Investigating individual differences in emotion recognition ability using the ERAM test

Laukka, Petri; Banziger Flykt, Tanja; Israelsson, Alexandra; Cortes, Diana Sanchez; Tornberg, Christina; Scherer, Klaus R.; Fischer, Håkan

How to cite

LAUKKA, Petri et al. Investigating individual differences in emotion recognition ability using the ERAM test. In: Acta psychologica, 2021, vol. 220, p. 103422. doi: 10.1016/j.actpsy.2021.103422

This publication URL: https://archive-ouverte.unige.ch/unige:166136

Publication DOI: 10.1016/j.actpsy.2021.103422

Acta Psychologica 220 (2021) 103422

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy

Investigating individual differences in emotion recognition ability using the ERAM test

Petri Laukka [a,*], Tanja Bänziger [b,1], Alexandra Israelsson [a], Diana Sanchez Cortes [a], Christina Tornberg [a], Klaus R. Scherer [c,d], Håkan Fischer [a]

[a] Department of Psychology, Stockholm University, Stockholm, Sweden
[b] Department of Psychology, Mid Sweden University, Östersund, Sweden
[c] Department of Psychology, University of Geneva, Geneva, Switzerland
[d] Department of Psychology, Ludwig-Maximilians-University of Munich, Munich, Germany

ARTICLE INFO

Keywords:
Emotion recognition test
Emotion understanding
Empathy
Meta-cognitive judgments
Multimodal expressions
Personality
Sex differences

ABSTRACT

Individuals vary in emotion recognition ability (ERA), but the causes and correlates of this variability are not well understood. Previous studies have largely focused on unimodal facial or vocal expressions and a small number of emotion categories, which may not reflect how emotions are expressed in everyday interactions. We investigated individual differences in ERA using a brief test containing dynamic multimodal (facial and vocal) expressions of 5 positive and 7 negative emotions (the ERAM test). Study 1 ($N = 593$) showed that ERA was positively correlated with emotional understanding, empathy, and openness, and negatively correlated with alexithymia. Women also had higher ERA than men. Study 2 was conducted online and replicated the recognition rates from Study 1 (which was conducted in lab) in a different sample ($N = 106$). Study 2 also showed that participants who had higher ERA were more accurate in their meta-cognitive judgments about their own accuracy. Recognition rates for visual, auditory, and audio-visual expressions were substantially correlated in both studies. Results provide further clues about the underlying structure of ERA and its links to broader affective processes. The ERAM test can be used for both lab and online research, and is freely available for academic research.

1. Introduction

Emotion recognition ability entails the skills needed to accurately recognize the emotional expressions conveyed by others' nonverbal facial, vocal, and bodily cues. Many studies have noted that individuals vary widely in emotion recognition ability (e.g., Elfenbein & Ambady, 2002), but the causes and correlates of these individual differences are not well understood (e.g., Hodges & Wise, 2016). In this article, we therefore investigate the associations between emotion recognition ability and broader affective processes across two studies. In doing so, we also introduce the Emotion Recognition Assessment in Multiple modalities (ERAM) test.

A majority of previous emotion recognition tests have investigated either facial or vocal expressions in isolation, often using still photographs of facial expressions, and have also focused on a limited number of emotion categories where happiness is often the only positive emotion (e.g., Gur et al., 2002; Matsumoto et al., 2000; Mayer et al., 2003; Nowicki & Duke, 1994; Palermo et al., 2013; Scherer & Scherer, 2011; Wilhelm, Hildebrandt, Manske, Schacht, & Sommer, 2014; Young et al., 2002; for reviews see Bänziger, 2016; Olderbak et al., 2021). However, in everyday life, affective expressions are not restricted to one modality only – rather, dynamic facial, vocal, and bodily cues together constitute the most effective means of communication (Bänziger et al., 2012; de Gelder, 2000). Recent studies also suggest that it is possible to communicate nonverbally a wide range of both positive and negative emotions (e.g., Cordaro et al., 2018; Cowen et al., 2019). The ERAM test was designed to provide a more ecologically valid measure of emotion recognition ability and thus contains dynamic multimodal expressions of 12 different emotions, including 5 positive and 7 negative states.

Although most previous emotion recognition tests are unimodal and include few emotion categories, there are also some exceptions. We especially note the work by Katja Schlegel and colleagues (e.g., Schlegel

* Corresponding author at: Stockholm University, Department of Psychology, 106 91 Stockholm, Sweden.
E-mail address: petri.laukka@psychology.su.se (P. Laukka).
[1] Deceased.

https://doi.org/10.1016/j.actpsy.2021.103422
Received 5 May 2021; Received in revised form 22 September 2021; Accepted 23 September 2021
Available online 27 September 2021
0001-6918/

et al., 2014; Schlegel & Scherer, 2016) in developing and validating the Geneva Emotion Recognition Test (GERT) which is a multimodal emotion recognition test that also contains a wide range of emotions. In contrast to the GERT, which only contains multimodal stimuli, the ERAM test presents dynamic stimuli in 3 different conditions: video only, audio only, and multimodal (audio-visual) expressions. This allows us to investigate whether recognition of emotions from specific modalities are separate abilities or if individuals who perform well on one modality also perform well on other modalities. Previous studies have provided mixed results and some studies report low correlations between recognition measures for different modalities (e.g., Scherer & Scherer, 2011), whereas others instead report substantial correlations between modalities (e.g., Bänziger et al., 2009). Here, we investigate the underlying structure of emotion recognition ability using correlation analyses (Studies 1 and 2) and confirmatory factor analysis (Study 1).

The ERAM test further shows similarities with another previous test, the Multimodal Emotion Recognition Test (MERT) developed by Tanja Bänziger and colleagues (Bänziger et al., 2009), which also presents stimuli in video only, audio only and multimodal conditions. However, compared to the MERT, the ERAM test is much shorter, which makes it more resistant to fatigue and thus suitable for inclusion in large test batteries.

The ERAM test is here used in two studies which investigate how emotion recognition ability correlates with various individual difference variables related to emotional processes. The question of how different emotional abilities correlate with each other is important for research on emotional intelligence (EI) (e.g., Elfenbein & MacCann, 2017; Roberts et al., 2010). Emotion recognition ability is considered to be an important part of EI, together with other abilities such as emotion understanding and aspects of emotion regulation. Evidence for positive correlations between different parts of EI is a key criterion for viewing EI as an intelligence, and not as a broad array of personality characteristics (see Elfenbein & MacCann, 2017). Meta-analyses indeed suggest that emotion recognition is positively correlated with other aspects of EI such as emotion understanding and emotion regulation (Elfenbein & MacCann, 2017; Joseph & Newman, 2010). Associations between emotion recognition and broader affective processes have also been investigated in research on the construct of interpersonal sensitivity (Rosenthal et al., 1979), in which emotion recognition ability plays an important part. Research in this tradition suggests that interpersonal sensitivity is positively correlated with positively valenced individual difference measures (e.g., openness and self-reported empathy) and negatively correlated with negatively valenced measures (e.g., neuroticism) (for a meta-analysis, see Hall et al., 2009).

In Study 1, we investigate how ERAM scores correlate with a wide range of individual difference measures – including sex and self-reports of Big-5 personality traits (McCrae & John, 1992), emotional competencies, and socio-emotional dysfunction – in a large sample tested in lab. Women consistently perform better than men in emotion recognition tasks, although the effect sizes are usually small (Hall et al., 2016; Thompson & Voyer, 2014), and we therefore hypothesized that women would perform better also on the ERAM test. Because women may also score higher than men on individual difference measures related to emotion, we exercised control for possible confounding by sex when investigating associations between emotion recognition and other variables (Hall et al., 2009). Regarding the other individual differences variables, we did not state specific hypotheses for each measure. However, in general we expected ERAM scores to be positively correlated with measures that tap on emotional competence, and negatively correlated with measures of emotional dysfunction (Elfenbein & MacCann, 2017; Hall et al., 2009). We did not have specific expectations about correlations between ERAM and Big-5 traits, because the previous literature has not shown consistent trends (e.g., Furnes, Berg, Mitchell, & Paulmann, 2019).

