ELSEVIER

Intelligence 29 (2001) 373–388

INTELLIGENCE

Predictive validity of the emotional accuracy research scale

Glenn Geher[a,*], Rebecca M. Warner[b], Andrew S. Brown[c]

[a]*Department of Psychology, State University of New York at New Paltz, New Paltz, NY 12561, USA*
[b]*University of New Hampshire, Durham, NH, USA*
[c]*Western Oregon University, Monmouth, OR, USA*

Received 20 March 1999; received in revised form 31 January 2000; accepted 15 March 2000

Abstract

This study was designed to assess the predictive validity of the Emotional Accuracy Research Scale [EARS; Intelligence 22 (1996) 89], a performance measure of emotional intelligence that has the benefit of being relatively practical to administer. Participants included 40 individually run undergraduate students who first completed the EARS and five self-report measures of trait empathy. Participants then engaged in a relatively realistic emotion-judgment task; they were asked to assess the emotions of six target individuals whose discussions of emotional issues were included in a videotape. Compared with their scores on the trait-empathy measures, participants' scores on the EARS were generally more predictive of accurately detecting the emotions of targets, thus providing evidence of the EARS' predictive validity as a measure of emotional intelligence. Measurement issues concerning emotional intelligence are discussed.

Keywords: Emotional intelligence; Empathy; Emotion defection

1. Introduction

Emotional intelligence (Salovey & Mayer, 1990) has been conceptualized as a kind of intelligence that includes the abilities both to accurately understand one's own emotions and to accurately interpret others' emotional states. Recently, some psychologists have found that

* Corresponding author. Tel.: +1-845-257-3091.
E-mail address: geherg@newpaltz.edu (G. Geher).

0160-2896/01/$ – see front matter
PII: S0160-2896(00)00045-3

emotional intelligence is a somewhat distinct form of intelligence (Mayer & Geher, 1996); in other words, it is not simply redundant with other conceptualizations of intelligence such as social intelligence (Cantor & Kihlstrom, 1987). Further, much recent popular work suggests that emotional intelligence is highly predictive of general functioning (Goleman, 1995) and of functioning in specific domains such as career performance (Goleman, 1998). Most recently, Mayer, Caruso, and Salovey (1999) have demonstrated that emotional intelligence has the characteristics of a "scientifically legitimate intelligence." Specifically, these authors found that emotional intelligence (a) is comprised of abilities that form a related set, (b) is related to extant intelligence constructs, and (c) develops with age between adolescence and adulthood.

The current research is primarily concerned with measuring the empathic component of emotional intelligence. Empathy has been defined in several different ways by past researchers (Wispe, 1986). Some researchers, such as Batson and Coke (1983), have defined empathy as sharing feelings with others and responding compassionately to others' feelings. Self-report measures of empathy have generally been designed to tap this component of empathy (e.g., Mehrabian & Epstein, 1972). Other empathy researchers have focused on empathy as the ability to know what another person is feeling (e.g., Levenson & Ruef, 1992). This research has focused on a component of empathy called "empathic accuracy" (Ickes, Stinson, Bissonette, & Garcia, 1990) as well as "perspective taking" (Batson & Coke, 1983).

Emotional intelligence has been defined in relatively broad terms that encompass several components of empathy. The current work is primarily concerned with the component of emotional intelligence relating to accurate perceptions of others' emotions. Different kinds of emotional intelligence measures have previously been employed to predict the ability to accurately perceive others' emotions.

2. Existing measures of emotional intelligence

Existing measures of emotional intelligence may be divided into (a) self-report measures and (b) performance measures.

2.1. Self-report measures

Some self-report measures of emotional intelligence have attempted to tap empathy by defining it as the sharing of feelings with others and responding compassionately to others' feelings (e.g., Mehrabian & Epstein, 1972). These measures rely on participants' accurate self-reports as a basis for whether they are empathic people who can accurately interpret emotional stimuli.

In addition to self-report empathy measures, other self-report measures have been designed to tap emotional intelligence more explicitly. For instance, the Trait Meta-Mood Scale (TMMS; Salovey, Mayer, Goldman, Turvey, & Palfai, 1995) has been employed as a measure of this general ability. The TMMS addresses the components of emotional intelligence associated with being in tune with one's own emotions. This scale requires that participants describe how well they believe they know their own emotions in terms of emotional clarity,

attention to self-emotions, and emotional regulation. Scores on this measure have been positively associated with outcomes such as life satisfaction (Martinez-Ponz, 1997).

