Multistudy Report

It Works Without Words

A Nonlinguistic Ability Test of Perceiving Emotions With Job-Related Consequences

Gerhard Blickle[1], Iris Kranefeld[1], Andreas Wihler[2], Bastian P. Kückelhaus[1], and Jochen I. Menges[3]

[1]Institute of Psychology, University of Bonn, Germany
[2]Business School, University of Exeter, UK
[3]Department of Business Administration, University of Zurich, Switzerland

Abstract: Emotion recognition ability of emotions expressed by other people (ERA-O) can be important for job performance, leadership, bargaining, and career success. Traditional personnel assessment tools of this ability, however, are contaminated by linguistic skills. In a time of global work migration, more and more people speak a language at work that is not their mother tongue. Consequently, we developed and validated the Face-Based Emotion Matching Test (FEMT), a nonlinguistic objective test of ERA-O in gainfully employed adults. We demonstrate the FEMT's validity with psychological constructs (cognitive and emotional intelligence, Big Five personality traits) and its criterion validity and interethnic fit.

Keywords: emotion recognition ability, GMA, emotional intelligence, Big Five, social astuteness, adaptive performance

Individuals who can accurately recognize the emotions in others may sometimes enjoy benefits in everyday life, for example, well-being during the COVID-19 pandemic (Schlegel et al., 2021). Such emotion recognition ability of emotions expressed by other people (ERA-O) can also be vital for job performance (e.g., Elfenbein & Ambady, 2002; for opposing results, see Joseph & Newman, 2010), negotiations (Elfenbein et al., 2007), leadership (Rubin et al., 2005), and career outcomes (Momm et al., 2015). Computer experts and labor-market economists expect that, although computer algorithms allow many non-routine tasks to be automated, jobs that demand complex social perception and emotional intelligence are unlikely to be managed by computers in the next two decades (Frey & Osborne, 2017).

Individuals differ in how accurately they can detect and decode the emotion expressions in others (Elfenbein et al., 2002; Schlegel et al., 2017). ERA-O is a basic part of a dimension of emotional intelligence (i.e., perceiving emotions), which is defined as a general set of abilities required to process emotions adaptively, that is, perceiving, understanding, and managing emotions and facilitating thoughts (Fiori & Vesely-Maillefer, 2018). A recent meta-analysis supports the notion that emotion recognition ability and other dimensions of emotional intelligence are within the Cattell-Horn-Carroll model of intelligence (Bryan & Mayer, 2020).

Researchers have used a variety of objective tests to assess ERA-O with implications for work outcomes and job performance (e.g., Diagnostic Analysis of Nonverbal Accuracy [DANVA]; Nowicki & Duke, 2001; Geneva Emotion Recognition Test [GERT]; Schlegel et al., 2014, Schlegel & Mortillaro, 2019). These tests usually consist of a presentation of an emotional expression (in a photograph, video clip, or audio recording) and then a request for the participants to label the emotional expression. For example, test takers viewing a picture of a smiling face are supposed to identify the displayed emotion as "happiness." The extent to which test takers accurately name the displayed emotion across a number of items indicates their level of ERA-O. Objective tests developed to assess ERA-O in the context of the broader construct of emotional intelligence usually follow a similar procedure (e.g., Mayer-Salovey-Caruso Emotional Intelligence Test [MSCEIT]; Mayer et al., 2002).

Individual differences in ERA-O scores diagnosed using these objective tests may result from different mechanisms: First, people may differ in the extent to which they *perceive*

the emotional expression of another person. The perception of an emotion is a basic process that involves the activation of neural areas that correspond to the expressed emotion (Enticott et al., 2008); this process requires no linguistic skills and no prior knowledge of verbal terms for emotions. For example, perceiving a smile can induce a sense of happiness even if the perceived emotion is not labeled as "happiness." Second, people may differ in the ability to *label* the emotional expression they perceive in another person. The labeling of an emotion is a cognitively demanding process that requires linguistic abilities and a vocabulary of emotion terms: People have to match a perceived emotion expression with the appropriate linguistic label.

Perceptual emotion recognition abilities can be well-developed even when the relevant knowledge or vocabulary for naming those emotions is absent (Sauter et al., 2011). Some researchers distinguish between the basic ability to process emotions and the more advanced ability to label emotions (Borod et al., 2000). And more recently, tasks for measuring the nonlinguistic perception and recognition of facial expressions of emotions have been developed (Herzmann et al., 2008; Palermo et al., 2013). These tasks separate two distinct but related dimensions: first, the *nonlinguistic* dimension of emotion recognition, which operates outside the bounds of verbal cognition, and second, the elaborative, *linguistic* dimension of emotion understanding (Sauter et al., 2011), which involves the linking of an emotion expression perceived in others (e.g., a smile) with the right label for that expression (e.g., "happy").

