

Emotional Intelligence and the Identification of Emotion

JOHN D. MAYER

GLENN GEHER

University of New Hampshire

This article is concerned with individual differences in the ability to connect thoughts to emotions. People who are good at connecting thoughts to feelings may better "hear" the emotional implications of their own thoughts, as well as understand the feelings of others from what they say. We had 321 participants read the writings of a target group of people and guess what those targets had felt. Several criteria were used to evaluate the participants' emotional recognition abilities, including agreement with the group consensus and agreement with the target. Participants who agreed more highly with the group consensus and with the target also scored higher than the other participants on scales of empathy and self-reported SAT scores, and lower on emotional defensiveness. Such results are interpreted to mean that some forms of emotional problem solving require emotional openness as well as general intelligence.

A person's *general* intelligence represents that individual's overall level of intellectual attainment and ability, and has often been used to successfully predict a person's academic and occupational achievement (e.g., Matarrazzo, 1972; Ree & Earles, 1992). Although general intelligence indicates a person's overall intellectual functioning, it says little about the more specific intelligences that comprise it (e.g., Detterman, 1986). Consequently, psychologists have sought to divide general intelligence in various ways. For example, general intelligence can be divided into more specific intelligences that represent either groups of abilities or specific abilities (e.g., Cattell, 1963; Gardner, 1983, 1993; Guilford, 1967; Sternberg, 1988; Thorndike, 1920; Wechsler, 1987). These more specific intelligences are viewed as intercorrelated but somewhat distinct from one another. Thus, even those who most strongly argue for the independence of the intelligences acknowledge that they are empirically correlated (Detterman, 1986; Gardner, 1993).

One of the earliest and most influential divisions of intelligence split it into

We gratefully acknowledge the assistance of Adam Geher, who assisted in the development of the stimuli used in this research, Kathleen Bauman and Dennis Mitchell, who read and commented on an earlier draft of this article, and of Craig Chamberlain, who entered the data on the computer.

Correspondence and requests for reprints should be sent to John D. Mayer, Department of Psychology, University of New Hampshire, Durham, NH 03824.

three broad classes of abilities (Thorndike, 1920). The first of these three classes of intelligence typically involves the abstract, analytic, and/or verbal intelligences. The second class involves the mechanical, performance, visual-spatial, and/or synthetic intelligences. The third, less-studied, class consists of the social and/or practical intelligences. Social intelligence has been less studied because it seems the hardest of the three broad classes of intelligence to distinguish from the others, both theoretically (e.g., Mayer & Salovey, 1993) and empirically (Cronbach, 1960). Interest in social intelligence, however, has recently undergone a revival (see Cantor & Kihlstrom, 1987; Ford & Tisak, 1983; Legree, 1995; Sternberg & Smith, 1985).

Rather than simply dropping the idea of social intelligence, it may make sense to more plainly distinguish it from other intelligences by subdividing portions of it into, say, emotional and motivational intelligences. Motivational intelligence would involve understanding motivations such as the need for achievement, affiliation, or power, as well as understanding tacit knowledge related to those motivations (e.g., Wagner & Sternberg, 1985) and the goal-setting related to them (Cantor & Kihlstrom, 1987). The contrasting emotional intelligence involves recognizing emotion, reasoning both with emotion and emotion-related information, and processing emotional information as a part of general problem-solving ability (Mayer & Salovey, in press; Salovey & Mayer, 1990). Either emotional, motivational, or both intelligences together could substitute, or at least help, in the better definition of social intelligence. Although we employ the term *emotional intelligence*, several closely related concepts exist under different terminology including *intrapersonal intelligence* (Gardner, 1983), *hot processing* (Mayer & Mitchell, in press), and *emotional creativity* (Averill & Thomas-Knowles, 1991).

This article is concerned with emotional intelligence and, more specifically, with understanding how people recognize emotions. The recognition of emotion may be the best starting place for the empirical measurement of emotional intelligence because there exist provisionally agreed-on ways to identify what someone is experiencing (Mayer & Salovey, in press). In contrast, more complex emotional problems require extremely careful consideration before emotional reasoning and its outcomes can be fairly evaluated (Mayer & Salovey, 1995, in press). Because the ability to recognize emotions is basic to a person's emotional well-being, considerable research on it already exists, as discussed later. Its potential importance to daily functioning has also been noted. For example, Reik (1952) associated mental health with the ability to recognize one's emotion, and mental illness with the inability to recognize it. Consider his example:

A patient . . . was having an affair with a married man . . . One day she asked [the married man] to promise her that he would not come from [his] home when he visited her and that he would not return home when he left her. She formulated what she expected from him more clearly the next day. "You must not come from

her or go to her when you see me." It is obvious that the wife of her lover was meant . . . She spoke of it as if it were an indifferent thought that had occurred to her, a convenient arrangement, yes, even a kind of amusing idea . . . [But the analyst could] put himself into the place of his patient . . . he got an inkling . . . of the emotions of his patient: her jealousy, her suffering from the thought that her lover left her to go home to his wife. (Reik, 1952, pp. 309–311)

A person like the aforementioned patient, who is unable to connect her thoughts to her own emotion, may find herself at a social disadvantage—and appear irrational and demanding. A person like the therapist who can "hear" the emotions in another's thoughts may excel at handling certain social demands. Sometimes the task of emotional identification requires considerable perspective taking, as in the preceding example. At other times, such inferences may be more direct. People simply may sense that pleasant thoughts indicate pleasant moods, or that unpleasant thoughts indicate unpleasant moods (Bower, 1981; Forgas, 1995; Mayer, Gaschke, Braverman, & Evans, 1992). In addition, they may recognize the correlations between thoughts of injustice and anger, and of perceptions of threat and fear, and so forth that stem from emotional appraisals of events (Ortony, Clore, & Collins, 1988; Roseman, 1984; Smith & Ellsworth, 1985).

This article is concerned with the ability to identify emotions from thoughts. Some people, like the aforementioned therapist, are presumably better at recognizing and/or producing appropriate thought–emotion combinations than others. This may be because their cognitive processing skills are generally better, because they are open to their own and others' emotional reactions, or because they have constructed expert knowledge concerning such connections (Mayer & Salovey, 1995). In this study, participants were presented with transcripts of real peoples' thoughts, and asked to infer their emotions or moods (for the purposes of this research, emotion and mood can be considered interchangeable). Participants were asked questions of the form, "Given that the target is now thinking about three emotion or mood related events: A, B, and C, how is that person feeling?" We believe the ability to successfully reason about emotions in this way is a central aspect of emotional intelligence. We hypothesize that the ability to know other people's emotions is related to other indices of emotional intelligence, such as empathy, openness, and general intelligence.

BACKGROUND

In the 1930s, the study of social intelligence was largely a study of how people made judgments regarding others and the accuracy of such judgments. By the 1950s, however, this work had become divided into an intelligence tradition that was interested in abilities of person perception, and a social psychological tradition that focused on the social determinants of person perception. The two areas

had diverged sufficiently that researchers in one area often were unaware of the work in the other (Walker & Foley, 1973). More recently, there has been a growing convergence among these and other areas, as intelligence researchers have become more interested in social intelligence, as social psychologists have become more interested in cognitive determinants of perceptions (e.g., Fiske & Taylor, 1991), and as a new group of evolutionary psychologists became interested in nonverbal behavior (Buck, 1984). The recent cognitive developments in person perception research are of particular importance to the work here; as we explain, a few recent social psychological studies have even reintroduced the use of intelligence scales.

