e., seldom exceeding .30) with personality scales. Furthermore, correlations tend to vary across the type of intelligence domain (i.e., primary mental ability or branch) assessed. Multiple regression analysis identified two relationships that seem both meaningful and consistent across scoring methods: (a) the relationship between Management and Extraversion and (b) that between Understanding and Openness.

Although the magnitude and direction of correlates was generally as expected, these data also raise problems. Again, there were discrepancies between expert and consensus scoring. Consensus scores were generally more predictive of personality in both bivariate and multivariate analyses. The multiple regressions revealed that some branch scores, especially emotion perception, showed significant relationships of opposing sign in analyses of expert- and consensus-based data. In consensus-scored data, the person adept at identifying emotions is somewhat unintelligent, agreeable, and extraverted, but in expert-scored data, these qualities are associated with poor emotion perception. Consensus-scored emotion management was a relatively strong predictor of Conscientiousness and Agreeableness, a result consistent with expectation, but these relationships were near zero in the expert-scored data. In some cases, the branch-score data showed a lack of coherent associations with criteria. In the consensus-scored data, it is unclear why Emotion Perception should be negatively related to general intelligence when assimilation and management show positive relationships, in the multiple regression analyses. Likewise, there is no obvious rationale for agreeable individuals being emotionally intelligent with respect to perception and management but unintelligent with respect to understanding. At the least, the ability model of EI requires a more detailed account of the role of personality.

Perspectives on EI

The data analyses raise both some immediate problems with assessment of EI using the MEIS and some more fundamental issues about the contribution of research on EI. It is likely that some of the psychometric problems we have indicated are soluble through the normal processes of test refinement. There seems to be no barrier in principle to the development of internally consistent scales whose intercorrelations support a replicable three- or four-branch factor solution. However, the data also highlight the more fundamental problem of differences between scoring methods. Dimensions derived from expert and consensus scores fail to correlate well, at either scale or factor levels, and it is especially disturbing that the general factors extracted from the two types of data correlated only .26. Furthermore, expert- and consensus-based scores give very different patterns of association with group and individual difference factors. The discrepancies are sufficiently large that they imply that one or the other scoring method should be discarded, in that it is hard to envisage modifications that would bring factors that are correlated at less than .50 into alignment.

An optimistic perspective on EI would consider the problems of different scoring protocols as surmountable. Indeed, they would appear as the type of conceptual problem that plagued Binet, Simon, Wechsler, and their predecessors historically, when they first tried to develop intelligence tests (see, e.g., Gregory, 1996). In the emotional domain, it may be inappropriate to insist that test items should have rigid, unequivocal right and wrong answers. Recently, Salovey et al. (2000) recommended a consensus-based approach to scoring the MEIS on the basis that large normative samples tend to be reliable judges of emotional questions. More generally, consensus scoring (or its derivatives) has had great survival value, witnessed in chronicles from the first century A.D. (e.g., gladiators in the Coliseum) to the present day (e.g., many of our political leaders are elected through this process).[16] Given that intelligence is often equated

[16] There is even now widespread interest in the entertainment value of consensus scoring in so-called "reality TV"—witness, for example, the international popularity of the CBS-produced *Survivor* series (Burnett, 2000).

with the ability of the organism to adapt to its environment (see, e.g., Sternberg, 1985; Thorndike et al., 1921), the ecological validity of this scoring procedure per se should not be underestimated.

However, there is a further twist to the role of consensus agreement. In general terms, we would expect a person whose beliefs match the group consensus to be better adapted to the social environment than someone whose beliefs deviate from the group. The conformist is likely to receive more positive reinforcement from others and to negotiate challenges shaped by social convention, such as finding a job or a life partner more easily. In other words, consensual EI may be adaptive not because it refers to any cognitive advantage, but because of the social advantages of conformity. Consistent with this suggestion, people with generally desirable personality characteristics such as conscientiousness and agreeableness seem to be perceived by others as better at emotion management, as evidenced by the consensus-based data here, but this link is not evident in the expert-scored data. A conformity construct is of real-world relevance, but it is highly misleading to label it as an intelligence, because it relates to person–environment fit rather than to any characteristic of the individual. Indeed, in some instances it is the nonconformist who should be deemed emotionally intelligent, for example, a writer or an artist who finds a new and original way of expressing an emotion.

Another possibility is to develop some hybrid scoring protocol. For example, deriving consensus scores from the corpus of experts–professionals who read this journal or who are members of organizations such as the International Society for Research on Emotions seems feasible and conceptually justifiable. Within this hybrid model, it is expert consensus that forms a theoretically defensible scoring criterion, assuming that problems of bias associated with the demographic characteristics of experts may be avoided.