However, this new approach has been mainly used in clinical populations or basic research (Palermo et al., 2013) but it has not yet been applied to the job performance of workers. Thus, we have developed a new objective nonlinguistic ability test (as opposed to a self-report emotion perception measure) that can be completed in a short-time, while demonstrating construct- and work-related criterion validity. Consequently, we merge the different lines of research in basic and applied psychology in the context of workplace behavior. In a time of global work migration, more and more people speak a language at work that is not their mother tongue (International Organization for Migration, 2019). Hence, traditional tests of emotion recognition ability rely too much on vocabulary and linguistic skills and suffer from a linguistic bias.

To counterbalance this linguistic bias, we apply a new approach to the outcomes of emotion recognition ability at work: We propose that ERA-O can be measured through a testing procedure that requires workers to match an emotional expression with another emotional expression (e.g., a smiling face is matched with another smiling face) rather than to match an emotion expression with a label for that expression (e.g., a smiling face is matched with the word "happy"). If employees can consistently match expressions

Figure 1. Two sample FEMT items. Subjects were asked whether the two presented images within each pair featured the *same emotion*, *somewhat similar emotions*, *somewhat different,* or *different emotions*, or to use the button *don't know.*

for the same emotion and distinguish different emotions, then this can be taken as an indication that they can recognize emotions accurately in a nonlinguistic way, independent of whether they would also be able to linguistically label the matched emotion accurately. This testing paradigm can effectively capture nonlinguistic ERA-O (see Figures 1 and 2).

Our paper seeks to address the following research questions:

> *Research Questions*: Does a new nonlinguistic measure of emotion recognition ability for faces designed for workers meet the standards of reliability and factorial, construct, and criterion validity, and will it display interethnic fit?

Thus, our studies contribute to the literature of emotional intelligence in general and ERA-O at work specifically.
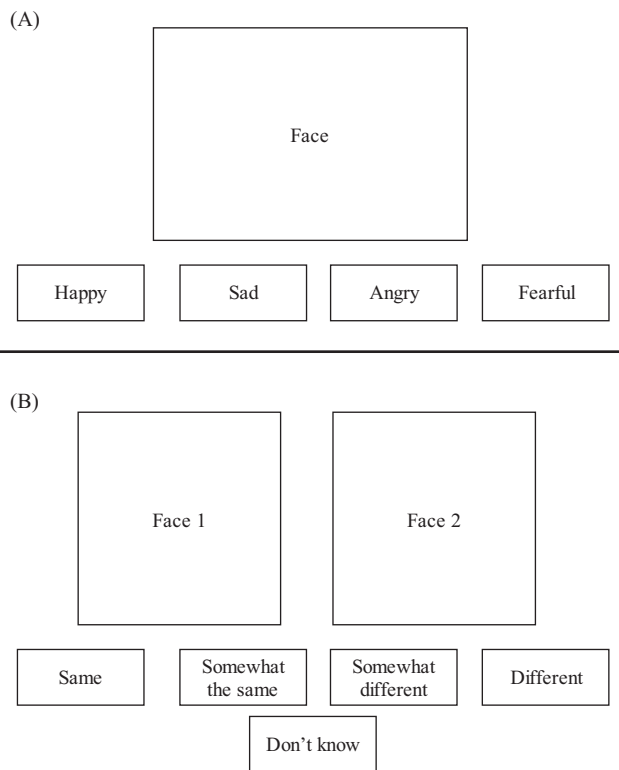
(A)

Face

| Happy | Sad | Angry | Fearful |

(B)

| Face 1 | Face 2 |

| Same | Somewhat the same | Somewhat different | Different |

Don't know

Figure 2. Traditional ERA-O Items (A) Compared with FEMT Items (B).

First, we enhance the traditional understanding of emotional abilities. We apply a new conceptual model of ERA-O that represents a nonlinguistic (i.e., language-independent; see Sauter et al., 2011) dimension, by distinguishing between nonlinguistic automatic emotion recognition abilities and controlled linguistic abilities that are linked to the labeling of emotions (Fiori, 2009). Second, we designed, validated, and applied a new instrument to assess ERA-O as a nonlinguistic ability – the Face-Based Emotion Matching Test (FEMT) with a duration of less than 6 min. This test, the first of its kind for adults in organizations and at work, offers a methodological and practical contribution to those who would like to apply an ERA-O measure that is independent of people's emotional vocabulary. We developed and validated the FEMT in studies with workplace samples (i.e., in samples with study participants who were active in gainful employment) in order to optimize the fit between the new test and people in the workplace. Our findings are relevant to the workplace because the FEMT scores are associated with social astuteness and adaptive performance. Finally, we expect that there will be an increased demand in personnel assessment of the ability for social perception in order to staff positions that demand complex social perception and emotional intelligence (Frey & Osborne, 2017).