Related Research

Recall that in this study, participants were asked to estimate a target person's emotions from that person's self-reported, emotion-related thoughts. Although this is the first study to use this specific procedure, a number of studies examine similar types of emotional identification. For example, researchers interested in the nonverbal communication of emotion have sometimes asked participants to judge what a target person is feeling on the basis of a visual depiction of that person's facial or postural features. Similarly, researchers examining dyadic relationships have asked pairs of people to interact with one another and then estimate what each individual was thinking and/or feeling during the interaction. Our study shares with these related studies of nonverbal emotion and dyadic relationship research a general interest in the accurate identification of emotion. Common among such studies are three related issues: (a) what is the best criterion of what the target is feeling, (b) what is the best language with which to describe emotionality, and (c) what sorts of personality variables may be related to the ability to identify emotion? We examine these three issues in turn.

The Criteria for the Target's Emotions

Target Agreement. The first and most central issue addressed by studies of emotional identification involves identifying valid criteria for what the person is feeling. In previous research, one acceptable criterion for an individual's emotional state has been his or her freely given self-report. This reasoning states that the self-reported language of feeling is "as 'close' as one can come to studying emotional experience" (Davitz, 1969, p. 2). Even theorists who have argued for the fallibility of self-report acknowledge that only the experiencing individual can have direct access to his or her internal feelings (e.g., Nisbett & Wilson, 1977, p. 255). Let us call the person experiencing the feeling the *target* and the person trying to estimate the feeling the *judge*. The degree to which a judge identifies the target's self-reported emotion will be referred to here as *target agreement*. Many studies have employed such a criterion, which has the virtue of being direct and convenient. For example, Buck, Miller, and Caul (1974) asked

participants to view slides of actors who displayed emotions and asked the participants to report what emotion the actor attempted to portray. Similarly, Ickes, Stinson, Bissonette, and Garcia (1990) had participants in mixed-gender dyads identify what their partners reported feeling during videotaped conversation. Levenson and Ruef (1992) had targets use a rating dial to express their affect on a continuum from positive to negative. Participants then used this same dial to try to match the rating given by the target. Although target agreement makes sense as a criterion of emotion identification, it is not without its potential problems. A target person may experience complex feelings that are difficult to communicate, may not be good at labeling feelings, or may distort reports of feelings in order to appear more socially desirable. For that reason, other criteria for what a person is feeling have been proposed, such as how external observers rate the person's feeling.

Consensus Agreement. In consensus agreement studies, the identification of a target's feeling is scored as correct when it is consensually agreed on by a number of judges who view similar information about the target. Note that this *consensus agreement* criterion is independent of the target's self-reported emotion. For example, Wagner, MacDonald, and Manstead (1986) studied judges who viewed the videotapes of participants and rated what emotions they thought the participants were expressing. The alternative selected by the plurality of judges was scored as correct, independent of the target's feelings. Similarly, Boucher and Carlson (1980) had judges rate a battery of photographs of Americans and Malaysians expressing emotions. The answer scored as correct was that which was consensual among the judges independent of the characteristics of the photograph. Mayer, DiPaolo, and Salovey (1990) had participants view ambiguous stimuli and rate what emotion those stimuli elicited. A response was considered correct when it was in agreement with the group consensus for a particular stimulus.

Relationship Between Target and Consensus Criteria. The relationship between the target and consensus criteria of emotion is complex. In limited instances, the two may agree. For example, in Ekman's work on facial expressions, people asked to model an angry expression are typically identified as doing so by judges (Ekman, Friesen, & Ellsworth, 1972, pp. 102–103). Although a standard in which target and consensus agree is plainly desirable, it may be more possible with simple, fixed, and stereotyped facial expressions than with more complex, dynamic internal processes. For example, Ekman et al. (1972) found that participants report internal emotions that correlate with their external facial expressions on average only around $r = .30$ (range = $-.10$ to $.60$) while viewing emotionally charged videotapes. Relations between self-reported feeling and highly specific facial expressions may increase as measurement techniques improve (Ekman, Davidson, & Friesen, 1990). Target–consensus agreement

may drop off, however, in more complex social interactions. Dyadic studies permit target perception to be correlated with observer perception, although the observer is in this case a single person rather than a group of observers. One recent study of dyadic interaction (Levenson & Ruef, 1992) defined emotional identification as occurring when the target and observer estimated the target's emotion within one unit normal deviate of one another on a positive-negative mood continuum. Agreement ranged between 28% and 43%, which was only moderately above chance levels (roughly 11%). In another instance from the dyadic research literature, target-observer correlations of thought and feeling content were reported as entirely nonsignificant (Ickes et al., 1990, p. 735). In our study, we examine both target and group consensus criteria to understand their level of agreement.

Target, Consensus, and Expert. It is worth noting that a third potential criterion is that of an expert's judgment—a specialist who understands the internal emotion of an individual better than the individual who is feeling the emotion, and better than a lay or nonspecialist judge attempting the same evaluation. Expert judgments have sometimes been studied when examining social intelligence (e.g., Legree, 1995; Wagner & Sternberg, 1985). It is possible that specialists in emotion, such as clinical psychologists, might serve as independent experts, although that was not examined here. Legree (1995) described a model in which the judgments of nonexperts are equivalent to those of experts except for being less reliable (and can therefore be aggregated to match the judgments of experts). And, as Legree noted, there are likely to be some circumstances where an expert is no more than a reliable indicator of the group, albeit a particularly sensitive one. Other circumstances exist, however, in which the expert's knowledge base may be different and more accurate than any group aggregation. (Consider the introductory example in which Reik described a patient's feelings about her lover, based on that patient's rather vague request that the lover not come to her from his home. A group consensus might miss the patient's feeling, and can be productively contrasted with the therapist's interpretation, which captures at least some intuitive validity.) Legree's approach, therefore, may apply only to some experts. In addition, it is unlikely to apply to the target criterion used here, in that the target has extra information (his or her own feelings) that are privileged internal events unavailable to the outside observer. In this sense, the target can be considered to be a special case of the aforementioned expert "whose knowledge extends beyond" what the group knows, and is possibly irreducible to the group consensus.

The Language of Emotional Reports

A second issue addressed by studies of emotional identification is the language used to label emotion. Individuals seem to employ related but distinct lexicons of

emotions, including pure emotion terms such as *happy*, physical terms such as *smiling*, or more cognitive terms such as *appreciated*, and people also employ terms that depict physical or mental acts closely related to emotion such as *dance around* (e.g., related to joy), or emotional management, such as *be open to the feeling*. The original mood scales of the 1960s use a broad collection of such emotion and emotion-related terms. From the late 1970s to the 1980s, however, many new emotion scales were developed that included only pure emotional terms (i.e., happy, angry, sad, etc). More recent research is returning to broader definitions of emotion (see Mayer, Salovey, Gomberg-Kaufman, & Blainey, 1991, for a review). Initial research into emotional identification was often limited to a narrow view of the emotional lexicon. For example, Rosenthal et al. (1979) asked participants only the degree of dominance-submission, or positive-negative emotionality modeled in a face or voice. More recent research has allowed participants to express their emotions in an open-ended format (e.g., Ickes et al., 1990). Although this latter procedure allows participants to express themselves in their own language, it can become unwieldy because judges are needed to identify agreement between participants and targets. There may be advantages, however, to having targets express what they are feeling in a variety of emotional languages so as to better sample the domain (e.g., emotion-related physical and action terms, as well as pure emotional terms). In this research we used scales of emotional experience that were closed ended, and yet sampled broadly from diverse emotion-related lexicons, including those drawn from domains of cognitive appraisals, physiological sensations, pure emotions, or emotion management. Such an approach allowed our participants to express what they themselves were feeling, or believed someone else was feeling, in a standard but widely sampled set of emotion and closely emotion-related languages.

