J Nonverbal Behav (2011) 35:305–326
DOI 10.1007/s10919-011-0115-4

ORIGINAL PAPER

Assessing the Ability to Recognize Facial and Vocal Expressions of Emotion: Construction and Validation of the Emotion Recognition Index

Klaus R. Scherer · Ursula Scherer

Published online: 24 July 2011

Abstract Despite extensive research on emotional expression, there are few validated tests of individual differences in emotion recognition competence (generally considered as part of nonverbal sensitivity and emotional intelligence). This paper reports the development of a rapid test of emotion recognition ability, the Emotion Recognition Index (ERI), consisting of two subtests: one for facial and one for vocal emotion recognition. The rationale underlying the test's construction, item selection, and analysis are described and a major validation study with more than 3,500 professional candidates, providing stable norms, is reported. Additional analyses concern differences for gender, age, and education, as well as correlations with cognitive intelligence and personality factors. Moreover, a separate validation study with a student sample reports the correlations of the ERI with some of the major published tests in this area, demonstrating satisfactory construct validity. Correlations between ERI scores and the position of candidates in the organizational hierarchy suggest that recognition competence might be might contribute to predicting career advancement.

Keywords Emotional intelligence · Emotional competence · Personality · Facial/vocal emotion expression · Emotion recognition · Test development

Electronic supplementary material The online version of this article (doi:10.1007/s10919-011-0115-4) contains supplementary material, which is available to authorized users.

K. R. Scherer (✉)
Swiss Centre for Affective Sciences, University of Geneva, rue des Battoirs 7, 1205 Geneva, Switzerland
e-mail: Klaus.Scherer@unige.ch

U. Scherer
SIGMA Assessment, Geneva, Switzerland

Introduction

The notion of emotional intelligence (EI) enjoys great popularity but has proven recalcitrant to objective measurement. While many EI self report instruments seem to assess social and emotional adjustment rather than emotion skills (see contributions in Matthews et al. 2007), the EI *ability model* proposed by Mayer and Salovey (1993) focuses on four specific skills (perceiving, using, understanding, and managing emotions). While emotion perception (the ability to detect and decipher emotions in faces, pictures, voices, and cultural artifacts) is considered as basic, much of the emphasis in the model is on *knowledge about emotion*, which can be construed as crystallized intelligence in a way that is comparable to many other types of cognitive knowledge, as is reflected in the nature of the tests proposed and the lively discussion, based on considerable data (see overviews in Schulze and Roberts 2005), as to the differential validity of EI and IQ tests and their ability to discriminate different subfactors of cognitive performance. Scherer (2007) has suggested to dissociate the study of emotion skills from cognitive conceptualizations of intelligence and to focus on the degree to which an individual's emotion mechanism optimally works regarding its evolutionary function and strategic aims in culturally defined social situations, proposing the term *emotional competence* (EC). Three major domains of EC can be identified: emotion production, emotion regulation, and emotion perception. *Emotion production competence* refers to the appropriateness of the total pattern of bodily and behavioral changes as an adaptive response to a relevant event, allowing the organism to successfully cope with its consequences. *Emotion regulation competence* reflects an individual's ability to monitor and manipulate his or her emotional state and its motor expression for organismic homeostasis, sociocultural norms and expectations, and strategic intentions. In contrast, *emotion perception competence* refers to the ability to accurately perceive and interpret the emotional state of other individuals to correctly infer their reactions to salient events and to predict their action tendencies (Scherer 1984, 2001, 2007, 2009). Clearly, the ability to accurately infer the emotions of others is a central socio-emotional competence as it provides important information on the reaction of significant others to recent events (including our behavior) and their likely actions in the future—information that is central to our strategic interaction management. Thus, in negotiations or board meetings it is essential to understand our opponents emotional reactions to our moves, in the workplace we need to be able to gauge the reaction of our collaborators and our subordinates to our decisions or changing situations, in family life smooth interactions and enduring positive relations are greatly furthered by empathy based on accurate emotion inference. Furthermore, the ability to correctly identify the emotional reactions of others is an essential professional skill for many different types of occupation – therapists, teachers, policemen, lawyers, salesmen, politicians, and many other professions. Past research has shown major individual differences in this ability (Bänziger et al. 2009; Hall and Bernieri 2001) and thus it seems reasonable to subsume this competence under an ability concept of personality. While, to our knowledge, there are no controlled studies of heritability or other constitutive factors that might determine the respective level of ability, most researchers seem to assume a relatively high degree of plasticity and possibilities for improvement with appropriate training (see Scherer 2007 on a discussion of competence, ability, and skills, and their respective plasticity).

Given the importance of this socio-emotional competence, and the apparent conduciveness to improvement with skills training, it is surprising that there are few established diagnostic instruments that allow a reliable and validated assessment of this essential competence, although it is generally acknowledged to be a central ability component of EI

(Matthews et al. 2007; Mayer and Salovey 1993). Recently, Joseph and Newman (2010) have discussed the issue from an applied perspective and use an integrative meta-analysis to propose a progressive (cascading) pattern among ability-based EI facets, in which emotion perception must causally precede emotion understanding, which in turn precedes conscious emotion regulation and job performance. In consequence, the measurement of this basic ability upon which other skills need to build should have absolute priority. Obviously, emotion recognition ability cannot be objectively assessed by self report EI questionnaires. Even performance-based EI tests that are claimed to study actual abilities do not allow assessment of this competence systematically according to established psychometric criteria. For example, the Mayer Salovey Caruso Emotional Intelligence Test (MSCEIT; Mayer et al. 2003), the leading instrument in this domain, contains only a handful of emotion recognition items to assess this component. Furthermore, the items in this task, consisting of abstract designs, landscapes, and ambiguous facial expressions, are not defined for the underlying emotions; answers are scored as "correct" by comparing them to the mean scores of reference or "expert" groups. Given the lack of an appropriate criterion to define accuracy of recognition, it is exceedingly difficult to ascertain exactly what is measured by these items. One suspects that what is measured is conformity to majority standards for the interpretation of ambiguous emotion representations in expressions or objects rather than the capacity to correctly recognize prototypical forms of expressions for major emotions (and thus, the capacity to infer sender reactions and action tendencies).

The lack of psychometrically sound and construct-validated test instruments capable of diagnosing individual differences in the central emotion perception ability component of EI (see Joseph and Newman 2010) is all the more surprising because emotion psychologists have extensively studied the capacity of individuals to recognize emotions from facial and vocal expressions (see Ekman 1972, 2007; Ekman and Rosenberg 2005; Scherer et al. 2003) and as there is evidence that this competence can be improved rather efficiently (e.g., see Ekman's micro-expression training tool,[1] Russell et al. 2006). Another research tradition in the area of nonverbal communication focuses on *nonverbal sensitivity*, defined as the ability to infer emotions or interpersonal attitudes from nonverbal cues in face, body, and/or voice (see Hall and Bernieri 2001, for a review). Although these two research strands have produced published instruments to measure emotion recognition in different modalities of expression (see the overview in the "Construct Validity" subsection), they are rarely used in EI research.

This article describes the development and validation of the Emotion Recognition Index (ERI), a test consisting of two subscales, namely, the Index of Facial Emotion Recognition (FACIAL-I) and the Index of Vocal Emotion Recognition (VOCAL-I). Based on a specific ability concept of EC defined in terms of emotion functions (see Scherer 2007, for a detailed discussion), the instrument assesses the ability of individuals to correctly infer target emotions from actor portrayals of vocal and facial emotion expressions. In this article, we describe, separately for each subscale, the construction of the test, item analyses, measures of central tendency and variation that can serve as comparison standards, and establishment of construct validity. In addition, we report first efforts to investigate gender differences, the effects of social and educational background, potential relationships with personality traits, and the predictive validity of the test.

