NATIONAL RESEARCH UNIVERSITY
HIGHER SCHOOL OF ECONOMICS

Dmitry Lyusin, Victoria Ovsyannikova

A NEW VIDEOTEST FOR MEASURING EMOTION RECOGNITION ABILITY

BASIC RESEARCH PROGRAM

WORKING PAPERS

SERIES: PSYCHOLOGY
WP BRP 16/PSY/2014

This Working Paper is an output of a research project implemented as part of the Basic Research Program at the National Research University Higher School of Economics (HSE). Any opinions or claims contained in this Working Paper do not necessarily reflect the views of HSE.

Dmitry Lyusin[1], *Victoria Ovsyannikova*[2]

A NEW VIDEOTEST FOR MEASURING EMOTION RECOGNITION ABILITY[3]

A new measure for emotion recognition abilities, the Videotest of Emotion Recognition, is described. Two aspects in emotion recognition are distinguished, accuracy of recognition of emotion types that constitute the emotional state of the observed person and sensitivity to the intensity of the observed emotions. The Videotest of Emotion Recognition allows obtaining the accuracy and sensitivity indices that reflect these two aspects of emotion recognition. Psychometric analysis showed satisfactory reliability and validity of the indices.

Keywords: emotion recognition, emotion intelligence, Videotest of Emotion Recognition

JEL Classification: Z

[1] National Research University Higher School of Economics. Scientific-Educational Laboratory for Cognitive Research; Leading Research Fellow. E-mail: ooch@mail.ru

[2] National Research University Higher School of Economics. Scientific-Educational Laboratory for Cognitive Research; Senior Research Fellow. E-mail: v.ovsyannikova@gmail.com

[3] The study was implemented in the framework of the Basic Research Program at the National Research University Higher School of Economics in 2014 with financial support from the Russian Foundation for the Humanities (12-36-01287a2) and the Russian Foundation for Basic Research (14-06-00393).

Introduction

Emotion recognition has been widely studied for decades in psychology. One of the important directions in this field is developing methods for measuring emotion recognition ability. In modern psychology, emotion recognition is often conceptualized and measured in the frame of emotional intelligence research. Broadly, emotion intelligence refers to the set of abilities that allows the understanding and managing of emotions. Emotion recognition is widely considered to be one of the basic emotional intelligence components. The emotion intelligence model proposed by Mayer and Salovey (Mayer, Salovey, Caruso, & Sitarenios, 2001) identifies four branches of emotional intelligence; two of them, Emotion Perception and Emotion Understanding, are related to emotion recognition. Emotion Perception includes skills concerned with accurate detection and identification of emotions in oneself and others. Emotion Understanding concerns the ability to understand relations between emotions, emotion language and signals conveyed by emotions. According to this model four branches are ordered hierarchically, the basic branch being Emotion Perception (Salovey & Grewal, 2005). It seems that distinguishing Emotion Understanding from Emotion Perception is artificial and has an intuitive rather than theoretical background.

Another approach to the conceptualization of emotion skills, proposed by Scherer (Scherer & Scherer, 2011), understands emotion perception as one of three major domains of emotional competence along with emotion production and emotion regulation. Emotion perception is considered a central socio-emotional competence essential for many different types of occupation.

Tests for Measuring Emotion Recognition Ability: Diversity and Problems

The number of studies on measuring emotion recognition ability has been growing in recent decades. Most of the new measurement instruments have been developed in the context of emotion intelligence assessment. Two types of assessment methods are traditionally distinguished in the emotional intelligence, objective tests and self-report questionnaires. They correspond to two types of emotional intelligence models that are usually called ability and mixed models (Mayer, Salovey, & Caruso, 2000). Ability models understand emotional intelligence as a set of cognitive abilities and competencies analogous to other types of intelligence such as verbal or spatial. Mixed models, also called trait models, define emotional intelligence more broadly, as an array of cognitive, personality, and motivational traits. For measuring emotional intelligence, proponents of ability models mostly use objective tests similar to traditional intelligence tests with answers that can be assessed as right or wrong. Mixed models proponents prefer self-report questionnaires similar to personality inventories. Some exceptions from this correspondence between the two types of models and approaches to measurement are possible. For example, the EmIn Questionnaire developed by Lyusin (2006a, 2006b) that will be described below is based on the ability model. The author claims that it measures perceived emotional intelligence, understood as a cognitive ability, rather than personality traits.

The limitations of self-report assessment are broadly known; hence this paper will focus on objective tests that evaluate emotion recognition ability independent of an individual's self-concept and beliefs about his or her behavior. There is a large diversity of such tests in modern psychology. They differ in stimuli, item formats, indices, and scoring procedures. For instance, stimuli can be photographs of facial expressions, videos with various types of behavior, voice recordings, vignettes describing emotional situations, and even thoroughly non-human stimuli such as geometric figures.