Schutte et al. (1998) also developed a self-report measure of emotional intelligence. Their scale requires people to describe how well they are able to "hear" the emotions of others. This scale was found to have high reliability. Also, these researchers demonstrated discriminant validity for this measure by showing that scores on their scale were negligibly related to indices of intellectual intelligence. Most recently, Mayer, Salovey, and Caruso (1999) have made available an on-line empathy measure (as part of the larger Emotional IQ Test). This measure also is comprised of a series of self-reported items (e.g., "The suffering of others deeply disturbs me.") whereby subjects rate their perceptions of their own levels of trait empathy.

While several self-report measures of emotional intelligence have demonstrated their usefulness in many domains, such measures have not generally been found to predict actual performance on realistic emotional detection tasks (Levenson & Ruef, 1992). In other words, reporting that one is capable of accurately knowing another's emotions does not necessarily translate into the actual ability to do so. Also, self-report measures of emotional intelligence have been found to be significantly related to the social desirability response set [e.g., in one study (Geher, 1994), scores on the Crowne and Marlowe (1960) social desirability scale were positively and significantly correlated with scores on the fantasy, empathic concern, and perspective taking subscales of the Davis (1983) empathy scale].

2.2. Performance measures

Several researchers have employed actual performance measures to tap constructs related to the empathic component of emotional intelligence. One such measure is the PONS (Rosenthal, Hall, DiMateo, Rogers, & Archer, 1979), which assesses nonverbal sensitivity to emotional content by measuring one's ability to know how others feel based on an interpretation of nonverbal behavior. This measure employed videotapes depicting actors displaying different intensities of several emotions (e.g., anger). In some tapes, the faces of the actors could be seen, while in others, actors' faces were partially covered; thereby forcing subjects to tap some other nonverbal cue (e.g., body language). Subjects completing the PONS are given the task of assessing qualitative and quantitative aspects of actors' emotions. High scores indicated a tendency to accurately determine the targets' intended emotions, an ability that is very similar to the emotion-perception component of Mayer, Caruso, et al.'s (1999) and Mayer, Salovey, et al.'s (1999) conceptualization of emotional intelligence.

A similar measure has been developed by Ickes et al. (1990). In this research, participants were asked to write down the emotion they thought a target felt. Then blind judges rated whether what these participants wrote was similar to corresponding sentences written by the targets themselves. Similarly, Levenson and Ruef (1992) had participants watch videotapes of others and then assess, using an affect-rating dial, how the targets of those videotapes were feeling using an affect-rating dial. Ratings were made on a negative-to-positive-affect continuum. Empathic accuracy was measured by how closely the dial ratings of a subject

matched the dial ratings of a target. Again, the measure of empathy was how well one can judge the feelings of others, independent of self-perception and response bias.

2.3. Rationale for the Emotional Accuracy Research Scale (EARS)

While self-report measures of emotional intelligence may be limited in terms of their validity as predictors of emotional-detection ability, they do have some advantages over laboratory measures of empathic accuracy. For instance, self-report measures are typically much easier to administer than laboratory measures. In addition, they are less dependent on specific apparatus, such as videotape recorders, monitors, and computers.

The EARS (Mayer & Geher, 1996) was designed to afford the benefits of both self-report and laboratory measures of emotional intelligence. The EARS is a paper-and-pencil measure of emotional intelligence. However, it is also a performance measure. The EARS consists of eight emotionally-laden vignettes followed by pairs of mood items. Both the vignettes and the mood items were obtained from actual target individuals. Participants who complete the EARS are asked to read the vignettes and then choose mood items that they believe the target of the vignette reported feeling. For each participant completing the EARS, two emotion-perception scores are obtained. The *consensus-agreement* score corresponds to the degree to which participants completing the EARS choose items that are most commonly chosen. The *target-agreement* score corresponds to the degree to which participants are able to accurately choose items that the targets of the vignettes endorsed themselves.

The EARS has the practical benefits of being easy to administer. Further, it seems to tap the actual (as opposed to self-reported) ability to accurately detect others' emotions. In a previous study, the EARS demonstrated convergent validity (scores on the EARS were positively related to scores on Davis' (1983) measure of trait empathy) and discriminant validity (scores on the EARS were found to be negligibly related to SAT scores). Similar discriminant validation data have been described recently in a paper by Mayer, Caruso, et al. (1999). Mayer et al. have created a new measure of emotional intelligence, the Multifactor Emotional Intelligence Scale (MEIS) that includes a "stories" component that is a derivative of the EARS. In this paper, Mayer et al. describe a study in which they administered the MEIS and a modified version of the Army Alpha test of intelligence (Yerkes, 1921) to address the issue of discriminant validity. Similar to the results of Mayer and Geher (1996), these researchers found a moderate positive relationship between these measures of cognitive intelligence and emotional intelligence, thus demonstrating some degree of discriminant validation for measures of the emotional perception component of emotional intelligence.