Against these proposals, we may wonder whether an intelligence may ever be satisfactorily defined by consensus, even expert consensus. We have previously discussed the difficulties in principle in deciding on the best or most adaptive response to an emotional encounter (Matthews & Zeidner, 2000; Matthews, Zeidner, & Roberts, in press). If there is no right answer, consensus is of no validity. Furthermore, because consensus methods are largely a function of the particular group assessed (cultural, ethnic, age, or gender), what may be the consensus and modal response for one group may not be so for another group, making scores relatively incomparable across different groups (e.g., ethnic, national) of examinees. There are also difficult issues about the legitimacy and ownership of expertise: Many disciplines and subdisciplines may claim primacy. It is also questionable whether well-educated, professional people should solely set standards for EI.

Conclusions

Mayer, Salovey, and Caruso (2000a) deserve much credit for formulating a novel, clearly articulated model of EI and seeking to operationalize its constructs as ability tests through careful construct validation studies. The view expressed here is that despite the merits of this project, there are significant measurement problems to be overcome. Perhaps inevitably, given the level of interest in EI, criticism of work on EI has already invoked strong emotions in many of its chief proponents and critics. For example, Salovey et al. (2000) equated the conclusions reached by Davies et al. (1998) with the French Academy of Sciences decision, at a time when logical positivism was growing, to destroy all meteorites housed in museums because they were “heavenly bodies” and heaven did not exist. We aim to conclude with a balanced appraisal of the promise and problems of ability-model EI research. From a positive perspective, EI is a real quality of the person, distinct from existing personality and ability factors and best measured by performance-based tests such as the MEIS. The problems raised here may be essentially technical problems to be solved by advances in test construction.

The sceptic may prefer another astronomical metaphor. In 1877, the Italian astronomer, Schiaparelli, announded that his telescope revealed linear channels on the surface of Mars. His observation inspired Percival Lowell to map hundreds of canals in fine detail, and to enthrall the general public with his popular lectures on the construction of the canals by a martian civilization. There are various suggestive parallels with EI. Respected scientists, who made important contributions to other areas of astronomy, reported the initial observations. The canals were not completely fanciful; Mars does have surface markings (of an irregular nature). An elaborate empirical and theoretical artiface was constructed from fairly modest beginnings, and popular interest was sparked by excessive speculation. It remains to be seen whether EI, like the canals of Mars, is the product of the tendency of even expert observers to see, in complex data, patterns that do not exist.

At this early stage of research, it is premature to label EI as either a real "meteorite" or an illusory "martian canal," but the present study has identified significant issues that require resolution, related to both reliability and validity. The most severe problems relate to scoring (i.e., reliability), including the difficulty of justifying both expert and consensus scoring, the limited psychometric convergence of these methods, and their differing relationships with other criteria. At the least, further progress requires a single scoring method with a strong rationale that what is being measured is a form of cognitive ability. In particular, it should be demonstrated that items have better and worse answers with respect to some objective criterion (although human judgment may be required to assess how well answers meet the criterion). The difficult problems of group differences and possible cultural and gender bias must also be addressed. The data provide some support for Mayer, Salovey, and Caruso's (2000a) *correlational criteria,* for example, by demonstrating only modest correlation between EI and general intelligence. The data also showed some overlap with personality constructs. Although performance-based EI measures appear to be free of the redundancy with existing personality scales that plagues questionnaire measures, the validity coefficients for the MEIS also appear to be typically small and often less than .30 (see Mayer, Salovey, & Caruso, 2000a). It is unclear whether the predictive validity of the MEIS would be maintained with personality and ability controlled statistically. Finally, it is not established that EI, as operationalized by the MEIS, is a major construct on a par with general intelligence as a determinant of individual differences in meaningful, real-world behavior. There may well be further primary ability dimensions such as emotion perception that are poorly assessed by existing instruments. However, to assert that these abilities are as important in human functioning as general intelligence is to go far beyond the available data. An urgent task for future research is to show real-world adaptive advantages for high scores on the MEIS over and above those they obtain from their higher general intelligence and their personality characteristics.

References

Abraham, R. (2000). The role of job control as a moderator of emotional dissonance and emotional intelligence: Outcome relationships. *The Journal of Psychology, 134,* 169–184.

American Psychological Association Public Affairs Office. (1997). *Intelligence: Knowns and unknowns.* Washington, DC: Author.

Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Upper Saddle River, NJ: Prentice Hall.

Arbuckle, J. L. (1997). AMOS user's guide (Version 3.6). [Computer Software]. Chicago: Smallwaters Corporation.

Australian Council of Educational Research (ACER). (1989). *Standard progressive matrices: Australian manual.* Hawthorne, Victoria, Australia: Author.

Bagby, R. M., Parker, J. D., & 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.

Bar-On, R. (1997). *The Bar-On Emotional Quotient Inventory (EQ-i): Technical manual.* Toronto, Ontario, Canada: Multi-Health Systems.

Bar-On, R. (2000). Emotional and social intelligence: Insights from the emotional quotient inventory (EQ-i). In R. Bar-On & J. D. A. Parker (Eds.), *Handbook of emotional intelligence* (pp. 363–388). San Francisco: Jossey-Bass.

Bar-On, R., Brown, J. M., Kirkcaldy, B. D., & Thome, E. P. (2000). Emotional expression and implications for occupational stress: An application of the Emotional Quotient Inventory (EQ-i). *Personality and Individual Differences, 28,* 1107–1118.

Brody, N. (1992). *Intelligence* (2nd ed.). San Diego, CA: Academic Press.

Burnett, M. (Producer). (2000). *Survivor.* New York: CBS.

Carretta, T. R., & Ree, M. J. (1995). Near identity of cognitive structure in sex and ethnic groups. *Personality and Individual Differences, 19,* 149–155.

Carretta, T. R., & Ree, M. J. (1997). Negligible sex differences in the relation of cognitive and psychomotor abilities. *Personality and Individual Differences, 22,* 165–172.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* New York: Cambridge University Press.

Carroll, J. B. (1997). Psychometrics, intelligence, and public perception. *Intelligence, 24,* 25–52.

Cattell, R. B. (1971). *Abilities: Their structure, growth, and measurement.* Boston: Houghton Mifflin.

Christal, R. E. (1994, November). Non-cognitive research involving systems of testing and learning. *Final Research and Development Status Report* (USAF Contract No. F33615-91-D-0010). Armstrong Laboratory, Brooks Air Force Base, United States Air Force.

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.

Ciarrochi, J. V., Chan, A. Y. C., Caputi, P., & Roberts,

R. D. (2001). Measuring emotional intelligence. In J. Ciarrochi, J. P. Forgas, & J. D. Mayer (Eds.), *Emotional intelligence in everyday life: A scientific inquiry* (pp. 25–45). Philadelphia: Psychology Press.

Cooper, R. K., & Sawaf, A. (1997). *Executive EQ: Emotional intelligence in leadership and organizations.* New York: Putnam.

Costa, P. T., Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI–R) and NEO Five-Factor Inventory (NEO–FFI).* Odessa, FL: Psychological Assessment Resources.

Cronbach, L. J. (1960). *Essentials of psychological testing.* New York: Harper.

Davies, M., Stankov, L., & Roberts, R. D. (1998). Emotional intelligence: In search of an elusive construct. *Journal of Personality and Social Psychology, 75,* 989–1015.

Dawda, D., & Hart, S. D. (2000). Assessing emotional intelligence: Reliability and validity of the Bar-On Emotional Quotient Inventory (EQ-i) in university students. *Personality and Individual Differences, 28,* 797–812.

Ekman, P. (1999). Facial expressions. In T. Dalgleish & M. J. Power (Eds.), *Handbook of cognition and emotion* (pp. 301–320). Chichester, UK: Wiley.

Elias, M. J., & Weissberg, R. (2000). Primary prevention: Educational approaches to enhance social and emotional learning. *Journal of School Health, 70,* 186–190.

Epstein, S. (1998). *Constructive thinking: The key to emotional intelligence.* Westport, CT: Praeger.

Fisher, C. D., & Ashkanasy, N. M. (2000). The emerging role of emotions in work life: An introduction. *Journal of Organizational Behavior, 21,* 123–129.

Flynn, J. R. (1987). Massive gains in 14 nations: What IQ tests really measure. *Psychological Bulletin, 101,* 171–191.

Foley, P., & Rucker, L. S. (1989). An overview of the Armed Services Vocational Aptitude Battery (ASVAB). In R. F. Dillon & J. W. Pellegrino (Eds.), *Testing: Theoretical and applied perspectives* (pp. 16–35). New York: Praeger.

