Citation:

Wilhelm, O. (2005). Measures of emotional intelligence: Practice and standards. In R. Schulze & R. D. Roberts (Eds.), *Emotional intelligence: An international handbook* (pp. 131-154). Hogrefe & Huber.

7

Measures of Emotional Intelligence: Practice and Standards

Oliver Wilhelm
Humboldt University of Berlin, Germany

Summary

In this chapter emotional intelligence (EI) is discussed from a psychometric perspective with a focus on ability measures. Prior research is used to demonstrate that in EI research, like in other psychological fields, measures addressing the same construct but being based on performance or self-report show little to no convergence. It is argued that performance based measures are better suited as indicators of EI. The Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT) as one such measure is presented and its validity is discussed. Whether or not ability measures of EI can be considered to be intelligence tasks is considered from several perspectives. This critique of EI measures tries to outline research questions warranting more attention in the future. The proposed recommendations include (a) trying to develop tasks with a strong background in Emotion Psychology, (b) using a broader variety of tasks in multivariate studies, and (c) using more appropriate criteria in validating EI.

7.1 INTRODUCTION

When new constructs of individual differences are introduced into psychology scientists are supposed and expected to react fairly skeptically, critically, and

conservatively. When new measures are associated with new constructs things get even tougher. There might be two causes for these defensive routine reactions. First, viewed historically, lay persons did not contribute valuable constructs and measurement instruments to individual differences research, and although it was psychologists investigating the idea of an emotional intelligence (EI) first (Mayer, DiPaolo, & Salovey, 1990), it was popularized—even within psychology—by lay persons (Goleman, 1995, 1998). Second, psychologists feel the need to legitimize why they make such a big fuss about *their* measures of dispositions of persons (i.e., what makes their personality measures any different from the ad hoc questionnaires in Cosmopolitan magazine). These routine reactions make good sense in order to avoid false positives when it comes to establishing new constructs and new measures—on the other side there is the threat of being overly cautious and rejecting new ideas and new measures even though they might be worth being further investigated, developed, and used in practical settings. Being overly conservative might cause an unacceptable high rate of false negatives. Slightly simplifying historical events (see Matthews, Zeidner, & Roberts, 2002, for an adequate description) EI intruded the quiet waters of individual differences research, testing, and assessment in the early 1990s (Mayer et al., 1990; Mayer & Salovey, 1993; Salovey & Mayer, 1990) and sparked strong public interest (Goleman, 1995, 1998) in the construct and its measurement subsequently. This public interest can be considered to be indicative of the demand that is more or less satisfied through measurement instruments developed within the scientific community. Some researchers are investigating the construct to the best of their knowledge and abilities while others turn both thumbs down and direct the construct and associated measures to psychology's unmarked grave of poor ideas.

This chapter will first focus on an important distinction between various instruments proposed for the measurement of EI: the assessment of typical versus maximal behavior. A brief evaluation of EI measures of typical behavior is followed by a more extensive discussion of measures of maximal behavior. The latter begins with a description of the Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT V.2; Mayer, Salovey, Caruso, and Sitarenios, 2002; Mayer, Salovey, Caruso, & Sitarenios, 2003), continues with requirements to classify a measure as an intelligence test, and concludes with a critique and some recommendations for future research.

7.2 TYPICAL AND MAXIMAL BEHAVIOR

A distinction between typical and maximal behavior was first drawn by Cronbach (1949). The distinction between measures of typical and maximal behavior is strongly associated with the content of a measure. Typical behavior is usually assessed with self-reports of preferences and valences. Sometimes life data are used to measure typical behavior. Maximal behavior is associated with measuring abilities, achievement, skills, and declarative knowledge. Situations in which maximal behavior is recorded are usually characterized by

(a) the assessed person being aware of the performance appraisal, (b) the assessed person being willing and able to demonstrate maximal performance, and (c) the standards for evaluating performance being adequate for assessment (Sackett, Zedeck, & Fogli, 1988).

The difference between performance based and self-report measures has several aspects. Performance based measurement procedures rely on maximal behavior, they are external appraisals of performances, they have minimal response bias, they are effortful and lengthy to administer and they are supposed to measure an "ability". On the other side, self-report based measures rely on reported typical behavior, they are internal appraisals of preferences, response bias can be substantial—specifically in high stakes testing, they are easy and quick to administer and they are supposed to measure personality-like constructs. Measures of typical behavior are used predominantly in personality psychology and measures of maximal behavior in individual differences in proficiencies, abilities, and achievement.

It is important to note that the distinction in typical and maximal performance leaves open how close to their maximal behavior people operate when behaving typically. Similarly, putting more effort into maximal behavior is not always possible or instrumental (Kahneman, 1973). Efforts to bridge the gap between constructs of maximal and typical behavior can be attempted from both sides. It is possible to assess personality constructs with measures of maximal behavior, and it is possible to assess abilities with measures of typical behavior (Riemann & Abels, 1994). There are several examples where the latter approach has been taken and it is possible to profit from considering these approaches when dealing with emotional intelligence.

First, in aging research there is a frequent use of self-reported memory complaints (Hertzog, Park, Morrell, & Martin, 2000) and these self-reports compete with objective measures of memory performance. Second, in clinical neuropsychology self-report measures have been developed that demonstrate a loss of insight into objectively measured performance decrements on measures of maximal behavior (McGlynn & Schacter, 1989; Seidenberg, Haltiner, Taylor, Hermann, & Wyler, 1994). Third, in cognitive psychology there are several self-report measures to assess attention slips and memory failures (Broadbent, Cooper, Fitzgerald, & Parkes, 1982; Herrmann, 1982; Reason, 1993) and these measures can be related to measures of maximal behavior of working memory, short-term memory, and attention (Oberauer, Süß, Schulze, Wilhelm, & Wittmann, 2000). Finally, in educational psychology, differential psychology, and social psychology there have been several self-report measures trying to capture typical academic and intellectual engagement (Cacioppo & Petty, 1982; Epstein, Pacini, Denes-Raj, & Heier, 1996; Goff & Ackerman, 1992; McCrae, 1990, 1996; Wilhelm, Schulze, Schmiedek, & Süß, 2003) and these questionnaires can be related to established intelligence tests.

In all of these areas researchers have not been successful in establishing substantial or high correlations. In fact, only for the last domain there are small to moderate relations between measures from both sides of the gap (i.e., between typical intellectual engagement and intelligence measures). It has been

argued in the past that traditional measures of maximal behavior are usually administered in controlled settings and that the laboratory context of the measurement keeps these tests from being useful predictors of relevant criteria (Dennis, Sternberg, & Beatty, 2000). Indeed, if one thinks about everyday activities the number and duration of situations in which humans behave to the best of their abilities might be quite limited. However, when it comes to prediction the power of measures of maximal behavior is soundly established and of substantial magnitude (Ones, Viswesvaran, & Dilchert, 2004).

Available evidence suggests that for EI the pattern of results found for comparable constructs summarized above is replicated—disregarding problems on the conceptual and empirical end for measures of both typical and maximal behavior. The pattern of results suggests that despite a considerable conceptual overlap of what constitutes EI, in the context of typical and maximal behavior there is little to no relation between measures from both ends (see e.g., O'Connor & Little, 2003). These zero correlations leave little room for alternative interpretations other than that both forms of measures assess distinct characteristics. Attributing the absence of a correlation to the relevance of method artifacts is not satisfactory if the goal is to establish a new construct that is associated with new measures (see Chapter 9 by Pérez, Petrides, & Furnham). In terms of multitrait-multimethod validation substantial correlations across methods and within a trait (monotrait-heteromethod) are required, and if measures of typical and maximal behavior are considered as different methods, these correlations are not of sufficient magnitude. On the other side, heterotrait-monomethod correlations should be low or zero, and they are typically not in the case of constructs assessed with self-reported EI measures. Ability measures of EI correlate modestly and meaningfully with other abilities. Similar results have been found in the domain of social intelligence (see Chapter 10 by Weis & Süß).