The present research was designed to elaborate on past research on the EARS by assessing its predictive validity. Specifically, this research was designed to assess whether scores on the EARS were predictive of accurate emotion detection in a more real-life situation. Additionally, the predictive power of the EARS was juxtaposed with the predictive power of two self-report measures of empathy to assess whether scores on performance measures such as the EARS do, indeed, better predict this ability.

3. Method

A preselection procedure was designed to select participants who scored either extremely high or extremely low on the EARS. This extreme-groups design was utilized as participants in the main part of this study were run individually in 1-h sessions. Participants were run individually to ensure that each participant had the opportunity to focus an optimal amount of attention on the stimuli that were presented. To assess the differential predictive value of self-report empathy measures vs. the EARS on a performance-based laboratory measure of emotional perception, participants first completed the EARS and two different self-report measures of empathy. Participants then engaged in a laboratory emotion-perception task.

3.1. Participants

A total of 124 undergraduate students (52 males, 72 females) enrolled in introductory psychology at a large state university in New England were pretested on the EARS. Of these 124 potential participants, the 10 highest scoring males, 10 highest scoring females, 10 lowest scoring males, and 10 lowest scoring females ($N=40$) were asked to engage in a subsequent laboratory test of emotional detection. For their participation, students received partial credit towards fulfillment of a requirement for their introductory psychology class.

3.2. Materials

3.2.1. Preselection materials

For the preselection process, participants completed the EARS (Mayer & Geher, 1996) and two self-report empathy measures (Davis, 1983; Mehrabian & Epstein, 1972). In completing the EARS, participants first read eight emotionally laden vignettes written by other students. Then, for each vignette, participants were presented with a series of forced-choice items that included a pair of mood-related terms. Participants completing the EARS are given the task of determining which item in each pair they believe the target reported feeling more strongly. Two scores are computed: (a) a *target* score corresponding to the degree to which participants were able to accurately choose items that the actual targets reported and (b) a *consensus* score corresponding to the degree to which participants' responses matched other participants' responses (for a more detailed description of the EARS, see Mayer & Geher, 1996).

3.2.2. Laboratory measures of emotion perception

Participants who were selected for the second phase of the study completed several empathic accuracy tasks in a laboratory setting. This laboratory measure was conceptually similar to the EARS. However, instead of asking participants to read stories about targets and then rate their emotions (as in the EARS), in completing this laboratory measure, participants watched videotapes of targets who discussed emotionally-laden topics. The videotape consisted of six graduate students (three males and three females), who talked about three situations that were affecting their moods at the time they were being taped. After describing each situation, the targets wrote down one adjective which they felt best captured their moods at that time. Finally, the targets completed the Present Reactions Scale (Mayer,

Salovey, Gomberg-Kaufman, & Blainey, 1991); a mood scale that targets of the EARS also completed. In completing this scale, targets were asked to rate their moods along several broad dimensions.

3.2.3. Scoring the laboratory measure

Four scores were calculated for the laboratory empathy measure: two scores corresponding to multiple-choice items, and two corresponding to forced-choice items.

3.2.3.1. Multiple-choice items. First, recall that for each of three emotionally laden situations, targets were asked to write one word that best captured their moods. These emotionally-laden words were juxtaposed with two similar words to create multiple-choice items. For instance, if a target wrote "frustrated" after discussing an emotionally-laden situation, "frustrated" was juxtaposed with "angry" and "depressed" to create a multiple-choice item. As three such situations existed for each of six targets, a total of 18 multiple-choice items were created.

For each participant who watched the targets in the videos, two multiple-choice scores were computed. Target-agreement scores were computed by summing items for which participants' answers agreed with targets' answers. To compute consensus-agreement scores, data regarding each multiple-choice item from all participants were analyzed to determine the modal score for each item. Then consensus-agreement scores were computed by summing instances in which participants chose this modal, or consensual, score.

3.2.3.2. Forced-choice items. Forced-choice items were computed with the same method used to compute target-agreement and consensus-agreement scores for the EARS. Specifically, for each target, 12 pairs of mood items from that target's Present Reactions Scale scores were juxtaposed. Pairings were based on the criterion that one item in each pair was rated by the target as describing his or her mood more accurately. Participants completing this measure were asked to make 72 (12 pairs of items × six targets) judgments by determining which item in each pair the target reported feeling more strongly.

As with the multiple-choice items, two scores were obtained for the forced-choice measure. Target-agreement scores were computed for each participant by summing instances in which that participant chose the alternative in each pair that the target reported feeling more strongly. Consensus-agreement scores were computed by summing instances in which participants agreed with the item within each pair that was chosen by a majority of other participants.