Personality Dimensions Related to Emotional Intelligence

The third issue addressed by studies of emotional identification concerns the search for correlates of emotional identification. Potential correlates of emotional identification have typically been drawn from emotion- or intellect-related domains. Higher emotional intelligence is frequently said to covary with greater internal openness, as indicated both by higher scores on empathy scales and lower scores on measures of defensiveness, as well as to covary with higher scores on intelligence scales (Mayer & Salovey, 1995). In a review of the research concerning nonverbal emotional identification scales, Buck (1984, Chapter 7) concluded that emotional identification has generally correlated positively with self-report measures of empathy. Although a few recent findings have failed to find the relationship (Ickes et al., 1990; Levenson & Ruef, 1992), others have continued to find it (Mayer, et al., 1990). Regarding intellect-related measures, Ickes et al. (1990, p. 736) reported low, positive correlations between grade-point average and measures of accuracy for thought and feeling content. Here,

we use a combination of measures of empathy, defense (which are predicted to perform opposite to empathy), and self-reported SAT scores, which are indicative of intelligence, to predict emotional identification.

INTRODUCTION TO THIS STUDY

In this experiment, participants read descriptions of situations faced by eight actual individuals and then tried to estimate what those eight individuals were feeling at the time. The thoughts about situations reported by the eight individuals were formed into a standard scale in which each thought sample was followed by a set of 12 dichotomous item choices. Participants read through each thought sample and indicated how they thought the target had felt on the 12 item choices following each sample. Then, the accuracy with which the participants identified the emotions of the thought sample (calculated in several ways) was correlated with a number of criterion measures (e.g., self-reported empathy). In this study, people predict from thoughts to emotions, which requires more propositional thinking than predicting from faces to emotion as done in nonverbal studies. The general skill of predicting emotion, however, may be similar in each case, and hence may be related to similar criteria (e.g., self-reported empathy).

The hypotheses to be tested fell into two related classes: those that concerned how one decides what a person is feeling, and those that concerned people's ability to identify the target's emotions. The first set of hypotheses concerned relations among the possible criteria for accurate emotional responding. Recall that two possible criteria for what a target felt included what the target reported feeling (the *target criterion*) and what a group of observers believed the target felt (the *group-consensus criterion*). Our hypotheses concerning these criteria were that (a) there will be some agreement between the target and group-consensus criteria, and (b) both the target and group-consensus criteria will possess reliable variance independent of the pleasantness or social desirability of the set of emotional terms.

The second set of hypotheses concerned performance at emotional identification. Performance at emotional identification was gauged in two ways: as the participant's agreement with the target criterion, and as his or her agreement with the group-consensus criterion. We hypothesized that both (c) a participant's *target agreement* and *group-consensus agreement* will correlate with criterion measures of emotional intelligence. That is, they will correlate positively with empathy and indicators of intelligence, and negatively with defensiveness. Lastly, we predict that (d) a participant's *target* and *consensus* agreement scores will be unrelated to his or her agreement with the most socially desirable or pleasant emotional alternatives, as emotionally intelligent individuals should be able to adequately filter out such information (to the extent it is irrelevant) when judging the emotions of others.

METHOD

Overview

In the main portion of the study, 321 participants studied eight thought samples and judged the emotions of the person experiencing those thoughts. To create the eight thought samples, we first had eight real people (the *targets*) construct written reports of actual events that were affecting their current moods. Each of these eight target individuals described three situations he or she believed were most strongly contributing to his or her mood; each target then completed a mood scale. We retyped their written descriptions of the three situations, editing them slightly for clarity and to ensure their anonymity. In the following excerpt of a retyped thought sample, a young man attending the university described his desire for a romantic relationship, and continued, "My lab partner has a girlfriend, he goes home to visit her every weekend and then comes back to tell me what they did . . . he persists in bringing it up." The complete version of this thought sample was followed by 12 emotion-related pairs of terms (as were the seven other thought samples), including in this instance, "act as if no problem—sharing another's joy." Each of the 12 emotion-related test pairs that followed a thought sample contained one alternative that had been strongly endorsed by the target and one that had been less strongly endorsed. Other than the difference in endorsement strength, there was no attempt to match items as to emotional form. Thus, actions, feelings, and mood management were intermixed with one another in the pairs. In the given example, the person indicated he was "acting as if no problem," more than he was "sharing another's joy." It was the job of the participants to judge which alternative the target had felt more strongly. This dichotomous reply format forced a choice as to which alternative was felt more strongly and as a consequence provided a straightforward basis for the calculation of item agreement (e.g., either the participant and target agreed or disagreed); no estimate was requested of the participant as to how strongly each alternative was felt independent of the other. A complete example of an item and its correct answers appears in Table 1.

The participants worked through the 96-item scale (8 thought samples \times 12 items) and made judgments as to which alternative within a given dichotomous item the target had felt more strongly. The participant's judgments on the scale were then evaluated according to several criteria: the aforementioned *target agreement* and *group-consensus agreement*, as well as two additional criteria: agreement with the most socially desirable alternative—*desirability agreement*—and with the most pleasant alternative—*pleasantness agreement*. The most socially desirable and pleasant alternatives in the paired terms of the test were identified by using independent judges (discussed later). Participants' four scores, calculated according to the four given criteria, were then correlated with a group of personality measures related to emotional intelligence, including self-report measures of empathy and defensiveness.

TABLE 1
Complete Example of a Test Item

Written by a 20-year-old woman:

My best friend's father died this weekend. He had diabetes for a long time, and as he got older his health grew worse and worse. I went to his funeral on Monday. Many of my friends from high school were also there because we all wanted to be there for our friend and because we all knew and liked her father. It made me realize how lucky I am to have younger, healthy parents when I saw my friend standing there crying. Just watching her huge family come pouring into the synagogue also made me sad.

* * *

My roommate has been kind of blowing off her boyfriend. She told him she did not want to see him until Spring Break. He is hurt because he thinks she does not like him anymore, and he wants to come up here to see her this weekend. I have been gone almost every weekend since school started, giving her plenty of opportunities to have him up here while I am gone, and now I'm finally getting to stay here for the weekend and he might be coming up. (Why can't she go visit him instead?!?!)

* * *

I got to know this girl a little because she was in one of my classes. We happened to be talking about housing for next year and I mentioned wanting to live in the Knollwood apartments. She asked me if I would like to try to get into Knollwood with her and some of her friends. Without giving it very much thought, I said "yes." I gave it more thought and I am not sure if I can handle living with a smoker. I don't like where I'm living this semester, and if I get stuck somewhere I don't like next year too, I'm GOING TO GO CRAZY.