If measures of typical and maximal EI are unrelated they should not have the same label. EI apparently is intended to be an ability construct. Hence, self-report measures of EI should not be given the label intelligence. Workaround labels like "trait EI" do not resolve the problem because "ability"-based EI is considered to be a trait too.

7.3 SELF-REPORTED AND SELF-RATED EMOTIONAL INTELLIGENCE

The attempts to measure emotional intelligence are clearly twofold (Mayer, Caruso, & Salovey, 2000). On the one side there are traditional self-reports of typical behavior, and on the other side there are measures conceptually related to traditional ability measures. The latter will be labelled "ability models" here although it is not yet established whether or not these measures unequivocally qualify as ability measures. This issue will be addressed below in Section 7.5. Given that both forms of EI measurement are basically unrelated and given that the term *intelligence* is associated with the use of measures provoking

maximal behavior, self-report measures of EI should not include the term *intelligence*.

More profound than this terminological problem is the status of corresponding self-report measures. Such measures have been developed based on various definitions of what constitutes emotionally intelligent behavior. Bar-On (1997, 2000) distinguishes some 15 components of successful emotional functioning. These 15 components are organized within 5 broader interrelated dimensions including intrapersonal EI, interpersonal EI, adaptability EI, stress management EI, and general mood EI. The test corresponding to this model is called the BarOn Emotional Quotient Inventory EQ-i (Bar-On, 1997). However, the proposed as well as alternative structures could not be supported empirically (Palmer, Manocha, Gignac, & Stough, 2003; Petrides & Furnham, 2000, 2001). Similarly, the Schutte et al. Emotional Intelligence Inventory (Schutte et al., 1998) and its extensions (Saklofske, Austin, & Minski, 2003) has been extensively used but no final decision about its internal structure can be made at this time. Amongst other available measures the Trait Emotional Intelligence Questionnaire (TEIQue) seems to be the most promising candidate in terms of available evidence and effort in validating the measure (see Chapter 9 by Pérez et al.). The TEIQue is a measure with 144 items assigned to ten scales: adaptability, assertiveness, emotion perception, emotion expression, emotion regulation, empathy, low impulsivity, relationship skills, social competence, and stress management.

One problem with the TEIQue and similar measures is that the items are mostly taken from existent measures such as Emotional Empathy (Mehrabian & Epstein, 1970), the Toronto Alexithymia Scale (Bagby, Parker, & Taylor, 1994a, 1994), and other self-report measures of emotional intelligence. Technically then the TEIQue is mostly an assembly of existent items and the constructs assessed by the questionnaire therefore can hardly be new. A second problem for the TEIQue, as well as for similar measures, is that no satisfying measurement model on the item level for the total test or individual scales has been established so far. A third problem for all self-report measures of EI is that redundancy with competing and established constructs emerging from self-report measures has not been adequately assessed as yet. The last point is very important. Within individual differences research abundant efforts have been devoted to establish the dimensionality of traditional self-reports. The five-factor model (Costa & McCrae, 1992) is the most prominent of these efforts and within this model several lower-order facets have been proposed and investigated for each of the factors. Additionally, there is a broad variety of other self-report dimensions that have been investigated in the past. When new constructs based on self-reports are established, unequivocal evidence that individual differences on the new measure cannot be reduced on individual differences as assessed with available self-report measures is required. After controlling for a broad battery of competing self-report dimensions, measures of the new construct should still be meaningfully and substantially related with each other. Additionally, the new measure should incrementally predict interesting criteria over and above competing self-report dimensions and other estab-

lished predictors. To date there seems to be no scientific evidence that support the unidimensionality, incremental validity, and utility of self-reported EI.

Therefore, the self-report measures for EI proposed so far should not be labelled EI. Available evidence does not prove that these measures assess something new. Considering these measures as indicators of a new construct—say emotional self-efficacy—requires more sophisticated embedding into related and established nomological nets.

Self-ratings of abilities fall in the "no-man's-land" between measures of typical and maximal behavior (Stankov, 1999). It is not unusual to find items that represent self-ratings of abilities in measures that are supposed to assess some self-report dimension. Items like "I am good in expressing my moods and feelings" are not very far from structured attempts to measure self-ratings of emotional intelligence. Such items do not properly reflect preferences for typically behaving emotionally intelligently but rather express insight into the relative standing on the ability to adequately express moods and feelings. For some abilities it is easier than for others to provide appropriate self-ratings. The more appropriate introspection and knowledge about abilities are, the higher the relation between the ability and self-ratings. It is important to note though that in traditional areas of intelligence the correlations between self-ratings and actual abilities usually are somewhere between .20 and .50 (Ackerman, Beier, & Bowen, 2002). Although there is some convergent and discriminant validity in the relations between various self-ratings of ability and knowledge and actual measurement of these traits, these numbers are surprisingly low because human lives are filled with feedback about how well they perform in a variety of fields. The correlation between self-rated EI and ability EI is not likely to be any higher. Based on correlations well below .50 it is certainly not appropriate to use self-ratings as proxies for ability EI. Additionally, it is unclear what self-ratings of emotional intelligence actually reflect. Preferences, valences, abilities, a bias to overestimate or underestimate actual abilities, and other personality constructs are the most salient candidates to account for self-ratings of abilities. In order to establish a new construct and new measurement procedures, self-ratings are of very limited use, both as criteria and predictors. Hence, the reminder of this chapter will be devoted to so-called ability EI.

7.4 THE MSCEIT: DESCRIPTION, STRUCTURE, AND VALIDITY

The Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT) is a shortened and improved version of the Multi-Factor Emotional Intelligence Scales (MEIS; Mayer et al., 2002; Mayer, Caruso, & Salovey, 1999). The major goals in developing the MSCEIT were to abbreviate the very lengthy MEIS and to improve the psychometric properties of individual scales and items. The MSCEIT is highlighted in this discussion because it represents the most recent and up-to-date development of the research group surrounding Mayer, Salovey, and