3.3. Procedure

Participants who fit the criteria for inclusion in the laboratory phase of this study were run individually. After informed consent was obtained, participants were given an answer sheet that included all the multiple-choice and forced-choice items from the laboratory measures. Participants were asked to watch the videotapes and to try to best assess the emotions of the targets. After each of the 18 emotionally-laden segments of the videotapes was presented, the tape was paused for 30 s. During this time, participants were asked to complete the multiple-

choice item corresponding to the appropriate segment of video. During the six segments corresponding to the completion of a particular target (i.e., after each target's third emotionally-laden segment), the video was paused for 3 min. During this time, participants responded to the 12 pairs of mood items for a given target by indicating the item within each pair they believed the target reported feeling most strongly across all three emotionally laden situations. Participants repeated this process for all six target subjects in the videotape. One whole session lasted approximately 55 min. When participants were finished, they were thanked and thoroughly debriefed.

4. Results

The primary hypothesis of the present study is that performance-based measures of emotion perception should be stronger predictors of behavioral indices of emotional intelligence than standard self-report measures of trait empathy. Consequently, the analyses performed assessed how well two scores on a performance measure of emotion perception, the EARS (corresponding to both consensus and target agreement), predicted scores on a behavioral measure. Further, for comparison purposes, analyses addressed how well the five scores on the self-report measures of empathy predicted scores on these same dependent variables.

Analyses were performed in the following order: First, to assess the internal reliability of the different measures employed, a Cronbach's α was computed for each measure. Next, intercorrelations among the predictor variables were determined to assess the degree of shared variance between these variables. Also, to address the primary hypothesis (that scores on the performance-based test of emotion perception should be more strongly associated with behavioral measures of emotion perception than self-report empathy measures), zero-order correlations between all predictor variables and all dependent variables were computed. Additionally, four multiple regressions were performed to assess how well each variable independently predicted empathy in the laboratory.

4.1. Conceptual issues surrounding scoring

In this research, the EARS and the laboratory empathy measures yield scores pertaining to target and consensus criteria. As is indicated in the Method section, these criteria offer conceptually different indices of emotional intelligence. Target scores generally speak to whether individuals tend to agree with targets' self-reported emotions while consensus scores assess whether individuals agree with others' judgments regarding the targets' emotions. While target scoring of this sort has been employed in previous research (e.g., Ickes et al., 1990), such a scoring procedure has potential problems. Specifically, targets who produce "correct answers" for such measures may employ response biases and/or they may not accurately know themselves. Consensus scoring procedures, on the other hand, are based on aggregated judgments and, as such, tend to be more reliable. It is important to note that these criteria are mathematically and conceptually different when considering this research.

4.2. Instrument reliabilities

Cronbach's α coefficients were calculated to determine the internal reliabilities of the EARS and of the laboratory measures of empathy. Previous research on the EARS reported relatively low alpha coefficients. Specifically, Cronbach's α for target agreement was .24 while Cronbach's α for consensus agreement was .53 (Mayer & Geher, 1996). In the current research, an item analysis was conducted for both EARS scores to improve these reliabilities. After deleting items that reduced α for both target- and consensus-agreement scores, α values increased considerably to .75 and .80, respectively. A similar item analysis was conducted for the four laboratory measures. After deleting items that reduced internal reliability, the four laboratory measures achieved reasonable levels of reliability. For the multiple-choice items, α for target agreement was .52 and α for consensus agreement was .55. For the forced-choice items, α for target agreement was .86 and α for consensus agreement was .77. While the reliabilities for the multiple-choice items are somewhat low, they are considerably higher than the initial reliabilities for the EARS.

It should be noted that the preselection procedure in the current research entailed creating extreme groups based on EARS scores. Overall, this selection process was chosen to optimize the EARS' predictive validity. However, this process may have affected instrument reliabilities of the measures employed. As such, the reliability estimates of the laboratory measures may be somewhat inflated.

4.3. Descriptive statistics

The means and standard deviations for all predictor and criterion variables are reported in Table 1. Recall that the EARS yields two separate scores representing both consensus and target agreement. Similarly, the two kinds of items on the laboratory empathy measures (multiple-choice and forced-choice pairs) each yield consensus- and target-agreement scores. For each of these three measures (the EARS and two laboratory empathy measures), the mean for consensus scoring was greater than the mean for target scoring. For the EARS, the mean for target scoring (i.e., the average number of items for which the subject agreed with the target as to which alternative the target felt more strongly) was 33.45 (S.D. = 6.93), whereas the mean for consensus scoring (i.e., the average number of items for which the subject agreed with the group consensus as to which alternative the target felt more strongly) was 40.30 (S.D. = 6.63). A dependent measures t test revealed that this difference was statistically significant [$t(39) = 4.56$, $P < .01$]. Similarly, for the two kinds of items on the laboratory empathy measure, the mean for consensus scoring was found to be significantly greater than the mean for target scoring (see Table 1).