Overview of the Studies

In this paper, based on Simmons et al. (2012), we present two studies concerning the development, construct and criterion validation, and interethnic fit of the new measure of nonlinguistic ERA-O (for data and codes, see Blickle et al., 2021, https://osf.io/sxtr3/?view_only=3a71cd86c6c64b7d-b9799bf33c1087cd). We report in the repository of the OSF link how we determined our sample size, all data exclusions (if any), all data inclusion/exclusion criteria, whether inclusion/exclusion criteria were established prior to data analysis, all measures in the study, and all analyses including all tested models. If we use inferential tests, we report exact p values, effect sizes, and 95% confidence or credible intervals.

Although emotions are conveyed not only through facial expressions but also through body posture and voice, for the development of the FEMT we focused solely on facial expressions because research has suggested that the face offers the most direct cues for felt emotions: Emotions displayed by faces are easier for people to grasp than emotions expressed vocally or through bodily movements (Momm et al., 2010).

In Study 1, we report on the development of the FEMT. In Study 2, we examined the validity of the new test in four additional non-overlapping samples. In Sample 1 (S1) of Study 2, we examine the construct validity of the new measure with reference to traditional measures of ERA-O such as the DANVA – Faces and Postures (Nowicki & Duke, 2001) and the MSCEIT (Mayer et al., 2002). In S2, we examine the relations between the FEMT and general cognitive ability using the General Aptitude Test Battery (GATB; Weiss, 1972). In addition, we test the relations between the FEMT and the Big Five personality traits (NEO-Five Factor Inventory [NEO-FFI]; Borkenau & Ostendorf, 1993). In S3, we examine the criterion validity of the new test and test its relations with coworker assessments of task performance, social astuteness, and adaptive performance. Finally, in S4, we compared Japanese (Matsumoto & Ekman, 1988) and Caucasian (Lundqvist et al., 1998) face stimuli and tested whether the FEMT would differently predict (or not) Western European employees' ability to perceive emotions better in Caucasian than in a sample of Japanese face stimuli.

For all samples used in Studies 1 and 2, university students recruited people active in the German labor force in partial fulfillment of their course requirements. All study participants were working adults; recruiters were instructed not to recruit school students, university students (including part-time students), retirees, or jobless persons. Seventy-six percent of the study participants were full-time workers.

Study 1

The test development comprised three steps: First, we generated items following our nonlinguistic approach by asking participants to match adult faces that display the same emotion and distinguish between adult faces that display different emotions. Second, we selected test items. Third, we validated the factorial validity of the new ERA-O item set in a merged overall sample of 1,567 study participants and conducted reliability and measurement invariance analyses. The larger the sample, the more stable the estimated relations (Schönbrodt & Perugini, 2013).

We developed 32 pairs of facial emotion expressions. Each pair comprised two images that we took from a conceptually and empirically validated database, the Radboud Faces Database (Langner et al., 2010), thereby ensuring the validity of our items. The database included images of models displaying facial expressions of eight emotions, adopting three gaze directions (frontal, left, right), and showing five head orientations (180°, 135°, 90°, 45°, 0°). We selected emotion comparisons that were intended to optimally assess individual differences, that is, approximately achieve a normal distribution of test results (Langner et al., 2010, Figure 4).

We created 16 pairs of images featuring the same emotions and 16 pairs featuring different emotions (Figure 1). In order to prevent influences on the test results resulting from the specific model as a person or the model's gender, head orientation, or gaze direction, we varied the gender, head orientations, and gaze directions of the models as distracting features, and used pictures of different persons displaying emotions. We combined, in pairs, the facial expressions of the following emotions which have a higher rate of misclassification according to previous research (Langner et al., 2010, Figure 4; Scherer & Ellgring, 2007, Table 5): contempt versus neutral, contempt versus disgust, disgust versus anger, surprise versus fear, anger versus neutral, sadness versus fear, sadness versus contempt, and contempt versus anger.

In contrast to traditional ERA-O tests, participants were not asked to linguistically label the expressed emotion but were instead asked to assess whether for each pair the two presented images featured: *the same emotion*, *somewhat the same emotion*, *somewhat different emotions*, or *different emotions* (see Figure 2). They could also use the button marked *don't know*. As the emotion expressions used in the FEMT had been picked from a validated database, they could clearly be categorized as either representing the same emotion or not. Hence, points were assigned only when participants correctly chose the labels *the same emotion* or *different emotions*. The answer options *somewhat the same emotion*, *somewhat different emotions*, and *don't know* were always classified as incorrect answers. Test takers were not informed about this. Using this sort of response format and coding we intended to minimize hits by chance. Participants were administered the FEMT items online.

Method

Participants and Procedures

Study 1 consisted of 348 working adults (see Table 1). We computed the item difficulty, that is, the average emotion recognition accuracy, for each item (see Table 2) in Study 1 and the overall sample ($N = 1,567$).