A wide range of emotional competence measures were included in Study 1 – including affective style, emotional expressivity, emotion regulation, emotional understanding, and empathy – in order to give a nuanced view of the connections between emotion recognition ability and broader affective processes. Measures of emotional dysfunction and maladaptive traits included measures of alexithymia, anxiety, depression, and traits related to autism and psychopathy. Emotion recognition difficulties are important symptoms of several psychiatric disorders (Phillips et al., 2003), and many of the above traits have previously been associated with lower emotion recognition ability in clinical populations, including, for example, alexithymia (e.g., Grynberg et al., 2012), autism (e.g., Uljarevic & Hamilton, 2013), and psychopathy (e.g., Dawel et al., 2012). In Study 1 we use the ERAM test to investigate if these associations can be observed also in a non-clinical sample.

Finally, in Study 2 we distribute the ERAM test online to see if results are consistent across different participant samples and distribution methods (in lab vs. online). In addition, we investigate if meta-cognitive skills are related to emotion recognition ability. Kelly and Metcalfe (2011) reported that individuals have metacognitive awareness of their interpersonal sensitivity in the sense that they were aware of when they did, or did not, recognize accurately the emotions conveyed by static pictures of facial expressions. Lausen and Hammerschmidt (2020) also showed that accurate recognition of vocally expressed emotions elicited increased confidence judgments. We attempt to replicate these findings using the dynamic and multimodal ERAM test.

2. Study 1

Study 1 presents recognition rates for the ERAM test in a large sample of participants ($N = 593$) tested in lab, and also investigates potential sex differences in emotion recognition ability. We further examined the underlying structure of emotion recognition ability using correlation analyses and confirmatory factor analysis. Finally, we investigated associations between emotion recognition ability and broader affective processes using self-report measures of emotional competencies, personality, and socio-emotional dysfunction.

2.1. Methods

2.1.1. Participants and procedure

Six hundred young adults participated in the study. One participant later wished to withdraw from the study, and data from 6 others were lost because of equipment failure, leading to a total sample size of 593 participants (226 men, mean age = 23.4 years, SD = 3.30, range = 18-36 years; 367 women, mean age = 22.9, SD = 3.15, range = 18-34 years). All participants had normal or corrected-to-normal vision and hearing, and no self-reported history of neurological psychological illness or drug or alcohol abuse. Participants were recruited through paper advertisements at universities in the Stockholm area and through an online recruiting site. Written informed consent was received from all participants, who also provided saliva samples for subsequent DNA analysis. Results from the genetic analyses are reported elsewhere (Hovey et al., 2018). Sample size was determined a-priori based on considerations of expected strength of phenotype-genotype correlations.

Participants first completed a multimodal emotion recognition task (described below) followed by an emotional memory task which is reported elsewhere (for details and results, see Cortes et al., 2017). The session concluded with a battery of self-report questionnaires (described below), and the length of one session was approximately 1.5 h. The study design was approved by the Stockholm area Regional Ethical Review Board (decision no. 2012/1511-31/2), and each participant received 3 movie vouchers as compensation for their voluntary participation.

2.1.2. Materials

2.1.2.1. Multimodal emotion recognition task. Emotion recognition accuracy was assessed using a new test specifically designed for this study – the Emotion Recognition Assessment in Multiple modalities (ERAM) test. The ERAM test is based on a selection of stimuli from the Geneva Multimodal Emotion Portrayal (GEMEP) corpus (Bänziger et al., 2012) – which is a large database of dynamic audio-video emotion expressions portrayed by professional actors. The actors were instructed to improvise interactions wherein they expressed emotions while pronouncing standardized pseudo-linguistic sentences (e.g., "ne kali bam sud molen!"). Video files of the selected portrayals consisted of close-up frontal views of the actor's face and upper torso, thereby providing facial, vocal and also some bodily cues to emotion. For a detailed description of the development and recording of the GEMEP corpus, we refer to Bänziger and Scherer (2010).

The ERAM test was designed to provide a sensitive and ecologically valid measure of emotion recognition ability, and contains 72 items selected from the GEMEP corpus based on existing recognition rates and believability ratings (Bänziger et al., 2012; Bänziger & Scherer, 2010). It includes portrayals conveying 12 different emotional expressions: hot anger, anxiety, despair, disgust, panic fear, happiness, interest, irritation, pleasure, pride, relief, and sadness. The selected expressions cover both basic and more subtle emotions, and both positive and negative states. In addition, anger, fear, and sadness were expressed with both low (irritation, anxiety, and sadness) and high (hot anger, panic fear, and despair) arousal levels. The 72 items were divided into 3 conditions – video (24 video clips presented without sound), audio (24 sounds presented alone), and audio-video (24 video clips presented with sound) – which allowed for separate assessment of visual, auditory, and audio-visual emotion recognition ability. Each presentation modality further consisted of items with varying levels of difficulty (2 items, one easier and one harder, for each presentation modality × expression cell) in order to avoid floor and ceiling effects. The duration of the items ranged between 1 and 5 s and sound level was normalized separately for each actor.

Experiments were conducted individually using Authorware software (Adobe Systems Inc., San Jose, CA) running on PCs to present stimuli and record responses. Video content was presented on 24″ LED monitors and audio content was presented through headphones (AKG K619, AKG Acoustics GmbH, Vienna, Austria) with sound level kept constant for all participants. Before the test started, there was a brief training session in which participants were presented with one sample item from each modality (sample items were not included among the test items). During the test, the participants first rated the video only stimuli, followed by the audio only and audio-video stimuli in a fixed order. They were instructed to choose one label which best represented the expression conveyed by each portrayal, and the alternatives they could choose from were the same as the 12 intended expressions. Responses were scored as correct if they matched the intended expressions of the emotion portrayals. On average, the participants required 15 min to complete the ERAM test.

2.1.2.2. Self-report measures. The participants filled in a comprehensive battery of self-report questionnaires related to emotional competencies, personality, and socio-emotional dysfunction. Questionnaire items were presented using LimeSurvey software (www.limesurvey.com). A more detailed presentation of the questionnaires is available in the supplementary material.

2.1.2.2.1. Emotional competencies. Emotion regulation was assessed using the Emotion Regulation Questionnaire (ERQ; Gross & John, 2003) and the Affective Style Questionnaire (ASQ; Hofmann & Kashdan, 2010). Emotional expressivity was assessed with the Berkeley Expressivity Questionnaire (BEQ; Gross & John, 1997), emotional understanding with the Situational Test of Emotional Understanding (STEU; MacCann & Roberts, 2008), and empathy with the Interpersonal Reactivity Index (IRI; Davis, 1983) and the short (26 item) version of the Empathy Quotient questionnaire (EQ; Allison et al., 2011; Baron-Cohen & Wheelwright, 2004).

2.1.2.2.2. Personality and socio-emotional dysfunction. Personality traits were assessed using the Big Five Inventory (BFI, John et al., 2008). We also administered the Toronto Alexithymia Scale (TAS-20; Bagby et al., 1994), the Youth Psychopathic Traits Inventory (YPI; Andershed et al., 2002), the Autism Spectrum Quotient questionnaire (AQ; Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001), the Hospital Anxiety and Depression scale (HAD; Zigmond & Snaith, 1983), and the Life History of Aggression scale (LHA; Coccaro et al., 1997) to measure dysfunctional traits and behaviors.

2.1.2.2.3. Other questionnaires. In addition to the above measures, we also administered the Positive and Negative Affect Schedule (PANAS; Watson et al., 1988) to assess current mood and the Social Network Index (SNI; Cohen et al., 1997) to assess participation in social relationships. Finally, a subset of the participants answered questions about their sleep habits (results for the sleep variables are reported in Holding et al., 2017).