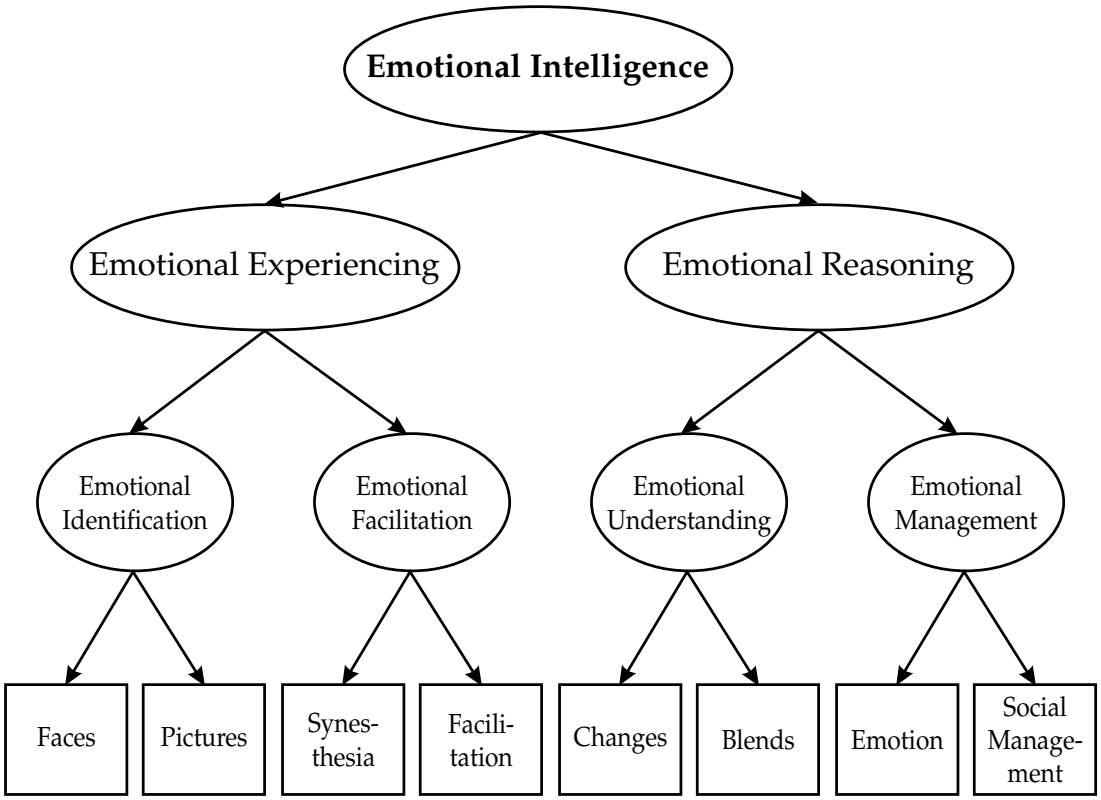

Figure 7.1 Subscales of the MSCEIT and its proposed structure.

Caruso, and it is the most widely used and best developed ability measure of emotional intelligence.

There are eight subscales of the MSCEIT (see the squares in Figure 7.1). These eight subscales combine in four pairs to represent four branches of emotional intelligence. The four branches combine to form two area-levels. These two area-levels in turn combine to make the MSCEIT total score. The MSCEIT thus represents a higher-order model of emotional intelligence. There are three levels in the model that are assigned ability status. Emotional intelligence at the top of the hierarchy, the two area-level scores of emotional experiencing and emotional reasoning, and the four branch-level scores of perceiving emotions, using emotions, understanding emotions, and managing emotions. Figure 7.1 additionally represents a fourth level—the specific tests as indicators of the MSCEIT.

Interpretation of test results is proposed down to the branch-level scores and can include interpretation of task-level scores in rare individual cases. Following the higher-order model there is a total of seven abilities that are measured by the MSCEIT. The interpretation of the four branches on the lowest level is:

Perceiving emotions: Participants with high scores are able to accurately identify and recognize their own and others' emotions. These participants are

also able to express feelings accurately and they are sensitive to faked and false emotional expression.

Using emotions: Participants with high scores are able to generate emotions to support problem solving. Those participants are also able to direct their attention to relevant changes, take several perspectives in considering emotions, and facilitate thinking by using different kinds of moods.

Understanding emotions: Participants with high scores understand causes and changes of emotions, both abstractly and in terms of relations. Those participants are also able to adequately recognize similarities amongst emotions of varying intensities and to reason about the dynamics of feelings in an interpersonal context.

Managing emotions: Participants with high values are successful in using their emotional awareness in drawing optimal decisions while assigning adequate importance to their emotions. Those participants manage to stay open to feelings, to engage and disengage when necessary and appropriate, and they are good in meta-evaluating their moods in terms of typicality, acceptability, and relevance.

The first two of these abilities can be aggregated into the ability of emotional experiencing. This ability is supposed to reflect accurately perceiving, responding to, and manipulating emotional information. The second pair of abilities combines to form emotional reasoning. This ability is expected to reflect understanding and managing emotions and how accurately a person understands the meaning of emotions and how those emotions can be managed in oneself and in relevant others.

On top of the proposed hierarchy is general emotional intelligence (Mayer & Salovey, 1997; Mayer, Salovey, & Caruso, 2000). It is computed as the mean of emotional experiencing and emotional reasoning. It is interpreted as the ability to perceive emotions, to use emotions so as to assist thought, to understand emotions, and to successfully regulate emotions.

There are many good descriptions of the tasks included in the MSCEIT (Mayer et al., 2002) and its predecessor—the MEIS (see also Chapter 2 by Neubauer & Freudenthaler). The scoring of the MSCEIT follows the same scheme for all subtests. The frequency distribution of the response options for a subscale is used to weight the response of an individual on that test. For example, in the "Faces" task individuals rate how much happiness a photograph of a face expresses and select one of the five options, ranging from no happiness to extreme happiness. Assume that for a certain photograph the five options in ascending order of happiness are endorsed by 10%, 20%, 40%, 20%, and 10% of participants, respectively. An individual endorsing response option 3 would thus be credited with a score of .40 while an individual endorsing response option 5 would only be credited with a score of .10. The same procedure is repeated for all responses and the scores on individual items of a subtest are averaged to express performance for this subtest. The rationale for such a scoring procedure is that for many important domains of human abilities and knowledge no universally accepted unequivocal standards of correctness are

available (see Chapter 8 by Legree, Psotka, Tremble, & Bourne). Consensus-based scoring is widely and successfully used in Situational Judgment Tests (McDaniel, Morgeson, Finnegan, Campion, & Braverman, 2001) and can be justified in domains of tacit or procedural knowledge. For the MSCEIT an empirical comparison between scores computed by application of a general consensus method—based on participants from the standardization group—and an expert consensus method—based on 21 experts from the International Society for Research on Emotions—reveals very high correlations of scores. The validity of the MEIS and the MSCEIT is under intense investigation and firm conclusions would certainly be premature. Rather than exhaustively presenting available evidence, the focus here will be on an eclectic summary of available prototypic investigations; specific attention will be devoted to (a) the MSCEIT as the momentary state of the art measure of EI and (b) some desiderata and standards for future research.

The structure of the MSCEIT in the data collected so far seems by and large to be robust. However, a structural model assuming four correlated factors can only be estimated if the covariance between the Identification and Facilitation factors and the covariance between the Understanding and Management factors are constrained to be equal to one another (Mayer et al., 2003). In exploratory factor analysis the proposed distinctions between factors are supported mostly. However, the loadings of tasks vary widely in both confirmatory and exploratory factor analysis, implying that factors are dominated by individual tasks. For example, the task "Synesthesia" has a much higher loading on the Facilitation factor than the second task "Facilitation" that is assigned to this factor. Consequently, the factors lack broadness in content. Content validity of the MSCEIT has not been demonstrated thoroughly so far. There is also a problematic mismatch between factor labels and tasks of the MSCEIT. For example, Branch 1 is labelled by the test authors "Perception and Expression of Emotion" but seemingly only perception of emotion is assessed.