To select the specific items making up the scale, we used a confirmatory factor model with two correlated factors, with one factor that covered the items that feature the same emotions (true emotion identification) and one factor that covered the ability to accurately distinguish between adult faces that display different emotions (true emotion discrimination). We dropped all items with standardized loadings on the respective factor below $\lambda = .45$ and retained for each factor the equal number of items, that is, nine items, using Mplus and weighted least square mean and variance adjusted (WLSMV) for estimations based on binary true-false responses (Muthén & Muthén, 2012).

For a test-retest analysis, we invited all participants who had completed the FEMT in the development sample (Study 1) and in S1 and S3 of the validation study to participate again in FEMT testing.

Statistical Analysis

We report the percentage of correct answers on the FEMT and its correlations with the other manifest test scores. We also assessed the relations between the FEMT and the other variables at the construct level beyond measurement error by using structural equation modeling (SEM; Muthén & Muthén, 2012).

We assessed our SEM models' goodness of fit by applying multiple criteria. The criteria for an acceptable fit were: $p(\chi^2) \geq .01$, Root Mean Square Error of Approximation (RMSEA) $< .08$, Comparative Fit Index (CFI) $> .95$, Tucker-Lewis Index (TLI) $> .90$, and Standardized Root Mean Square Residual (SRMR) $< .10$. Schermelleh-Engel et al. (2003, p. 36), however, noted: "The usual test of the null hypothesis of exact fit is invariably false in practical situations and will almost certainly be rejected if sample size is sufficiently large." Therefore, they recommended assessing whether the model fits approximately well in the population. The RMSEA is a measure of approximation in the population. It is relatively independent of sample size and favors parsimonious models. We therefore used RMSEA and p(RMSEA) to assess model fit. An RMSEA $\leq$ .05 indicates good fit with values for p(RMSEA) ranging between $0.10 < p \leq 1.00$ (Schermelleh-Engel et al., 2003).

Table 1. Overview of the sample and study characteristics of all study samples

Sample	Function	Final N of study	Age M	Age SD	Gender Male (%)	Gender Female (%)	Weekly working hours M	Weekly working hours SD	Total work experience M	Total work experience SD	Response rate (%)	Participants not considered for main statistical analyses	Incentives for study participation	Targets with a coworker rating	Design
Study 1															
1	Test development	348	35.43	11.66	39	61	38.67	6.99	–	–	68	48 participants not meeting inclusion criteria[a]	Feedback on ERA-O	–	Online field study
Study 2															
1	Construct validation	344	42.25	12.10	44	56	38.81	9.83	15.99	12.51	52	16 participants not meeting inclusion criteria[a]	Feedback on ERA-O, chance to win Amazon voucher or donation to charity	–	Online field study
2	Construct validation	182	36.95	10.92	46	54	35.35	8.19	14.41	10.32	55	6 participants not meeting inclusion criteria[a]	Feedback on ERA-O and GMA; allowance (€15) or donation to charity	–	Proctored study
3	Criterion validation	310	43.48	10.69	47	53	38.09	9.07	20.69	11.68	61	5 participants not meeting inclusion criteria[a]	Feedback on ERA-O, chance to win Amazon voucher or donation to charity	310 with a coworker rating	Online field study
4	Test of interethnic fit	383	41.60	12.92	38	62	38.15	9.71	19.19	13.04	66	27 participants not meeting inclusion criteria[a]	Feedback on ERA-O	–	Online field study

Note. [a]Participants had to be gainfully working at least 18 hr per week.

Table 2. Test items, item difficulties, and factor loadings in Study 1 and the overall sample

Emotions	Item in FEMT or not	# in item presentation	Test items		Item difficulty		Item loadings	
			Emotion 1	Emotion 2	Study 1	Overall Sample	Study 1	Overall Sample
Same	Yes	6	Happy	Happy	.75	.70	.48	.58
Same	Yes	11	Fearful	Fearful	.68	.68	.53	.61
Same	Yes	14	Neutral	Neutral	.69	.70	.55	.57
Same	Yes	18	Neutral	Neutral	.42	.39	.60	.52
Same	Yes	20	Angry	Angry	.65	.62	.65	.63
Same	Yes	21	Sad	Sad	.53	.54	.74	.66
Same	Yes	25	Angry	Angry	.34	.32	.69	.64
Same	Yes	26	Fearful	Fearful	.60	.66	.62	.63
Same	Yes	30	Disgusted	Disgusted	.64	.66	.58	.60
Same	No	2	Contemptuous	Contemptuous	.16	.13		
Same	No	3	Surprised	Surprised	.47	.44		
Same	No	7	Disgusted	Disgusted	.42	.45		
Same	No	12	Sad	Sad	.56	.55		
Same	No	15	Surprised	Surprised	.63	.59		
Same	No	17	Contemptuous	Contemptuous	.64	.65		
Same	No	22	Happy	Happy	.88	.89		
Different	Yes	8	Fearful	Sad	.62	.54	.54	.62
Different	Yes	9	Angry	Neutral	.73	.69	.45	.61
Different	Yes	16	Contemptuous	Angry	.31	.28	.51	.58
Different	Yes	23	Sad	Contemptuous	.49	.43	.75	.71
Different	Yes	24	Contemptuous	Disgusted	.91	.85	.79	.81
Different	Yes	28	Angry	Neutral	.35	.30	.63	.67
Different	Yes	29	Angry	Contemptuous	.81	.72	.75	.78
Different	Yes	31	Fearful	Sad	.69	.65	.67	.73
Different	Yes	32	Contemptuous	Disgusted	.67	.60	.61	.66
Different	No	1	Contemptuous	Neutral	.19	.15		
Different	No	4	Sad	Contemptuous	.65	.54		
Different	No	5	Disgusted	Angry	.48	.41		
Different	No	10	Disgusted	Angry	.14	.12		
Different	No	13	Surprised	Fearful	.16	.18		
Different	No	19	Surprised	Fearful	.24	.22		
Different	No	27	Neutral	Contemptuous	.04	.03		