[1] See http://face.paulekman.com/aboutmett2.aspx.

We aim to test the ability to correctly recognize intentionally produced full-blown emotional expressions as prototypical communicative signals of an individual's emotional reaction to an event and of his or her consequent action tendencies. Specifically, the ERI is designed to reliably measure individual differences in emotion recognition ability in a reasonably short period, given the targeted application as a rapid screening instrument—compared to lengthy (45–60 min.) assessment instruments like the MERT (Bänziger et al. 2009).

Method

Test Construction

General Design

Because most research has been devoted to facial and vocal expression of emotion, we decided to create two subtests for measuring emotion recognition ability, one from photographs of facial expressions and the other from voice recordings of vocal expressions. Expressions of joy/happiness, sadness, fear, anger, and disgust were chosen because these emotions play a central role in human behavior in social contexts and are considered by most theorists as "basic" and thus universal (see Ekman 1992; Izard 1977), which is an important prerequisite for being able to use the test on a global scale. Because the test is intended to be used in applied settings, such as assessment in the human resource domain, its duration was not to exceed approximately 20 min, instructions included. We therefore decided to limit the test to 60 items, 30 for each subtest, which still allows the inclusion of a sufficient number of items for each of the five emotions.

Facial Subtest Items

The test stimuli for the five chosen emotions were selected from the series of 65 still photographs published by Ekman and Friesen (1976) as Pictures of Facial Affect (POFA), a corpus that has been extensively used in research on facial expression and for which the universality of recognition has been solidly established (Ekman et al. 1987). The POFA series contains both prototypical examples of pure emotions and photographs of blends of two different emotions. The ERI contains mainly photographs showing single prototypical emotions. A few photographs of blends of two emotions (as contained in the POFA) were included to increase the difficulty of the item set (given the relatively high intensity and prototypicality of the portrayals). Given the fact that the POFA items contain the prototypical action units that are widely considered to be associated with the respective emotions and that have been empirically shown to occur in spontaneous expressions of the respective emotions (Ekman and Rosenberg 2005) these stimuli can be considered as valid representation of the emotion-specific expressions.

The 30 items selected from the POFA series included more items for the emotions that are more difficult to detect and fewer for those that are very easily recognized (as based on the accuracy percentages reported in Ekman et al. 1987). As shown in Table 1, the resulting number of items per emotion was as follows: joy, 4 (one blended item); disgust, 4; anger, 7 (two blends); sadness, 7; fear, 8 (two blends).

Table 1 Confusion matrices as part of the item analysis (professional sample) for facial and vocal subtests

Stimulus no.	Emotion encoded	Emotion decoded				
		Anger	Fear	Joy	Sadness	Disgust
Facial subtest						
2	Anger/contempt	*0.03*	0.11	0.36	0.47	0.04
5	Anger/sadness	*0.35*	0.02	0.00	0.46	0.16
10	Anger	*0.91*	0.04	0.00	0.02	0.02
15	Anger	*0.89*	0.02	0.00	0.03	0.05
20	Anger	*0.55*	0.33	0.00	0.00	0.12
23	Anger	*0.57*	0.06	0.03	0.15	0.19
27	Anger	*0.35*	0.30	0.02	0.26	0.07
Mean	Anger	*0.55*	0.15	0.07	0.15	0.08
4	Fear/Surprise	0.01	*0.84*	0.03	0.02	0.10
9	Fear	0.07	*0.46*	0.00	0.32	0.15
12	Fear/Anger	0.64	*0.25*	0.00	0.01	0.10
16	Fear	0.02	*0.94*	0.00	0.00	0.03
18	Fear	0.01	*0.94*	0.00	0.00	0.05
22	Fear	0.02	*0.86*	0.00	0.01	0.10
25	Fear	0.01	*0.87*	0.01	0.01	0.11
28	Fear	0.08	*0.31*	0.06	0.35	0.21
Mean	Fear	0.17	*0.67*	0.01	0.07	0.08
1	Joy	0.00	0.01	*0.98*	0.00	0.01
3	Joy	0.00	0.02	*0.97*	0.00	0.01
14	Joy/Contempt	0.00	0.01	*0.97*	0.01	0.00
17	Joy	0.00	0.00	*0.98*	0.01	0.00
Mean	Joy	0.00	0.01	*0.98*	0.01	0.01
7	Sadness	0.02	0.42	0.01	*0.54*	0.01
8	Sadness	0.02	0.02	0.00	*0.76*	0.20
13	Sadness	0.06	0.03	0.00	*0.81*	0.09
21	Sadness	0.02	0.12	0.00	*0.79*	0.07
24	Sadness	0.04	0.02	0.07	*0.81*	0.06
26	Sadness	0.11	0.08	0.01	*0.70*	0.09
29	Sadness	0.01	0.01	0.00	*0.91*	0.07
Mean	Sadness	0.08	0.09	0.01	*0.72*	0.10
6	Disgust	0.03	0.00	0.00	0.00	*0.96*
11	Disgust	0.14	0.01	0.00	0.01	*0.84*
19	Disgust	0.19	0.01	0.00	0.00	*0.80*
30	Disgust	0.10	0.00	0.00	0.01	*0.89*
Mean	Disgust	0.12	0.01	0.00	0.00	*0.87*
Vocal subtest						
5	Anger	*0.79*	0.03	0.06	0.08	0.04
8	Anger	*0.90*	0.05	0.03	0.00	0.02
10	Anger	*0.42*	0.27	0.01	0.30	0.01

Table 1 continued

Stimulus no.	Emotion encoded	Emotion decoded				
		Anger	Fear	Joy	Sadness	Disgust
13	Anger	*0.92*	0.01	0.05	0.00	0.01
19	Anger	*0.89*	0.07	0.01	0.00	0.03
26	Anger	*0.46*	0.08	0.05	0.01	0.40
30	Anger	*0.59*	0.03	0.06	0.00	0.32
Mean	Anger	*0.71*	0.08	0.04	0.06	0.12
3	Fear	0.02	*0.82*	0.01	0.15	0.01
4	Fear	0.05	*0.70*	0.12	0.04	0.08
15	Fear	0.00	*0.97*	0.02	0.00	0.01
18	Fear	0.05	*0.69*	0.12	0.03	0.10
22	Fear	0.04	*0.07*	0.03	0.18	0.68
Mean	Fear	0.03	*0.65*	0.06	0.08	0.18
7	Joy	0.06	0.00	*0.63*	0.00	0.31
12	Joy	0.00	0.00	*0.54*	0.01	0.44
14	Joy	0.00	0.06	*0.83*	0.04	0.06
16	Joy	0.05	0.01	*0.21*	0.07	0.67
24	Joy	0.06	0.09	*0.73*	0.11	0.01
25	Joy	0.01	0.13	*0.57*	0.29	0.01
Mean	Joy	0.03	0.05	*0.59*	0.09	0.25
1	Sadness	0.00	0.02	0.00	*0.52*	0.46
2	Sadness	0.00	0.06	0.00	*0.50*	0.44
6	Sadness	0.00	0.03	0.01	*0.88*	0.08
17	Sadness	0.00	0.01	0.06	*0.41*	0.53
20	Sadness	0.01	0.06	0.01	*0.81*	0.10
21	Sadness	0.00	0.03	0.00	*0.90*	0.07
27	Sadness	0.00	0.01	0.00	*0.92*	0.07
29	Sadness	0.00	0.10	0.00	*0.89*	0.01
Mean	Sadness	0.00	0.04	0.01	*0.73*	0.22
9	Neutral	0.00	0.00	0.02	0.07	*0.90*
11	Neutral	0.01	0.04	0.03	0.28	*0.64*
23	Neutral	0.11	0.03	0.05	0.05	*0.76*
28	Neutral	0.10	0.03	0.02	0.02	*0.83*
Mean	Neutral	0.06	0.03	0.03	0.11	*0.78*