For each of the twelve pairs below, choose the word or phrase within that pair which best describes the reported feelings of the person who wrote the above passage across all of the situations she described.^a

- | | |
|---|---------------------------------------|
| 1. Be by myself*—Kick something | 7. Hostile—Unhappy for another* |
| 2. Stomping feet—Alone* | 8. Fearful—Apart from others* |
| 3. Pretend everything is okay*—Threaten a fight | 9. Cheated*—My teeth clenched |
| 4. Angry for someone else—Help a friend* | 10. Withdraw—Scared for someone else* |
| 5. Evade feeling*—Defiant | 11. Attacked—Isolate myself* |
| 6. Sharing another's anger*—Threatened with death | 12. Mad*—Delighted |

^aThe correct answers are identified with an asterisk.

Participants and Participant Protection

Participants in the main study were 321 undergraduates (204 women, 114 men, 3 gender unreported; age: 18–40 years) who participated either in order to fulfill a requirement for their introductory psychology course or in order to gain extra credit for an upper division psychology course. In the scale-construction phase of the study, 40 additional undergraduates (23 women, 17 men, age: 18–45) wrote thought samples and reported their current emotions. The information provided by 8 of these 40 targets then formed the basis of a scale presented to the 321 participants in the main study.

The 40 scale-construction participants completed a standard consent form at

the beginning of the study as well as a special release form at the conclusion of their work. The release form stated that, if signed, the participant's written responses would be read and evaluated in subsequent studies by other participants. Placement of the release form last ensured both that the participants would not consider the possibility of other groups of people reading what they had written until they were finished, and that the participants knew exactly what would be released (because they had completed writing it). Participants were assured that should we use their material (a) all descriptions of personal events and reactions would be retyped to eliminate identifying handwriting, spelling, and punctuation; (b) any identifying information would be changed; and (c) no information from any of the personality scales would be released. All 40 participants were willing to release their information under these conditions.

After the scale construction was finished, 12 new participants served as independent raters of the social desirability ($N = 10$) and pleasantness ($N = 2$) of the emotion-related alternatives that followed each of the eight thought samples. Two judges provided sufficiently high rater reliabilities for the pleasantness dimension that more raters were not required.

Materials

This materials section describes (a) the development of the performance scale for estimating emotion and (b) the various self-report criterion measures.

Emotional Accuracy Research Scale (EARS). This scale was constructed as a performance measure of a participant's accurate identification of others' emotions. A preliminary survey study was conducted in which participants revealed their present feelings and the thoughts that elicited them. These reports were then employed to form a scale of emotion identification.

Construction of the EARS. In a preliminary study, 40 participants were asked to report their thoughts about those situations that most strongly contributed to their moods at the time of the study. Participants were asked to describe three situations that most strongly contributed to their mood, because a person's mood is frequently the result of more than one mental set. For each situation about which they were thinking, participants were to describe (a) what led to the situation, (b) what the situation was, and (c) what happened in the situation that brought about the present emotion or mood. Responses were made in four blank lines below each question. Next, the participants reported their mood on a 78-item mood scale, called the *Present Reaction Scale*, described in greater detail in the Criterion Scales section.

Creation of the EARS. After the 40 sets of descriptions were completed, two judges selected eight descriptions from the 40 on the basis that descriptions appeared relatively emotionally expressive and well written, that they represented a

range of emotional situations, and that passages were drawn equally from men and women so as to equally represent both sexes. These eight descriptions were then lightly edited and retyped as the thought samples. Any potentially identifying information such as names of friends, dormitories, specific college majors, or hometowns, were changed so as to assure anonymity. An example of one finished three-part thought sample is as follows:

I have had a stressful week and I have had very little sleep or time to relax. I am stressed out about tests and work that has to be completed. I feel a little depressed because I have a feeling I failed two tests I took in the last two days. My day is also never ending with one thing after the other that has to be done.

* * *

Someone was harassing my girlfriend. A guy was hitting on my girlfriend and trying to get some play off her. I got a hold of him and told him to lay off because it was upsetting her and pissing me off. She and I talked about the situation last night and I still have this in mind. This may also be adding to my stressful week.

* * *

I have been trying to decide whether or not to go camping this weekend. I would like to go camping but I can't drive my car because I have no insurance to cover anyone other than myself. I would do it but my parents said not to and if anything happened I would be in deep trouble. I keep thinking about it and I can't decide whether or not to go or stay here at the university for the weekend. I am feeling pressure from right and wrong.

Each thought sample was then followed by 12 pairs of mood items drawn from the target person's responses to the Present Reactions Scale (described later). In each pair, an item that was rated more highly by the target was juxtaposed with a less highly rated item. Otherwise, phrases were intermixed independent of whether they involved emotions, emotional management, or emotion-related actions. These thought samples were followed by 12 such items including:

dared—isolate myself*
act as if no problem*—lively
chuckling—angry for someone else*

The asterisked alternatives were the ones within each pair that this particular target reported feeling more strongly. This thought sample and its 12 alternative pairs, along with 7 other thought sample/alternative pair sets formed the Emotional Accuracy Research Scale (EARS). (See Table 1 for a complete item.)

Scoring of the EARS. When a participant completes the EARS, he or she reads each of the 8 thought samples and then responds to the 12 response alternatives that follow it by circling the alternative within each item that he or she

believes the target felt more strongly. Four scores are calculated. *Target agreement* is the number of times the participant's judgment agreed with the original target's mood report over the 96 items. *Consensus agreement* is a weighted sum of the proportion of people with which the participants agreed for each of the 96 selections the participants made. For example, if the participant chose "isolate myself" in the first set of alternatives in the given example, and 90% of the group chose the same alternative, then .9 would be added to the score of the participant. If the participant chose "lively" for the second alternative in the given example, and if 15% of the group chose the same alternative, then .15 would be added to the participant's score. A participant's proportional agreements with each of the 96 items were summed over the test to obtain the weighted consensus score for that participant.¹

Two checks for socioemotional biases were also employed. The *Desirability agreement* score indicated the number of times the participant's choice agreed with the socially desirable alternative of the pair. Socially desirable alternatives were identified by having 10 additional participants (5 male, 5 female) read each alternative pair in context and judge which alternative was more socially desirable. The coefficient alpha reliability of the raters was $r(96) = .83$. Similarly, the *Pleasantness agreement* score indicated the number of times the participant's choice agreed with the more pleasant alternative of the pair. The more pleasant alternatives within each pair were identified by employing two raters (coefficient alpha $r(96) = .96$).

Criterion Scales. After the main group of 321 participants completed the EARS, they completed a number of self-report criterion scales potentially related to emotional intelligence.

Present Reaction Scale (PRS). The first criterion scale was a self-report mood scale constructed with the aim of measuring pleasant or unpleasant mood, emotional openness (later discarded, discussed later), and empathic openness (cf. Mayer et al., 1991). (This was also the scale employed by targets to report their mood, for the EARS.) The specific version of the scale employed 78 mood items representing six classes of emotional reactions (Mayer, Turner, & Thayer, 1994). The classes of terms relevant to the present three scores included (a) 12 pure emotions, including *happy* or *furious*, (b) 12 empathic reactions including *sharing another's joy* or *unhappy for another*, and (c) 12 mood management items such as being *open to my feelings* or *distance self from feeling*. Also in-

¹We also calculated unweighted group consensus scores. They performed much the same way as the weighted scores, attaining statistical significance in most or all of the same instances as did the weighted scores. The weighted scores, however, possessed greater reliability and appeared to perform in a slightly superior fashion in each case. Because they are also theoretically meaningful, they are employed here.

cluded were additional items related to (d) cognitive appraisals, such as *admired* or *cheated*, (e) plans of action, such as to *celebrate* or *slam a door*, and (f) bodily experiences such as *chuckling* or *trembling*. Each of these classes contained equal numbers of happy-related, angry-related, and fear-related items. Subjects responded to each individual mood(-related) item on a 5-point scale anchored at 1 (*definitely do not feel*) and 5 (*definitely do feel*). Pleasant or unpleasant mood was calculated by summing the responses to positive emotions (e.g., *happy*) and subtracting those responses to negative pure emotion terms (e.g., *fear* and *anger*). State empathy was calculated by summing all items indicating empathy for another, whether happy, fearful, or angry (e.g., *happy for another*, or *afraid for another*). For this sample, these two scales possessed alpha reliabilities of $r(193) = .68$ and $.61$, respectively. The mood management scale is not further reported due to its low reliability, $r(193) = .27$.