Note. Study 1, $N = 348$; overall sample (all study samples combined), $N = 1,567$; standardized item loadings, all $p < .05$. FEMT = Face-Based Emotion Matching Test.

Results

The goodness of fit indices of the model with two correlated factors in Study 1 (Figure 3A) was good: $\chi^2(134) = 153.30$ ($p = .122$), CFI = .98, TLI = .98, SRMR = .07, RMSEA = .02 ($p = 1.00$). The sample size ($N = 348$) was sufficient for stable parameter estimates (Schönbrodt & Perugini, 2013). The two factors correlated at $\rho = .40$ ($p < .01$). The goodness of fit indices of the model with two correlated factors was also good in the overall sample (Figure 3B): $\chi^2(134) = 325.32$ ($p < .001$), CFI = .97, TLI = .97, SRMR = .05, RMSEA = .03 ($p = 1.00$); the two factors correlated at $\rho = .42$ ($p < .01$). Zero values of skewness and kurtosis represent perfectly normal distributions, whereas skewness > ±2 and kurtosis > ±7 are indicative of nonnormal distributions (Curran et al., 1996). The accuracy score based on the 18 items was 57.30% in the overall sample ($SD_{FEMT} = 21.70\%$; skewness$_{FEMT} = -.39$, kurtosis$_{FEMT} = -.17$) indicating a normal distribution. The item difficulties (ID) ranged between $.28 \leq ID \leq .85$ (see Table 2). Cronbach's α of the 18 FEMT items was α = .79.