Fox, S., & Spector, P. E. (2000). Relations of emotional intelligence, practical intelligence, general intelligence, and trait affectivity with interview outcomes: It's not all just 'G'. *Journal of Organizational Behavior, 21,* 203–220.

Garcia, A., & Roberts, R. D. (2000, April) *Emotional 'intelligence', cognitive abilities, and personality.* Paper presented at *the 27th Annual Conference of the Australasian Experimental Psychology Society,* Novotel Twin Waters Resort, Queensland, Australia.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences.* New York: Basic Books.

Gibbs, N. (1995, October 2). The EQ factor. *Time,* 60–68.

Goleman, D. (1995). *Emotional intelligence.* New York: Bantam Books.

Gregory, R. J. (1996). *Psychological testing: History, principles, and applications* (2nd ed.). Needham Heights, MA: Allyn & Bacon.

Guilford, J. P. (1959). *Personality.* New York: McGraw-Hill.

Guttman, L. (1965a). A faceted definition of intelligence. In R. Eiferman (Ed.), *Studies in psychology. Scripta Hierosolymitana* [Jerusalem Scripts in Greek] (Vol. 14, pp. 166–181), Jerusalem: Magnes Press.

Guttman, L. (1965b). The structure of intercorrelations among intelligence tests. In *Proceedings of the 1964 Invitational Conference on Testing Problems* (pp. 53–65). Princeton, NJ: Educational Testing Service.

Guttman, L., & Levy, S. (1991). Two structural laws for intelligence. *Intelligence, 15,* 79–103.

Hammer, D. (1997, September). *Air Force operational viewpoint.* Paper presented at the United States Air Force Office of Scientific Research (AFOSR) Chronobiology and Neural Adaptation Conference, Colorado Springs, CO.

Herrnstein, R. J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life.* New York: Free Press.

Horn, J. L. (1988). Thinking about human abilities. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (2nd ed., pp. 645–685). New York: Plenum.

Horn, J. L., & Hofer, S. M. (1992). Major abilities and development in the adult period. In R. J. Sternberg & C. Berg (Eds.), *Intellectual development* (pp. 44–99). New York: Cambridge University Press.

Horn, J. L., & Noll, J. (1994). System for understanding cognitive capabilities: A theory and the evidence on which it is based. In D. K. Detterman (Ed.), *Current topic in human intelligence: Volume IV* (pp. 151–203). New York: Springer-Verlag.

Horn, J. L., & Stankov, L. (1982). Auditory and visual factors of intelligence. *Intelligence, 6,* 165–185.

Hunt, E. (1995). *Will we be smart enough? A cognitive analysis of the coming workforce.* New York: Russell Sage Foundation.

Jensen, A. R. (1985). Armed Services Vocational Aptitude Battery. *Measurement and Evaluation in Counseling and Development, 18,* 32–37.

Jensen, A. R. (1998). *The g factor: The science of mental ability.* Westport, CT: Praeger.

Kaplan, R. M., & Saccuzzo, D. P. (1997). *Psychological testing: Principles, applications, and issues* (4th ed.). Pacific Grove, CA: Brooks/Cole.

Kihlstrom, J. F., & Cantor, N. (2000). Social intelligence. In R. J. Sternberg (Ed.), *Handbook of intelligence* (2nd ed.). New York: Cambridge University Press.

Krause, H., Lee, S.-K., Stankov, L., Pallier, G., & Roberts, R. D. (in press). *Time management, personality, and individual differences.* Manuscript submitted for publication.

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

Mabe, P. A., & West, S. G. (1982). Validity of self-evaluation of ability: A review and meta-analysis. *Journal of Applied Psychology, 67,* 280–296.

Matthews, G., & Zeidner, M. (2000). Emotional intelligence, adaptation to stressful encounters, and health outcomes. In R. Bar-On & J. D. A. Parker (Eds.), *Handbook of emotional intelligence* (pp. 459–489). San Francisco: Jossey-Bass.

Matthews, G., Zeidner, M., & Roberts, R. (in press). *Emotional intelligence: Science and myth.* Cambridge, MA: MIT Press.

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

Mayer, J. D., Caruso, D., & Salovey, P. (2000). Selecting a measure of emotional intelligence: The case for ability scales. In R. Bar-On & J. D. A. Parker (Eds.), *The handbook of emotional intelligence* (pp. 320–342). New York: Jossey-Bass.

Mayer, J. D., & Cobb, C. D. (2000). Educational policy on emotional intelligence: Does it make sense? *Educational Psychology Review, 12,* 163–183.