4.4. Intercorrelations among the predictor variables

The predictor variables in the present study included two scores for the EARS (separate consensus and target agreement scores), one global measure of trait empathy (Mehrabian & Epstein, 1972), and four self-report measures of more specific kinds of empathy from Davis' (1983) empathy scale including the self-reported ability to take the perspective of others, the

Table 1
Means and standard deviations of the predictor and criterion variables

	M	S.D.
Predictor variables		
EARS:		
Target agreement	33.45_a	6.93
Consensus agreement	40.30_b	6.63
Self-reported empathy:		
General empathy[a]	120.33	11.63
Empathic concern[b]	29.53	3.67
Fantasy scale[b]	25.85	5.00
Personal distress[b]	20.23	5.44
Perspective taking[b]	26.95	4.18
Criterion variables		
Multiple-choice items:		
Target agreement	9.70_a	1.38
Consensus agreement	11.03_b	1.69
Forced-choice items:		
Target agreement	26.48_a	7.60
Consensus agreement	35.08_b	5.39

Subscripts (a, b) indicate significant differences between consensus- and target-agreement scores for a given measure.

[a] Mehrabian and Epstein (1972).

[b] Davis (1983).

tendency to be empathically concerned about others, the inclination to engage in fantasy, and the propensity to be overly concerned with one's own personal distress. In general, both target and consensus scores on the EARS were found to be uncorrelated with scores on the five trait measures of empathy (see Table 2). Interestingly, the target and consensus scores on the EARS were found to be uncorrelated with each other as well [$r(40)=.02$, ns]. In other words, the degree to which participants agreed with the targets on the EARS was found to be unrelated to the degree to which participants agreed with the group consensus. This result was also found in previous research on the EARS and will be addressed in the Discussion. Several of the self-report measures of empathy were found to be significantly related to each other in theoretically meaningful ways. For instance, scores on the empathic concern scale were positively related to scores on the general empathy scale [$r(40)=.55$, $P<.01$].

4.5. Zero-order correlations between the predictor and dependent variables

To address the primary hypothesis that the EARS should better predict behavioral indices of emotional intelligence than self-report empathy measures should zero-order correlations were computed between the predictor variables and four criterion variables. The criterion variables involved in the present analyses include scores from the three-item multiple-choice section of the laboratory empathy measure scoring for both target and consensus agreement as well as scores from the forced-choice-pairs section of the laboratory empathy measure scoring for both target and consensus agreement.

Table 2
Intercorrelations among predictor variables

	EARS		Empathy variables				
	Target score	Consensus score	General empathy[a]	Empathic concern[b]	Fantasy[b]	Personal distress	Perspective taking[b]
EARS							
Target	–						
Consensus	.02	–					
Empathy							
General empathy	−.21	−.02	–				
Empathic concern	−.12	−.11	.55**	–			
Fantasy	−.07	.02	.39*	.28	–		
Personal distress	−.54**	.05	.30	−.03	−.02	–	
Perspective taking	.18	−.06	.24	.40**	.22	−.18	–

[a] Mehrabian and Epstein (1972).
[b] Davis (1983).
* $P<.05$.
** $P<.01$.

The zero-order correlations were computed to provide a general sense of the relationship between the different predictor measures of emotional intelligence and the behavioral measures of emotional intelligence. Interestingly, none of the 20 correlations between the trait measures of empathy and the four laboratory empathy scores were significant (see Table 3). These correlations ranged from $r(40)=-.28$ to .28.

The pattern of correlations between scores on the EARS and the criterion variables revealed generally positive relationships between the EARS and the behavioral indices of emotional intelligence. Of eight correlations computed between scores on the EARS and scores on the criterion variables, three were positive and significant (see Table 3). Specifically, these correlations included the correlation between consensus scores on the EARS and target scores on the three-item multiple-choice section of the laboratory empathy measure [$r(40)=.37$, $P<.01$], the correlation between consensus scores on the EARS and consensus scores on the three-item multiple-choice section of the laboratory-empathy measure [$r(40)=.36$, $P<.01$], and the correlation between the target scores on the EARS and the target scores on the forced-choice-pairs section of the laboratory-empathy measure [$r(40)=.43$, $P<.01$]. Of the other five correlations computed between scores on the EARS and scores on the laboratory-empathy measure, four were nonsignificant while one significant negative correlation existed between consensus scores on the EARS and target scores on the forced-choice section of the laboratory-empathy measure ($r=-.41$, $P<.01$).