Note: The numbers in italics represent the proportions of accurate judgment

Vocal Subtest Items

The vocal stimuli were selected from a vocal expression corpus that had been created for a large cross-cultural study of vocal emotion recognition (the International Study of Vocal Emotion Expression [ISVEE]; see Scherer et al. 2001). Vocal portrayals of joy/happiness, sadness, fear, anger, and disgust by four professional actors (two male, two female;

regularly employed in radio and television productions) were recorded in a professional broadcasting studio of the West German Radio (Westdeutscher Rundfunk, WDR) in Cologne, Germany. The recording sessions were directed by a professional radio producer and the actors were paid for their services.

The actors were instructed to imagine prototypical emotion scenarios (based on results from cross-cultural experience sampling research; Scherer et al. 1986) and act them out as if they were experiencing them (for further details, see Scherer et al. 1991). To eliminate the potential effects of semantics, actors were asked to produce two "standard utterances" for their portrayals: "Hat sundig pron you ventsy" and "Fee gott laish jonkill gosterr" (chosen from a set of meaningless utterances created by a phonetician who had selected two meaningless syllables from each of six European languages and randomly arranged them into several seven-syllable sequences).

The elicitation of speech samples followed a design of four actors × two scenarios × five emotions × two sentences, resulting in 80 emotional utterances. In addition, each of the four actors uttered the two sentences in a neutral, non-emotional fashion, yielding eight neutral stimuli. Further details and the acoustic analyses of these stimuli, as well as the results of differences in acoustic parameters across emotions, are reported in Scherer et al. (1991).

From the results of a series of judgment studies reported by Scherer et al. (1991), a set of 30 stimuli was selected that satisfied both the criterion of high naturalness and discriminant identification of the target emotion, yielding the following number of stimuli per emotion: anger, 7; fear, 5; joy/happiness, 6; and sadness, 8. We decided not to include disgust portrayals in the ERI, as this emotion is typically recognized with much lower accuracy in vocal recognition studies (see Scherer 1999; Scherer et al. 1991). Instead, four neutral items were included because they represented sufficient difficulty to add to test discrimination.

Sixteen items were encoded by female actors and 14 by male actors. One utterance was represented 18 times and the other 12 times.

As in the case of the facial items the validity of the stimuli has been empirically confirmed. Scherer et al. (1991) performed digital extraction of the major acoustic parameters used in the literature on the vocal expression of emotion and showed that these (shown in Table 4 of the publication) largely correspond to theoretical predictions (shown in Table 3). Furthermore, these patterns correspond to widely replicated acoustic profiles for the different emotions as reviewed by Scherer et al. (2003).

Test Administration

In the initial item analysis studies, the test items were presented via slide and loudspeaker or earphone presentation (see Appendix A of the Electronic Supplementary Material), but the current version of the ERI is computer administered. The facial expression samples were digitized and are presented as jpg files and the voice samples are presented as wav files.

For each of the two subtests, instructions are presented on a series of pages on the computer screen. They provide information on the aims of the test (recognize the emotions expressed as correctly as possible) and on the type of actor portrayals used. For the vocal subtest, the construction of the "sentences" is explained and candidates are asked to focus on the emotion expressed rather than on trying to understand the utterances. Candidates are explicitly told that the five emotions do not necessarily occur with equal frequency in the set of items. Candidates are requested to answer in a fast and spontaneous manner.

In each subtest, first, two examples are presented followed by the 30 items. The participants report their responses on the following screen by clicking on one of the answer alternatives on the screen (without time limitation). For the facial subtest, each photo is shown for 3 s. In the vocal subtest, the duration of the presentation depends on the length of the voice sample. After the participant's response, it is announced that the next stimulus will be presented in 3 s. The data are stored and analyzed automatically (a response corresponding to the target is scored as 1 and otherwise as 0).

Using back-translation procedures, different language versions of the ERI subtests were produced to allow administration in different languages. Currently, computer administration versions are available in English, French, German, Italian, and Spanish.

Item Analysis

The following initial item analysis studies were conducted:

1. GVA-96: 154 students in classroom settings for the facial and vocal subtests in French-speaking Switzerland
2. GVA-99: 79 students in classroom settings for the vocal subtest in French-speaking Switzerland
3. ISVEE: 390 students in eight countries in small group laboratory settings for the vocal subtest (as part of a cross-cultural research project to establish the universality of vocal emotion recognition; see Scherer et al. 2001). These studies are described in detail in Appendix A of the Electronic Supplementary Material.
4. Computer Assessment of Personal Potential_1 (CAPP_1): 848 candidates, professionals; about one third in upper-management positions, one third in middle or lower management positions, and one third in non-management positions; 29.7% female; modal age approximately 40 years; individual administration by human resource consultants as part of a computerized assessment system mainly for career development and, to a lesser degree, for selection purposes. Candidates varied widely in cultural and linguistic background and chose one of the five languages available for the CAPP package (see "Test Administration" subsection).

The results of the item analyses are documented in Appendix A of the Electronic Supplementary Material and described in the Results section.

Norm Establishment

Test norms are based on two types of samples: psychology students and professional assessment candidates. The student sample (Multimodal Emotion Recognition Test [MERT]) consists of 72 psychology students who served in a construct validation study (for details see the "Construct Validation" subsection). For the professional sample (CAPP_2), 3,681 assessment candidates (25% females) were used. The detailed background information for this large norming sample of professionals is shown in Table 2.

Construct Validation

Tests Used

Remarkably few attempts have been made to construct psychometrically satisfactory tests of emotion recognition ability, and very few instruments can be considered reasonably

Table 2 Background characteristics (gender, age, education) of the professional norming sample (CAPP_2)

Education	Gender	Age (years)		
		Under 40	Over 40	Total
Obligatory school level	Male	79	102	181
	Female	21	26	47
	Total	100	128	228
Secondary school level	Male	203	323	526
	Female	107	97	204
	Total	310	420	730
Higher education level	Male	934	1,038	1,972
	Female	388	261	649
	Total	1,322	1,299	2,621

Lack of age and education responses account for the missing data among the 3,681 participants

close to the aims of ERI and can therefore be used for construct validation purposes. From these instruments, only those that have been validated were chosen as criterion tests that focus on assessing the ability to identify the meaning of prototypical expression configurations for modal emotions as portrayed by actors (see the following paragraphs).