2.2. Results

2.2.1. Emotion recognition rates for the ERAM test and sex differences

Participants' emotion judgments are detailed in Table 1 in the form of a confusion matrix. The intended emotions are plotted in the columns and percentage of judgments is plotted in the rows, separately for each emotion (across presentation modalities). Percentage of correct recognition is shown in the diagonal and the other cells show percentage of incorrect answers by emotion category. Recognition rates varied between 41% (sadness) and 72% (pleasure) and the overall recognition rate was 56%, which is 6.7 times higher than chance responding (the chance level in a 12-alternative forced-choice task is 1 out of 12 = 8.33%). All emotions were recognized with accuracy above chance as indicated by 95% confidence intervals (also shown in Table 1). Confusions were most common between conceptually similar emotion categories such as anxiety being mistaken for fear, despair for sadness, and sadness for fear. Some confusions also occurred between various positive emotions (e.g., pride and happiness). Confusion patterns for facial, vocal, and multimodal expressions were similar to the ones observed for overall accuracy and are available in online supplemental Table S1.

The distribution of recognition rates across conditions was examined with a mixed analysis of variance (ANOVA). Emotion recognition accuracy (percentage of correct responses) served as the dependent variable, and independent variables were: participant sex (between-subjects variable: male, female), presentation modality (within-subjects variable: video only, audio only, audio-video) and emotion (within-subjects variable: 12 emotions), see Table 2. Significant main effects of sex, modality and emotion showed that women ($M = 56.71$, $SD = 9.97$) achieved higher recognition accuracy than men ($M = 54.38$, $SD = 10.30$, Cohen's $d = 0.23$), that multimodal expressions ($M = 65.82$, $SD = 12.91$) were better recognized than facial ($M = 53.56$, $SD = 12.61$) and vocal ($M = 48.13$, $SD = 12.23$) expressions ($ps < .001$, dependent t-tests, Bonferroni corrected alpha $= 0.017$), and that recognition accuracy varied across emotions (see diagonal values in Table 1). Significant two-way interactions indicated that women performed better than men for facial ($p = .010$) and multimodal expressions ($p = .002$; independent t-tests, Bonferroni corrected alpha $= 0.017$) (see Table 3), and for the emotions despair, relief and sadness ($ps \leq .001$; independent t-tests, Bonferroni corrected alpha $= 0.004$) (see online supplemental Table S2). Finally, recognition rates also varied across emotions and modalities (see online supplemental Table S1). For example, happiness was the best recognized emotion from facial expressions, but was not as well recognized from vocal expressions.

2.2.2. Underlying structure of emotion recognition ability

The factor structure of the ERAM test was investigated with confirmatory factor analysis (CFA) using the 'lavaan' package for R (Rosseel, 2012). The model incorporated a higher-order general emotion

Table 1
Confusion matrix showing the overall percentage of judgments in the ERAM task (Study 1).

Intended emotion												
Judgment	Hot anger	Anxiety	Despair	Disgust	Panic fear	Happiness	Interest	Irritation	Pleasure	Pride	Relief	Sadness
Hot anger	**69.81** [68, 72]	1.21	1.01	0.98	6.16	1.88	0.34	2.87	0.06	1.77	0.20	1.01
Anxiety	2.11	**43.56** [42, 45]	10.48	6.13	11.33	1.57	4.78	6.58	1.01	0.62	2.42	16.08
Despair	1.88	8.01	**48.37** [46, 50]	7.73	26.03	9.25	1.77	2.61	0.39	0.28	1.88	9.42
Disgust	2.05	2.22	0.34	**52.56** [51, 54]	0.06	0.20	3.32	8.68	0.67	2.22	1.60	11.07
Panic fear	3.34	26.64	6.46	2.11	**51.21** [50, 53]	2.84	0.62	2.25	0.03	0.28	0.65	6.18
Happiness	0.17	0.14	1.12	0.28	0.03	**61.38** [60, 63]	2.64	0.34	10.29	22.20	2.02	0.25
Interest	1.94	6.63	0.31	0.62	0.22	2.33	**59.92** [58, 62]	13.41	1.07	6.49	1.46	2.28
Irritation	17.96	5.68	1.63	3.46	2.50	1.43	4.67	**51.71** [50, 54]	0.31	3.65	3.37	5.03
Pleasure	0.34	1.10	0.20	4.50	0.00	1.55	8.26	2.33	**72.09** [70, 74]	5.82	14.02	2.47
Pride	0.22	0.39	0.20	5.93	0.06	3.57	6.75	4.69	4.41	**52.95** [51, 55]	5.23	2.61
Relief	0.06	1.74	0.34	2.61	0.06	8.71	6.49	3.71	8.66	3.68	**65.94** [64, 68]	3.06
Sadness	0.11	2.67	29.54	13.10	2.36	5.28	0.45	0.82	1.01	0.03	1.21	**40.53** [39, 42]

Note. Recognition rates for which the expression portrayed is the same as the expression judged are shown in the diagonal cells (marked in bold). Numbers in brackets indicate 95% confidence intervals for the diagonal values.

Table 2
Results from ANOVA analysis of emotion recognition rates (Study 1).

Effect	F	df	*p*	Partial eta-squared
Modality	538.23	2, 1180	<.001	0.477
Emotion	135.75	11, 6490	<.001	0.187
Sex	7.47	1, 590	.006	0.013
Modality×Emotion	91.16	22, 12,980	<.001	0.134
Modality×Sex	3.17	2, 1180	.042	0.005
Emotion×Sex	3.63	11, 6490	<.001	0.006
Modality×Emotion×Sex	1.28	22, 12,980	.171	0.002

Table 3
Emotion recognition rates for facial, vocal, and multimodal expressions as a function of sex (Study 1).

Modality	Women M (SD)	Men M (SD)	*t* (590)	*p*	Cohen's *d*
Facial expression	54.62 (12.74)	51.86 (12.26)	2.60	.010	0.22
Vocal expression	48.41 (11.56)	47.60 (13.22)	0.78	.438	0.07
Multimodal expression	67.12 (12.60)	63.68 (13.17)	3.17	.002	0.27

recognition factor and subfactors representing each of the 12 included emotion categories. For each subfactor, the indicators consisted of all items expressing the emotion in question ($N = 6$ per emotion) and they were restricted to only load on that specific factor. The indicators with the highest loadings for each subfactor were fixed to 1 and served as reference. Diagonal weight least squares (DWLS) were used as estimators, because the analysis was run on the binary scores (correct = 1, incorrect = 0). The model fit was reasonably good, $\chi^2(2472) = 2936.15$, $p < .001$, CFI = 0.908, TLI = 0.905, RMSEA = 0.018 (CI 90% [0.015, 0.020]). Estimated factor loadings on the general factor were significant for all first-order emotion factors ($ps < .001$) and were relatively strong (loadings ≥0.68) for all emotions except relief (0.59), fear (0.54) and happiness (0.33). All factor loadings are available in online supplemental Table S3.

The internal consistency of the ERAM test was investigated by calculating omega hierarchical (ω_h) which indicates the proportion of test variance due to a general factor, and omega total (ω_t) which indicates the proportion of test variance explained by all factors (Revelle & Zinbarg, 2009). These coefficients were calculated using the 'psych' package in R (Revelle, 2020). Results showed that the contribution of one general factor was strong ($\omega_h = 0.69$), although the first-order factors explained more variance ($\omega_t = 0.78$). Cronbach's alpha was 0.74.

The dimensionality of emotion recognition ability was further explored by investigating the correlations between overall recognition rates for facial, vocal, and multimodal expressions. All recognition rates were substantially correlated: $r_{facial/vocal} = 0.41$, $r_{facial/multimodal} = 0.52$, and $r_{vocal/multimodal} = 0.50$ ($ps < .001$). In addition, recognition rates for facial ($r = 0.80$), vocal ($r = 0.78$), and multimodal expressions ($r = 0.84$) were highly correlated with overall recognition accuracy ($ps < .001$). Because there were only 2 items per emotion for each modality it was not possible to conduct CFA analyses that incorporated both emotions and modalities as subfactors.