Predictive validity of the MSCEIT has been assessed by correlating the scores with a variety of criteria. Correlations with fluid intelligence are generally small and correlations of some tasks with crystallized intelligence are substantially higher. MacCann, Roberts, Matthews, and Zeidner (2004) report correlations of individual EI tasks with broad visualization (Gv) tasks in the range of .20. A recent meta-analysis provides evidence that the MEIS—unlike self-report measures of EI—is associated with general mental ability ($\rho = .33$; $SD\rho = .093$) (Van Rooy & Viswesvaran, 2004). In various samples correlations between the MSCEIT and self-reports of empathy were found ranging from .17 to .52. Correlations with other self-report scales are mostly small although significant in several cases. Ciarrochi, Chan, and Caputi (2000) report a correlation of .31 with self-esteem. Correlations with Life Satisfaction vary widely but coefficients for larger studies are around .20 (Mayer et al., 2002). Emotional intelligence is substantially negatively associated with peer-nominated aggression and positively associated with prosocial behavior (Mayer et al., 2002). EI in general and Emotional Experiencing in particular are associated negatively with illegal drug use, alcohol use, and deviant behavior and these correlations

are due to the male subgroup exclusively (Brackett, Mayer, & Warner, 2004). Correlations between the four branches and the total score of the MSCEIT and self-report measures of EI do not exceed .28 and are mostly substantially smaller (Brackett & Mayer, 2003). In this study the MSCEIT was predictive of social deviance, even after controlling for the Big-five and verbal SAT scores. Despite this evidence a variety of controversies and problems remain and these issues will be discussed in the following sections.

7.5 IS "ABILITY" EI AN INTELLIGENCE?

The term *intelligence* in the construct label EI has caused considerable discussion. What are the reasons to subsume a new construct under the rubric of intelligence? First, measuring intelligence is a shortcut for success in applied settings (Ones et al., 2004). Intelligence is the single best predictor psychology has invented. Intelligence tests are widely used and integrating a new construct into such a context might facilitate acquiring some of the fame and credibility of an established construct.

Besides such marketing considerations it can also be argued that measures of emotional intelligence require effortful information processing and people are more or less apt at this processing. If information processing is less effortful *ceteris paribus* there will be poorer performance. The Levels of Emotional Awareness Scale (LEAS; Lane, Quinlan, Schwartz, Walker, & Zeitlin, 1990) is a cognitive-developmental measure of emotion that distinguishes between five levels of increasing complexity, thus distinguishing between easier and more sophisticated information processing. The five levels of complexity are bodily sensations, action tendencies, single emotions, blends of emotions, and combinations of blends (Lane & Schwartz, 1987). Participants are presented with twenty vignettes and write their responses to the two questions "How would you feel?" and "How would the other person feel?". The LEAS has been carefully developed and validated (Ciarrochi, Scott, Deane, & Heaven, 2003; Lane et al., 1998) and—due to its explicit consideration of information processing—represents an interesting and possibly better way of assessing emotional intelligence than measures relying on consensus or expert scoring. At present, the notion of information processing is not explicitly discussed for the MSCEIT. Developing alternative and additional measures that apply the model of five levels of emotional complexity to assess the generality of emotional awareness would be interesting. The role of aspects of information processing is even more strongly pronounced in several experimental measures of emotion like the Emotional Stroop Test for example (Coffey, Berenbaum, & Kerns, 2003; Matthews et al., 2002). It is presently unclear, however, how these measures of emotional information processing relate to other ability measures of emotional intelligence and how coherent such measures are. A third reason to label EI an intelligence might be that unlike self-report measures, EI measures, like other ability measures, can be scored by agreement with some external correctness standard (Guttman, 1965). Such a standard usually classifies indi-

vidual responses as right or wrong. Sometimes partial credit is given introducing degrees of correctness into assessment. These correctness standards apply to ability tests only. In most attitude measures, for example, participants are asked to provide us with information about how favorable their evaluation of some object is. In assessing personality traits participants are usually asked to provide us with information about how typical a behavior, thought, and the like is for them. In ability tests responses are compared to some explicit rule and classified as more or less correct. There are various evaluation standards that can be used in classifying response behavior in ability tests (Nevo, 1993). Performance can be assessed as number of correct responses, as latency per correct response, as variety of responses generated, or as agreement with authority. In assessing emotional intelligence the standards that have been used predominantly so far in assessing response behavior of an individual are target scoring, general consensus scoring, and expert consensus scoring—all varieties of the "authority" type of performance assessment. The response norms used in EI measures can be conceptualized as correctness standards. Although not desirable, it is common that various measures of a specific construct apply a single correctness standard. For example, reasoning tasks usually apply logical standards, mental speed tasks apply standards of work rate, and measures of emotional and practical intelligence most frequently apply consensus standards.

A fourth reason that is put forward in supporting a classification of EI measures as intelligence tasks is their relation with other intelligence measures. Starting from the positive manifold found amongst tests classified as intelligence measures, it is argued that if measures of EI represent an intelligence they must be correlated with other measures of intelligence. However, there are other indicators associated with intelligence that scientists would not be willing to classify as intelligence tests. For example, parents' education might be correlated with offspring intelligence and it would be very unusual to use parental education as an indicator of offspring intelligence. From a perspective endorsing a general factor and a positive manifold emanating from it the correlation between any two established intelligence tests is primarily or exclusively a function of their correlation with the general factor. Still, positive manifold is not itself the cause of observed relations amongst intelligence measures. Intelligence measures are correlated positively with each other because they tap the same underlying abilities. It has been argued repeatedly that measures of emotional intelligence should be correlated with general intelligence (Mayer, Salovey, Caruso, & Sitarenios, 2001; Roberts, Zeidner, & Matthews, 2001; Zeidner, Matthews, & Roberts, 2001). The reason such a correlation is expected is that it is a well replicated finding that there is a positive manifold between all measures that have been labelled intelligence tests. Hence, if emotional intelligence qualifies as an aspect of general intelligence there should be a positive correlation between indicators of emotional and "traditional" intelligence.

There are extensions and elaborations of this argument that go beyond a mere statistical necessity. Specific measures of emotional intelligence were expected and found to be correlated with some aspects of intelligence but not

others (MacCann et al., 2004). Psychologically, there is no problem in specific measures of emotional intelligence being unrelated with specific or general aspects of intelligence. If there is no overlap in the causes of individual differences there is no need that two measures be correlated. In most cases, however, there will be some overlap. For example, some reasoning ability is involved in measures that are subsumed under the "Understanding" branch of the MSCEIT. Similarly, some measures do have demands on visual processing and hence they might be related to broad visualization. Other measures of emotional intelligence require basic knowledge and hence they might be associated with crystallized intelligence. On the other side, some of the relations that have been found might represent artifacts. If, for example, a measure of emotional intelligence heavily relies on extensive verbal descriptions in the vignette, reading comprehension might be a necessary but not sufficient condition to actually get to the part dealing with the emotional content of an item. As a result, there might be an artifactual relation between performance on such an EI measure and reading comprehension, verbal intelligence or even general intelligence.

On the other hand, it has been argued that it supports the validity of performance measures of EI if they are unrelated to personality scales. Although high correlations would certainly be a cause for concern, moderate correlations could very well be meaningful. If, for example, openness to aesthetics, a facet of openness to new experiences, is correlated with performance on the "Designs" task from the MSCEIT, such a relationship could reflect aesthetic engagement of participants as expressed in their preference for behaviors that include an openness towards aesthetics. By mere exposition time, or by intellectual elaboration, persons with high openness for aesthetics might be better at performing on tasks like "Designs" because they are more familiar with the stimuli and have a more elaborated knowledge base of what various designs could actually express. There are other similar personality constructs that could be meaningfully correlated with performance on measures of emotional intelligence. Thus, ability models of emotional intelligence cannot simply be validated convergently by showing positive relationships with other ability measures and discriminantly by showing zero relationships with personality measures. What constitutes convergent and discriminant evidence is a psychological question in need of substantiation in every case.