The ERI construct validation was performed as part of a study to obtain validation data for a new test of emotion recognition using dynamic stimuli, the MERT. This test requires participants to identify 10 emotions (hot anger, cold anger, panic fear, anxiety, despair, sadness, elation, happiness, disgust, and contempt) portrayed by professional theater actors (30 portrayals) in four different modes: audio/video, video only, audio only, and picture only (a single still extracted from the videos), resulting in 120 test items presented in one of two fixed random orders. A comprehensive description of this test and the details of the construct validation study are provided in an article by Bänziger, Grandjean, and Scherer (2009). It also includes the detailed descriptions of the tests used for construct validation. In consequence, in the present article we provide only rudimentary information concerning these tests:

1. The Diagnostic Analysis of Nonverbal Accuracy (DANVA; Nowicki and Duke 1994) consists of 24 audio recordings of vocal expressions and 24 photographs of facial expressions of four emotion categories (anger, fear, joy/happiness, and sadness) with two intensities (weak or strong). The emotions are portrayed by several amateur posers (the vocal stimuli uses the sentence “I’m going out of the room now, and I’ll be back later”).
2. The Profile of Nonverbal Sensitivity (PONS; Rosenthal et al. 1979) is composed of 20 audio/video recordings in which one sender (a young female research collaborator) portrays 20 attitudes (such as “trying to seduce someone,” “saying a prayer,” “admiring nature,” and “expressing jealous anger”) classified as dominant versus submissive and positive versus negative attitudes. Each recording is shown in 11 different modes (channels): (1) face alone; (2) body (from neck to knees); (3) face and torso (head/face and body down to the waistline, showing hand gestures); (4) low-pass filtered speech alone (no picture); (5) randomized-spliced speech (Scherer 1971) alone (no picture); and (6–11) combinations of the three visual recordings with the two manipulated voice/speech recordings. The 220 portrayals are presented in fixed order

and the test participants are required to select the correct choice of two potential attitudes.

3. The Japanese and Caucasian Facial Expressions of Emotion (JACFEE; Biehl et al. 1997) consists of 56 photos of facial expressions posed by 14 Caucasian males, 14 Caucasian females, 14 Asian males, and 14 Asian females, portraying one of seven basic emotions (surprise, sadness, anger, happiness, fear, disgust, and contempt). Responses are timed, each expressive picture is displayed for 200 ms between two 500-ms exposures of the same face without expression, and the candidate is required to select one of the seven alternative answers.

Participants and Procedure

Seventy-two undergraduate psychology students (87.5% female) participated in the study in exchange for course credit. They were also promised and given personalized feedback on their results for all emotion recognition tests. All the tests were administered on individual computers in a computer laboratory for groups of 10 (minimum) to 20 (maximum) participants.

Computation of Accuracy Scores

Accuracy was computed as the proportion of correct answers given by a participant. Each participant obtained a global score in each test (based on all items in the test). The two ERI subscales have 30 items each. For MERT, the overall accuracy score is computed on 120 items (30 emotion portrayals × four presentation modes). For DANVA, the score is computed on 48 items (24 facial portrayals and 24 vocal portrayals). For PONS, the global accuracy score is computed on 220 items and for JACFEE on 56 items.

Results

Item Analysis

Item Difficulty

The variation of item characteristics is limited because the expression portrayals that serve as items need to be valid communicative messages. One of the criteria for the selection of portrayals for recognition studies is that a sizeable proportion of judges recognize the target emotion because the actor portrayal might otherwise not be valid. Thus, items must be chosen that, on the one hand are not too difficult, that is, sufficiently recognizable, and on the other hand are not too easy, that is, recognized by most participants. In constructing the ERI, our criterion was to keep items that had obtained the highest percentage of accurate responses for the target emotion and for which the nontarget responses were relatively evenly distributed over the nontarget response alternatives. To evaluate the items chosen on this criterion, we calculated confusion matrices, as is good practice in recognition studies (see Banse and Scherer 1996).

The results of the item analysis studies with psychology students in classroom settings (GVA-96, GVA-99; participants had to indicate the perceived intensity for each emotion alternative) are shown in Appendix A of the Electronic Supplementary Material, Table A2.

In almost all cases, the target emotion was indeed rated with the highest intensity. The distribution of intensities for the nontarget emotions varied over items, as some emotion alternatives are semantically closer to each other than to others. The same pattern is found for the confusion matrices listing the accuracy proportions that were computed for the group administrations in the cross-cultural study with students (Appendix A of the Electronic Supplementary Material, Table A3).

The results of the item analyses for professional participants (CAPP_1; see "Method" section) are shown in Appendix A of the Electronic Supplementary Material, Table A4. As in the student samples, the target emotion is virtually always recognized with the highest accuracy and the confusions are distributed according to the similarity between the emotions (see the detailed discussion of empirically found confusion patterns in emotion recognition in Banse and Scherer 1996, and Bänziger et al. 2009). We decided that, given the response distributions shown in Tables A2, A3, and A4, it was not necessary to exclude or replace any item, and consequently all items were retained.

The item difficulty of the ERI also compares very favorably with that of the tests used for construct validation (see Bänziger et al. 2009; Table 1). The respective table, reproduced as Table C1 in the Appendix C of the Electronic Supplementary Material, shows that the proportion of items recognized by more than 80% of the participants for the ERI is 53% compared to 28% for the MERT, 60% for the PONS, 52% for the JACFEE, 62% for the DANVA. In conclusion, the item difficulty of the ERI can be considered as in line with other published instruments in this field.

Internal Consistency

The habitual psychometric reliability analyses for scales consisting of verbal items are not applicable to emotion recognition tests. The reason is that highly reliable verbal scales basically measure the content of only one item via various formulations and thus cannot represent more than an extremely narrow facet of a response domain. This issue is closely related to the problem of "bloated specific" described early on by Cattell (1978). Boyle (1991) has convincingly shown that the term "internal consistency," as measured by Cronbach's alpha, is a misnomer because a high estimate of internal item consistency or homogeneity may also reflect a high level of item redundancy, wherein essentially the same item is rephrased in several different ways. Variation of a theme, as is possible in constructing verbal or mathematical test items, is not feasible in expression recognition tests. This is partly because (a) different encoders need to be used to generate the stimuli, (b) actors cannot systematically vary specific expression features without affecting recognizability, and (c) as amply shown in studies of emotional expression, many emotions can be expressed in widely different ways (see Scherer and Bänziger 2010) and yet will be well recognized (even though there are strong individual differences in recognizing specific emotions and the expressions of specific encoders; see Scherer and Ellgring 2007). Hall (2001) has suggested that the standard psychometric model, which assumes that random error causes low internal consistency, may not be applicable to nonverbal sensitivity tests. Rather, she argues, it may be that such tests actually gain in validity by including items that represent a number of different, albeit related, skills or different channels (see also Bollen and Lennox 1991).

Thus, standard item consistency analysis is not feasible. The use of test–retest reliability is also problematic. A pilot study of test–retest reliability for the two ERI subtests showed that prior exposure to a whole set of expression portrayals generates strong learning effects because, after the first presentation of the test items, participants know the complete set of stimuli and response alternatives and can easily use discrimination and exclusion strategies.

As a consequence, scale scores in recognition tests used to examine individual differences in recognition ability tend to be organized around the cumulative model described by Bollen and Lenox (1991): testing how many of possibly different portrayals of an emotion the testee can recognize.