The problem of scoring is one of the hardest in performance-based assessment of socio-emotional abilities. Unlike traditional intelligence tests, there are no obvious logical foundations for establishing the correct answers in most emotion recognition tests. Three major approaches to scoring have been suggested, namely expert, consensus, and target scoring. Expert scoring is based on expert opinions about the correct or best choice among the suggested answers. The main difficulty is to decide who has expertise in this case. In most cases, emotion researchers are suggested for this role, but it is often questioned if they or any other professionals such as psychotherapists, counselors, actors qualify as emotion experts. Some authors even claim that the emotion domain is one of those ill-defined knowledge domains where no objective standards for verification exist and, therefore, no qualified experts can be suggested (Legree, Psotka, Tremble, & Bourne, 2005). Consensus scoring is based on the opinion of the majority of the participants about the correct answers. It is often supposed that consensus scoring reflects cultural biases in beliefs about emotions. Moreover, it is regarded logically unacceptable to establish correct answers to the intelligence tests items, especially to the difficult ones, on the basis of a consensus opinion. In target scoring, the correct response is set by a target person who creates the stimuli. These target persons can be actors portraying emotions for photographs or voice recordings, authors of the vignettes who define a priori which emotion should be experienced by a certain character, etc. Target scoring can be applied only to a limited range of stimuli, and it can always be questioned if the target emotion was adequately portrayed or expressed in the stimuli. All three approaches have their own limitations, but they are used in psychological research and assessment for the lack of better solutions.

An important feature of emotion recognition items, and of any emotional and social abilities items, is the difficulty in establishing one correct response. Several responses to the same item can often be regarded as correct with different levels of confidence. This situation is quite normal for the psychological content being measured since emotional states are often ambiguous and constitute a mixture of various emotion types. The stimuli cannot represent all individual and situational features that result in a certain emotional state. Two important consequences result from this. First, it makes sense to use rate-the-extent format of responses similar to the Likert-type scales, rather than just to classify responses as correct and incorrect. Secondly, the unidimensional format of responses when a participant estimates the presence of only one emotion in the stimulus is less appropriate as compared to the multidimensional format that allows estimating the presence of an array of emotions in the stimulus.

Different approaches to scoring and different response formats (unidimensional or multidimensional) are used in modern emotion ability tests. The following brief review of emotion recognition tests summarizes the main tendencies in this field.

One of the most prominent early techniques for emotion recognition is the Profile of Nonverbal Sensitivity (PONS; Rosenthal, Hall, DiMatteo, Rogers, & Archer, 1979). It consists of twenty audio/video recordings in which one female person represents twenty attitudes (such as expressing jealousy, asking for a favor). The participant must assess the attitude expressed by the character. Attitudes are set initially by the test developer and are classified as dominant versus submissive and positive versus negative. Each recording is represented by eleven channels of expression (face, speech, etc.). The 220 portrayals are presented to the participant in a fixed order. For each portrayal, the participant is required to select one of two alternative answers. The accuracy index is calculated as the percentage of correct answers of the total number of test stimuli.

The Diagnostic Analysis of Nonverbal Accuracy was designed to assess the sensitivity to nonverbal expressions of emotions (DANVA; Nowicki & Duke, 1994). Twenty-four photographs of facial expressions and 24 voice recordings of four emotions (anger, fear, joy, sadness) are used as stimuli. Each emotion category is presented in two intensities, low and high. The emotions are portrayed by professional actors. The participant has to choose one of the four emotion categories for each stimulus. The accuracy scores are calculated as the percentage of correct responses separately for both types of stimuli and for the whole test.

A notable feature of the Japanese and Caucasian Brief Affect Recognition Test is the use of the images of people of different races as stimuli (JACBART; Matsumoto et al., 2000). The photographs of European and Japanese facial expressions of seven basic emotions (anger, joy, sadness, contempt, disgust, fear, and surprise) are presented to participants who have to assess the presence of each of the seven emotions in the portrayals by means of nine-point scales. The average values for each emotion category obtained in the American sample are considered to be standard. Accuracy scores are calculated as correlations between the participants' responses and the standard estimates. An interesting feature of the technique is the possibility to calculate different accuracy scores separately for each emotion category, for different races and sexes.

The most famous measure of emotion recognition is the Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT; Mayer, Salovey, Caruso, & Sitarenios, 2003). The test is based on Salovey and Mayer's model of emotional intelligence that regards it as a set of hierarchically organized cognitive abilities. The MSCEIT consists of four subtests. The first and the third subtest, Emotion Perception and Emotion Understanding, measure abilities related to emotion recognition. The Emotion Perception subtest includes two types of tasks with photographs of facial expressions and pictures of landscapes and abstract designs as stimuli. The participant must assess the degree of several emotions in each stimulus using Likert five-point scales. The Emotion Understanding subtest consists of the Blends task and the Changes task. In the Blends tasks, the participant must identify which emotions will result from the blend of several other emotions and select one of the response options. In the Changes tasks, the participant must select the emotion from the list of emotions that may result from the situation described. The weights based on expert and consensus ratings are attributed to each response option. Accuracy index is calculated by averaging the weights of the responses selected by the participant.

Recently, the Ability Emotional Intelligence Measure (AEIM; Warwick, Nettelbeck, & Ward, 2010) was developed, which is, actually, a revised version of the MSCEIT. The two scales, Emotion Perception and Emotion Management, have been changed. The principles of stimuli selection and scoring methods are similar to the MSCEIT.

The Situational Test of Emotional Understanding (STEU; MacCann & Roberts, 2008) consists of the descriptions of situations related to different emotions. The STEU items were developed according to Roseman's appraisal theory of emotions (Roseman, 2001). On the basis of this theory, the test authors set the correct answers. The accuracy index is calculated as the percentage of correct answers.

The Multimodal Emotion Recognition Test (MERT; Bänziger, Grandjean, & Scherer, 2009) consists of the presentations of expressions of five emotion families in four formats, video with sound, video without sound, audio without image, photo taken from video. Emotional expressions are demonstrated by professional actors. In total there are 120 stimuli. The participant has to select one of ten emotion categories; accuracy is calculated as the percentage of correct answers. It is possible to obtain the accuracy scores separately for different types of emotions and different formats of stimuli.