2.2.3. Emotion recognition ability and broader affective processes

Analyses of individual difference correlates of emotion recognition ability focused on overall recognition accuracy. This is the most reliable measure because it is based on judgments of all of the 72 items included in the ERAM test. Table 4 shows descriptive statistics (M, SD) and Cronbach alpha for the questionnaire variables and their correlations with emotion recognition accuracy. Significant positive correlations were observed between ERAM and emotional understanding (STEU, $r = 0.38$), the personality factor openness ($r = 0.17$), and empathy (EQ, $r = 0.16$). A significant negative correlation between ERAM and alexithymia (TAS-20, $r = -0.19$) was also observed ($ps < .001$, Bonferroni corrected alpha level = 0.0017). These associations were not explained by sex, because including sex as a covariate left the correlations practically unchanged: STEU (partial $r = 0.37$), openness (partial $r = 0.18$), EQ (partial $r = 0.14$), and TAS-20 (partial $r = -0.19$) ($ps \leq .001$).

Table 4
Descriptive statistics (M, SD) and Cronbach alpha for self-report questionnaires related to emotional competencies, personality, and socio-emotional dysfunction, and their correlations (Pearson *r*) with emotion recognition ability (Study 1).

Questionnaire	Possible range of scores	Mean (SD)	Cronbach alpha	r (with ERAM)
Emotional competencies				
Emotion Regulation Questionnaire (ERQ)				
Reappraisal	1–7	4.77 (1.12)	0.82	−0.02
Suppression	1–7	3.32 (1.21)	0.72	−0.12**
Affective Style Questionnaire (ASQ)				
Concealing	8–40	24.66 (6.76)	0.83	−0.07
Adjusting	7–35	23.07 (5.50)	0.81	−0.10*
Tolerating	5–25	18.34 (3.60)	0.62	0.02
Berkeley Expressivity Questionnaire (BEQ)				
Total score	1–7	4.59 (0.88)	0.82	0.10*
Emotional Understanding (STEU)				
Total score	0–1	0.59 (0.11)	0.60	**0.38*****
Interpersonal Reactivity Index (IRI)				
Perspective taking	0–28	18.15 (4.64)	0.74	0.08
Fantasy	0–28	18.92 (5.38)	0.79	0.13**
Empathic concern	0–28	20.58 (4.29)	0.74	0.04
Personal distress	0–28	12.70 (4.62)	0.73	−0.03
Empathy Quotient (EQ)				
Total score	0–52	27.74 (7.51)	0.82	**0.16*****
Personality and socio-emotional dysfunction				
Big Five Inventory (BFI)				
Extraversion	1–5	3.36 (0.74)	0.84	−0.02
Agreeableness	1–5	3.80 (0.54)	0.69	−0.02
Conscientiousness	1–5	3.55 (0.64)	0.78	−0.05
Neuroticism	1–5	2.82 (0.70)	0.79	0.04
Openness	1–5	3.77 (0.67)	0.81	**0.17*****
Toronto Alexithymia Scale (TAS-20)				
Total score	20–100	46.06 (11.20)	0.83	**−0.19*****
Youth Psychopathic Traits Inventory (YPI)				
Grandiose/ manipulative	1–4	1.88 (0.48)	0.89	−0.05
Callous/unemotional	1–4	1.81 (0.41)	0.79	−0.11*
Impulsive/ irresponsible	1–4	2.27 (0.47)	0.80	−0.01
Autism Spectrum Quotient (AQ)				
Total score	0–50	16.39 (5.17)	0.67	−0.12*
Hospital Anxiety and Depression Scale (HAD)				
Anxiety	0–21	7.33 (3.31)	0.74	0.01
Depression	0–21	3.39 (2.28)	0.63	−0.10*
Life History of Aggression (LHA)				
Total score	0–55	10.57 (6.71)	0.77	0.01
Other				
Positive and Negative Affect Schedule (PANAS)				
Positive affect	10–50	32.35 (6.49)	0.83	−0.08*
Negative affect	10–50	14.45 (4.59)	0.81	−0.11**
Social Network Index (SNI)				
Network diversity	0–12	5.17 (1.52)	na	0.05
N people on social network	0–84	17.52 (8.27)	na	0.06

Note. Asterisks indicate uncorrected *p*-values: * $p < .05$, ** $p < .01$, *** $p < .001$. Values in bold indicate correlations that remain significant after correcting for multiple comparisons (Bonferroni corrected alpha level = 0.0017).

2.3. Discussion

2.3.1. Emotion recognition rates for the ERAM test and sex differences

Results provided a detailed description of the distribution of recognition rates in the ERAM test in a large sample of young adults which can be used as baseline values in future research. The ERAM test had an overall level of accuracy (56%) which is optimal for assessing individual differences without floor or ceiling effects. Effects of sex on recognition accuracy were small but significant, and women overall performed better than men as hypothesized. We note that the overall effect size (Cohen's d = 0.23) was of a similar magnitude as in previous studies on sex differences, although previous work has mainly used unimodal stimuli and fewer emotion categories (e.g., Thompson & Voyer, 2014). Effects of sex varied as a function of expression modality, and the largest sex differences were observed for the multimodal condition and the smallest for the audio-only condition. Multimodal expressions contain a larger number of expressive cues than unimodal expressions, which was reflected in higher recognition rates for multimodal expressions. Multimodal stimuli also better reflect how emotions are conveyed in everyday interactions (Bänziger et al., 2012). We speculate that women may have been better able to make use of the additional information present in multimodal expressions, and may perhaps also have been preferentially motivated by the higher ecological validity of multimodal vs. unimodal expressions. Future studies could investigate these possible explanations.

2.3.2. Underlying structure of emotion recognition ability

The results from the CFA suggest that the ERAM test has good internal consistency and can be viewed as a largely unidimensional measure of general emotion recognition ability. Recognition rates for facial, vocal, and multimodal expressions were substantially correlated with each other and highly correlated with overall recognition accuracy, in line with previous research (e.g., Bänziger et al., 2009; Borod et al., 2000; Schlegel et al., 2012). This pattern of results suggests that recognition of emotions from specific modalities appear to be related yet distinct aspects of an underlying modality-independent emotion recognition skill (Lewis et al., 2016). However, while modality-specific recognition skills may converge into a higher-order modality-independent skill, the sizes of the correlations were not so large that the modality-specific skills might be seen as redundant.

2.3.3. Emotion recognition ability and broader affective processes

Correlations between emotion recognition ability and questionnaire variables were overall small. The finding that ability measures do not correlate highly with self-report measures is consistent with previous research (e.g., Schlegel et al., 2019), and it has been suggested that ability and questionnaire tests measure fundamentally different constructs (Roberts et al., 2010). The test of emotion understanding (STEU; MacCann & Roberts, 2008) was the only ability test among the questionnaires, and the positive association between ERAM and STEU was also the strongest observed. This corroborates previous research (e.g., Schlegel et al., 2019; Thingujam et al., 2012; see Elfenbein & MacCann, 2017, for a review) and suggests convergence between emotion perception and emotion understanding – which are considered two main branches of the emotional intelligence construct (Roberts et al., 2010).

We observed a positive correlation between ERAM scores and the personality factor openness, and note that similar findings have been reported in the literature (e.g., Matsumoto et al., 2000; Schlegel et al., 2019; Terracciano et al., 2003). It has been proposed that persons who score high on openness are more interested in stimulation and pay more attention to others' emotions, and thus would perform better on emotion recognition tests (Matsumoto et al., 2000). However, we also note that the previous literature on associations between emotion recognition ability and personality shows many conflicting results and no clear trends can be discerned (for a review, see Furnes, Berg, Mitchell, & Paulmann, 2019). This indicates that individual differences in emotion recognition ability may not be best explained at the level of broad Big-5 personality factors.