Mayer, J. D., DiPaulo, M., & Salovey, P. (1990). Perceiving affective content in ambiguous visual stimuli: A component of emotional intelligence. *Journal of Personality Assessment, 54,* 772–781.

Mayer, J. D., & Geher, G. (1996). Emotional intelligence and the identification of emotion. *Intelligence, 22,* 89–113.

Mayer, J. D., & Salovey, P. (1993). The intelligence of emotional intelligence. *Intelligence, 17,* 443–450.

Mayer, J. D., & Salovey, P. (1997). What is emotional intelligence? In P. Salovey & D. Sluyter (Eds.), *Emotional development and emotional intelligence: Implications for educators* (pp. 3–31). New York: Basic Books.

Mayer, J. D., Salovey, P., & Caruso, D. (2000a). Competing models of emotional intelligence. In R. J. Sternberg (Ed.), *Handbook of human intelligence* (2nd ed., pp. 396–420). New York: Cambridge University Press.

Mayer, J. D., Salovey, P., & Caruso, D. (2000b). Emotional intelligence as zeitgeist, personality, and as a mental ability. In R. Bar-On & J. D. A. Parker (Eds.), *The handbook of emotional intelligence* (pp. 92–117). New York: Jossey-Bass.

McCrae, R. R. (2000). Emotional intelligence from the perspective of the five-factor model of personality. In R. Bar-On & J. D. A. Parker (Eds.), *The handbook of emotional intelligence* (pp. 263–276). New York: Jossey-Bass.

Mehrabian, A. (2000). Beyond IQ: Broad-based measurement of individual success potential or "emotional intelligence." *Genetic, Social, and General Psychology Monographs, 126,* 133–239.

Mehrabian, A., & Epstein, N. (1970). A measure of emotional empathy. *Journal of Personality, 40,* 525–543.

Most, B., & Zeidner, M. (1995). Constructing personality and intelligence test instruments: Methods and issues. In D. Saklofske & M. Zeidner (Eds.), *International handbook of personality and intelligence* (pp. 475–503). New York: Plenum.

Murphy, K. (1985). Armed Services Vocational Aptitude Battery. In D. J. Keyser & R. C. Sweetland (Eds.), *Test critiques* (Vol. 1). Kansas City, MO: Test Corporation of America.

Murphy, K. R., & Davidshofer, C. O. (1998). *Psychological testing: Principles and applications* (4th ed.). Upper Saddle River, NJ: Prentice Hall.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

O'Sullivan, M., Guilford, J. P., & de Mille, P. (1965). *The measurement of social intelligence.* (Rep. No. 34). Los Angeles, CA: University of Southern California, Psychometric Laboratory.

Pallier, G., Wilkinson, R., Danthiir, V., Kleitman, S., Knezevic, G., Stankov, L., & Roberts, R. D. (2001). *The role of question format and individual differences in the realism of confidence judgments.* Manuscript submitted for publication.

Paul, A. M. (1999). Promotional intelligence. *Salon.com* [On-line]. Available Internet: http://www.salon.com/books/it/1999/06/28/emotional/index3.html.

Paulhus, D. L., Lysy, D. C., & Yik, M. S. (1998). Self-report measures of intelligence: Are they useful as proxy IQ tests? *Journal of Personality, 66,* 525–554.

Pavot, W., Diener, E., & Suh, E. (1998). The Temporal Satisfaction with Life Scale. *Journal of Personality Assessment, 70,* 340–354.

Petrides, K. V., & Furnham, A. (2000). On the dimensional structure of emotional intelligence. *Personality and Individual Differences, 29,* 313–320.

Raven, J. C., Court, J. H., & Raven, J. (1979). *Manual for Raven's Progressive Matrices and Vocabulary Scales.* London: H. K. Lewis & Co.

Ree, M. J., & Carretta, T. R. (1994). Factor analysis of

ASVAB: Confirming a Vernon-like structure. *Educational and Psychological Measurement, 54,* 457–461.

Ree, M. J., & Carretta, T. R. (1995). Group differences in aptitude factor structure on the ASVAB. *Educational and Psychological Measurement, 55,* 268–277.

Roberts, R. D. (in press). Review of the handbook of emotional intelligence. *Intelligence.*

Roberts, R. D., Goff, G. N., Anjoul, F., Kyllonen, P. C., Pallier, G., & Stankov, L. (2000). The Armed Services Vocational Aptitude Battery: Little more than acculturated learning (Gc)!? *Learning and Individual Difference, 12,* 81–103.