Table 1 summarizes the characteristic features of these tests with an emphasis on the methods of obtaining accuracy indices. The majority of techniques use a unidimensional response format, and the test scores are calculated as the percentage of correct responses. As mentioned above, we claim that a unidimensional response format is not quite adequate for emotion recognition tasks because of ambiguous nature of most emotional states; ignoring this fact reduces the ecological validity of the measurement.

Table 1 Measures of Emotion Recognition Ability and Methods of Obtaining Indices of Emotion Recognition Accuracy

Measure	Stimuli	Method of scoring	Response format	Calculation of the accuracy index
PONS (Rosenthal, Hall, DiMatteo, Rogers, & Archer, 1979)	Video recording of emotion expression and its components (only faces, only speech, etc.)	Target	Unidimensional	Proportion of correct responses
DANVA (Nowicki & Duke, 1994)	Photographs of faces and voice recordings	Target	Unidimensional	Proportion of correct responses
JACBART (Matsumoto et al., 2000)	Photographs of faces	Consensus	Multidimensional	Correlation between the standard estimates and a participant's responses
MSCEIT (Mayer, Salovey, Caruso, & Sitarenios, 2003); AEIM (Warwick, Nettelbeck, & Ward, 2010)	Photographs of faces and other images (Emotion Perception subtest); descriptions of situations and other verbal tasks (Emotion Understanding subtest)	Expert and consensus	Multidimensional (Emotion Perception subtest); Unidimensional (Emotion Understanding subtest)	The averaged weights of a participant's responses
STEU (MacCann & Roberts, 2008)	Descriptions of situations	Target	Unidimensional	Proportion of correct responses
MERT (Bänziger, Grandjean, & Scherer, 2009)	Video recordings with or without sound, audio recordings, photographs	Target	Unidimensional	Proportion of correct responses

Emotion Recognition Accuracy Indices for Multidimensional Response Format of Test Items

If the multidimensional response format has been chosen by a test developer, a test score should reflect the degree of similarity between the participant responses and the correct responses. The similarity index can be obtained in different ways. For further discussion, we will, as an example, take a typical test item that requires assessment of the stimulus (e.g., the emotional state of a video character) using several scales representing different emotion categories. In this case there is a set of a participant responses and a set of the standard estimates that are regarded as correct responses.

Fig. 1 demonstrates hypothetical responses of a participant to an item that consists of the fifteen Likert six-point scales. The solid line represents the profile of correct responses; the dashed line represents the profile of responses of Participant 1. What is the best way to assess the degree of similarity between these two profiles? A simple and often used measure of similarity for non-metric data is the so-called 'city-block metric' (Reis & Judd, 2000) which is calculated as the sum of the absolute values of deviations of participant responses from the correct responses on each scale. It can be defined as

$D = \Sigma |Q_i - R_i|$

where Q_i is a participant's response on Scale_i, R_i is a correct response on the same scale.

The greater the D value, the less accurate the participant's evaluations of the character's emotional state. For the data presented in Fig. 1, D = 27.

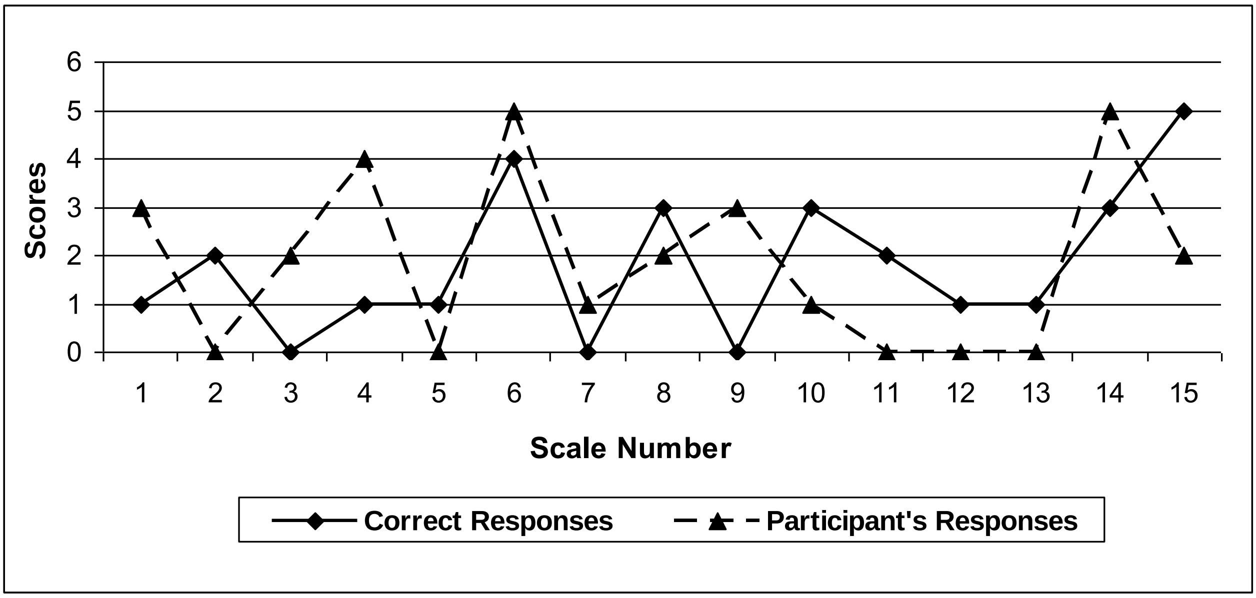

Fig. 1. Hypothetical responses of Participant 1. D = 27.