The αs in the five samples were satisfactory: .76, .72, .78, .78, and .83. A group of 250 persons took the FEMT for a second time. The mean time interval between the first and the second test was 6.19 months ($SD = 1.24$). The test-retest reliability was $r = .69$. The length of the time interval did not moderate the test-retest reliability.
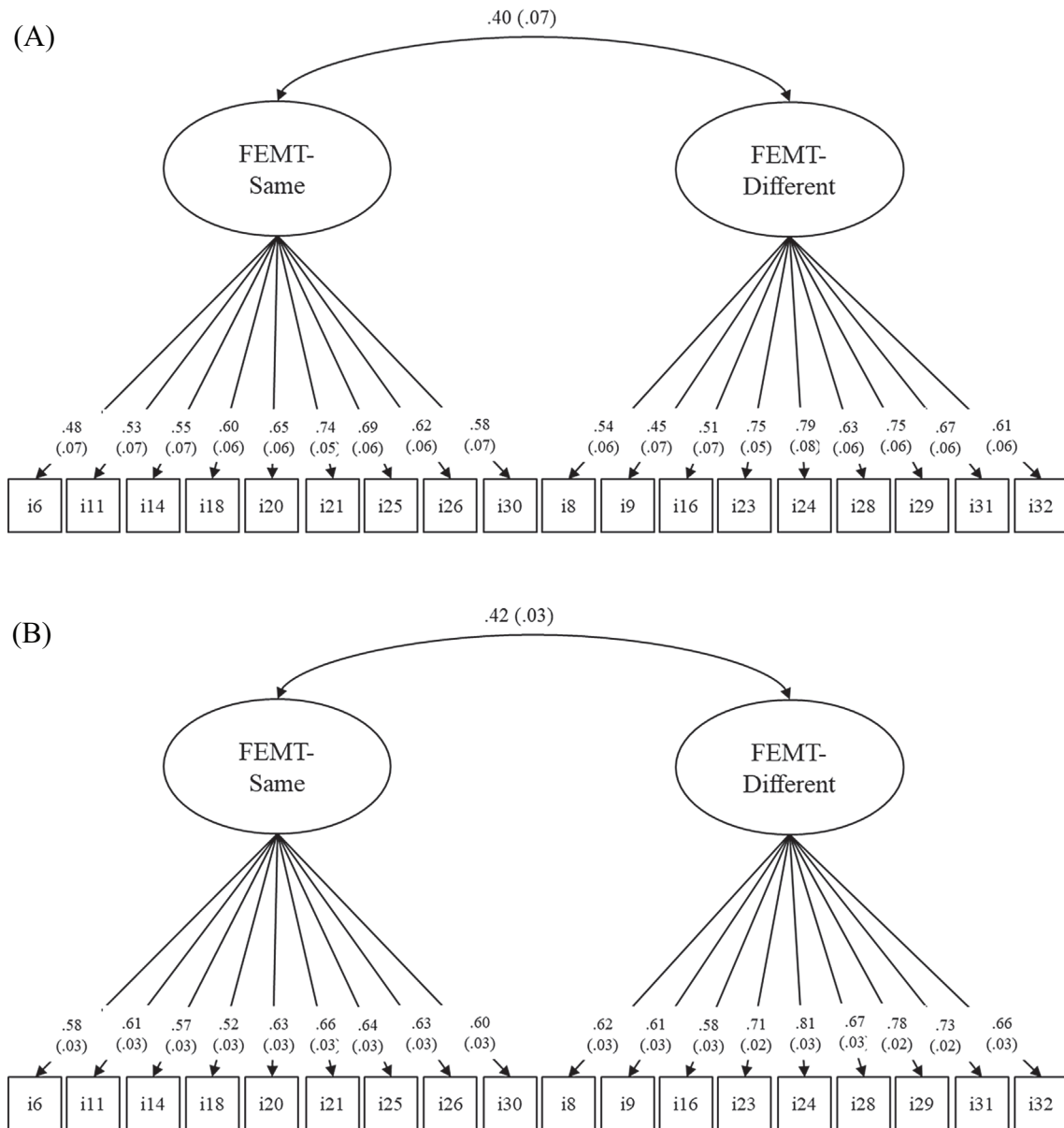
Figure 3. (A) Study 1, $N = 348$; standardized loadings (all $p < .05$) and standard errors. (B) Overall Sample, $N = 1,567$; standardized loadings (all $p < .05$) and standard errors.

Next, we conducted measurement invariance analyses for categorical data (Svetina et al., 2020). Using χ^2-difference tests and ΔRMSEA, we confirmed invariance of slopes and thresholds both when comparing the development sample ($N = 348$) with the other samples ($N = 1,219$; $\Delta\chi^2 = 18.13$, $\Delta df = 16$, $p = .317$, ΔRMSEA $< .01$) and when comparing women ($N = 905$) and men ($N = 604$, $\Delta\chi^2 = 20.94$, $\Delta df = 16$, $p = .181$, ΔRMSEA $< .01$).

Discussion

The FEMT encompasses a balanced number of same and different emotion comparisons, it has an acceptable goodness of fit, scores are normally distributed, αs and test-retest reliability are acceptable, and the items were sample- and gender-invariant. Next, we sought to investigate the FEMT's nomological network.

Study 2

We examined (S1) the convergent and discriminant validity of the FEMT in relation to emotional intelligence, emotion recognition ability in faces and postures, cognitive intelligence, and the Big Five personality traits (S1 and S3). Additionally, we examined criterion validity (S3) with reference to task performance, social astuteness, and adaptive performance at work. Finally (S4), we tested the FEMT's

interethnic fit by comparing emotion recognition accuracy in Caucasian and Japanese faces.

Construct Validation

An important element of construct validation is ensuring that the FEMT measures emotion recognition in faces. We approached this problem in Study 2 by comparing convergent and discriminant validity. We compared the FEMT's relations with performance in accurately labeling emotions in the DANVA-Faces (Nowicki & Duke, 2001) and the Perceiving Emotions in Faces tests (MSCEIT; Mayer et al., 2002) as opposed to the Perceiving Emotions in Pictures in an art or landscape test (MSCEIT; Mayer et al., 2002), that is, we compared convergent and discriminant validity.

Schlegel et al. (2020) found a positive relation between emotion recognition ability and general mental ability (GMA) and concluded that it is a sensory-cognitive ability that is substantially related to cognitive intelligence. Therefore, and in line with Bryan and Mayer (2020), we expected a positive correlation between ERA-O as assessed with the FEMT and general cognitive ability.

Furthermore, we expected that ERA-O as assessed with the FEMT will be positively related to conscientiousness. Joseph and Newman (2010, p. 58) suggested that "conscientious individuals may use the emotional cues from others to guide their need for controlled behavior." Based on linguistic labeling, Joseph and Newman (2010) found a measurement error-corrected correlation of $r_c = .25$ for this relation in their meta-analysis.