Empathy was measured as a personality trait by the Mehrabian and Epstein empathy scale (1972), which yields one global score of emotional responsiveness, and by the Davis empathy scale (1983), which yields four measures of empathy including: (a) empathic concern, (b) fantasy, (c) personal distress, and (d) perspective taking.

The first 197 participants completed scales of defensiveness, measured by the Marlowe-Crowne scale of social desirability (1960) and the Kolm scale of authoritarianism (1972), which yield one score each. Defensiveness should be negatively correlated with emotion identification because defenses divert or foreshorten the processing necessary to make correct decisions about feeling (Mayer & Salovey, 1995). In addition, the final 123 participants also were requested to report their SAT scores.²

Procedure

Participants in the main study were tested in groups. They first filled out the EARS, followed by the PRS, and the remaining self-report scales related to emotional intelligence.

RESULTS

Examination of Criteria for Emotional Intelligence

Recall that at least two criteria exist for what a given target is feeling: the target's report and the group consensus. We first examined these two criteria and their interrelation.

²Self-reported SAT scores provide only a very rough index of intellectual aptitude or attainment, but to the degree that they correlate with a criterion as expected they probably reflect reliable, valid variance reflecting intellectual ability. For example, Trice (1990) found that self-reported SATs did correlate with actual grade point average, but far less than actual SATs (self-reported $r = .37$, vs. actual $r = .58$, for the same group). Reported correlations between self-reported and actual SATs ranged from $r = .63$ to $.70$; elsewhere, ranges between $r = .74$ to $.85$ have been reported (Goldman, Flake, & Matheson, 1990).

We first report some characteristics of the group consensus scores. The reliability of the participant-generated consensus criterion can be directly estimated from the intercorrelations among the participants. To do this, it was necessary to transpose our participant (rows) \times variable (columns) subject matrix. Because of the large data matrix size, we compromised by transposing and calculating the coefficient alpha reliability based on 35 of the 321 participants (equivalent to estimating a test's reliability on only 35 of its 321 items, assuming homogeneity of items). The alpha reliability of the consensus score based on these 35 participant-evaluators alone was $r(96) = .92$, $p < .001$. The estimated reliability of $r(96) = .92$ implies an estimated reliability for an individual participant of $r(96) = .247 \approx .25$, and for the whole consensus group of $r(96) = .99$, both estimated from the Spearman-Brown prophecy formula based on the coefficient alpha. Application of the correction for attenuation would mean that the average correlation between a participant and the group consensus should be $r(96) = .50$. In fact, a calculation of this value made directly from the data for all 321 participants yielded a value of $r(96) = .50$ (this close congruence indicates both the validity of our estimates and the precision of psychometric theory).

We next examined the target criterion. It is not possible to compute the targets' reliability because different item pairs from the original emotion scale had been employed to represent each targets' responses on the EARS. Under such conditions, each target's single report provides no method of computing its internal consistency. The target's reports did, however, correlate with certain criteria (e.g., $r(96) = .27$, $p < .01$, with the social desirability of the item). This suggests that the target's report did possess some reliability (because correlations are not possible absent some reliability in both scores), and places the lower bounds of that reliability at $r < .10$, with no definite upper bound.

We next asked to what extent the target and the group consensus agreed as to what the target was feeling. (Note that this is different than studying individual differences among the group members' feeling estimates, which is discussed shortly). Chance would predict that the targets' reported feelings and those judged by the group would overlap 50% of the time (because each item was dichotomous). Consistent with the dyadic research literature, actual target-group consensus agreement was nonsignificantly higher than chance, 55%, $z(96) = .69$, *ns*). Another way to represent the same data is to examine the correlation between the item endorsements of the targets and those of the weighted group consensus; this was $r(96) = .18$, *ns*. Assuming a target's responses are about as reliable as an individual member of the consensus group (e.g., $r = .25$), the targets should correlate $r = .50$ with the consensus criterion just as did the consensus-group members, unless the target is using a different approach to making the ratings, or the assumption that their reliability was equal to or greater than $r = .25$ was too optimistic.

We considered two hypotheses beyond extremely low target reliability as to why targets and the group consensus might have diverged: (a) that only some of our targets were particularly competent at reporting their emotions, and (b) that

certain types of alternatives were more likely to be chosen by targets than by observers. If only some target individuals were competent in reporting their emotions, it should show up in greater consensus agreement for some targets than others. A one-way analysis of variance (ANOVA) examining percentage agreement across the eight targets with the consensus judgments indicated that there were no significant individual differences among targets as to how difficult their reactions were to guess, $F(7, 95) = .73, ns$. There was therefore no indication that particular targets were more inscrutable than others.

The target–consensus criterion discrepancy may also have occurred because the targets and the group used different strategies for reporting what the person felt. We found some evidence for this hypothesis. Recall that we had rated alternatives according to their pleasantness and their social desirability. A z test indicated that more socially desirable alternatives were chosen by the targets than by the group consensus (64% vs. 43%; $z(96) = 2.9, p < .01$), as were more pleasant alternatives (57% vs. 42%; $z(96) = 2.3, p < .01$). This fact suggests that the targets were more highly concerned with what others would think of them, whereas the group consensus was relatively insensitive to this social desirability bias. Later, in the section on individual differences, we further examine the possibility that only a minority of participants are good at judging the target's feelings.

Summary Comments. The first portion of the results indicates that there were considerable differences between the target's emotion reports and the group consensus. Some divergence may be due to lower target reliability. Other divergence may exist because targets chose more socially desirable alternatives than did members of the group. The results suggest that target and consensus criteria should be treated as separate criteria for what the target felt in the individual-differences hypotheses to be examined next.

Individual Differences in Emotional Intelligence

Recall that in this study, members of the main group of participants were asked to estimate what the target person felt. When participants in our study estimate what the target person felt—as reported on the 96 dichotomous items of the EARS—their responses can be scored according to four different performance criteria. These four scores represent the degree to which the participant agreed (a) with the target, (b) with the weighted group consensus, (c) with the more socially desirable alternative, or (d) with the more pleasant alternative. Each score will have different reliabilities and different correlations associated with it. A summary of the reliabilities and intercorrelations among the scores are found in Table 2.