This measure of similarity is often used, but it seems that two essentially different aspects of emotion recognition are mixed, which can be illustrated by the hypothetical responses of Participant 2 presented in Fig. 2. The D value is equal to 27 as it was the case with Participant 1.

However, Participant 2 identifies emotions very accurately in a certain sense. He or she gives higher estimates on Scales 6, 8, 10, 14, 15 and lower estimates on Scales 3, 7, 8. Thus, the shape of the participant's responses profile perfectly corresponds to the correct responses profile. The only difference concerns the average level of these two profiles, the participant profile being noticeably higher.

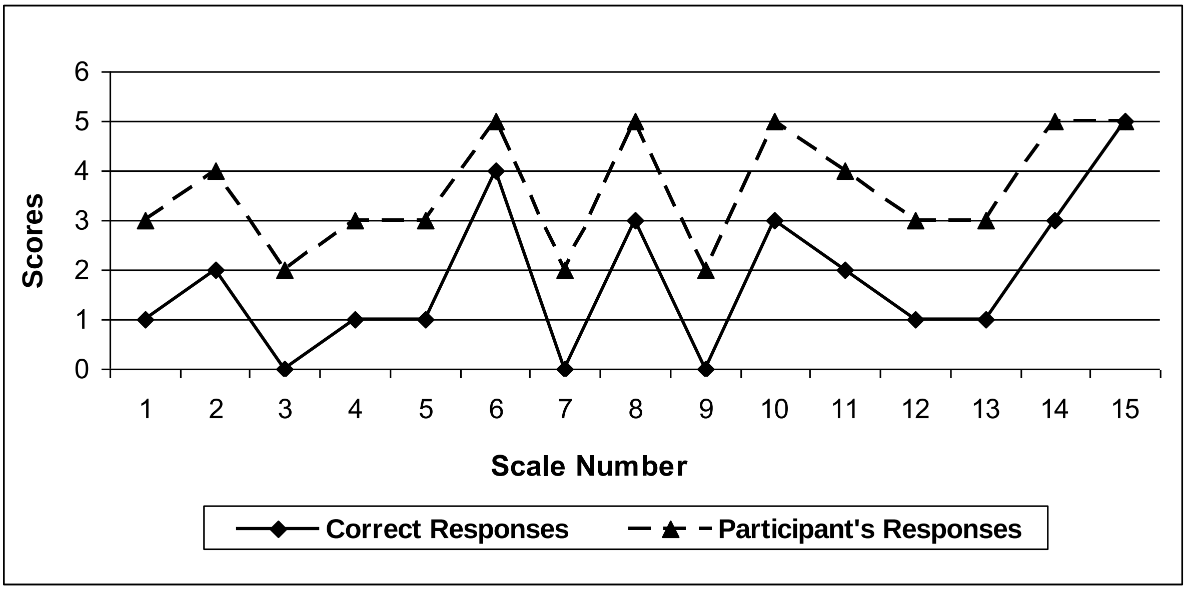

Fig. 2. Hypothetical responses of Participant 2. *D* = 27, *C* = 0.99, *S* = 1.80.

Therefore, it is important to introduce two different indices. The first indicates the accuracy of recognition of various emotion types that constitute the emotional state of the observed person. The second indicates the observer's sensitivity to the intensity of the observed emotions. The Spearman's correlation between participant responses and the correct responses can be used as the first index; it is designated by *C*. The sensitivity index, *S*, is calculated as the sum of deviations of the participant responses from the correct responses on each scale; unlike in the formula for the *D* index, signs of the deviations are taken into account. Dividing this sum by the number of scales would result in putting its range into limits defined by the number of points of the chosen Likert scales. The *S* index can be defined as

$$S = \Sigma(Q_i - R_i)/m$$

where Q_i is a participant's response on Scale_i, R_i is a correct response on the same scale, *m* is the number of scales, i.e., of emotion categories used for assessment.

The theoretical range of the *C* index would lie within the limits of -1 and +1. The maximum value +1 would mean the perfect accuracy of recognition of emotion types constituting the emotional state. A zero value would mean that the participant is thoroughly inaccurate in recognition. The negative values of the *C* index can emerge only in special cases such as if the participant misunderstood the instructions. The theoretical range of the *S* index would lie within the limits defined by the chosen Likert scales. The *S* values would be positive if the participant overestimates the intensity of the observed person emotions, and negative if the participant underestimates the intensity of emotions. The indices *C* and *S* are mathematically independent which allows assessing accuracy and sensitivity as two independent aspects of emotion recognition. It can be illustrated by the hypothetical data presented in Fig. 2 where *C* = 0.99 which means a nearly perfect accuracy in emotion recognition; however, *S* = 1.80 which means the obvious tendency to overestimate emotion intensity, raising its estimates on almost two scores out of possible five. The hypothetical responses presented in Fig. 3 show the opposite case. Participant 3 is extremely inaccurate, actually, anti-accurate in emotion recognition (*C* = -0.89), however, this participant does not over-estimate or under-estimate the emotion intensity (*S* = -0.07). It is remarkable that the *D* values are identical for Participants 2 and 3. Their emotion recognition ability could be erroneously regarded as similar without the use of the suggested index of accuracy (*C*) and index of sensitivity (*S*). Nevertheless, the data presented in Fig. 2 and 3 clearly show that there are two different ability structures beyond the identical *D* indices.