Criterion Validation and Interethnic Fit

In their meta-analysis, Joseph and Newman (2010) found a low measurement error-corrected correlation of $r_c = .10$ between labeling-based measures of emotion recognition ability in others and job performance as rated by coworkers. In order to improve predictor-criterion validity, researchers have called for narrowing the criterion so that the relevance and breadth of predictor and criterion match. We suggest that the following two criteria are more relevant and better match the specificity of ERA-O than overall task performance: social astuteness and adaptive performance.

Social astuteness is a social skill that involves the ability to read people and situations and to understand social interactions at work (Ferris et al., 2005). We expect a positive relation between FEMT and social astuteness. Adaptive performance comprises behaviors necessary for an individual's effectiveness in complex, uncertain, and dynamic situations (Griffin et al., 2007). According to affective

events theory (Weiss & Cropanzano, 1996), a situation can spontaneously change on the basis of others' emotional reactions to a specific event. Individuals with better ERA-O understand the emotion causing these dynamics and can adapt their behavior accordingly. We propose that ERA-O should have a positive effect on individuals' adaptive performance. Finally, to ensure that our new test approach is not confounded by ethnic similarity/dissimilarity, we compared Japanese and Caucasian face stimuli in S4 of Study 2.

Method

Participants and Procedures
An overview of all samples in Study 2 is presented in Table 1. In three online field samples, prospective test takers received an e-mail from us with a link to the study. In the proctored study-design (S2), participants were invited to a standardized test environment. In order to meet the standards of intelligence testing, examiners were trained in the administration of the intelligence test. In the multisource design (S3), they were asked at the end of the survey to nominate one coworker who knew them well and would thus be able to assess their behavior in the workplace. The coworkers subsequently received an e-mail with a link to a survey that asked them to assess the target's behavior and performance in the workplace. Of the coworkers, 87% were peers, 8% were supervisors, 3% were subordinates, and 2% were others, for example, HR workers.

In the test design for interethnic fit (S4), in addition to the FEMT we administered two other tests with the same response format but different stimulus material: We created one test with 27 comparisons of Caucasian faces (Lundqvist et al., 1998), while the other test comprised 35 items with Japanese faces (Matsumoto & Ekman, 1988) set. All study participants in S4 likely were Caucasian. The order of presentation of the three tests in S4 was completely balanced. The FEMT was assessed as in Study 1 in all four samples.

Measures
DANVA2 – Faces and Postures (S1–S3)
Emotion recognition of faces and postures based on linguistic labels was measured with the corresponding two scales of the DANVA2 (adult version; Nowicki & Duke, 2001). Both scales comprise 24 items (i.e., photographs of adult faces or postures). Of the 24 items per scale, there are six items each for the emotions of anger, fear, happiness, and sadness. Higher scores indicate higher accuracy in emotion recognition. The DANVA2 test has been comprehensively validated (e.g., Bechtoldt et al., 2011). We administered the DANVA-F and DANVA-P test in S1, and the DANVA-F

Table 3. Confirmatory factor analyses of all measures used in Studies 1 and 2

Measure	Study	Sample	N	$\chi^2(df)$	$p(\chi^2)$	RMSEA	p(RMSEA)	CFI	TLI	SRMR
FEMT[a]	1		348	153.30 (134)	.122	.03	1.00	.98	.98	.07
	2	1	344	197.30 (134)	.001	.04	0.98	.94	.93	.08
	2	2	182	155.38 (134)	.100	.03	0.97	.97	.96	.10
	2	3	310	182.21 (134)	.004	.03	0.99	.96	.96	.08
	2	4	383	173.11 (134)	.013	.03	1.00	.98	.98	.07
	1, 2	Overall	1,567	325.32 (134)	.001	.03	1.00	.97	.97	.05
DANVA-F	2	1[e]	344	280.06 (230)	.013	.03	1.00	.77	.75	.12
	2	2[f]	182	229.22 (209)	.161	.02	1.00	.80	.78	.14
	2	3	310	357.08 (252)	.001	.04	1.00	.50	.45	.15
	2	1, 3	654	403.34 (252)	.001	.03	1.00	.66	.63	.12
DANVA-P	2	1	344	335.78 (252)	.001	.03	1.00	.55	.51	.11
MSCEIT[bd]	2	1	344	7.66 (12)	.811	.00	0.99	1.00	1.00	.01
GATB-GMA[cd]	2	2	182	2.94 (1)	.086	.10	0.16	.98	.95	.15
Neuroticism[d]	2	1, 3	654	214.98 (54)	.001	.07	0.01	.93	.92	.04
Extraversion[d]	2	1, 3	654	427.89 (54)	.001	.10	0.01	.79	.74	.07
Openness[d]	2	1, 3	654	235.34 (54)	.001	.07	0.01	.84	.81	.05
Agreeableness[d]	2	1, 3	654	198.63 (54)	.001	.06	0.01	.85	.81	.05
Conscientiousness[d]	2	1, 3	654	413.60 (54)	.001	.10	0.01	.77	.72	.07
Task Performance[d]	2	3	310	13.44 (5)	.020	.07	0.17	.99	.98	.02
Social Astuteness[d]	2	3	310	15.06 (5)	.010	.08	0.12	.98	.95	.03
Adaptive Performance[d]	2	3	310	93.50 (5)	.001	.24	0.01	.88	.75	.07
KDEF[a]	2	4	383	578.55 (323)	.001	.05	0.89	.92	.92	.12
JACFEE[a]	2	4	383	951.46 (559)	.001	.04	1.00	.93	.92	.12