Measures of Central Tendency and Variation: Norms

The accuracy scores for the emotions tested by the facial and vocal subtests of the ERI, as well as the total scores for each subtest, are listed in Table 3 for both the professional and student groups (see also Fig. 1). Given the extremely large size of the professional norming

Table 3 Measures of central tendency and variation for two types of populations

	Facial subtest			Vocal subtest	
	Professionals ($N = 3{,}505$)	Students ($N = 72$)		Professionals ($N = 3{,}469$)	Students ($N = 68$)
Anger			*Anger*		
Mean	0.54	0.51	Mean	0.72	0.69
SD	0.18	0.14	*SD*	0.18	0.16
Max	1.00	0.86	Max	1.00	1.00
Min	0.00	0.29	Min	0.00	0.43
Fear			*Fear*		
Mean	0.68	0.72	Mean	0.65	0.61
SD	0.15	0.13	*SD*	0.17	0.17
Max	1.00	1.00	Max	1.00	0.80
Min	0.13	0.50	Min	0.00	0.20
Joy			*Joy*		
Mean	0.98	0.99	Mean	0.58	0.53
SD	0.08	0.05	*SD*	0.23	0.22
Max	1.00	1.00	Max	1.00	1.00
Min	0.00	0.75	Min	0.00	0.00
Sadness			*Sadness*		
Mean	0.77	0.81	Mean	0.74	0.75
SD	0.17	0.16	*SD*	0.15	0.14
Max	1.00	1.00	Max	1.00	1.00
Min	0.00	0.29	Min	0.00	0.38
Disgust			*Neutral*		
Mean	0.87	0.90	Mean	0.78	0.76
SD	0.19	0.16	*SD*	0.22	0.22
Max	1.00	1.00	Max	1.00	1.00
Min	0.00	0.50	Min	0.00	0.25
Total			*Total*		
Mean	0.73	0.75	Mean	0.69	0.67
SD	0.08	0.06	*SD*	0.11	0.09
Max	0.97	0.90	Max	0.93	0.83
Min	0.27	0.63	Min	0.10	0.47

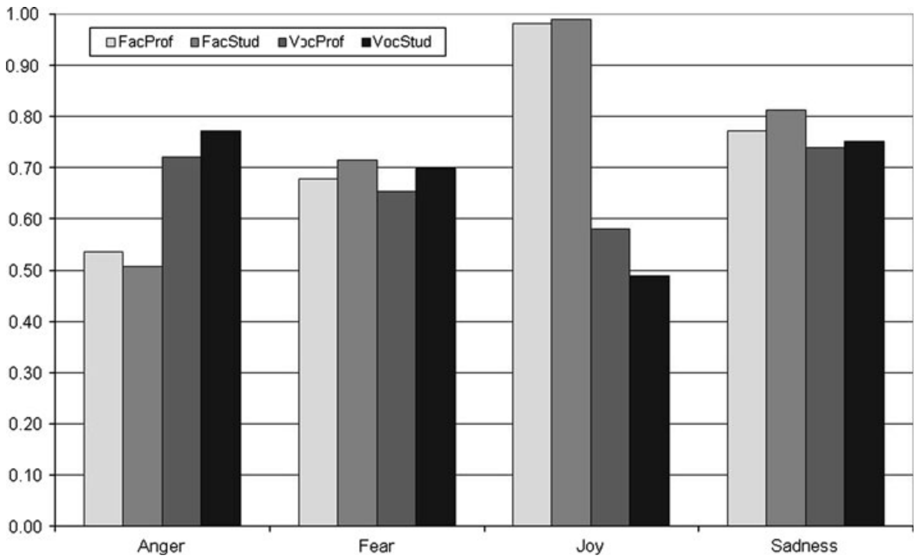

Fig. 1 Proportions of correct responses across four emotions. *Note*. FacProf—Facial test, Professionals, FacStud—Facial test, Students, VocProf—Vocal test, Professionals, VocStud—Vocal test, Students

group, the means reported can be considered as very stable. Testifying to the stability of the means and the distribution of the scores is the fact that the student group, although much smaller than the professional group, shows virtually the same means. Given this high degree of stability, we suggest using the statistics for the professional groups as test norms for the ERI, as it seems to make little sense to propose different norms for professionals and students.

Table 4 shows the complete list of measures of central tendency and variation, as well as the percentiles for the ERI total score, which is computed as the average of the facial and vocal subscores. Figure 2 shows a boxplot of the distribution of the accuracy scores. The results show that most individuals have no major problems in recognizing prototypical expressions in the face and the voice, a central requisite for smooth social interaction, but that substantial variation nevertheless results from individual differences in this central component of EC, demonstrating that the overall difficulty of the ERI is acceptable for psychometric testing.

Correlations Between Accuracy Scores for Individual Emotions and Total Facial and Vocal Scores

Table 5 shows the intercorrelations of the accuracy scores for the different emotions and the total vocal and facial accuracy scores (showing the contribution to the total scores). These results confirm the assumption outlined in the Internal Consistency subsection: The accuracy scores for the different emotions are not strongly correlated because the emotions are not interchangeable items measuring the same content. In consequence, normal procedures to test the subscale components of an overarching scale, using principal components analysis, are not applicable in the present case. This is because the emotion portrayals studied are discrete entities that are similar on some dimensions, but vary on many others. Research shows that there are strong individual differences for sender encoding of the

Table 4 Measures of central tendency, variation, and distribution for the ERI total score (in %)

	Statistic	*SE*
a) Central tendency		
Mean	71.44	0.12
95% Confidence interval for mean		
Lower bound	71.21	
Upper bound	71.68	
5% Trimmed mean	71.71	
Median	71.67	
Variance	50.54	
SD	7.11	
Minimum	20.00	
Maximum	90.00	
Range	70.00	
Interquartile range	10.00	
Skewness	−0.76	0.04
Kurtosis	1.91	0.08

	Percentile						
b) Distribution							
Scores	5.00	10.00	25.00	50.00	75.00	90.00	95.00
	60.00	61.67	66.67	71.67	76.67	80.00	81.67