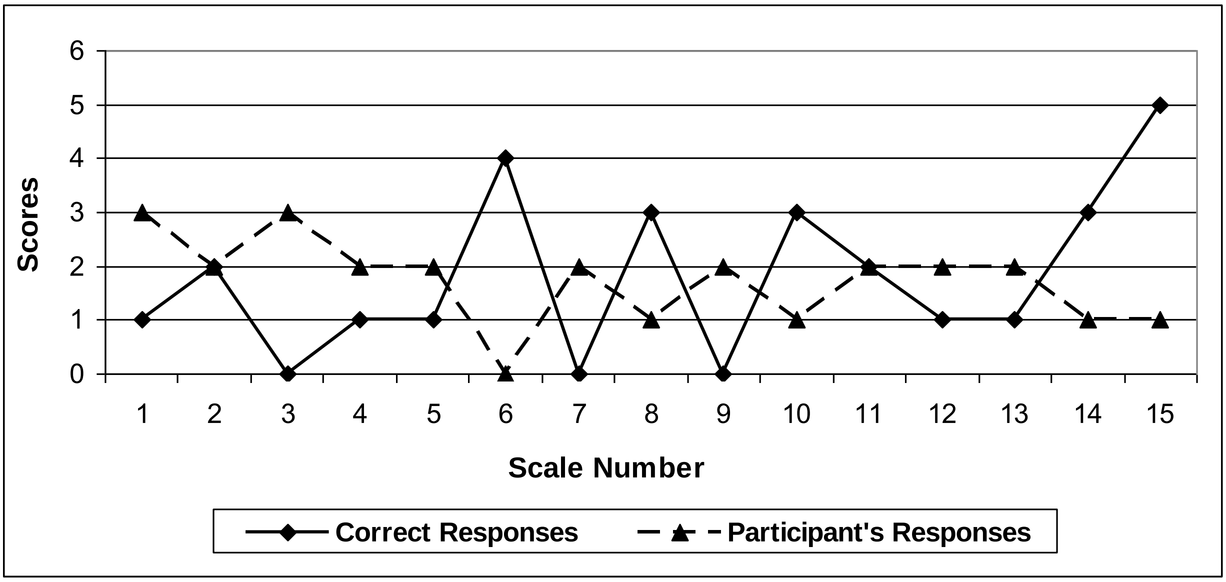

Fig. 3. Hypothetical responses of Participant 3. *D* = 27, *C* = -0.83, *S* = -0.07.

The Videotest of Emotion Recognition

The review of modern emotion recognition measures revealed their typical limitations. The present study aims to develop a new emotion recognition test that would overcome some of them. First, the stimuli used in emotion recognition assessments often lack ecological validity. In real life, people rarely identify emotions on the basis of separate aspects of human behavior, such as only mimics or voice sound. Most typically, human behavior is perceived holistically and within a certain situation. A rich diversity of information sources is usually involved in emotion recognition. The knowledge of the situational context allows the understanding of factors influencing the person, rules restricting or prescribing possible emotion expressions, etc. Which is why we decided to use video recordings showing various aspects of the character's behavior including facial expressions, movements, speech, and the situational context that should be understandable at least roughly. To further improve the ecological validity of the stimuli, excerpts from the natural behavior should be presented as video recordings. Emotional behavior portrayed by the actors should not be used, since their emotional expressions are often either exaggerated or too much tuned to the cultural standards which makes them not natural enough.

Secondly, as demonstrated above, many limitations of the existing measures can be overcome by using the multidimensional response format that gives the participants the opportunity to estimate the intensity of different emotions in the stimulus.

Thirdly, we wanted to use two indices of emotion recognition, the *C* index (accuracy of recognition of various emotion types in the stimulus) and the *S* index (sensitivity to the intensity of emotions in the stimulus) that assess two different aspects of emotion recognition.

In line with these ideas, the Videotest of Emotion Recognition has been developed. Video recordings of natural behavior in various real-life situations were taken as stimuli. The selection of these recording was based on the following criteria.

1. Each video recording must represent human behavior in natural situations, rather than in a laboratory setting.

2. The target person in the video must be in some emotional state. However, this emotional state should not be too intense in order not to make it obvious which emotion is being experienced. We also avoided test items with intense emotions because they could be so simple that the variability of the responses would be too low.

3. Diverse types of information must be available from the videos, including facial expressions, movements, speech, and reactions of other people. The situational context of the behavior should also be comprehensible at least in a general way.

The video recordings were between 10 and 60 seconds long, the targets were both males and females.

Participants had to assess the characters' emotional states with a set of 15 scales representing different emotion categories. The categories were selected in the pilot study so that they corresponded to the characters' emotions. The selection procedure is described in detail in Ovsyannikova (2007). Each scale is a unipolar Likert-type six-point scale with points from 0 to 5, where '0' means that this emotion category does not correspond to the character's state at all, ‘1’ means that this emotion category corresponds to the character’s state minimally, ‘5’ means that this emotion category describes the character’s emotional state perfectly. The list of the scales is presented in Fig. 4.

Anger	0 1 2 3 4 5
Relaxation	0 1 2 3 4 5
Surprise	0 1 2 3 4 5
Contempt	0 1 2 3 4 5
Shame	0 1 2 3 4 5
Anxiety	0 1 2 3 4 5
Disgust	0 1 2 3 4 5
Interest	0 1 2 3 4 5
Displeasure	0 1 2 3 4 5
Arousal	0 1 2 3 4 5
Suffering	0 1 2 3 4 5
Happiness	0 1 2 3 4 5
Fear	0 1 2 3 4 5
Calmness	0 1 2 3 4 5
Guilt	0 1 2 3 4 5

Fig 4. List of the Videotest scales.