The decision of whether or not tasks such as the ones from the MSCEIT should be labelled as intelligence tests has several conceptual aspects. Intelligence itself is so imprecisely defined that it is impossible to draw a clear line that allows for assigning the status of an intelligence test or withdrawing such a status. Assigning the status of cognitive ability measures to tasks as the ones used in the MSCEIT seems to be an option. A cognitive ability measure should certainly possess some features. For example, performance on a measure of cognitive ability should decrease if less time is available for working on the problems. Performance on cognitive ability measures—except measures of knowledge—should *ceteris paribus* also decrease if less effortful processing is warranted from participants; that is, if participants are asked to perform on a

typical level performance should be worse than when they are asked to perform at a maximal level. Similarly, instructions to fake good performance (i.e., to demonstrate better performance) will usually not work with an ability measure. If participants get a chance to work a second time on the same problems of a traditional intelligence test, they will improve substantially. A profound understanding of the ability involved in a cognitive ability measure implies to have some good ideas about how to manipulate the difficulty of problems. These and similar possibilities have not been thoroughly tested with measures of EI so far. With respect to retesting, Caruso, Mayer, and Salovey (2002) report a decrease of performance in the retest for nine of the twelve measures of the MEIS, the remaining three tests showing no change in performance level. Currently available evidence does not allow firm conclusions about whether or not EI measures from ability models qualify as cognitive ability measures.

7.6 CRITIQUE AND RECOMMENDATIONS

To be totally explicit, past experience in individual differences research, current evidence in research on EI, and the hope for a prosperous future for the construct all indicate that EI should not be investigated on the level of self-reports or self-ratings. Simply annotating the term *emotional intelligence* with some arbitrary addition will not do the job of clearly expressing that self-reported and performance appraised measures labelled emotional intelligence are conceptually fundamentally different and empirically by and large independent from each other. Furthermore, self-report measures are easy to develop and collect. Hence, there are many self-report measures and there is a large body of research exploring as well as testing the structure of individual differences on such self-reports. Every attempt to establish a new construct that is assessed solely or exclusively by relying on self-reports must establish the distinctiveness of these measures from established measures. With a broadly defined construct like emotional intelligence it will also be necessary to investigate the internal structure of the proposed indicators and to thoroughly check whether or not there is enough coherence amongst the various indicators to be summarized under one label. A collection of indicators from which one best fits to self-reported extraversion and another one to self-reported agreeableness is not satisfying. It is desirable to demonstrate at least some level of independence from the methods used for investigation. For example, a relation between corresponding life-data and self-report data is desirable. Substantial convergence of self-reports and peer-reports on the same participants is another example of demonstrating some method independent trait variance. Finally, in order to be worth pursuing it is eventually necessary that the new measure demonstrates some incremental validity of non-trivial magnitude. All of these steps are essential in establishing a construct of EI conceptualized as typical behavior. It would remain, however, that a construct of this sort should be labelled differently from the construct assessed by ability

measures of EI. A term like *typical emotional engagement* might be a good label for such a prospective and elusive construct.

On the ability side the MEIS and the MSCEIT represent the broadest assessment of EI. The MEIS and the MSCEIT are the EI measures that have been subject to most of the validation efforts undertaken so far and they have gained the largest proportion of attention in research and application. The critique and recommended research strategy presented below thus focuses on the MEIS and the MSCEIT—alternative measures should meet a similar set of requirements and challenges.

7.6.1 Scoring

For EI as an ability it is theoretically assumed that all participants from the intended application population possess this ability in varying degrees, and that this ability has some stability over time. The required psychometric properties of measurement instruments for EI should follow established standards. In proceeding through these psychometric requirements it is important to bear in mind that the psychometric evaluation of a measurement instrument is usually started after the assignment of numbers to specific responses. The process of this assignment—the scoring of a measure—is in need of justification itself (see Chapter 8 by Legree et al., for a detailed description of consensus based scoring). Consensus based scoring is one procedure used to assign numbers to responses. It can be defended for use with measures assessing tacit or procedural knowledge (Chapter 8 by Legree et al.), but it does not seem as if proponents of EI have adopted the idea that EI assesses such implicit knowledge. Interestingly, consensus based scoring is used in two other domains of psychology. Situational Judgment Tests (SJT) describe a methodology to assess job relevant implicit knowledge using consensus based scoring (McDaniel et al., 2001; McDaniel & Nguyen, 2001) and Practical Intelligence (PI) is a recently proposed highly controversial construct (Gottfredson, 2003; Sternberg, 2003) that is intended to measure success in real-life contexts (Wagner, 1987). A critical question prevalent in research on PI and SJT is whether or not PI can be assessed without relying on tests using consensus based scoring. The same question can be asked for EI: is there a coherent construct of EI and does the collected evidence on the validity of the construct and associated measures transfer to other measures of EI that do not rely on consensus based scoring? There is currently not sufficient evidence to answer this question.

A more technical but possibly critical difficulty is that various procedures of consensus based scoring do not sufficiently converge (MacCann et al., 2004). The question thus arises, which scoring procedure is the most appropriate one. Relying on psychometric results to pick the procedure that produces the most reliable or consistent scores is not an adequate solution. The procedure selected to score ability measures must be rationally appropriate too. The procedures compared by MacCann et al. are not very different on the rational end. There are thus competing and not converging scoring procedures for tests like

the MEIS and the MSCEIT. Satisfactory convergence between expert and consensus scoring is not yet sufficient to justify the MSCEIT scoring.

7.6.2 Available Validity Studies

Within the domain of psychometric measures, be they self-report or ability measures, correlational evidence can be pretty hard to assess. This is mostly due to a somehow arbitrary interpretation of associations. This problem is not new. Whenever a new intelligence test is constructed it is validated by correlating it with established measures. It is usually assumed that the correlations should be high but not perfect. If the correlations were perfect, there would be no point in establishing a new measure. If the correlation is high, there is some room for the new measure to be better than existent measures. However, high but not perfect correlations in no way imply that scenario. It could as well be the case that the new measure is psychometrically deficient and if it would be better the correlations with existent measures would be perfect. The situation is similar with measures of EI. If, for example, it is found that a new self-report measure of EI is correlated .70 with a measure of happiness and $-.50$ with a measure of neuroticism, is this indicating the validity of the measure? It could be argued that this provides strong evidence for the convergent validity of the self-reported EI questionnaire. However, it could also be argued that this result leaves little to no place for uniqueness of measures of self-reported EI and that apparently the construct is completely redundant with established constructs. Similarly, small to moderate correlations between a measure of emotional abilities and an established ability measure, say verbal intelligence, can be said to demonstrate discriminant validity—the EI measure is likely to measure something not captured by verbal intelligence measures. On the other hand, it could be argued that the small to medium correlation expresses an artifact of the test medium and that participants with high verbal intelligence are advantaged when taking measures of EI. Finally, it can be argued that the small correlation expresses some shared variance that can be attributed to general intelligence. Given that there are several explanations for the same result, the interpretation is necessarily arbitrary. If the truth be told: this scenario is not very different from the situation that exists for any specific intelligence test and its relations with other established ability measures. However, for most decent intelligence measures there is additional and replicated evidence demonstrating their embeddedness in a nomological net, their incremental utility in practical settings, their theoretically predicted redundancy with other, similar and dissimilar forms of tests, and much more. Although the majority of the studies on traditional intelligence tests are merely conceptual replications of each other there are many studies left that exclude alternative explanations for correlative results, thereby strengthening the interpretation and validity of the results.