Reliability and Intercorrelations of Performance Scores. The four performance scores represent the degree to which individual participants (from the group of 321) agreed with each of the foregoing criteria (as opposed to agreement among the four scores themselves). The four performance scores had some-

TABLE 2
Raw Correlations Among Participant Agreement Scores (Lower left),
and Reliabilities for Agreement Scores (Diagonal)

	Targets' Reports	Consensus (Weighted)	Social Desirability	Pleasantness
Targets' reports	.24**	—	—	—
Weighted consensus	.14*	.53**	—	—
Social desirability:	.51**	-.10	.52**	—
Pleasantness	.40**	-.19**	.81**	.56**

* $p < .10$; * $p < .05$; ** $p < .01$, two-tailed tests.

what different reliabilities; all were calculated across the 96 items (disregarding scenario) and all were significant at $p < .01$. The coefficient alpha for agreement with the target was $r(321) = .24$; for agreement with the group consensus it was $r(321) = .53$; for agreement with social desirability it was $r(321) = .52$; and for agreement with pleasantness it was $r(321) = .56$. The reliabilities can be found on the diagonal of the matrix in Table 2.

These criterion measures are somewhat independent of each other. A person who agrees with the group consensus does not necessarily agree more with the target report, $r(321) = .14, ns$, which might be expected from the lack of agreement between the target's emotional report and the group consensus criterion reported earlier. On the other hand, participants who agreed with the targets also (by necessity) chose more socially desirable responses, $r(321) = .51, p < .001$, because targets often chose the more socially desirable alternatives when reporting their feelings. Finally, socially desirable responding was also closely related to pleasantness of response, $r(321) = .81, p < .001$. These correlations may be found in the lower left of Table 2. The fact that, in some instances, an intercorrelation between two scores far exceeds the reliability of either score is because the individual agreement scores, in those instances, have been calculated from criteria that are themselves highly correlated (e.g., pleasantness, social desirability).

General Ability to Estimate Target and Consensus Criteria

Although participants had been instructed to estimate the feelings of the targets, in general they performed better at predicting the group consensus than the target report, $\bar{x}(321) = 66.4$ vs. 51.2, $t = 38.7, p < .001$. The superiority at judging the consensus criterion cannot be accounted for by the fact that the judge's response in part made up the criterion, for the judge was only 1/321 (.3%) of the consensus score. A Participant-Sex by \times Target-Sex ANOVA indicated that there was greater group consensus for what the four women targets felt in comparison to what the four men targets felt, $\bar{x}(318) = 63\%$ vs. 47%; $F(1, 316) = 4728.0, p < .001$, suggesting that the women targets were better at expressing their emotions. In addition, women participants were slightly better at predicting consen-

sus feelings than were men, $\bar{x} = 55\%$ vs. 54% ; $F(1, 316) = 3.91, p < .05$. Finally, the participants as a group tended to choose socially desirable and pleasant responses at about chance levels, $\bar{x}(321) = 44.1, 42.6$, respectively. $z = -.55, -.83, ns$, respectively). We now proceed to examine individual differences in agreement with the various criteria and their correlation with certain other self-reported personality measures.

Correlation of Performance Scores With Criteria. Our central hypothesis was that participants who had higher target and consensus agreement scores would also have higher openness (i.e., high empathy, low defensiveness) profiles and higher indications of intelligence on the criterion measures. Table 3 reports the correlations of the four performance scores with a variety of personality criteria; the central hypotheses were upheld. Higher target and consensus agreement scores did correlate positively with many measures of empathy and showed a trend toward negative correlations with measures of defense. Agreement with the target correlated significantly with Epstein–Mehrabian empathy, and with the Empathic Concern and Fantasy subscales of the Davis empathy scales, $r(321) = .13, .16$, and $.11, p$'s $< .05$. Agreement with the group consensus also correlated

TABLE 3
Correlation Among the Performance and Self-Report Measures ($N = 321$,
Except Where Noted)

Participant Self-Report: State and Trait Scales	Participant Performance: Agreement With Criteria			
	Target	Consensus (Weighted)	Social Desirability	Pleasantness
<i>State measures</i>				
Pleasant mood ^a	.06	-.03	.18**	.17**
State empathy ^a	.02	-.18**	.17**	.22***
<i>Trait measures</i>				
<i>Empathy</i>				
Epstein/Mehrabian	.13**	.24***	-.02	-.06
<i>Davis</i>				
Empathic concern	.16***	.13**	.01	.00
Fantasy scale	.11**	.23***	.05	-.02
Personal distress	-.06	.00	-.05	-.11*
Perspective taking	.07	-.06	.05	.07
<i>Defensiveness</i>				
Authoritarianism ^a	-.01	-.14*	.02	-.02
Marlowe–Crowne ^a	-.07	-.14*	-.04	-.01
<i>Academic aptitude</i>				
Reported SAT scores ^b	-.06	.26**	-.19*	-.23**

^a $N = 197$ for these scales; ^b $N = 92$.

* $p < .10$; ** $p < .05$; *** $p < .01$, two-tailed tests.

significantly with Epstein–Mehrabian empathy, and with the Empathic Concern and Fantasy Scales of the Davis empathy scales (e.g., $r(321) = .24, .13, .23, p$'s $< .05$). The Consensus Agreement scores showed tendencies toward negative correlations with the two scales closely associated with psychological defensiveness, Kohn's Authoritarian-Rebellion scale and the Marlowe–Crowne social desirability scale (e.g., $r(178) = -.14, -.14, p$'s $< .10$). For the subsample that was requested to report their SAT scores, general intelligence, as indicated by self-reported SAT scores, was found to be positively correlated with Consensus Agreement scores, $r(92) = .26, p < .01$. Moreover, the self-reported SAT scores were unrelated to the scales of empathy (with Epstein–Mehrabian, and the first two Davis Scales, respectively, $r(92) = .04, .06$, and $.08$, all ns , indicating that the empathic and intelligence predictions are independent of one another.³ These results held true whether one employed women or men as targets. For example, the correlation between weighted consensus scores and Epstein–Mehrabian empathy was $r(213) = .24, p < .01$ using all eight targets, $r(213) = .20, p < .01$ using the four women targets alone, and $r(213) = .19, p < .01$ using the four men targets alone. There were no significant (or noticeable) differences between men and women in their correlations with the various criteria.

Some other points of interest—and anomalies—also appear in the results. The state empathy scale (e.g., responding on the mood scale that one is currently feeling “anger on behalf of another” or “fear for another”) does not correlate with the traditional Empathy scales (highest $r(321) = .10, ns$, with Davis empathic concern), and unlike the other Empathy scales, was unrelated to emotion identification. This anomalous result with the “state” empathy scale may have occurred because people reporting state empathy may have been distracted from the task by their thoughts of nontargeted people they were empathizing with at the time. Some other points of interest include the fact that people who were in more pleasant moods chose more pleasant alternatives, $r(197) = .17, p < .05$, which is consistent with research on mood-congruent judgment (cf., Mayer et al., 1992), and that the trait measure of social desirability was unrelated to choosing socially desirable emotional alternatives.

In conclusion, the main findings were twofold: that both measures of target and consensus agreement correlated positively and significantly with measures of trait empathy, and showed trends toward negative correlations with defensiveness. In addition, consensus agreement correlated with indicators of intellectual ability.

³The correlations between empathy and emotional identification were unchanged when SAT was partialled out of those relationships within the subsample for which reported SAT scores were taken. For that subsample, the raw correlation between Epstein–Mehrabian empathy and weighted consensus was lower than that for the sample as a whole, at $r(79) = .07, ns$, versus $r(321) = .24, p < .01$, but rose to $r(79) = .14, p < .1$, when the influence of SAT was removed.