The Videotest consists of seven video recordings selected from a large number of recordings on the basis of judges’ estimates. Judges were seven counseling psychologists with more than ten years of professional experience. The judges assessed the target emotional states using the set of fifteen scales described above. The internal consistencies of their estimates of each recording were assessed with Cronbach’s alphas. The recordings selected for the final version had alphas in the range from .82 to .95. For each scale in each recording the medians of the judges’ estimates were calculated. It yielded standard estimates that were considered correct responses.

The testing procedure consists of playing the video recordings in a fixed order. Before each recording, the testee is informed who the target is. After each recording, the testee assesses the character’s emotional state by using the set of fifteen scales. Two indices of emotion recognition ability are calculated, accuracy of recognition of various emotion types in the stimulus (*C* index) and sensitivity to the intensity of emotions in the stimulus (*S* index).

The Videotest of Emotion Recognition as well as some other ability and personality measures were administered to a rather large sample (N = 645). We expected the two suggested indices of accuracy and sensitivity (1) to be independent or, at least, not highly intercorrelated and (2) to yield different correlation patterns with other psychological measures. Such a result would confirm our understanding of accuracy and sensitivity as two different aspects of emotion recognition ability.

Method

Participants

A total of 645 of young adults (427 female), with an average age of 21.1 (*SD* = 5.5), participated in the study. They were undergraduate students, high school students, and adults of different occupations.

Measures and procedure

All participants were administered the Videotest of Emotion Recognition. In addition, subsamples of different sizes completed two emotional intelligence measures, an intelligence test, and two personality questionnaires.

Emotional intelligence measures

The first measure was the Russian adaptation of the Emotion Perception branch of the MSCEIT that consists of the Faces and Pictures subtests (Mayer, Salovey, & Caruso, 2002; Sergienko & Vetrova, 2010). It was chosen because it measures practically the same construct as the Videotest does. Emotionally laden stimuli, such as faces, landscapes, and geometric designs, were administered to 45 participants who had to assess which emotions were present in these stimuli. The second measure was the EmIn Questionnaire, a Russian self-report measure of emotional intelligence that allows for the assessment of people's beliefs about their emotional abilities (Lyusin, 2006a, 2006b). It consists of 46 items with 4-point Likert scale response format, from “completely disagree” to “completely agree”. These items form four scales: Interpersonal EI (e.g., “I understand other people’s inner states without words”), Intrapersonal EI (e.g., “I know what to do to improve my mood”), Emotion Comprehension (e.g., “Often, I can’t find the words to describe my feelings to my friends”), and Emotion Management (e.g., “If I hurt somebody’s feelings, I don’t know how to restore the relationship with them”). The EmIn Questionnaire was completed by 239 participants.

Intelligence measure

Two-hundred and thirty participants completed the Raven’s Advanced Progressive Matrices (with a 40 minute time limit) as a measure of general intelligence (Raven, Raven, & Court, 1998).

Personality questionnaires

The Russian adapted version of the NEO Five-Factor Inventory (Costa & McCrae, 1989) was used as a measure of personality traits (41 participants). Also, the Russian adapted version of Mehrabian and Epstein's Questionnaire Measure of Emotional Empathy (Mehrabian & Epstein, 1972; Tutushkina, 1996) was completed by 55 participants.

The Videotest of Emotion Recognition, the Emotion Perception subtests of the MSCEIT, and Raven's Advanced Progressive Matrices were administered individually. The questionnaires were administered either individually, or in small groups.

Results

Descriptive statistics for the accuracy and sensitivity indices are presented in Table 2. Since the distributions of both indices did not match the normal distribution (Kolmogorov-Smirnov's test), nonparametric statistical tests were used in further data analysis.

Table 2

Descriptive Statistics for the Videotest Indices.

	Accuracy Index (*C*)	Sensitivity Index (*S*)
Mean	0.61	0.20
Standard Deviation	0.15	0.42
Skewness	-1.29 (SE = 0.10)	0.55 (SE = 0.10)
Kurtosis	2.35 (SE = 0.19)	0.14 (SE = 0.19)
Minimum	-0.04	-0.73
Maximum	0.88	1.56

To assess the reliability of the accuracy and sensitivity indices, the internal consistency and test-retest reliability coefficients were calculated (Table 3). The reliability of *C* index is lower than for the *S* index; in general, however, the reliability of both scores is satisfactory.

Table 3

Reliability of the Videotest Indices.

Reliability Measures	n	Accuracy Index (*C*)	Sensitivity Index (*S*)
Internal consistency (Cronbach's α)	645	.74	.93
Test–retest reliability (Spearman's correlation)	48	.55	.86

The reliability coefficients are comparable to those of other emotional intelligence tests. For example, Cronbach's alphas reported for the Emotion Perception branch of the MSCEIT, the most similar by content to the Videotest, were .68 for the Faces subtest and .80 for the Pictures subtest (Roberts et al., 2006). The authors of the MSCEIT (Mayer et al., 2003) obtained higher internal consistency indices for these subtests, .82 and .87 respectively. Cronbach's alpha for the STEU is .71 (MacCann & Roberts, 2008); it varies from .86 to .92 for the JACBART (Matsumoto et al., 2000).

The Spearman's correlation between the accuracy and sensitivity scores is -.14 ($p < .01$). Due to the large sample, this correlation is statistically significant, but very low. It gives grounds to claim that the accuracy and sensitivity indices reflect different aspects of emotion recognition.