N = 3,505. *ERI* Emotion Recognition Index

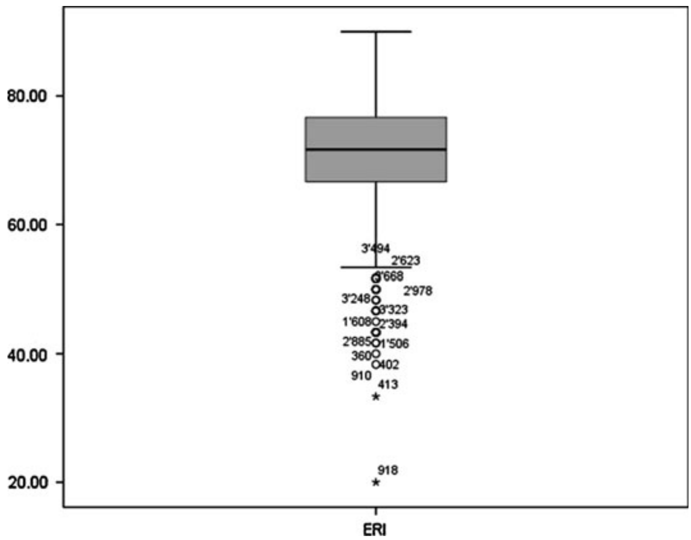

Fig. 2 Exploratory statistics *boxplot* for ERI total scores

Table 5 Correlations between emotion subscales for the facial and vocal subtests

	Facial anger	Facial fear	Facial joy	Facial sadness	Facial disgust	Facial total	Vocal anger	Vocal fear	Vocal joy	Vocal sadness	Vocal neutral	Vocal total
Facial anger	1.000	−.026	.057**	.056**	.006	.531**	.150**	.051**	.122**	.068**	.025	.156**
Facial fear	−.026	1.000	.048**	.048**	.220**	.586**	−.050**	−.013	−.062**	.004	.012	−.044**
Facial joy	.057**	.048**	1.000	.037*	.073**	.216**	.051**	.017	.029	.028	.000	.047**
Facial sadness	.056**	.048**	.037*	1.000	.062**	.557**	.044**	−.011	−.018	.075**	.031	.044**
Facial disgust	.006	.220**	.073**	.062**	1.000	.459**	.008	.029	−.054**	.070**	.026	.022
Facial total	.531**	.586**	.216**	.557**	.459**	1.000	.081**	.025	.009	.097**	.041*	.090**
Vocal anger	.150**	−.050**	.051**	.044**	.008	.081**	1.000	.275**	.299**	.298**	.008	.706**
Vocal fear	.051**	−.013	.017	−.011	.029	.025	.275**	1.000	.167**	.254**	.012	.535**
Vocal joy	.122**	−.062**	.029	−.018	−.054**	.009	.299**	.167**	1.000	.149**	.030	.641**
Vocal sadness	.068**	.004	.028	.075**	.070**	.097**	.298**	.254**	.149**	1.000	.008	.625**
Vocal neutral	.025	.012	.000	.031	.026	.041*	.008	.012	.030	.008	1.000	.295**
Vocal total	.156**	−.044**	.047**	.044**	.022	.090**	.706**	.535**	.641**	.625**	.295**	1.000

* $p < .05$, ** $p < .01$

same emotion and the ability to recognize the respective emotion in different modalities of expression (Banse and Scherer 1996; Bänziger et al. 2009; Scherer and Ellgring 2007). We advise against using "emotion subscales" for diagnostic purposes because the number of items ranges generally from four to seven, which seems too small a basis for a reliable diagnostic judgment. The statistics shown in Table 3 should be used for information only, for example, for the relative difficulty of recognizing certain emotions in the facial and vocal modalities, but not as norms. We strongly suggest using only the ERI total score and the facial and vocal subscores for diagnostic purposes.

Construct Validity

Table 6 shows the correlations of the facial and vocal subtest total scores, with subscores for the recognition of expression in still photos or of dynamic vocal expressions in the construct validation tests. The ERI facial subscore correlates significantly with the respective MERT and JACFEE scores, but not with the DANVA scores. The ERI vocal subscore correlates significantly with the respective MERT score but not with the DANVA and PONS scores. In interpreting this pattern, note that generally the DANVA had rather low correlations with all other tests in the construct validation package (see Bänziger et al. 2009) and that the PONS measures social attitudes rather than emotions and uses filtered voices. Thus, the MERT vocal score is the most directly comparable to the ERI vocal score. In fact, if one uses the total MERT emotion family recognition score (MERT_FR, which does not distinguish between two members of an emotion family, e.g., hot anger vs. cold anger, and thus is the more appropriate comparison because the ERI uses only the family label), the correlation with the ERI total score is $r = .383$ ($p < .001$, one-tailed; see Bänziger et al. 2009, Table 6), indicating rather satisfactory construct validity. On the whole, these results support the claim that the ERI measures the ability of emotion recognition as a component of EC.

Relationship to Biographical Variables, Cognitive Capacity, and Personality

We first examined the extent to which there are differences in the total ERI score related to gender, age, and education level of the professionals tested. According to earlier reviews of

Table 6 Correlations of Scores Based on (a) Still Pictures of Facial Portrayals or (b) Vocal Portrayals

	MERT photo	DANVA facial	ERI facial
(a)			
DANVA facial	.224* ($N = 70$)		
ERI facial	.296** ($N = 70$)	.146 ($N = 72$)	
JACFEE	.325** ($N = 67$)	.278* ($N = 67$)	.300** ($N = 67$)
(b)			
DANVA vocal	.323** ($N = 70$)		
ERI vocal	.274* ($N = 66$)	.013 ($N = 68$)	
PONS vocal	.252* ($N = 68$)	.235 ($N = 70$)	.032 ($N = 68$)

N varies between 65 and 72 (see "Methods"); *MERT* Multimodal Emotion Recognition Test, *DANVA* Diagnostic Analysis of Nonverbal Accuracy, *ERI* Emotion Recognition Index, *JACFEE* Japanese and Caucasian Facial Expressions of Emotion, *PONS* Profile of Nonverbal Sensitivity. Reproduced from Bänziger et al. 2009

* $p < .05$ (one-tailed), ** $p < .001$ (one-tailed)

the literature (see Hall 1978) one would expect gender differences, specifically a superior score for women (possibly because women might be more motivated to succeed in interpersonal interactions and thus are more interested in smooth social relationships). The results of a univariate analysis of variance (ANOVA) with these three factors to test for main and interaction effects is shown in Table 7. As expected, there is a significant main effect for gender, even though the effect size is rather small (partial $\eta^2 = .009$). A comparison of the means shows that the difference is about 3 percentage points (see also Fig. 3). Neither age nor education level show significant main effects. Age effects have been described in the literature (Mill et al. 2009). However, less than 1% of the professionals making up the ERI norming sample are 60 years of age or older, whereas the age groups for which reduced competence is reported in the literature are much older. Thus, if there is a decrease in emotion recognition ability with age, the statistically noticeable effect is likely to be situated beyond the 60-year limit.

Education effects had not been expected. Interestingly, as shown in Fig. 3, there is a marginally significant ($p = .08$) interaction between gender and education level, suggesting that gender differences might be somewhat smaller at the university education level. This effect might be linked to a potential relationship between cognitive capacity or intelligence and EC. To test this hypothesis, we combined the scores of four subtests of cognitive intelligence used as part of the CAPP package—ability to perform deductions, to understand implications, and to distinguish between facts and opinions, and short-term memory performance—to an index of cognitive capacity. This index correlates, $r = .132$ ($p < .01$), with the ERI total score, suggesting that there might indeed be a potential effect of a “g factor.” However, the effect size is small and whether it is really strong enough to account for a smaller gender effect at the level of university education needs to be examined.

We then examined to what extent the ERI total score correlates with personality variables. All professional candidates who took the ERI as part of the CAPP package also responded to instruments designed to measure personality, coping strategies, stress resistance, emotional expressiveness, emotion regulation, and self-image. A principal

Table 7 Analysis of variance of the eri total score for gender, age groups, and education level

Source	Type III sum of squares	*df*	Tests of between-subjects effects			
			Mean square	*F*	Sig.	Partial η^2
Corrected model	5320.21	23	231.31	5.003	0	0.033
Intercept	3768529.38	1	3768529.38	81511.56	0	0.96
Sex	1399.98	1	1399.98	30.281	0	0.009
Age	184.63	3	61.54	1.33	0.262	0.001
Edu	144.93	2	72.46	1.57	0.21	0.001
Sex × Age	123.77	3	41.26	0.89	0.44	0.001
Sex × Edu	232.44	2	116.22	2.51	0.08	0.001
Age1 × Edu	256.19	6	42.70	0.92	0.48	0.002
Sex × Age × Edu	403.79	6	67.30	1.46	0.19	0.003
Error	156267.77	3380	46.23			
Total	1.76E + 07	3404				
Corrected total	1.62E + 05	3403				