Good emotion recognition skills are crucial for being able to adopt another person's perspective and to respond empathically, and thus impairment in emotion recognition is thought to underlie impairment in empathy (Blair, 2005). Our observation that ERAM scores were positively correlated with empathy is in line with previous studies (e.g., Besel & Yuille, 2010; Lockwood, Ang, Husain, & Crockett, 2017; Olderbak & Wilhelm, 2017; Schlegel et al., 2019), and provides new data from a large sample of healthy young adults. ERAM scores were further negatively correlated with self-reported alexithymia – a personality construct which is characterized by difficulties in identifying and describing one's feelings (Sifneos, 1973). This observation is in line with previous studies as reviewed by Grynberg et al. (2012), who also argue that alexithymia may drive socio-emotional problems found in clinical disorders. Most previous studies have investigated clinical samples, whereas our data suggests that the negative correlation between alexithymia traits and emotion recognition can be observed also in a non-clinical sample (see also Lane et al., 1996; Parker et al., 2005; Schlegel et al., 2019).

A problem with including many variables in a single study is that the risk for false positive findings increases. We therefore focused only on the strongest associations between emotion recognition ability and questionnaires – which were not explained by sex and survived a stringent correction for multiple comparisons. However, also some of the other correlations in Table 4, although not discussed here, may still provide useful data for future meta-analytic reviews. These include, for example, the positive correlation between ERAM and the Fantasy subscale of the Interpersonal Reactivity Index (Davis, 1983), and the negative correlations between ERAM and expressive suppression and self-reported autism spectrum traits.

The above associations between emotion recognition accuracy and broader affective processes speak in favor of the convergent validity of the ERAM test, although it must be noted that we did not directly test how ERAM scores relate to other emotion recognition tests. However, Flykt et al. (2021) recently reported that scores on the ERAM test were substantially correlated with scores on the Diagnostic Analysis of Nonverbal Accuracy (DANVA) test (Nowicki & Duke, 1994), which is a widely used test of unimodal recognition of basic emotions from static pictures and vocal expressions. In addition, Döllinger et al. (2021) reported that ERAM scores were substantially correlated with scores on the Patient Emotion Cue Test (PECT; Blanch-Hartigan, 2011), wherein emotions are expressed through both nonverbal and verbal channels in a medical patient interaction context.

3. Study 2

Study 2 investigated if ERAM can be distributed online and if recognition rates remain stable across different participant samples. In addition, we investigated meta-cognition of emotion recognition (see Kelly & Metcalfe, 2011; Lausen & Hammerschmidt, 2020). More specifically, we wanted to know if participants were aware of when they did (or did not) recognize the expression that was conveyed by the items in the ERAM test, and if the accuracy of such meta-cognitive judgments was correlated with actual emotion recognition accuracy.

3.1. Methods

3.1.1. Participants

106 young adults took part in the study (24 men, mean age = 24.5 years, SD = 3.48, range = 20-31 years; 82 women, mean age = 26.4 years, SD = 6.35, range = 20-53 years). The sample consisted of students from three universities in Middle Sweden who participated in exchange for course credits. The sample size was determined based on a-priori power calculation for a medium effect size ($r = 0.30$), a power level of 0.80, and an alpha of 0.05 (which gives a minimum required size of $N = 84$).

3.1.2. Procedure and materials

Emotion recognition accuracy was assessed using the ERAM test, as previously described. In contrast to Study 1, the test was administered online using Qualtrics software (Qualtrics, Provo, UT) to present items and collect responses. Participants were also asked to give confidence ratings for each judgment. Confidence ratings were given retrospectively after each emotion judgment on a scale from 0% to 100% using a slider that moved in intervals of 10%.

Each participant received a personal invitation with a link to the Qualtrics survey, a short presentation of the aim of the study, information about technical requirements (i.e., respond to the survey using a computer rather than a mobile phone) and recommendations about suitable browsers and use of headphones. They were also instructed that they should take the test in a quiet room where they will not be disturbed. All participants were asked for their informed consent online by either selecting "Yes, I consent" to continue or "No, I do not consent" to withdraw from participation. We used settings in Qualtrics to prevent the participants from taking the test more than once (this setting places cookies on the browser when a response is first submitted, and on subsequent attempts to take the test the cookies are recognized and access is denied). Participants took 15-20 min to complete the ERAM test, including the confidence ratings.

3.2. Results

3.2.1. Emotion recognition rates in the online version of the ERAM test

Overall percentage of correct responses was 57.35%, which is very similar to the recognition rate observed in Study 1. The distribution of recognition rates across conditions was examined with a repeated measures ANOVA with presentation modality (video only, audio only, audio-video) and emotion (12 emotions) as within-subjects variables. Unlike Study 1, participant sex was not included in the analyses due to the small number of male participants. A significant main effect of modality [$F(2, 210) = 87.21$, $p < .001$, partial eta-squared = 0.45] showed that multimodal expressions (M = 65.82, SD = 12.91) were better recognized than facial (M = 53.56, SD = 12.61) and vocal (M = 48.13, SD = 12.23) expressions ($ps < .001$, dependent t-tests, Bonferroni corrected alpha = 0.017). A significant main effect of emotion [$F(11, 1155) = 24.11$, $p < .001$, partial eta-squared = 0.19] indicated that recognition accuracy varied across emotions, and a significant modality × emotion interaction [$F(22, 2310) = 19.93$, $p < .001$, partial eta-squared = 0.16] indicated that recognition rates also varied across emotions and modalities. These effects are illustrated in the online supplemental Table S4 which shows the confusion matrices for overall recognition accuracy, as well as separately for each individual presentation modality. Both recognition rates and patterns of misclassifications were highly similar across Studies 1 and 2.

3.2.2. Underlying structure of emotion recognition ability

The online version of the ERAM test showed good internal consistency with Cronbach alpha = 0.80 (based on all 72 items). Recognition rates for facial, vocal and multimodal expressions were again substantially correlated: $r_{facial/vocal} = 0.37$, $r_{facial/multimodal} = 0.63$, and $r_{vocal/multimodal} = 0.58$ ($ps < .001$). Recognition rates for facial ($r = 0.81$), vocal ($r = 0.79$), and multimodal expressions ($r = 0.89$) were also highly correlated with overall recognition accuracy ($ps < .001$). In addition, we calculated the correlation between participants' overall recognition rates for each item of the ERAM test across Studies 1 and 2. This correlation was very high ($r = 0.93$, $p < .001$, $N = 72$). CFA was not performed in Study 2 because of the relatively small sample size.

3.2.3. Meta-cognition of emotion recognition

Following the procedure in Kelly and Metcalfe (2011), gamma correlations were calculated, separately for each participant, between the retrospective confidence ratings for each stimulus in the ERAM test and whether the emotion judgment was correct (coded as 1) or incorrect (coded as 0). Gamma correlations were significantly greater than zero (mean gamma = 0.45, SD = 0.19) as indicated by an independent *t*-test, $t(105) = 24.42$, $p < .001$.

We also calculated the Pearson correlations between overall emotion recognition accuracy and the retrospective gamma correlations. These correlations were significant for overall recognition accuracy across presentation modalities ($r = 0.41$, $p < .001$) as well as for each individual presentation modality: video ($r = 0.34$, $p < .001$), audio ($r = 0.19$, $p = .049$), and multimodal expressions ($r = 0.49$, $p < .001$) ($N = 106$ for all correlations).

3.3. Discussion

Results showed acceptable internal consistency for the online version of the ERAM test and the distribution of recognition rates was very similar to that obtained in Study 1. The high correlation between recognition rates for ERAM items for Study 1 and 2 indicates that the average recognition rates for each item are stable across participant samples and distribution methods. Multimodal emotion recognition tests that can be distributed online can be a useful asset for the research community, as online tests make it easier to collect data from more diverse participant demographics and to use larger sample sizes.

Results from the metacognition analyses replicated and extended previous findings (Kelly & Metcalfe, 2011; Lausen & Hammerschmidt, 2020) by using dynamic expressions in several modalities and a larger range of emotions. The finding that gamma correlations were significantly greater than zero indicates that relative meta-accuracy was good and suggests that participants had insight into how accurate their performance on the ERAM test was. The positive correlation between each participant's ERAM scores and gamma correlations further suggests that participants who performed better at emotion recognition were also more accurate in their meta-cognitive judgments about their own accuracy. The association between wider meta-cognitive and emotional abilities remains an interesting topic for further study.