Gender differences for the Videotest scores were analyzed by using the Mann-Whitney U test; the results are shown in Table 4. Women were more accurate than men in recognition emotion types, mean *C* indices are 0.63 and 0.58, respectively. There were no gender differences in sensitivity to the intensity of emotions.

Table 4

Sex Differences for the Videotest Indices.

	Accuracy Index (*C*)	Sensitivity Index (*S*)
Men (N = 218)	0.58	0.22
Women (N = 427)	0.63	0.19
Mann-Whitney's U	37530.50 ($p < .001$)	44871.00 ($p = .460$)

To assess the validity of the Videotest, Spearman's correlations of its indices with emotional intelligence, general intelligence and personality traits were calculated. We expected that the accuracy and sensitivity indices would yield different correlation patterns with other cognitive and personality variables. The results of this analysis are presented in Table 5.

Table 5

Spearman's Correlations Between Videotest Indices and Cognitive and Personality Variables.

Measure	N	Subtest or Scale	Accuracy Index (*C*)	Sensitivity Index (*S*)
MSCEIT	45	Emotion Perception: Faces	.40**	-.21
		Emotion Perception: Pictures	.12	-.34*
		Emotion Perception: Total Score	.37*	-.32*
EmIn Questionnaire	239	Interpersonal EI	.16*	.15*
		Intrapersonal EI	.09	.04
		Emotion Comprehension	.13*	.10
		Emotion Management	.11	.08
Raven's APM	230	Total score	-.09	.08
NEO Five-Factor Inventory	41	Neuroticism	.10	-.11
		Extraversion	.07	.20
		Openness	.11	.08
		Agreeableness	.22	-.21
		Conscientiousness	.07	0.00
Questionnaire Measure of Emotional Empathy	55		.25†	0.27*

Note. † p < .10, * p < .05, ** p < .01.

The accuracy and sensitivity indices correlate with the Emotion Perception branch of the MSCEIT in different ways: accuracy is related positively to Emotion Perception (r = .37, p < .05), whereas sensitivity provided a negative correlation with Emotion Perception (r = -.32, p < .05). Both indices positively correlate yet with the scales 'Interpersonal Emotional Intelligence' and 'Emotion Comprehension' of the EmIn Questionnaire. This result suggests that self-confidence in the field of emotion understanding and management is associated with the greater recognition accuracy of the general profile of other people's emotional states and also with the overestimation of the intensity of other people's emotions. Positive correlations were also found between both Videotest indices and empathy. General intelligence and personality traits did not provide any statistically significant correlations with the Videotest indices.

Discussion

This study provides evidence supporting the possibility and necessity of distinguishing two different indices of emotion recognition, namely accuracy and sensitivity. On the one hand, reliability coefficients of these two indices are satisfactory. On the other hand, being independent, they measure two different aspects of emotion recognition. Their independence was confirmed by three facts. First, their intercorrelation is very low (r = .14). Secondly, sex differences were found for *C* index (it is significantly higher for women), but not for *S* index. Finally, indices of accuracy and sensitivity provided different correlation patterns with other individual characteristics. The most dramatic difference was obtained in correlations with Emotion Perception subtest of the MSCEIT; the accuracy index has a positive correlation (r = .37), whereas the sensitivity index has a negative correlation (r = -.32).

The analysis of the relations between the accuracy and sensitivity indices and other variables show the existence of positive correlations with certain scales of the emotional intelligence questionnaire and with the empathy questionnaire. However, no significant correlations with general intelligence and Big Five personality traits have been found. Let us compare these results with the evidence obtained in other studies that used different emotion recognition measures. Traditionally, the construct validity of emotional intelligence tests is assessed by their correlations with intelligence tests and personality traits questionnaires. The largest body of evidence concerns the MSCEIT. The subtest scores and the total score of the MSCEIT provide low or moderate correlations with intellectual abilities scores. The relations between the Emotion Understanding subtest and crystallized intelligence, in particular, verbal intelligence are the most stable (Roberts et al., 2006). Many studies report positive, though low, correlations between the total score of the MSCEIT and GPA (r = .16, p < .05, for Brackett & Mayer, 2003). The meta-analysis conducted by Roberts et al. (Roberts, Schulze, & MacCann, 2008) showed that MSCEIT scores yield almost no correlations with personality traits, with the exception of the Emotion Management subtest that correlates positively with Agreeableness. It partially agrees with our results, since among all Big Five scales, Agreeableness has the highest correlations with the Videotest scores (.22 for accuracy and -.21 for sensitivity). These correlations are not statistically significant, however, they were obtained on a small sample (N=41) and may prove to be significant on larger samples. The MSCEIT scores also provide low positive correlations with some other emotion recognition tests, e.g., correlations with the JACBART scores range from .03 to .18 (Mayer, Salovey, & Caruso, 2012).

The main discrepancy between our results and those described above is the absence of correlations with general intelligence. It calls into question the cognitive nature of the constructs measured by the Videotest, since it is usually expected that cognitive abilities tests should correlate with each other. More precisely, it could be suggested that those cognitive processes that account for the general intelligence level do not play an essential role in emotion recognition. Indirectly, this suggestion is supported by the evidence that the relation between the MSCEIT subtest ‘Emotion Understanding’ and general intelligence is the most reproducible in different studies. The material of this subtest is thoroughly verbal, therefore, it uses mostly verbal abilities of the participant. However, the Emotion Perception subtest (the least verbal in the MSCEIT and the most similar to the Videotest in this sense) does not provide any stable relations with general intelligence.