ERI Emotion recognition index, *Sig.* significance level, *Edu* education

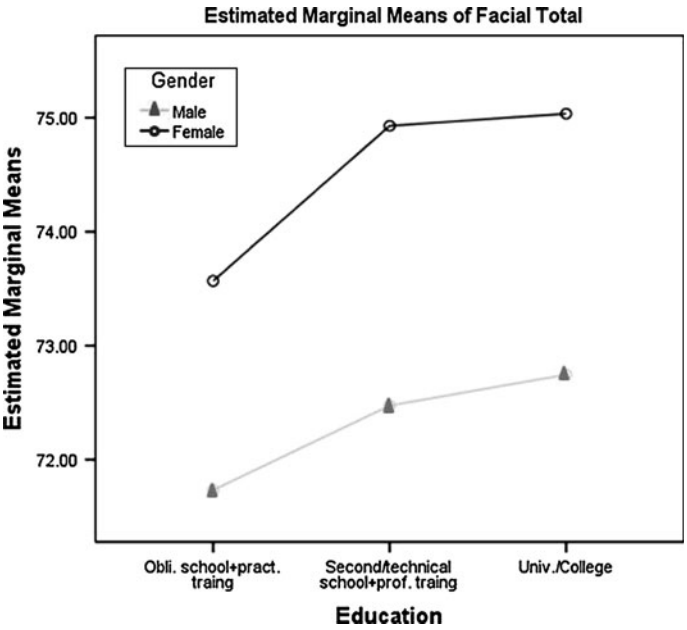

Fig. 3 Interaction effect result for sex × education (ERI total score)

components analysis of the traits measured by these instruments yields (according to Scree test criteria) five major factors (see Appendix B in Supplement material, Table B1): (a) Emotional Stability versus Lability/Neuroticism (self-assurance, emotional stability, emotion regulation, stress resilience, stress resistance, need for self-assertion, problem solving vs. emotional lability, apprehensiveness, anxiety, pessimism, irritability); (b) Dominance versus Altruism (aggressiveness, dominance, need for power, autonomy, excitability vs. warmth, social interaction skills, altruism); (c) Extraversion (extraversion, need for affiliation, emotional expressiveness, impulsivity); (d) Repressive Coping versus Seeking Social Support (emotion repression, problem repression, self-concept bolstering, repression confirmation, external attribution vs. seeking empathy, seeking social support, internal attribution); and (e) Efficiency versus Creativity (conscientiousness, methodical approach, managerial efficacy, need for achievement vs. creativity). Table 8 shows the correlations of the ERI total score and its two subtest scores with the factor scores of the

Table 8 Correlations with personality variables

	ERI total	Facial total	Vocal total
Emotional stability-neuroticism	.048**	.01	.057**
Dominance-altruism	.009	−.033	.038*
Extraversion	.166**	.064**	.171**
Repression-need for social support	−.160**	−.061**	−.166**
Efficiency-creativity	.001	−.013	.012

$N = 3{,}313$. *ERI* Emotion Recognition Index

* $p < 0.05$, ** $p < 0.01$, two-tailed

respective personality factors. As one might have expected, extraverted individuals and those who, in coping with major life events, privilege internal attribution of responsibility and search for empathy and social support are more competent in recognizing the emotional state of interaction partners than are individuals characterized by introversion or affect repression. This disposition seems explicable in that individuals who seek contact with others and rely on their support may also be more skilled in recognizing the emotional state of their interaction partners (see also Matsumoto et al. 2000). Furthermore, a somewhat smaller effect suggests that emotionally stable, self-assured individuals might be more competent in recognizing emotions than are anxious, neurotic, and emotionally labile persons. Anxiety and depression possibly lead an individual to withdraw and pay less attention to the emotional signals of others.

Predictive Validity

At this point, no hard data on the prediction of important behavioral differences or achievements from ERI scores are available. However, we can obtain indirect evidence by examining how much variance the ERI score explains for the type of position the respective individual occupies within a company and the department in which the person works. The assumption is that greater emotion recognition skills further advancement on the career ladder and that individuals who are better at recognizing emotions in others work mostly in contexts in which individuals need to engage in frequent interactions with other people. As to position in the company, there is indeed a significant effect of ERI in an ANOVA ($F = 5.208$, $p < .001$, partial $\eta^2 = 0.006$), but it is not the case that high competence in this domain leads to more rapid advances in the levels of management (means: upper management 71.35, middle management 71.38, sales/administration position 72.06, technical position 69.87, other 72.53). However, individuals in technical positions, who work less in social contexts, seem to have somewhat lower recognition ability than do those in sales/administration positions, who often interact with other people. As to the department of a company in which people work, there is also a significant effect ($F = 7.753$, $p < .01$, partial $\eta^2 = 0.011$), with the following means per group: marketing/sales 71.82, research/development 70.68, production/logistics/technical services 70.47, finances/bookkeeping 71.0, administration/personnel/legal services 72.67, other 71.80. Again, employees working in administration and personnel services seem to have a higher competence in emotion recognition. The current results do not allow determination of whether these effects are due to selective recruitment of individuals with this EC to positions requiring frequent social interactions or whether frequent interaction experiences have a learning effect, boosting recognition skill.

Discussion and Conclusion

This article describes the development and validation of a computer-administered test measuring the ability to recognize prototypical facial and vocal expressions of emotion as portrayed by actors. The test is straightforward in that candidates immediately understand what is at stake and generally perform well, as is expected given the importance of this skill for social interaction and maintaining successful social relationships. Yet it was possible to select items from leading research efforts in the area of motor expression of emotion that present a sufficient degree of difficulty to yield a distribution of test scores between approximately 50 and 90% accuracy, with the region between the 25th and 75th percentile situated between 67 and 78% accuracy. Thus, this test is capable of diagnosing sizeable individual variation in the central aspect of EC or EI. In addition, test administration is rapid and efficient, rarely exceeding 15 min. Of particular importance is the high stability of the test norms. The ERI test is administered online (in a Flash version) and is available free of charge for non-commercial academic research (see details under http://www.affective-sciences.org/eri).

A study of construct validity, using major published tests that are currently available in this area, showed significant correlations of the ERI and its subscales with the most pertinent tests or subtests of published instruments such as the JACFEE or MERT. A limitation of such instruments is that all of the current tests use actor portrayals as test items; the development of tests is desirable in which items that reflect more authentic expressions are included (although it is exceedingly difficult, for ethical and practical reasons, to generate systematic test material on the basis of recorded emotion expressions in everyday contexts).

It should also be noted that the ERI has some limitations. Thus, the black and white photographs of facial expression have been produced over 30 years ago and thus the posers' hairstyles look a bit dated. However, this has not caused any problems in test administration so far. Also, the small number of items, designed for a brief testing period, do not allow to compute subscores for the recognition of different emotions. The small test size is also the reason for the lack of complete balance in poser identity. In addition, the comparison of the facial and vocal subscores is contaminated by the fact that the posers are different in the two conditions and the difference in the static versus dynamic nature of the respective stimuli. The limited number of emotions also reduces the possibility of studying frequent confusions. For example, the absence of surprise (both with respect to stimulus type and answer alternative), which is frequently confused with fear due to an overlap of action units and possibly also voice cues, might lead to an overestimation of accuracy for some emotions. However, a larger number of emotions and answer categories would have negatively affected the ease and rapidity of administration which was a major aim in constructing the instrument.

Furthermore, the ERI measures only one aspect of emotion recognition—the ability to infer emotions that are expressed via prototypical cue configurations in the face and the voice, as they are used for meaningful social communication. Thus, the ERI does not measure the ability of an individual to see through strategic or deceitful use of expressive cues to mask underlying emotions the sender feels but tries to conceal. However, it can be argued that the ability to perceive and analyze the prototypical cue configurations and to correctly infer their emotional meaning is a major condition for the ability to decode masked expressions.