4. Summary and general discussion

Individual differences in emotion recognition ability were investigated in two studies using a brief test containing dynamic facial and vocal expressions of a wide range of emotions presented in video only, audio only, and audio-visual conditions (the ERAM test). Study 1 was conducted in lab and investigated associations between emotion recognition ability and self-report measures of emotional competencies, personality, and socio-emotional dysfunction. Results showed that women achieved higher accuracy than men, and emotion recognition was further positively correlated with emotional understanding, empathy, and openness, and negatively correlated with alexithymia. Study 2 instead used an online version of the ERAM test and replicated the recognition rates observed in Study 1. In addition, Study 2 also showed that participants were aware of when they did (or did not) recognize the expression that was conveyed by the items in the ERAM test, and that the accuracy of such meta-cognitive judgments was positively correlated with actual emotion recognition accuracy. These results provide further clues about how emotion recognition ability is associated with broader affective processes. Recognition rates for visual, auditory, and audio-visual expressions were substantially correlated in both studies, which suggests that modality-specific emotion recognition skills may be related yet distinct aspects of a general modality-independent emotion recognition skill. These results, in turn, provide further clues about the underlying structure of emotion recognition ability.

The ERAM test, which can be distributed both in lab and online, is available for academic research upon request. A unique feature of the ERAM test is that it provides an overall measure of general emotion recognition ability as well as separate measures of emotion recognition from dynamic facial, vocal, and multimodal expressions. In addition, it is a relatively brief test that only takes 15-20 min to complete and it also contains a relatively wide range of both positive and negative emotions. It can thus provide an alternative to previous unimodal tests that use, for example, static pictures of facial expressions of a small number of emotion categories. It also complements previous dynamic emotion recognition tests that contain only multimodal expressions such as the GERT (Schlegel et al., 2014; Schlegel & Scherer, 2016) – which is also based on the same set of expressions as the ERAM test (Bänziger et al., 2012). The brief ERAM test also complements previous longer tests that provide separate measures of facial, vocal and multimodal recognition such as the MERT (Bänziger et al., 2009).

We note that the ERAM test is language free in the sense that the actors vocalize in a pseudo-language, and it can therefore be used in many different cultural settings (for example, the present study was conducted with participants living in Sweden). However, the stimuli were portrayed by Swiss-French actors, so it is likely that the items also contain some culture-specific cues to emotion (see e.g., Laukka et al., 2016; Scherer et al., 2011). This means that recognition rates may differ between nations, with possibly higher accuracy in nations that are culturally closer to the actors' residence. How ERAM scores, and their associations with various individual difference variables, may vary across cultures presents an interesting topic for future studies. The stimuli in ERAM were produced using a natural acting procedure so that the expressions would be as authentic as possible (see Bänziger et al., 2012). However, it is possible that actor portrayals still differ in some important aspects from spontaneous emotion expressions (e.g., Juslin et al., 2018), so future studies could compare performance on the ERAM test with tasks that utilize spontaneous expressions (e.g., Israelashvili et al., 2021). Future studies could also assess possible effects of item presentation order (e.g., fixed vs. random order) and different response formats (e.g., language based labelling vs. non-linguistic perceptual matching) (e.g., Palermo et al., 2013).

Another limitation of the ERAM test is that it contains relatively few items per emotion, especially when it comes down to looking at individual emotions for separate expression modalities. This is a consequence of including 3 different modalities and 12 emotions, while keeping the number of items as low as possible. This may lead to decreased generalizability and increased measurement error for individual emotions compared to the overall recognition rate based on all items. Future studies of individual differences in emotion recognition ability could develop measures that allow investigation of individual emotions as a function of expression modality. Future studies could also investigate individual facial cues to emotion (e.g., facial action unit activations) together with vocal cues (e.g., pitch, loudness, speech rate) to get a sense of the mechanisms underlying individual differences in emotion recognition ability (see e.g., Leitman et al., 2010). Studies of individual differences would in general benefit from moving away from only demonstrating differences to also addressing the potential causes for the observed differences. Nonverbal emotion recognition tests, such as the ERAM test, may contribute to this quest.

Declaration of competing interest

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Acknowledgements

This study was funded by the Swedish Research Council through a grant to PL (grant no. 2012-801). Open access publication fees were covered by Stockholm University. PL, TB and HF contributed to the design of Study 1. PL and AI contributed to the design of Study 2. TB and KS contributed materials. DC and CT (Study 1) and AI (Study 2) contributed to data collection. PL and AI contributed to data analysis. PL wrote the paper with input from DC, CT, AI, KS and HF. We thank Marc Covents for programming the online version of the ERAM test.

The two studies that are presented in this article were not preregistered. Data and code are available on the Open Science Framework (https://osf.io/rf6md/). The ERAM test is freely available for academic research upon request (for more information, see https://www.unige.ch/cisa/emotional-competence/home/research-tools/).

Appendix A. Supplementary information

Supplementary materials to this article can be found online at https://doi.org/10.1016/j.actpsy.2021.103422.

References

Allison, C., Baron-Cohen, S., Wheelwright, S. J., Stone, M. H., & Muncer, S. J. (2011). Psychometric analysis of the empathy quotient (EQ). *Personality and Individual Differences, 51*, 829–835. https://doi.org/10.1016/j.paid.2011.07.005

Andershed, H., Kerr, M., Stattin, H., & Levander, S. (2002). Psychopathic traits in non-referred youths: A new assessment tool. In E. Blaauw, & L. Sheridan (Eds.), *Psychopaths: Current international perspectives* (pp. 131–158). The Hague, Netherlands: Elsevier.

Bagby, R. M., Parker, J. D., & Taylor, G. J. (1994). The twenty-item Toronto alexithymia scale: I. Item selection and cross-validation of the factor structure. *Journal of Psychosomatic Research, 38*, 23–32. https://doi.org/10.1016/0022-3999(94)90005-1

Bänziger, T. (2016). Accuracy of judging emotions. In J. A. Hall, M. Schmid Mast, & T. V. West (Eds.), *The social psychology of perceiving others accurately* (pp. 23–51). Cambridge, United Kingdom: Cambridge University Press.

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. https://doi.org/10.1037/a0017088

Bänziger, T., Mortillaro, M., & Scherer, K. R. (2012). Introducing the Geneva multimodal expression corpus for experimental research on emotion perception. *Emotion, 12*, 1161–1179. https://doi.org/10.1037/a0025827

Bänziger, T., & Scherer, K. R. (2010). Introducing the Geneva Multimodal Emotion Portrayal (GEMEP) corpus. In K. R. Scherer, T. Bänziger, & E. B. Roesch (Eds.), *Blueprint for affective computing: A sourcebook* (pp. 271–294). New York: Oxford University Press.

Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient (EQ): An investigation of adults with asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders, 34*, 163–175. https://doi.org/10.1023/B:JADD.0000022607.19833.00

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders, 31*, 5–17. https://doi.org/10.1023/A:1005653411471

Besel, L. D. S., & Yuille, J. C. (2010). Individual differences in empathy: The role of facial expression recognition. *Personality and Individual Differences, 49*, 107–112. https://doi.org/10.1016/j.paid.2010.03.013

Blair, R. J. R. (2005). Responding to the emotions of others: Dissociating forms of empathy through the study of typical and psychiatric populations. *Consciousness and Cognition, 14*, 698–718. https://doi.org/10.1016/j.concog.2005.06.004

Blanch-Hartigan, D. (2011). Measuring providers' verbal and nonverbal emotion recognition ability: Reliability and validity of the patient emotion cue test (PECT). *Patient Education and Counseling, 82*, 370–376. https://doi.org/10.1016/j.pec.2010.11.017

Borod, J. C., Pick, L. H., Hall, S., Sliwinski, M., Madigan, N., Obler, L. K., Welkowitz, J., Canino, E., Erhan, H. M., Goral, M., Morrison, C., & Tabert, M. (2000). Relationships among facial, prosodic, and lexical channels of emotional perceptual processing. *Cognition and Emotion, 14*, 193–211. https://doi.org/10.1080/026999300378932

Coccaro, E. F., Berman, M. E., & Kavoussi, R. J. (1997). Assessment of life history of aggression: Development and psychometric characteristics. *Psychiatry Research, 73*, 147–157. https://doi.org/10.1016/S0165-1781(97)00119-4

Cohen, S., Doyle, W. J., Skoner, D. P., Rabin, B. S., & Gwaltney, J. M., Jr. (1997). Social ties and susceptibility to the common cold. *Journal of the American Medical Association, 277*, 1940–1944. https://doi.org/10.1001/jama.1997.03540480040036

Cordaro, D. T., Sun, R., Keltner, D., Kamble, S., Huddar, N., & McNeil, G. (2018). Universals and cultural variations in 22 emotional expressions across five cultures. *Emotion, 18*, 75–93. https://doi.org/10.1037/emo0000302

Cortes, D. S., Laukka, P., Lindahl, C., & Fischer, H. (2017). Memory for faces and voices varies as a function of sex and expressed emotion. *PLoS One, 12*, Article e0178423. https://doi.org/10.1371/journal.pone.0178423

Cowen, A. S., Elfenbein, H. A., Laukka, P., & Keltner, D. (2019). Mapping 24 emotions conveyed by brief human vocalization. *American Psychologist, 74*, 698–712. https://doi.org/10.1037/amp0000399

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology, 44*, 113–126. https://doi.org/10.1037/0022-3514.44.1.113

Dawel, A., O'Kearney, R., McKone, E., & Palermo, R. (2012). Not just fear and sadness: Meta-analytic evidence of pervasive emotion recognition deficits for facial and vocal expressions in psychopathy. *Neuroscience and Biobehavioral Reviews, 36*, 2288–2304. https://doi.org/10.1016/j.neubiorev.2012.08.006

de Gelder, B. (2000). Recognizing emotions by ear and by eye. In R. D. Lane, & L. Nadel (Eds.), *Cognitive neuroscience of emotion* (pp. 84–105). New York: Oxford University Press.

Döllinger, L., Laukka, P., Högman, L. B., Bänziger, T., Makower, I., Fischer, H., & Hau, S. (2021). Training emotion recognition accuracy: Results for multimodal expressions and facial micro expressions. *Frontiers in Psychology, 12*, Article 708867. https://doi.org/10.3389/fpsyg.2021.708867

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin, 128*, 203–235. https://doi.org/10.1037/0033-2909.128.2.203

Elfenbein, H. A., & MacCann, C. (2017). A closer look at ability emotional intelligence (EI): What are its component parts, and how do they relate to each other? *Social and Personality Psychology Compass, 11*, Article e12324. https://doi.org/10.1111/spc3.12324

Flykt, A., Hörlin, T., Linder, F., Wennstig, A.-K., Sayeler, G., Hess, U., & Bänziger, T. (2021). Exploring emotion recognition and the understanding of others' unspoken thoughts and feelings when narrating self-experienced emotional events. *Journal of Nonverbal Behavior, 45*, 67–81. https://doi.org/10.1007/s10919-020-00340-4

Furnes, D., Berg, H., Mitchell, R. M., & Paulmann, S. (2019). Exploring the effects of personality traits on the perception of emotions from prosody. *Frontiers in Psychology, 10*, Article 184. https://doi.org/10.3389/fpsyg.2019.00184

Gross, J. J., & John, O. P. (1997). Revealing feelings: Facets of emotional expressivity in self-reports, peer ratings, and behavior. *Journal of Personality and Social Psychology, 72*, 435–448. https://doi.org/10.1037/0022-3514.72.2.435

Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology, 85*, 348–362. https://doi.org/10.1037/0022-3514.85.2.348

Grynberg, D., Chang, B., Corneille, O., Maurage, P., Vermeulen, N., Berthoz, S., & Luminet, O. (2012). Alexithymia and the processing of emotional facial expressions (EFEs): Systematic review, unanswered questions and further perspectives. *PLoS One, 7*, Article e42429. https://doi.org/10.1371/journal.pone.0042429

Gur, R. C., Sara, R., Hagendoorn, M., Marom, O., Hughett, P., Macy, L., Turner, T., Bajcsy, R., Posner, A., & Gur, R. E. (2002). A method for obtaining 3-dimensional facial expressions and its standardization for use in neurocognitive studies. *Journal of Neuroscience Methods, 115*, 137–143. https://doi.org/10.1016/S0165-0270(02)00006-7

Hall, J. A., Andrzejewski, S. A., & Yopchick, J. E. (2009). Psychosocial correlates of interpersonal sensitivity: A meta-analysis. *Journal of Nonverbal Behavior, 33*, 149–180. https://doi.org/10.1007/s10919-009-0070-5

Hall, J. A., Gunnery, S. D., & Horgan, T. G. (2016). Gender differences in interpersonal accuracy. In J. A. Hall, M. S. Mast, & T. V. West (Eds.), *The social psychology of perceiving others accurately* (pp. 309–327). Cambridge, United Kingdom: Cambridge University Press.

Hodges, S. D., & Wise, A. A. P. (2016). Interpersonal accuracy: Real and perceived links to prosocial behavior. In J. A. Hall, M. S. Mast, & T. V. West (Eds.), *The social psychology of perceiving others accurately* (pp. 350–375). Cambridge, United Kingdom: Cambridge University Press.

Hofmann, S. G., & Kashdan, T. B. (2010). The affective style questionnaire: Development and psychometric properties. *Journal of Psychopathology and Behavioral Assessment, 32*, 255–263. https://doi.org/10.1007/s10862-009-9142-4

Holding, B. C., Laukka, P., Fischer, H., Bänziger, T., Axelsson, J., & Sundelin, T. (2017). Multimodal emotion recognition is resilient to insufficient sleep: Results from crosssectional and experimental studies. *Sleep, 40*, Article zsx145. https://doi.org/10.1093/sleep/zsx145

Hovey, D., Henningsson, S., Cortes, D. S., Bänziger, T., Zettergren, A., Melke, J., Fischer, H., Laukka, P., & Westberg, L. (2018). Emotion recognition associated with polymorphism in oxytocinergic pathway gene ARNT2. *Social Cognitive and Affective Neuroscience, 13*, 173–181. https://doi.org/10.1093/scan/nsx141

Israelashvili, J., Pauw, L. S., Sauter, D. A., & Fischer, A. H. (2021). Emotion recognition from realistic dynamic emotional expressions cohere with established emotion recognition tests: A proof-of-concept validation of the Emotional Accuracy Test. *Journal of Intelligence, 9*, Article 25. https://doi.org/10.3390/jintelligence9020025

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative big five trait taxonomy: History, measurement, and conceptual issues. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (3rd ed., pp. 114–158). New York: Guilford Press.