This negative correlation is inconsistent with predictions regarding the EARS' relationship with the laboratory empathy measures. This troubling correlation may derive from several sources. This correlation is between a target index of the laboratory measure and a consensus

Table 3
Zero-order correlations between the predictor and criterion variables[a]

	Laboratory-empathy measures			
	Multiple-choice items		Forced-choice items	
Predictor variables	Target score	Consensus score	Target score	Consensus score
EARS				
Target	−.04	.05	.43*	−.09
Consensus	.37*	.36*	−.41*	.15
Empathy				
General empathy	.23	.25	−.06	.15
Empathic concern	−.22	−.28	.13	−.07
Fantasy	.07	.06	−.03	.28
Personal distress	.07	.18	−.21	.12
Perspective taking	−.19	−.01	.05	−.03

[a] Note that target and consensus scoring procedures are conceptually, mathematically, and empirically distinct.
* $P < .01$.

index of the EARS. It may be that consensus measures and target measures are very different conceptually, and, accordingly, relationships between such measures may not always be positive in nature. Further, this negative correlation may derive from Type I error. In any case, while this finding may be meaningful, it is inconsistent with the basic pattern of results. In general, scores on the EARS were positively correlated with scores on the laboratory-empathy measure while scores on the trait measures of empathy were uncorrelated with scores on the laboratory-empathy measure, supporting the hypothesis that a performance-based measure of emotional intelligence should be more strongly associated with behavioral measures of emotional intelligence than self-report measures.

4.6. Multiple regression analyses

To address the question of how well scores on the predictor variables differentially predict scores on the criterion (laboratory-empathy measure) variables, four separate multiple regression analyses were conducted. In each of these analyses, one of the four laboratory-empathy measure variables was used as the criterion variable while the two EARS scores (consensus and target agreement) and the five self-reported empathy scores served as predictor variables. To treat each predictor variable equally, a standard multiple regression was performed (i.e., all variables were entered on the same step). Thus, in each analysis, each predictor variable was assessed in terms of how well it independently accounted for variance in the dependent variable taking shared variance with other variables into account.

In one of these four multiple regression analyses, no combination of predictor variables accounted for a significant amount of variance in the dependent variable. This particular analysis employed the consensus-agreement-forced-choice score from the laboratory-empathy measure as a criterion variable.

In each of the other three regression analyses, a significant amount of variance in the criterion variable was accounted for by some combination of predictor variables. The first of these analyses used target-agreement-forced-choice scores as the criterion variable. Of the seven predictor variables in the present analysis, only two contributed significantly to the prediction of scores on the dependent variable. Consistent with the primary hypothesis of this study, these variables were the consensus and target-agreement EARS variables. None of the self-report empathy variables contributed significantly to predicting scores on the dependent variable. The most significant predictor [in terms of the semi-squared partial correlations (sr^2), which represents the amount of unique variability in the criterion variable predicted by a specific predictor variable] was the target-agreement EARS variable (sr^2=.18) followed by the consensus-agreement EARS variable (sr^2=.16). Altogether, 40% (27% adjusted) of the variability in this dependent variable was accounted for by scores on the target and consensus-agreement scores of the EARS.

In the remaining two analyses, scores from the consensus EARS scale, the Davis empathic concern scale, and the Mehrabian and Epstein trait-empathy scale all contributed significantly in accounting for variability in the laboratory empathy measure multiple-choice scale scoring for *both* target and consensus agreement.

The results from the analysis employing the laboratory-empathy-measure-multiple-choice scale scoring for target agreement are summarized in Table 4. The Multiple R was significantly different from zero [$F(7,32)=2.60$, $P<.05$]. The results from the analysis employing the laboratory-empathy-measure-multiple-choice scale scoring for consensus agreement were virtually identical to the results from the previous analysis described in Table 4. For this final regression analysis, a significant amount of variability was explained by the set of predictor variables [R=.64; $F(7,32)=3.24$, $P<.05$]. The three significant

Table 4
Multiple regression predicting target agreement for multiple-choice laboratory items from two EARS and five trait-empathy variables[a]

Independent variables	*B*	*b*	sr^2	*t*
EARS				
Target agreement	−.01	−.06	.00	−0.34
Consensus agreement	.07	.33	.11	20.33*
Self-report empathy measures				
General empathy[b]	.07	.55	.17	2.88**
Empathic concern[c]	−.16	−.44	.11	−2.35*
Fantasy engagement[c]	−.00	−.00	.00	−0.03
Personal distress[c]	−.05	−.18	.02	−1.01
Perspective taking[c]	−.05	−.15	.02	−0.91
			R^2=.36	
			R^2 (adjusted)=.22	
			R=.60	

[a] Note that target and consensus scoring procedures are conceptually, mathematically, and empirically distinct.
[b] Mehrabian and Epstein (1972).
[c] Davis (1983).
* $P<.05$.
** $P<.01$.

predictor variables in this analysis were the EARS scoring for consensus agreement, the Davis empathic-concern scale, and the Mehrabian and Epstein trait-empathy scale.