DISCUSSION

Reviewing the Aims of the Study and Comparison With Past Research

At the outset of this article we suggested that one basic skill related to emotional intelligence involves the ability to infer a person's emotions from his or her thoughts. The research reported here operationalized this skill by having individuals read the thoughts of others and then judge the emotion those others were feeling at the time. A number of other studies have been conducted previously that examined the ability to identify the emotions in others. This study focused more specifically on the internal thought-emotion linkage than have previous studies.

Summary of Results

Two criteria were chosen to represent the accuracy of participants' judgments: the degree to which they agreed with the *target* individual (who first reported his or her thoughts and feelings), and the degree to which they agreed with the group consensus. As it turned out, these two criteria were largely unrelated to one another. Nevertheless, the degree to which a participant approximated *either* criterion correlated positively with self-report measures of empathy and negatively with measures of defensiveness such as the Marlowe-Crowne scale of Social Desirability. In addition, and in keeping with our emotional-intelligence framework, the consensus agreement criterion also correlated positively with (reported) SAT scores. In fact, a multiple regression that included measures of empathy, defensiveness, and intelligence as predictors of consensus agreement explained a significant amount of variance, $R = .31$, $F(6, 151) = 2.77$, $p < .05$,⁴ although the amount of variance explained by a regression using the same predictors and the less reliable target agreement as the dependent measure was not significant, $R = .19$, $F(6, 151) = .98$, *ns*. This latter result may have been due to the lower reliability of the target agreement score, or due to the lower validity of the score. The target agreement score may have possessed lowered validity relative to the consensus if the targets had omitted crucial information from their thought samples or if they provided otherwise invalid emotional reports (e.g., due to emotional defensiveness, etc.).

Performance Versus Self-Reported Measures of Emotional Identification

Performance scales in general are known to be superior to self-report measures. For example, performance on intelligence tests predicts a variety of important criteria, whereas self-reported intelligence is a less powerful predictor. One ad-

vantage of this particular performance scale is its straightforward nature. Real people wrote the stories, and we had actual records of their self-reported moods at the time; this material gave us all that was needed to measure emotional identification ability. Because no complex laboratory setup was required (as it sometimes is with nonverbal emotion studies), it was possible to employ a large subject sample to ensure greater confidence in the obtained results.

Because this scale is a performance measure, as opposed to a self-report measure, it may ultimately serve as a more reliable criterion than the scales against which it was validated. A performance test is a more direct operationalization of the ability to perceive emotion than is self-reported empathy. Self-reported empathy, which is filtered through the individual's self-concept, may be distorted due to incomplete or faulty self-knowledge, defensiveness, or peculiarities of scale usage. Thus an outside observer who is dependent on self-report is one step removed from actual ability at understanding another person's feelings. Evidence of the value of this scale comes from Geher (1994), who videotaped real people talking about their present life situations, and then had experimental participants try to guess what those people were feeling. Skill at judging the feelings of such videotaped individuals—even more ecologically valid than this study in certain ways—correlated at roughly the $r = .50$ level with the scale developed here (and $r = 1.00$, when corrected for attenuation). That is, these performance measures correlate among themselves nearly as highly as do self-report scales of empathy. If we are correct in this contention, then future research with such scales should have better correlates with real-world criteria than do self-report empathy measures (which in turn may correlate more highly with other self-report instruments).

Group-Target Discrepancy as to How a Person Feels

One of the bases of emotional intelligence is that there are means for discriminating correct from incorrect answers to questions such as what a person is feeling (Mayer & Salovey, 1995, in press). In the work reported here, however, there were apparent differences between the target person and group consensus (in fact there was only a 55% agreement between the two criteria compared to 50% by chance, which was not significantly different). Some of this discrepancy was likely due to the low reliability of the self-report measure used for targets. At the same time, there is disturbing evidence that a person's feelings and the way a group consensually judges those feelings are different. Evidence for this comes in particular from the fact that the to-be-judged target individuals, but not the judges, endorsed a larger number of apparently socially desirable responses. There seems to be some imbalance between the target's socially desirable responses, and the judges, who may be estimating more directly what the target is feeling. Future research would need to be directed toward this discrepancy. In this work we instructed judges to assess "which alternative [mood item] the target had felt more strongly." In the future, researchers may want to make a dis-

⁴Pairwise correlations were used in the regression analysis and the N reported here represents a weighted mean of the pairs employed.

tion in such judgments between what the person *felt* more strongly, or would say in a public situation that they felt. The fact that 100% of our to-be-judged targets agreed to have their responses made public for the study suggests that they may have participated in the study by trying to appear nice or good to the experimenters, even though there was no explicit demand to do so (recall that permission for sharing of the protocols came at the end of the study).

Other technical issues need to be addressed as well. Reliability of target measurement needs to be improved. In subsequent research, target-reported feelings should be collected with some measure of consistency so as to better assess reliability. There are also good reasons to concentrate more centrally on consensus data: People may have found guessing what the target wrote an overly complicated exercise and reassigned themselves the task of deciding what an average person would feel in a given situation; these individuals would therefore not concern themselves with adjusting their responses to reflect the target's potential attempts to appear socially desirable. Evidence in support of this hypothesis comes from the fact that a few participants asked for clarifications of the instructions in such a way as to suggest they found the perspective taking quite difficult. Another interesting variation on this study would be to have targets rate themselves a few weeks later, as if external observers, to see whether their reported emotion would be more like their earlier selves, or more like group-consensus judges. If the target remained in the target role, this might be one way to assess the person's reliability. Presumably over time, however, the targets might gradually change their responses to that of an outside observer. By tracking the person's change of role from target to observer, it might be possible to understand some of the social influences that may be influencing socially desirable responding.

Despite these problems, both target agreement and consensus agreement did relate to the criterion measures of emotional intelligence, although unrelated to each other. This may mean that there are two ways by which to be emotionally intelligent, both of which should be measured.

Future Research

Earlier, we argued that the ability to predict emotions from thought will deliver a social advantage to an individual. High scorers on the EARS should therefore have advantages in certain life tasks. They may choose occupations that we suspect depend on emotional intelligence, including professions such as psychotherapy, social work, and teaching, or business careers including sales, academic or military recruiting, and personnel. In addition, we would predict they would have better, longer term intimate relationships, and better work histories within their occupation. If so, then it may be possible to educate those who are low in this skill to raise their ability levels and therefore better recognize the feelings of others. Exactly how demanding such a learning process is remains unknown. A number of school-based training programs potentially relevant to this skill indi-

cate that gains in interpersonal relations are possible through such training (Salovey & Sluyter, in press). It may well be worth the cost to obtain such positive social outcomes; the costs and benefits of such changes can be better evaluated by developing improved measures of emotional intelligence such as the one here, and studying the relation of such measures to the desired criteria.