7.6.3 Unavailable Validity Studies

Although there is a range of validity evidence that has been collected so far there is a surprising gap when it comes to exploring the relationship between the MEIS or the MSCEIT and related tasks. For example, the work on EI has rarely used indicators of performance measures of social intelligence as correlates. Similarly, there has been little research including experimental paradigms, for example, standard procedures used in face recognition research or the Emotional Stroop task. Closely related approaches to the investigation of individual differences in emotion related abilities—like the LEAS described above—have also been rarely used as correlates. An eclectic effort including many more than the standard MSCEIT tasks and representing a much broader variety of emotional tasks, including distinct scoring procedures, would provide us with a lot of invaluable information for further development of the investigated fields. With respect to construct validity it is crucial to learn more about how emotional intelligence is embedded in the nomological net of related constructs and measures. Besides established human cognitive ability constructs (Carroll, 1993) it is also relevant to discuss EI and its relation to social intelligence, empathic accuracy, PI, interpersonal abilities, intrapersonal abilities, and emotional awareness. Unfortunately, most of the above mentioned constructs are of dubious value.

7.6.4 Alternative Models

Not enough emphasis is given to possible alternative models of the data. Figure 7.2 shows just three of such alternative models in Panels A, B, and C (see Schulze, 2005, for a discussion of various model architectures).

The models in Figure 7.2 describe structures that are pretty similar to the one adopted in the MSCEIT. However, there are also important discrepancies. Some of the models do without a general factor (see Panels A and B). In other words, assuming the models provide a decent fit to empirical data, adequate explanations of the covariances between tasks can be established without postulating something like a general emotional intelligence factor. The structure of individual differences on available EI measures is not well established. Eclectic research applying a great bandwidth of available measures is warranted and necessary in order to compare various structural models of emotional intelligence with each other.

Continuing to stress this point the number of abilities that are supposed to be assessed with the MSCEIT is very high. Based on only eight tasks, participants receive feedback on seven abilities. This is the case because the scores of each test are used three times. The first time in computing values for the four branches, the second time when these four branches are combined into the two area scores, and the third time when the two area scores are combined into the MSCEIT total value. Doing poorly on a specific task will thus hurt you three times. The redundancy of this use of information is not sufficiently explicated to the participants and implications from the higher-order structure

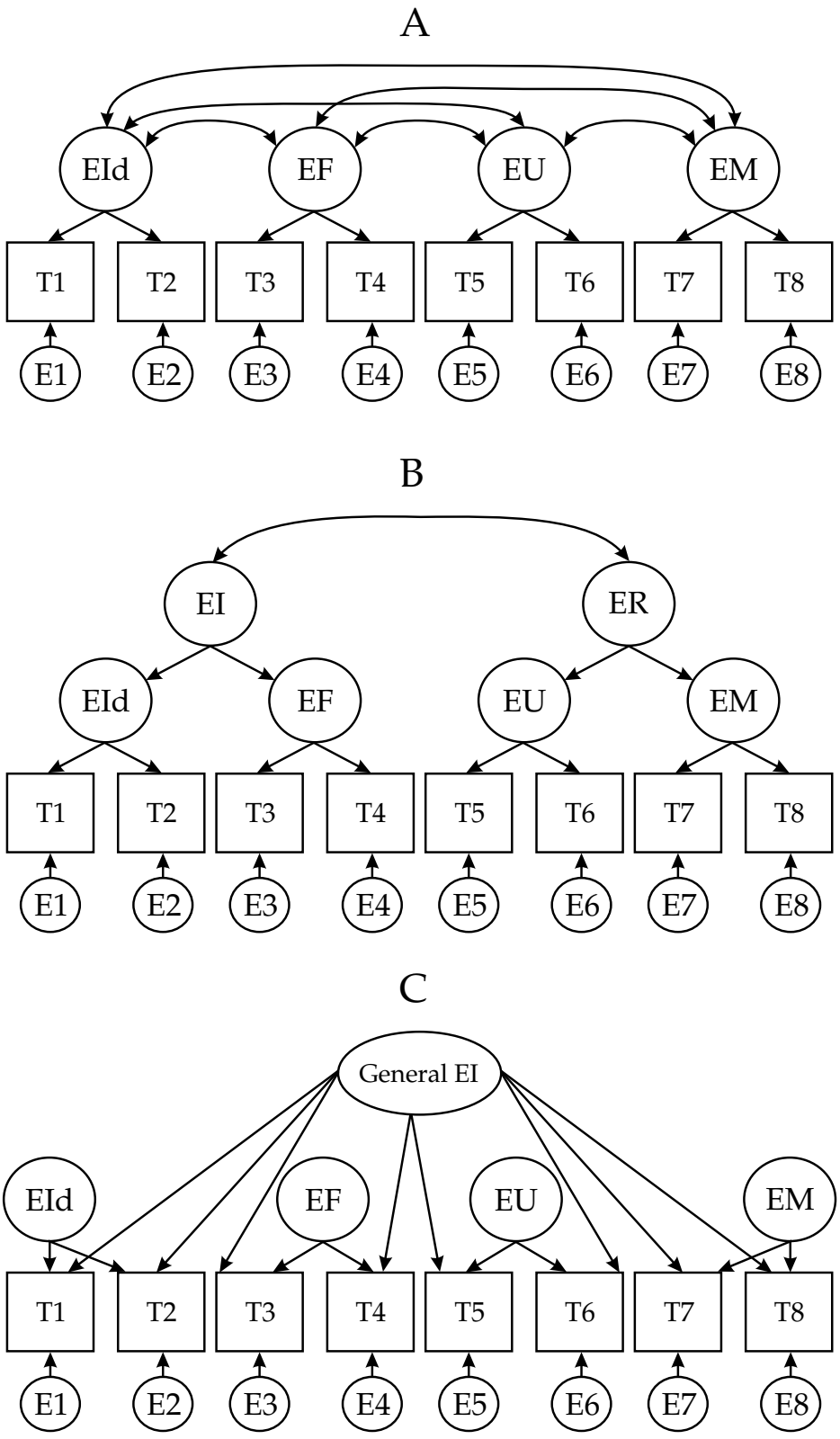

Figure 7.2 Alternative models for the MSCEIT.

of the model supposedly underlying the MSCEIT are not sufficiently considered. In order to avoid repeatedly analyzing the same information for the same data the variance of the indicators could be fractionated into variance due to general EI and lower-order factors.

7.6.5 Number of Tasks per Factor and Generality of Factors

In the MSCEIT, not enough tasks are used to assess the branches of emotional intelligence. For the branch level scores only two task types are used per ability. The interpretation of the factors goes far beyond the contents actually included in the test. A much broader variety of task types should be used before postulating abilities. For example, there are decades of research demonstrating over and over again that matrices tasks are decent measures of a construct labelled fluid intelligence. There is thus abundant evidence to use the task as an indicator of such a construct. Of course, it is a characteristic of good psychological measurement to rely on more than a single task type. On the other hand, each indicator used should qualify as a decent measure of the construct. It can easily be tested whether or not a specific indicator is fitting within a measurement model of an ability. What is required are four or more indicators that are all purported to measure the same ability. With only three indicators a measurement model would be just identified—no adequate test of such a model would thus be possible. With only two tasks it is necessary to extend the model to include additional factors and variables in order to be meaningful. The status of the four branches of the MSCEIT cannot be tested adequately. More tasks for each of the proposed constructs are required.