5. Discussion

Emotional intelligence has garnered much attention both in academic (e.g., Mayer & Salovey, 1990) and popular (e.g., Goleman, 1995) domains. Some researchers have found that emotional intelligence is predictive of a variety of positive outcomes (e.g., Martinez-Ponz, 1997). However, valid and accepted measures of emotional intelligence are still needed for researchers in this field. The development of such measures should, potentially, allow researchers to develop a more coherent conceptualization of emotional intelligence. The EARS was designed to be such a measure.

Several self-report measures of emotional intelligence have been developed (e.g., Schutte et al., 1998). These measures have demonstrated validity. However, as they are self-report measures, they are somewhat limited. For instance, self-report measures of emotional intelligence may be particularly subject to social desirability response biases. The EARS, on the other hand, is a performance measure of the empathic component of emotional intelligence. Being a performance measure gives the EARS the benefit of tapping an actual ability while, concurrently, being relatively free of social desirability response bias.

In previous research (Mayer & Geher, 1996), the EARS demonstrated both convergent and discriminant validity. However, that previous study did not adequately address the internal reliability nor the predictive validity of the EARS. The current research primarily addressed these issues. After completing a careful item analysis, target-agreement and consensus-agreement scores for the EARS increased from 0.24 to 0.75 and from 0.53 to 0.80, respectively. Thus, this revised version of the EARS is clearly more internally reliable.

Further, the current data suggest that the EARS has demonstrated reasonable predictive validity. Scores on the EARS were generally more predictive of empathy in the laboratory task compared with scores on the self-report empathy measures. Thus, the EARS seems to tap how well people can understand the emotions of others in a context that includes information across several sensory modalities; i.e., these data suggest that the EARS is predictive of emotional detection in a relatively real-life setting.

The current data also raise questions regarding the differential importance of consensus agreement and target agreement as criteria of the empathic component of emotional intelligence. The multiple regression analyses in the current research have several implications regarding distinctions between consensus agreement and target agreement. Except for in the first such analysis, where none of the predictor variables accounted for a significant amount of variability in the criterion variable, the EARS consensus-agreement score consistently appeared as a significant predictor of the different criterion variables. Additionally, in predicting one of the target-agreement variables, both the target-agreement and consensus-agreement EARS accounted for a significant amount of variability in the criterion variable. Further, in predicting the multiple-choice scale scoring for target agreement and in predicting this same scale scoring for consensus agreement, a combination of the EARS

consensus-agreement score and the two general-empathy-trait scores accounted for significant amounts of variability in the criterion variables.

These findings imply that the ability to agree with a consensus as to an individual's emotions may be beneficial in understanding emotions from the perspectives of both actors and observers. However, the ability to accurately assess an actor's own self-report seems to be, in general, less related to understanding one's emotion from the perspective of both actor and observer. Additionally, the tendency for a person to report that he or she is an empathic person may be related to understanding the emotions of a given target in terms of both how that target describes his or her own emotion as well as how other observers judge that emotion. Thus, self-reported measures of emotional intelligence clearly have demonstrated utility. However, as is most clearly indicated by the zero-order correlations reported in Table 3, this tendency to report high levels of empathy in oneself seems to be, in general, less related to accurate emotion detection than the tendency to score high on the EARS.

The fact that the EARS was more predictive of the performance-based-emotion-perception measures than the self-report empathy scales may, to some degree, result from shared method variance. Importantly, the fact that some of the measures in this research are performance-based and some of the measures in this research are self-report-based is by design. The fundamental impetus for this research, initially, was a skepticism regarding self-report empathy measures as predictors of relevant behaviors (such as being able actually to discern others' emotions). For such self-report measures, which are very widely employed, to be useful they would need to demonstrate predictive validity. In other words, for such measures to be useful, they would need to, ultimately, predict scores on some performance-based dependent measure. Accordingly, the shared method variance between the EARS and the dependent measures in the current research exists by design. The main point here is that performance measures of the emotion-discernment component of empathy can be just as easy to administer as self-report empathy measures with the added luxury of actually predicting how well people can perceive others' emotions. Ultimately, the current reasoning goes, work in this field would benefit from researchers using performance-based measures of constructs such as emotion perception whenever possible.