The correlations with biographical and personality background variables reported here support some central hypotheses in the field and produce several interesting new hypotheses. We did indeed find the predicted advantage for women in accuracy of emotion recognition, with a difference of about 3 percentage points. Although the effect size is small, it does seem to be stable across studies (Hall 1978). Our data also speak to the presumed link between cognitive intelligence and EI (Mayer et al. 2003), showing a significant but modest correlation of the ERI score with a summary measure of facets of cognitive intelligence. Finally, we reported some intriguing and rather strong relationships between the ERI score and a number of personality factors such as Extraversion, Need for Social Support, and Emotional Stability. Although the correlation coefficients (and the resulting effect sizes) are small, note that the *N* for our professional sample is extremely large, which normally leads to very small correlation coefficients.

In conclusion, the ERI is a very efficient rapid-screening instrument for emotion recognition ability, a central component of socio-emotional competence, now available for web administration in research or human assessment settings. As the results reported here show, its nomological network with other personality constructs yields a number of promising avenues for future research and first indications point to potential predictive validity for objective performance outcomes.

Acknowledgments The authors gratefully acknowledge important contributions and comments by Paul Ekman, Katja Schlegel, and Olivier Rosset (who programmed the Flash version as part of the Swiss Center of Affectice Science's web experimentation program).

References

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology, 70*, 614–636.

Bänziger, T., Grandjean, D., & Scherer, K. R. (2009). Emotion recognition from expressions in face, voice, and body: The multimodal emotion recognition test (MERT). *Emotion, 9*, 691–704.

Biehl, M., Matsumoto, D., Ekman, P., Hearn, V., Heider, K., Kudoh, T., et al. (1997). Matsumoto and Ekman's Japanese and Caucasian Facial Expressions of Emotion (JACFEE): Reliability data and cross-national differences. *Journal of Nonverbal Behavior, 21*, 3–21.

Bollen, K. A., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin, 110*, 305–314.

Boyle, G. J. (1991). Does item homogeneity indicate internal consistency or item redundancy in psychometric scales? *Personality and Individual Differences, 12*, 291–294.

Cattell, R. B. (1978). *Scientific use of factor analysis in behavioral and life sciences*. New York, NY: Plenum Press.

Ekman, P. (1972). Universals and cultural differences in facial expression of emotion. In J. R. Cole (Ed.), *Nebraska symposium on motivation* (Vol. 19, pp. 207–283). Lincoln: University of Nebraska Press.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion, 6*, 169–200.

Ekman, P. (2007). *Emotions revealed* (2nd ed.). New York: Owl Books.

Ekman, P., & Friesen, W. V. (1976). *Pictures of facial affect*. Palo Alto, CA: Consulting Psychologists Press.

Ekman, P., Friesen, W. V., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., et al. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology, 53*, 712–717.

Ekman, P., & Rosenberg, E. L. (2005). *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)* (2nd ed.). New York, NY: Oxford University Press.

Hall, J. A. (1978). Gender effects in decoding nonverbal cues. *Psychological Bulletin, 85*, 845–857.

Hall, J. A. (2001). The PONS test and the psychometric approach to measuring interpersonal sensitivity. In J. A. Hall & F. J. Bernieri (Eds.), *Interpersonal sensitivity: Theory and measurement* (pp. 143–160). Mahwah, NJ: Erlbaum.

Hall, J. A., & Bernieri, F. J. (2001). *Interpersonal sensitivity: Theory and measurement*. Mahwah, NJ: Erlbaum.

Izard, C. E. (1977). *Human emotions*. New York, NY: Plenum Press.

Joseph, D. L., & Newman, D. A. (2010). Emotional intelligence: An integrative meta-analysis and cascading model. *Journal of Applied Psychology, 95*, 54–78.

Matsumoto, D., LeRoux, J., Wilson-Cohn, C., Raroque, J., Kooken, K., Ekman, P., et al. (2000). A new test to measure emotion recognition ability: Matsumoto and Ekman's Japanese and Caucasian Brief Affect Recognition Test (JACBART). *Journal of Nonverbal Behavior, 24*, 179–209.

Matthews, G., Zeidner, M., & Roberts, R. D. (Eds.). (2007). *The science of emotional intelligence: Knowns and unknowns*. New York, NY: Oxford University Press.

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

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

Mill, A., Allik, J., Realo, A., & Valk, R. (2009). Age-related changes in the ability to recognize sadness and anger: A cross-sectional study. *Emotion, 9*, 619–630.

Nowicki, S., & Duke, M. P. (1994). Individual differences in the nonverbal communication of affect: The diagnostic analysis of nonverbal accuracy. *Journal of Nonverbal Behavior, 18*, 9–35.

Rosenthal, R., Hall, J. A., DiMatteo, M. R., Rogers, P. L., & Archer, D. (1979). *Sensitivity to nonverbal communication: The PONS test*. Baltimore, MD: John Hopkins University Press.

Russell, T. A., Chu, E., & Phillips, M. (2006). A pilot study to investigate the effectiveness of emotion recognition remediation in schizophrenia using the micro-expression training tool. *British Journal of Clinical Psychology, 45*, 579–583.

Scherer, K. R. (1971). Randomized splicing: A simple technique for masking speech content. *Journal of Experimental Research in Personality, 5*, 155–159.

Scherer, K. R. (1984). On the nature and function of emotion: A component process approach. In K. R. Scherer & P. Ekman (Eds.), *Approaches to emotion* (pp. 293–317). Hillsdale, NJ: Erlbaum.

Scherer, K. R. (1999). Universality of emotional expression. In D. Levinson, J. Ponzetti, & P. Jorgenson (Eds.), *Encyclopedia of human emotions* (Vol. 2, pp. 669–674). New York, NY: Macmillan.

Scherer, K. R. (2001). Appraisal considered as a process of multi-level sequential checking. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 92–120). New York and Oxford: Oxford University Press.

Scherer, K. R. (2007). Component models of emotion can inform the quest for emotional competence. In G. Matthews, M. Zeidner, & R. D. Roberts (Eds.), *The science of emotional intelligence: Knowns and unknowns* (pp. 101–126). New York, NY: Oxford University Press.

Scherer, K. R. (2009). The dynamic architecture of emotion: Evidence for the component process model. *Cognition and Emotion, 23*, 1307–1351.

Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology, 32*, 76–92.

Scherer, K. R., Banse, R., Wallbott, H. G., & Goldbeck, T. (1991). Vocal cues in emotion encoding and decoding. *Motivation and Emotion, 15*, 123–148.

Scherer, K. R., & Bänziger, T. (2010). On the use of actor portrayals in research on emotional expression. In K. R. Scherer, T. Bänziger, & E. B. Roesch, (Eds.), *Blueprint for affective computing: A sourcebook* (pp. 166–178). Oxford: Oxford University Press.

Scherer, K. R., & Ellgring, H. (2007). Are facial expressions of emotion produced by categorical affect programs or dynamically driven by appraisal? *Emotion, 7*, 113–130.

Scherer, K. R., Johnstone, T., & Klasmeyer, G. (2003). Vocal expression of emotion. In R. J. Davidson, K. R. Scherer, & H. Goldsmith (Eds.), *Handbook of the affective sciences* (pp. 433–456). New York, NY: Oxford University Press.

Scherer, K. R., Wallbott, H. G., & Summerfield, A. B. (Eds.). (1986). *Experiencing emotion: A cross-cultural study*. Cambridge, England: Cambridge University Press.

Schulze, R., & Roberts, R. D. (Eds.). (2005). *Emotional intelligence: An international handbook*. Cambridge, MA: Hogrefe & Huber.