7.6.6 Alternative Tasks (Number and Variety)

There is considerable arbitrariness when deriving a task from a construct description. Technically, the definitions of most individual differences constructs allow for an unlimited number of tasks to be derived from the construct description. Much specificity of ability indicators is considered to be irrelevant in measuring the ability we are interested in. The person administrating the test, the test medium, the specific stimuli used in a measure (for example, which faces are displayed in the Faces task) are all considered to be irrelevant for the measured construct. Variations in the task instructions and in the response scales used should have no substantial influence on what is measured with a specific test. The description of EI abilities allows for many more variations. For example, the perception of emotions can be assessed with music, prose, with videos of facial expression, with artificial stimuli, and so forth. Without variations in the form of measurement care must be taken to not overgeneralize the results from tests. For the MEIS and the MSCEIT substantially more and more diverse indicators are warranted before concluding that the four branch model is a sufficient and appropriate model of EI. In other words, the MSCEIT provides a very general interpretation of emotional intelligence but uses very specific tasks.

7.6.7 Test Construction

Test construction should proceed as deductive derivation from theory whenever possible. The measurement intention for EI measures should be inspired by experimental and neuropsychological research whenever possible. A critical and important issue is to create and maintain a strong relationship between psychometric research on measures for individual differences and general theories of emotion. The use of a measure should be justified by what it measures. After a precise description of the measurement intention and operationalization psychometric criteria are important but the test content is crucial. Although the tasks of the MEIS and the MSCEIT seem to be good indicators for the proposed branches little is known about alternatives and—for the tests themselves—nothing about the emotion background. It is desirable to create and maintain stronger links between individual indicators and the constructs they are supposed to measure.

7.7 CONCLUSIONS

Despite the need to consider the points raised above it is important to note that the MSCEIT represents the most ambitious and, to date, most appropriate approach to the broad assessment of emotion related capabilities. There are many challenges, both methodologically and psychologically. While the field is at an early stage in debating the validity and utility of the concept and proposed measures it is necessary to be very careful in applications of the measures proposed so far. Taken together, the proponents of performance based measures of EI have done a decent job. While it is too early for reification of a simple model the field has made considerable progress in the last decade. Besides the adverse impacts of premature use of measures and inflation of the concept as a whole, public attention to EI has had a major beneficial effect too: it has directed scientific efforts into an important and neglected area of human abilities.

On the other side, the enthusiastic uptake of the initial proposals of the construct EI has blurred sight for the state-of-the-art procedures used when investigating new ideas. One threatening consequence from this enthusiasm is that measures and interventions based on EI are underway and used in practice before crucial questions have been answered empirically. In fact, we might not even be able to spell out the right questions yet.

To conclude optimistically with some research prospects, one promising approach that is motivated and inspired by neuropsychological and experimental work on face recognition will be highlighted. Good face perception and face recognition allow humans to infer information about age, sex, mood, and identity of a person. Consequently, face recognition can be considered to be a limiting factor for some aspects of EI. Individual differences for these aspects of EI can thus be attributed to individual differences in face recognition. There is decent physiological evidence that there are two distinct components

of face recognition. The first of these components has to do with the encoding of faces. Successful and unsuccessful learning of unfamiliar faces are associated with neurophysiological differences (Schweinberger, Pfütze, & Sommer, 1995; Sommer, Komoss, & Schweinberger, 1997). The second component has to do with the retrieval of familiar and unfamiliar faces. The so-called early repetition effect is different for personally familiar persons, famous persons and celebrities, and unfamiliar persons (Herzmann, Schweinberger, Sommer, & Jentzsch, 2004). The promise of this and similar research is that there is convincing evidence for individual differences that can be attributed to the encoding and retrieval of faces (Pfütze, Sommer, & Schweinberger, 2002). It is thought provoking to think about options in this area. How about developing measures that assess perception and recognition of changes in facial expression, or measures that address just noticeable differences in facial expression?

REFERENCES

Ackerman, P. L., Beier, M. E., & Bowen, K. R. (2002). What we really know about our abilities and our knowledge. *Personality and Individual Differences*, *33*, 587–605.

Bagby, R. M., Parker, J. D. A., & Taylor, G. J. (1994a). The Twenty-Item Toronto Alexithymia Scale–II: Convergent, discriminant, and concurrent validity. *Journal of Psychosomatic Research*, *38*, 33–40.

Bagby, R. M., Parker, J. D. A., & Taylor, G. J. (1994b). 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). *BarOn Emotional Quotient Inventory (EQ–i): Technical manual*. Toronto, Canada: Multi-Health Systems.

Bar-On, R. (2000). Emotional and social intelligence: Insights from the emotional quotient inventory. In R. Bar-On & J. D. A. Parker (Eds.), *The handbook of emotional intelligence: Theory, development, assessment, and application at home, school, and in the workplace* (pp. 363–388). San Francisco, CA: Jossey-Bass.

Brackett, M. A., & Mayer, J. D. (2003). Convergent, discriminant, and incremental validity of competing measures of emotional intelligence. *Personality and Social Psychology Bulletin*, *29*, 1147–1158.

Brackett, M. A., Mayer, J. D., & Warner, R. M. (2004). Emotional intelligence and its relation to everyday behaviour. *Personality and Individual Differences*, *36*, 1387–1402.

Broadbent, D. E., Cooper, P. J., Fitzgerald, P. F., & Parkes, K. R. (1982). The Cognitive Failures Questionnaire (CFQ) and its correlates. *British Journal of Clinical Psychology*, *21*, 1–16.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *42*, 116–131.

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

Caruso, D. R., Mayer, J. D., & Salovey, P. (2002). Relation of an ability measure of emotional intelligence to personality. *Journal of Personality Assessment*, *79*, 306–320.

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

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

Coffey, E., Berenbaum, H., & Kerns, J. G. (2003). The dimensions of emotional intelligence, alexithymia, and mood awareness: Associations with personality and performance on an emotional stroop task. *Cognition & Emotion*, *17*, 671–679.

Costa, P. T., Jr., & McCrae, R. R. (1992). *Revised NEO personality and five factor inventory professional manual*. Odessa, FL: Psychological Assessment Resources.

Cronbach, L. J. (1949). *Essentials of psychological testing*. New York: Harper & Row.

Dennis, M. J., Sternberg, R. J., & Beatty, P. (2000). The construction of "user-friendly" tests of cognitive functioning: A synthesis of maximal– and typical–performance measurement philosophies. *Intelligence*, *28*, 193–211.

Epstein, S., Pacini, R., Denes-Raj, V., & Heier, H. (1996). Individual differences in intuitive–experiential and analytical–rational thinking styles. *Journal of Personality and Social Psychology*, *71*, 390–405.

Goff, M., & Ackerman, P. L. (1992). Personality–intelligence relations: Assessment of typical intellectual engagement. *Journal of Educational Psychology*, *84*, 537–552.

Goleman, D. (1995). *Emotional intelligence: Why it can matter more than IQ*. New York: Bantam Books.

Goleman, D. (1998). *Working with emotional intelligence*. New York: Bantam.

Gottfredson, L. S. (2003). Dissecting practical intelligence theory: Its claims and evidence. *Intelligence*, *31*, 343–397.