Joseph, D. L., & Newman, D. A. (2010). Emotional intelligence: An integrative meta-analysis and cascading model. *Journal of Applied Psychology, 95*, 54–78. https://doi.org/10.1037/a0017286

Juslin, P. N., Laukka, P., & Bänziger, T. (2018). The mirror to our soul? Comparisons of spontaneous and posed vocal expression of emotion. *Journal of Nonverbal Behavior, 42*, 1–40. https://doi.org/10.1007/s10919-017-0268-x

Kelly, K. J., & Metcalfe, J. (2011). Metacognition of emotional face recognition. *Emotion, 11*, 896–906. https://doi.org/10.1037/a0023746

Lane, R. D., Sechrest, L., Reidel, R., Weldon, V., Kaszniak, A., & Schwartz, G. E. (1996). Impaired verbal and nonverbal emotion recognition in alexithymia. *Psychosomatic Medicine, 58*, 203–210. https://doi.org/10.1097/00006842-199605000-00002

Laukka, P., Elfenbein, H. A., Thingujam, N. S., Rockstuhl, T., Iraki, F. K., Chui, W., & Althoff, J. (2016). The expression and recognition of emotions in the voice across five nations: A lens model analysis based on acoustic features. *Journal of Personality and Social Psychology, 111*, 686–705. https://doi.org/10.1037/pspi0000066

Lausen, A., & Hammerschmidt, K. (2020). Emotion recognition and confidence ratings predicted by vocal stimulus type and prosodic parameters. *Humanities and Social Sciences Communications, 7*, Article 2. https://doi.org/10.1057/s41599-020-0499

Leitman, D. I., Laukka, P., Juslin, P. N., Saccente, E., Butler, P., & Javitt, D. C. (2010). Getting the cue: Sensory contributions to auditory emotion recognition impairments in schizophrenia. *Schizophrenia Bulletin, 36*, 545–556. https://doi.org/10.1093/schbul/sbn115

Lewis, G. J., Lefevre, C. E., & Young, A. W. (2016). Functional architecture of visual emotion recognition ability: A latent variable approach. *Journal of Experimental Psychology: General, 145*, 589–602. https://doi.org/10.1037/xge0000160

Lockwood, P. L., Ang, Y.-S., Husain, M., & Crockett, M. J. (2017). Individual differences in empathy are associated with apathy-motivation. *Scientific Reports, 7*, Article 17293. https://doi.org/10.1038/s41598-017-17415-w

MacCann, C., & Roberts, R. D. (2008). New paradigms for assessing emotional intelligence: Theory and data. *Emotion, 8*, 540–551. https://doi.org/10.1037/a0012746

Matsumoto, D., LeRoux, J., Wilson-Cohn, C., Raroque, J., Kooken, K., Ekman, P., … Goh, A. (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. https://doi.org/10.1023/A:1006668120583

Mayer, J. D., Salovey, P., Caruso, D. R., & Sitarenios, G. (2003). Measuring emotional intelligence with the MSCEIT V2.0. *Emotion, 3*, 97–105. https://doi.org/10.1037/1528-3542.3.1.97

McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality, 60*, 175–215. https://doi.org/10.1111/j.1467-6494.1992.tb00970.x

Nowicki, S., Jr., & Duke, M. (1994). Individual differences in the nonverbal communication of affect: The diagnostic analysis of nonverbal accuracy scale. *Journal of Nonverbal Behavior, 18*, 9–35. https://doi.org/10.1007/BF02169077

Olderbak, S., Riggenmann, O., Wilhelm, O., & Doebler, P. (2021). Reliability generalization of tasks and recommendations for assessing the ability to perceive facial expressions of emotion. *Psychological Assessment*. https://doi.org/10.1037/pas0001030

Olderbak, S., & Wilhelm, O. (2017). Emotion perception and empathy: An individual differences test of relations. *Emotion, 17*, 1092–1106. https://doi.org/10.1037/emo0000308

Palermo, R., O'Connor, K. B., Davis, J. M., Irons, J., & McKone, E. (2013). New tests to measure individual differences in matching and labelling facial expressions of emotion, and their association with ability to recognize vocal emotions and facial identity. *PLoS One, 8*, Article e68126. https://doi.org/10.1371/journal.pone.0068126

Parker, P. D., Prkachin, K. M., & Prkachin, G. C. (2005). Processing of facial expression of negative emotion in alexithymia: The influence of temporal constraint. *Journal of Personality, 73*, 1087–1107. https://doi.org/10.1111/j.1467-6494.2005.00339.x

Phillips, K. L., Drevets, W. C., Rauch, S. L., & Lane, R. (2003). Neurobiology of emotion perception II: Implications for major psychiatric disorders. *Biological Psychiatry, 54*, 515–528. https://doi.org/10.1016/s0006-3223(03)00171-9

Revelle, W. (2020). psych: Procedures for psychological, psychometric, and personality research (R package version 2.0.12) [Computer software]. https://CRAN.R-project.org/package=psych.

Revelle, W., & Zinbarg, R. E. (2009). Coefficients alpha, beta, omega, and the glb: Comments on Sijtsma. *Psychometrika, 74*, 145–154. https://doi.org/10.1007/s11336-008-9102-z

Roberts, R. D., MacCann, C., Matthews, G., & Zeidner, M. (2010). Emotional intelligence: Toward a consensus of models and measures. *Social and Personality Psychology Compass, 4*, 821–840. https://doi.org/10.1111/j.1751-9004.2010.00277.x

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

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*, 1–36. https://doi.org/10.18637/jss.v048.i02

Scherer, K. R., Clark-Polner, E., & Mortillaro, M. (2011). In the eye of the beholder? Universality and cultural specificity in the expression and perception of emotion. *International Journal of Psychology, 46*, 401–435. https://doi.org/10.1080/00207594.2011.626049

Scherer, K. R., & Scherer, U. (2011). Assessing the ability to recognize facial and vocal expressions of emotion: Construction and validation of the emotion recognition index. *Journal of Nonverbal Behavior, 35*, 305–326. https://doi.org/10.1007/s10919-011-0115-4

Schlegel, K., Fontaine, J. R. J., & Scherer, K. R. (2019). The nomological network of emotion recognition ability: Evidence from the Geneva emotion recognition test. *European Journal of Psychological Assessment, 35*, 352–363. https://doi.org/10.1027/1015-5759/a000396

Schlegel, K., Grandjean, D., & Scherer, K. R. (2012). Emotion recognition: Unidimensional ability or a set of modality- and emotion-specific skills? *Personality and Individual Differences, 53*, 16–21. https://doi.org/10.1016/j.paid.2012.01.026

Schlegel, K., Grandjean, D., & Scherer, K. R. (2014). Introducing the Geneva emotion recognition test: An example of Rasch-based test development. *Psychological Assessment, 26*, 666–672. https://doi.org/10.1037/a0035246

Schlegel, K., & Scherer, K. R. (2016). Introducing a short version of the Geneva emotion recognition test (GERT-S): Psychometric properties and construct validation. *Behavior Research Methods, 48*, 1383–1392. https://doi.org/10.3758/s13428-015-0646-4

Sifneos, P. E. (1973). The prevalence of alexithymic characteristics in psychosomatic patients. *Psychotherapy and Psychosomatics, 22*, 255–262. https://doi.org/10.1159/000286529

Terracciano, A., Merritt, M., Zonderman, A. B., & Evans, M. K. (2003). Personality traits and sex differences in emotion recognition among African Americans and Caucasians. *Annals of the New York Academy of Sciences, 1000*, 309–312. https://doi.org/10.1196/annals.1280.032

Thingujam, N. S., Laukka, P., & Elfenbein, H. A. (2012). Distinct emotional abilities converge: Evidence from emotion understanding and emotion recognition through the voice. *Journal of Research in Personality, 46*, 350–354. https://doi.org/10.1016/j.jrp.2012.02.005

Thompson, A. E., & Voyer, D. (2014). Sex differences in the ability to recognise non-verbal displays of emotion: A meta-analysis. *Cognition and Emotion, 28*, 1164–1195. https://doi.org/10.1080/02699931.2013.875889

Uljarevic, M., & Hamilton, A. (2013). Recognition of emotions in autism: A formal meta-analysis. *Journal of Autism and Developmental Disorders, 43*, 1517–1526. https://doi.org/10.1007/s10803-012-1695-5

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*, 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Wilhelm, O., Hildebrandt, A., Manske, K., Schacht, A., & Sommer, W. (2014). Test battery for measuring the perception and recognition of facial expressions of emotion. *Frontiers in Psychology, 5*, Article 404. https://doi.org/10.3389/fpsyg.2014.00404

Young, A., Perret, D., Calder, A., Sprengelmeyer, R., & Ekman, P. (2002). *Facial expression of emotion-stimuli and tests (FEEST)*. Bury St Edmunds, United Kingdom: Thames Valley Test Company.

Zigmond, A. S., & Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica, 67*, 361–370. https://doi.org/10.1111/j.1600-0447.1983.tb09716.x