Another possible explanation for the absence of relations between the Videotest and intelligence scores could be the response format of test items. According to Roberts and MacCann’s evidence (MacCann & Roberts, 2008), correlations between emotional intelligence and general intelligence depend on the response format of the emotional intelligence test items. Items with the same content provide higher correlations with general intelligence if a multiple-choice response format is used instead of Likert-scales format. This regularity holds for any emotional abilities including emotion understanding and emotion management. Modification of the Videotest with multiple-choice items could provide higher correlations with intelligence tests.

The Videotest of Emotion Recognition described in this paper can be further developed and improved, first of all, through the selection of a larger set of the video recordings that would represent a more diverse array of emotional states. Another direction of future research could be the use of the two suggested indices of emotion recognition reflecting its two aspects, accuracy and sensitivity, in other emotion abilities measures.

References

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.

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

Costa, P. T., & McCrae, R. R. (1989). *The NEO-PII/NEO-FFI manual supplement*. Odessa, FL: Psychological Assessment Resources.

Legree, P. J., Psotka, J., Tremble, T., & Bourne, D. R. (2005). Using consensus based measurement to assess emotional intelligence. In R. Schulze, & R. D. Roberts (Eds.), *Emotional intelligence: An international handbook* (pp. 155–179). Cambridge, MA: Hogrefe.

Lyusin, D. (2006a). Emotional intelligence as a mixed construct: Its relation to personality and gender. *Journal of Russian and East European Psychology, 44*, 54 - 68.

Lyusin, D. (2006b). A new measure for emotional intelligence: EmIn Questionnaire. *Psikhologicheskaya Diagnostika*, 3-22. (In Russian)

MacCann, C., & Roberts, R. D. (2008). New Paradigms for Assessing Emotional Intelligence: Theory and Data. *Emotion, 8*, 540–551.

Matsumoto, D., LeRoux, J. A., Wilson-Cohn, C., Raroque, J., Kooken, K., Ekman, P., Yrizarry, N., Loewinger, S., Uchida, H., Yee, A., Amo, L., & 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.

Mayer, J. D., Salovey, P., & Caruso, D. (2000). Competing models of emotional intelligence. In R. J. Sternberg (Ed.), *Handbook of human intelligence* (2nd ed.) (pp. 396–420). New York: Cambridge University Press.

Mayer, J. D., Salovey, P., & Caruso, D. R. (2012). The validity of the MSCEIT: Additional Analyses and Evidence. *Emotion Review, 4*, 403–408.

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

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

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

Mehrabian, A., & Epstein, N. (1972). A measure of emotional empathy. *Journal of Personality, 40*, 525-543.

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.

Ovsyannikova, V. (2007). *Role of cognitive factors in emotional states recognition.* Doctoral thesis. Institute of Psychology of the Russian Academy of Sciences, Moscow. (In Russian)

Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's Progressive Matrices and Vocabulary Scales: Section 4. The Advanced Progressive Matrices.* San Antonio, TX: Harcourt Assessment.

Reis, H. T., & Judd C. M. (Eds.). (2000). *Handbook of Research Methods in Social and Personality Psychology*. New York: Cambridge University Press.

Roberts, R. D., Schulze, R., & MacCann, C. (2008). The measurement of emotional intelligence: A decade of progress? In G. Boyle, G. Matthews, & D. Saklofske (Eds.), *The Sage handbook of personality theory and assessment* (pp. 461–482). New York: Sage.

Roberts, R. D., Schulze, R., O'Brien, K., MacCann, C., Reid, J., & Maul, A. (2006). Exploring the Validity of the Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT) with Established Emotions Measures. *Emotion, 6,* 663–669.

Roseman, I. J. (2001). A model of appraisal in the emotion system: Integrating theory, research, and applications. In K. R. Scherer & A. Schorr (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 68–91). New York: Oxford University Press.

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

Salovey, P., & Grewal, D. (2005). The science of emotional intelligence. *Current Directions in Psychological Science*, *14,* 281–285.

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.

Sergienko, E. A., & Vetrova, I. I. (2010). *J. Mayer, P. Salovey, and D. Caruso's emotional intelligence test (MSCEIT v. 2.0). Manual.* Moscow: Institute of Psychology of the Russian Academy of Sciences. (In Russian)

Tutushkina, M. K. (Ed.). (1996). *Practical psychology for managers.* Moscow: Filin. (In Russian)

Warwick, J., Nettelbeck, T., & Ward, L. (2010). AEIM: A new measure and method of scoring abilities-based emotional intelligence. *Personality and Individual Differences, 48,* 66–71.

This Working Paper is an output of a research project implemented as part of the Basic Research Program at the National Research University Higher School of Economics (HSE). Any opinions or claims contained in this Working Paper do not necessarily reflect the views of HSE.

Authors
Dmitry Lyusin
National Research University Higher School of Economics (Moscow, Russia), Scientific-Educational Laboratory for Cognitive Research.
Leading Research Fellow
E-mail: ooch@mail.ru, tel. +7 9152493908

Victoria Ovsyannikova
National Research University Higher School of Economics (Moscow, Russia), Scientific-Educational Laboratory for Cognitive Research.
Senior Research Fellow
E-mail: v.ovsyannikova@gmail.com, tel. +7 9260425345

Any opinions or claims contained in this Working Paper do not necessarily reflect the views of HSE.

© Lyusin, Ovsyannikova, 2014