Guttman, L. (1965). A faceted definition of intelligence. *Scripta Hierosolymitana*, *14*, 166–181.

Herrmann, D. J. (1982). Know thy memory: The use of questionnaires to assess and study memory. *Psychological Bulletin*, *92*, 434–452.

Hertzog, C., Park, D. C., Morrell, R. W., & Martin, M. (2000). Ask and ye shall receive: Behavioral specificity in the accuracy of subjective memory complaints. *Applied Cognitive Psychology*, *14*, 257–275.

Herzmann, G., Schweinberger, S. R., Sommer, W., & Jentzsch, I. (2004). What's special about personally familiar faces? A multimodal approach. *Psychophysiology*, *41*, 688–701.

Kahneman, D. (1973). *Attention and effort*. Hillsdale, NJ: Prentice-Hall.

Lane, R., Quinlan, D., Schwartz, G. E., Walker, P., & Zeitlin, S. (1990). The levels of emotional awareness scale: A cognitive–developmental measure of emotion. *Journal of Personality Assessment*, *55*, 124–134.

Lane, R., Reiman, E., Axelrod, B., Yun, L.-S., Holmes, A. H., & Schwartz, G. E. (1998). Neural correlates of levels of emotional awareness: Evidence of an interaction between emotion and attention in the anterior cingulate cortex. *Journal of Cognitive Neuroscience*, *10*, 225–235.

Lane, R., & Schwartz, G. E. (1987). Levels of emotional awareness: A cognitive-developmental theory and its application to psychopathology. *American Journal of Psychiatry*, *144*, 133–143.

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

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

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

Mayer, J. D., Caruso, D. R., & Salovey, P. (2000). Selecting a measure of emotional intelligence: The case for ability testing. In R. Bar-On & J. D. A. Parker (Eds.), *The handbook of emotional intelligence: Theory, development, assessment, and application at home, school, and in the workplace* (pp. 320–342). San Francisco: Jossey-Bass.

Mayer, J. D., DiPaolo, 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., & Salovey, P. (1993). The intelligence of emotional intelligence. *Intelligence*, *17*, 433–442.

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

Mayer, J. D., Salovey, P., & Caruso, D. R. (2000). Models of emotional intelligence. In R. J. Sternberg (Ed.), *The handbook of intelligence* (pp. 396–420). New York: Cambridge University Press.

Mayer, J. D., Salovey, P., & Caruso, D. R. (2002). *The Mayer, Salovey, and Caruso Emotional Intelligence Test: Technical manual*. Toronto: Multi-Health Systems.

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

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

McCrae, R. R. (1990). Traits and trait names: How well is openness represented in natural languages. *European Journal of Personality*, *4*, 119–129.

McCrae, R. R. (1996). Social consequences of experiential openness. *Psychological Bulletin*, *120*, 333–337.

McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A., & Braverman, E. P. (2001). Use of situational judgment tests to predict job performance: A clarification of the literature. *Journal of Applied Psychology*, *86*, 730–740.

McDaniel, M. A., & Nguyen, N. T. (2001). Situational judgment tests: A review of practice and constructs assessed. *International Journal of Selection & Assessment*, *9*, 103–113.

McGlynn, S. M., & Schacter, D. L. (1989). Unawareness of deficits in neuropsychological syndromes. *Journal of Clinical and Experimental Neuropsychology*, *11*, 143–205.

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

Nevo, B. (1993). In search of a correctness typology for intelligence. *New Ideas in Psychology*, *11*, 391–397.

Oberauer, K., Süß, H.-M., Schulze, R., Wilhelm, O., & Wittmann, W. W. (2000). Working memory capacity—Facets of a cognitive ability construct. *Personality and Individual Differences*, *29*, 1017–1045.

O'Connor, R. M., & Little, I. S. (2003). Revisiting the predictive validity of emotional intelligence: self–report versus ability–based measures. *Personality and Individual Differences*, *35*, 1893–1902.

Ones, D. S., Viswesvaran, C., & Dilchert, S. (2004). Cognitive ability in selection decisions. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence* (pp. 431–463). London: Sage.

Palmer, B. R., Manocha, R., Gignac, G., & Stough, C. (2003). Examining the factor structure of the Bar-On Emotional Quotient Inventory with an Australian general population sample. *Personality and Individual Differences*, *35*, 1191–1210.

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

Petrides, K. V., & Furnham, A. (2001). Trait emotional intelligence: Psychometric investigation with reference to established trait taxonomies. *European Journal of Personality*, *15*, 425–448.

Pfütze, E.-M., Sommer, W., & Schweinberger, S. R. (2002). Age–related slowing in face and name recognition: Evidence from event–related brain potentials. *Psychology and Aging*, *17*, 140–160.

Reason, J. T. (1993). Self–report questionnaires in cognitive psychology: Have they delivered the goods? In A. D. Baddeley & L. Weiskrantz (Eds.), *Attention: Selection, awareness, and control* (pp. 152–170). Oxford: Clarendon Press.

Riemann, R., & Abels, D. (1994). Personality abilities: Construct validation. In B. de Raad, W. K. B. Hofstee, & G. L. M. van Heck (Eds.), *Personality psychology in europe* (Vol. 5, pp. 201–215). Tilburg: Tilburg University Press.

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

Sackett, P. R., Zedeck, S., & Fogli, L. (1988). Relations between measures of typical and maximal job performance. *Journal of Applied Psychology*, *73*, 482–486.

Saklofske, D. H., Austin, E. J., & Minski, P. S. (2003). Self–reported emotional intelligence: Factor structure and evidence for construct validity. *Personality and Individual Differences*, *34*, 707–721.

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

Schulze, R. (2005). Modeling structures of intelligence. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence* (pp. 241–263). London: Sage.

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

Schweinberger, S. R., Pfütze, E.-M., & Sommer, W. (1995). Repetition priming and associative priming of face recognition: Evidence from event–related potentials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 722–736.

Seidenberg, M., Haltiner, A., Taylor, M. A., Hermann, B. B., & Wyler, A. (1994). Development and validation of a multiple ability self–report questionnaire. *Journal of Clinical and Experimental Neuropsychology*, *14*, 93–104.

Sommer, W., Komoss, E., & Schweinberger, S. R. (1997). Differential localization of brain systems subserving memory for names and faces in normal subjects with event–related potentials. *Electroencephalography and Clinical Neurophysiology*, *102*, 192–199.

Stankov, L. (1999). Mining on the "no man's land" between intelligence and personality. In P. L. Ackerman, P. C. Kyllonen, & R. D. Roberts (Eds.), *Learning and individual differences: Process, trait, and content determinants* (pp. 315–337). Washington, DC: American Psychological Association.

Sternberg, R. J. (2003). Our research program validating the triarchic theory of successful intelligence: Reply to Gottfredson. *Intelligence*, *31*, 399–413.

Van Rooy, D. L., & Viswesvaran, C. (2004). Emotional intelligence: A meta-analytic investigation of predictive validity and nomological net. *Journal of Vocational Behavior*, *65*, 71–95.

Wagner, R. K. (1987). Tacit knowledge in everyday intelligent behavior. *Journal of Personality and Social Psychology*, *52*, 1236–1247.

Wilhelm, O., Schulze, R., Schmiedek, F., & Süß, H.-M. (2003). Interindividuelle Unterschiede im typischen intellektuellen Engagement [Individual differences in typical intellectual engagement]. *Diagnostica*, *49*, 49–60.

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