REFERENCES

- Averill, J.R., & Thomas-Knowles, C. (1991). Emotional creativity. In K.T. Strongman (Ed.), *International review of studies on emotion* (Vol. 1, pp. 269-299). London: Wiley.
- Boucher, J.D., & Carlson, G.E. (1980). Recognition of facial expressions in three cultures. *Journal of Cross-Cultural Psychology*, *11*, 263-280.
- Bower, G.H. (1981). Mood and memory. *American Psychologist*, *36*, 129-148.
- Buck, R. (1984). *The communication of emotion*. New York: Guilford.
- Buck, R., Miller, R.E., & Caul, D.F. (1974). Sex, personality, and physiological variables in the communication of emotion via facial expression. *Journal of Personality and Social Psychology*, *30*, 587-596.
- Cantor, N., & Kihlstrom, J.F. (1987). *Personality and social intelligence*. Englewood Cliffs, NJ: Prentice-Hall.
- Cattell, R.B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, *54*, 1-22.
- Cronbach, L.J. (1960). *Essentials of psychological testing* (2nd ed.). New York: Harper & Row.
- Crowne, D.P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, *24*, 349-354.
- Davis, M.H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, *44*, 113-126.
- Davitz, J.R. (1969). *The language of emotion*. New York: Academic.
- Detterman, D.K. (1986). Human intelligence is a complex system of separate processes. In R.J. Sternberg & D.K. Detterman (Eds.), *What is intelligence? Contemporary viewpoints on its nature and definition* (pp. 57-61). Norwood, NJ: Ablex.
- Ekman, P., Davidson, R.J., & Friesen, W.V. (1990). The Duchenne smile: Emotional expression and brain physiology II. *Journal of Personality and Social Psychology*, *58*, 342-353.
- Ekman, P., Friesen, W.V., & Ellsworth, P. (1972). *Emotion in the human face*. New York: Pergamon.
- Ford, M.E., & Tisak, M. (1983). A further search for social intelligence. *Journal of Educational Psychology*, *75*, 196-206.
- Fiske, S.T., & Taylor, S.E. (1991). *Social cognition* (2nd ed.). New York: McGraw-Hill.
- Forgas, J.P. (1995). The affect infusion model (AIM): Review and an integrative theory of mood effects on judgment. *Psychological Bulletin*, *117*, 39-66.
- Gardner, H. (1983). *Frames of mind*. New York: Basic Books.
- Gardner, H. (1993). *Frames of mind: The theory of multiple intelligences* (10th Anniversary ed. New York: Basic Books.
- Geher, G. (1994). *Assessing the validity of a scale designed to measure empathic accuracy*. Paper submitted as a partial requirement for the Master's degree, University of New Hampshire, Durham.
- Goldman, B.A., Flake, W.L., & Matheson, M.B. (1990). Accuracy of college students' perceptions of their SAT scores, high school and college grade point averages relative to their ability. *Perceptual and Motor Skills*, *70*, 514.
- Guilford, J.P. (1967). *The nature of human intelligence*. New York: McGraw-Hill.

- Ickes, W., Stinson, L., Bissonette, V., & Garcia, S. (1990). Naturalistic social cognition: Empathic accuracy in mixed-sex dyads. *Journal of Personality and Social Psychology, 54*, 730-742.
- Kohn, P.M. (1972). The authoritarianism-rebellion scale: A balanced F scale with left wing reversals. *Sociometry, 35*, 176-189.
- Legree, P.J. (1995). Evidence for an oblique social intelligence factor established with a likert-based testing procedure. *Intelligence, 21*, 247-266.
- Levenson, R.W., & Ruef, A.M. (1992). Empathy: A physiological substrate. *Journal of Personality and Social Psychology, 63*, 234-246.
- Matarazzo, J.D. (1972). *Wechsler's measurement and appraisal of adult intelligence* (5th ed.). New York: Oxford University Press.
- Mayer, J.D., DiPaolo, M., & Salovey, P. (1990). Perceiving the affective content in ambiguous stimuli: A component of emotional intelligence. *Journal of Personality Assessment, 50*, 772-781.
- Mayer, J.D., Gaschke, Y., Braverman, D.L., & Evans, T. (1992). Mood-congruent judgment is a general effect. *Journal of Personality and Social Psychology, 63*, 119-132.
- Mayer, J.D., & Mitchell, D.C. (in press). Intelligence as a subsystem of personality: From Spearman's g to contemporary models of hot-processing. In W. Tomic & J. Kingma (Eds.), *Reflections on the concept of intelligence*. Greenwich, CT: JAI.
- Mayer, J.D., & Salovey, P. (1993). The intelligence of emotional intelligence. *Intelligence, 17*, 433-442.
- Mayer, J.D., & Salovey, P. (1995). Emotional intelligence and the construction and regulation of feelings. *Applied and Preventive Psychology, 4*, 197-208.
- Mayer, J.D., & Salovey, P. (in press). What is emotional intelligence? In P. Salovey & D. Sluyter (Eds.), *Emotional development and emotional intelligence: Implications for educators*. New York: Basic Books.
- Mayer, J.D., Salovey, P., Gomberg-Kaufman, S., & Blainey, K. (1991). A broader conception of mood experience. *Journal of Personality and Social Psychology, 40*, 100-111.
- Mayer, J.D., & Turner, J., & Thayer, M. (1995). *Mood organization across experiential domains*. Manuscript submitted for publication.
- Mehrabian, A., & Epstein, N. (1972). A measure of emotional empathy. *Journal of Personality, 40*, 525-543.
- Nisbett, R.E., & Wilson, T.D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*, 231-259.
- Ortony, A., Clore, G.L., & Collins, A. (1988). *The cognitive structure of emotions*. Cambridge: Cambridge University Press.
- Ree, M.J., & Earles, J.A. (1992). Intelligence is the best predictor of job performance. *Current Directions in Psychological Science, 1*, 86-89.
- Reik, T. (1952). *Listening with the third ear: The inner experience of a psychoanalyst*. New York: Farrer, Straus.
- Roseman, I.J. (1984). Cognitive determinants of emotions: A structural theory. In P. Shaver (Ed.), *Review of personality and social psychology: Vol. 5. Emotions, relationships, and health* (pp. 11-36). Beverly Hills, CA: Sage.
- Rosenthal, R., Hall, J.A., DiMatteo, M.R., Rogers, P.L., & Archer, D. (1979). Sensitivity to non-verbal communication: The PONS test. Baltimore, MD: Johns Hopkins University Press.
- Salovey, P., & Mayer, J.D. (1990). Emotional intelligence. *Imagination, Cognition, and Personality, 9*(3), 185-211.
- Salovey, P., & Sluyter, D. (Eds.). (in press). *Emotional development and emotional intelligence: Implications for educators*. New York: Basic Books.
- Smith, C.A., & Ellsworth, P.C. (1985). Patterns of appraisal in emotion. *Journal of Personality and Social Psychology, 48*, 813-838.

- Sternberg, R.J. (1988). *The triarchic mind: A new theory of human intelligence*. New York: Viking Penguin.
- Sternberg, R.J., & Smith, C. (1985). Social intelligence and decoding skills in nonverbal communication. *Social Cognition, 3*, 168-192.
- Thorndike, E.L. (1920). Intelligence and its uses. *Harper's Magazine, 140*, 227-235.
- Trice, A.D. (1990). Reliability of students' self-reports of scholastic aptitude scores: Data from juniors and seniors. *Perceptual and Motor Skills, 71*, 290.
- Wagner, H.L., MacDonald, C.J., & Manstead, A.S.R. (1986). Communication of individual emotions by spontaneous facial expression. *Journal of Personality and Social Psychology, 50*, 737-743.
- Wagner, R.K., & Sternberg, R.J. (1985). Practical intelligence in real-world pursuits: The role of tacit knowledge. *Journal of Personality and Social Psychology, 49*, 436-458.
- Walker, R.E., & Foley, J.M. (1973). Social intelligence: Its history and measurement. *Psychological Reports, 33*, 839-864.
- Wechsler, D. (1987). *Wechsler Memory Scale—Revised*. New York: Psychological Corporation.