Roberts, R. D., & Kyllonen, P. C. (1999). Morningness–eveningness and intelligence: Early to bed, early to rise will likely make you anything but wise. *Personality and Individual Differences, 27,* 1123–1133.

Roger, D., & Najarian, B. (1989). The construction and validation of a new scale for measuring emotion control. *Personality and Individual Differences, 10,* 845–853.

Ryback, D. (1998). *Putting emotional intelligence to work: Successful leadership is more than IQ.* Stoneham, MA: Butterworth-Heineman.

Saarni, C. (1999). *The development of emotional competence.* New York: Guilford Press.

Salovey, P., & Mayer, J. D. (1990). Emotional intelligence. *Imagination, Cognition and Personality, 9,* 185–211.

Salovey, P., Mayer, J. D., Goldman, S., Turvey, C., & Palfai, T. (1995). Emotional attention, clarity, and repair: Exploring emotional intelligence using the Trait Meta-Mood Scale. In J. W. Pennebaker (Ed.), *Emotion, disclosure, and health* (pp. 125–154). Washington, DC: American Psychological Association.

Salovey, P., Woolery, A., & Mayer, J. D. (2000). Emotional intelligence: Conceptualization and measurement. In G. Fletcher & M. Clark (Eds.), *The Blackwell handbook of social psychology.* London: Blackwell.

Scherer, M. (1997). Perspectives: Heart start. *Educational Leadership, 54,* 4.

Schutte, N., Malouff, J., Hall, L., Haggerty, D., Cooper, J., Golden, C., & Dornheim, L. (1998). Development and validation of a measure of emotional intelligence. *Personality and Individual Differences, 25,* 167–177.

Spearman, C. (1927). *The abilities of man.* New York: Macmillan.

Stankov, L., Boyle, G. J., & Cattell, R. B. (1995). Models and paradigms in personality and intelligence research. In D. Saklofske & M. Zeidner (Eds.), *International handbook of personality and intelligence. Perspectives on individual differences* (pp. 15–43). New York: Plenum Press.

Stauffer, J. M., Ree, M. J., & Carretta, T. R. (1996). Cognitive-components tests are not much more than *g*: An extension of Kyllonen's analyses. *Journal of General Psychology, 123,* 193–205.

Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence.* New York: Cambridge University Press.

Sternberg, R. J. (2000a). The concept of intelligence. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 3–15). New York: Cambridge University Press.

Sternberg, R. J. (Ed.). (2000b). *Handbook of intelligence.* New York: Cambridge University Press.

Stoeber, J. (1998). Reliability and validity of two widely-used worry questionnaires: Self-report and self-peer convergence. *Personality and Individual Differences, 24,* 887–890.

Taylor, G. J. (2000). Recent developments in alexithymia theory and research. *Canadian Journal of Psychiatry, 45,* 134–142.

Thorndike, E. L. (1920). Intelligence and its use. *Harper Magazine, 140,* 227–235.

Thorndike, E. L., Terman, L. M., Freeman, F. M., Colvin, S. S., Pintner, R., & Pressey, S. L. (1921). Intelligence and its measurement: A symposium. *Journal of Educational Psychology, 12, 3,* 123–147.

Thurstone, L. L. (1931). Multiple factor analysis. *Psychological Reveiw, 38,* 406–427.

Thurstone, L. L. (1938). *Primary mental abilities.* Chicago: University of Chicago Press.

U.S. Department of Defense. (1982). *Profile of American youth: 1980 Nationwide administration of the Armed Services Vocational Aptitude Battery.* Washington, DC: U.S. Department of Defense, Office of the Assistant Secretary of Defense (Manpower, Reserve Affairs, and Logistics).

U.S. Department of Defense. (1984). *Test Manual for the Armed Services Vocational Aptitude Battery* (DoD No. 1340.12AA). North Chicago, IL: U.S. Military Entrance Processing Command.

Weinberger, D. A., Schwarz, G. E., & Davidson, R. J. (1979). Low-anxious, high-anxious, and repressive coping styles: Psychometric patterns and behavioral and physiological responses to stress. *Journal of Abnormal Psychology, 88,* 369–380.

Weisinger, H. (1998). *Emotional intelligence at work: The untapped edge for success.* San Francisco: Jossey-Bass.

Zeidner, M., & Feitelson, D. (1989). Probing the validity of intelligence tests for preschool children: A smallest space analysis. *Journal of Psychoeducational Assessment, 7,* 175–193.

Received September 27, 2000
Revision received April 4, 2001
Accepted April 4, 2001 ■