The findings in the current research also speak to the differential *meaning* of target and consensus scoring criteria. The correlation between target scores on the EARS and consensus scores on the EARS was .02. This finding suggests that the ability to know a person as he or she knows him or herself and the ability to agree with others as to what a target is feeling are largely independent dimensions.

Interestingly, this finding that emotion perception based on target scoring and emotion perception based on consensus scoring are somewhat independent, seems to pertain to Nisbett and Wilson's (1977) work on our ability to report on our own mental processes. Nisbett and Wilson argue that, more often than we believe, people are not privy to the actual reasoning underlying their decisions. To the extent that this ignorance regarding people's own mental processes exists, it makes sense that knowing what an individual thinks he or she feels may be quite different from knowing what others think that same individual feels. In other words, to the extent that humans are not adept at knowing the actual reasoning underlying their cognitive processes, it makes sense that consensus and target measures would yield different kinds of results and, further, that such measures would be orthogonal to one another.

With regard to target and consensus scoring, both kinds of scoring procedures have their costs and benefits. Target scoring makes sense in that it theoretically gets at one of the fundamental aspects of emotional intelligence: How well does one know the feelings of another. However, target scoring has its problems. Targets may not be accurate reporters of their own emotions due to response biases and/or poor self-understandings. Consensus scoring procedures, on the other hand, are more reliable. Further, in the current research, the EARS consensus score was most predictive of other indices of emotional intelligence suggesting that consensus scoring may be more useful than target scoring as well.

5.1. Future research

While the current research demonstrates internal reliability and predictive validity of the EARS, this research also raises questions for future research. For instance, these findings are somewhat unclear regarding the relationship between target and consensus agreement. In the current sample, target agreement and consensus agreement were uncorrelated. In other words, being able to accurately assess a target's self-report may be unrelated to being able to accurately assess others' perceptions of that target's self-report. This finding suggests two very different abilities that may underlie emotional intelligence.

However, interestingly, the target and consensus scores of the EARS were not predictably related to target and consensus scores of the laboratory-empathy measure. Specifically, target agreement on the EARS was not consistently related to the target-empathy scores while, similarly, consensus agreement on the EARS was not consistently related to consensus-empathy scores (see Table 3). For instance, while consensus agreement on the EARS was positively related to target agreement on the multiple-choice section of the laboratory measure, target agreement on the EARS was uncorrelated with this same criterion variable. Further research needs to more specifically address the differential correlates of target-agreement ability and consensus-agreement ability.

Acknowledgments

We gratefully acknowledge the assistance of Kathleen Bauman, Katherine Black, Christopher Hakala, Andrew McGarva, and Dennis Mitchell for participating as the targets in the latter half of this study. We also wish to thank Randolph Grace, Todd Gurry, and Richard Lyczak for their assistance with writing the computer programs used in the data collection. Finally, we wish to recognize the assistance of John Mayer and Victor Benassi who also played integral roles in the development of this project.

References

Batson, C. D., & Coke, J. S. (1983). Empathic motivations of helping behavior. In: J. T. Cacciopo, & R. E. Petty (Eds.), *Social psychophysiology: a sourcebook* (pp. 417–433). New York: Guilford.

Cantor, N., & Kihlstrom, J. F. (1987). *Personality and social intelligence*. Englewood Cliffs, NJ: Prentice-Hall.

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.

Geher, G. (1994). *Assessing the validity of a scale designed to measure empathic accuracy.* Paper submitted as partial requirement for the Master's degree, University of New Hampshire, Durham, NH.

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

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

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.

Levenson, R. W., & Ruef, A. M. (1992). Empathy: a physiological substrate. *Journal of Personality and Social Psychology, 63*, 234–246.

Martinez-Ponz, M. (1997). The relation of emotional intelligence with selected areas of personal functioning. *Imagination, Cognition, and Personality, 17*, 3–13.

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

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

Mayer, J. D., Salovey, P., & Caruso, D. R. (1999). The emotional IQ test. (on-line) Available at: http://www.virtent.com/eq.html..

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.

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.

Rosenthal, R., Hall, J. A., DiMateo, M. R., Rogers, P. L., & Archer, D. (1979). *Sensitivity to non-verbal communication: the PONS test.* Baltimore: Johns Hopkins Univ. Press.

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

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

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

Wispe, L. G. (1986). The distinction between sympathy and empathy: to call forth a concept, a word is needed. *Journal of Personality and Social Psychology, 50*, 314–321.

Yerkes, R. M. (1921). Psychological examining in the United States. *Memoirs of the National Academy of Sciences* (15:890 pp. entire).