Note. [a]Two correlated factors; [b]five correlated factors, that is, perceiving emotions in faces and pictures were modeled as two different factors; [c]three correlated factors, figural and numerical with equal loadings; [d]ML estimator (all others WLSMV); [e]without item 23 which had zero variance; [f]without item 23 and 24 which had zero variance.

Table 4. Descriptive statistics and correlations of the variables in Study 2, Sample 1 (construct validation)

	M	SD	1	2	3	4	5	6	7	8
1. FEMT	58.01	19.50	.72	.89**	.58**	.53**	.28**	.41**	.45**	.41**
2. DANVA-F	76.03	11.19	.34**	.62[a]	.71**	.31**	.23**	.35**	.49**	.36**
3. DANVA-P	66.90	11.28	.23**	.36**	.40	.14	.20**	.24**	.28**	.24**
4. Perceiving Emotions (Faces)	49.09	10.12	.31**	.23**	.09	.87	.36**	.50**	.27**	.33**
5. Perceiving Emotions (Pictures)	47.52	11.48	.17**	.17**	.14**	.36**	.90	.62**	.34**	.43**
6. Facilitating Thought	45.20	6.93	.18**	.22**	.13*	.39**	.48**	.73	.75**	.90**
7. Understanding Emotions	52.05	6.99	.23**	.29**	.16**	.23**	.29**	.50**	.74	.78**
8. Managing Emotions	40.69	6.86	.20**	.21**	.13*	.27**	.35**	.57**	.54**	.73
9. MSCEIT total	46.56	5.75	.30**	.31**	.18**	.58**	.67**	.83**	.74**	.78**

Note. N = 344. DANVA = Diagnostic Analysis of Nonverbal Accuracy, DANVA-F (Faces) and DANVA-P (Postures); MSCEIT = Mayer-Salovey-Caruso Emotional Intelligence Test. Correlations of the manifest variables are presented below the diagonal. Cronbach's αs in the diagonal (for MSCEIT total α = .92); Correlations of the 2nd order factor of the FEMT above the diagonal; [a]ω (Rodriguez et al., 2016); Goodness of fit: $\chi^2(380)$ = 435.68 (p = .025), RMSEA = .02 (p = 1.00), CFI = .97; *p < .05; **p < .01.

test additionally in S2 and S3 (see Tables 3–7 for the descriptive statistics of the DANVA2).

Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT)

In S1, we administered the official German translation and adaptation of the MSCEIT V 2.0 (Steinmayr et al., 2011) with 141 items. The Perceiving Emotions branch comprises two different tasks, Perceiving Emotions in Faces (PEI-Faces) and Perceiving Emotions in Pictures (PEI-Pictures). The American version of the MSCEIT Pictures Task uses pictures of art or a landscape and asks the participant to use cartoon faces of emotion to rate each picture. In the German version, however, participants are asked to use numbers instead of smileys to rate each picture. We used the consensus scoring key of the MSCEIT. Additional

Table 5. Descriptive statistics and correlations of the variables in Study 2, Sample 2 (construct validation)

	M	SD	1	2	3	4	5	6
1. FEMT	61.48	21.22	.78	.57**	.29**[b]	.30**[c]	.03[c]	.23**[c]
2. DANVA-F	76.42	10.70	.31**	.76[a]	.37**[b]	.41**[c]	.07[c]	.25**[c]
3. GATB-GMA	103.16	11.08	.17*	.21*				
4. Figural	103.70	12.28	.22**	.29**	.72**		.32**[c]	.46**[c]
5. Numerical	102.62	13.26	.01	.05	.79**	.31**		.57**[c]
6. Verbal	103.15	15.94	.17*	.18*	.88**	.46**	.57**[c]	

Note. N = 182. FEMT = Face-Based Emotion Matching Test; DANVA-F = Diagnostic Analysis of Nonverbal Accuracy – Faces; GATB-GMA = General Aptitude Test Battery – General Mental Ability. Correlations of the manifest are presented under the diagonal. Cronbach's αs in the diagonal. Correlations of the 2nd order factor of the FEMT are presented above the diagonal; [a]ω; [b]Goodness of fit indices of latent factors models with FEMT (2nd order factor) subscales (FEMT-Equal, FEMT-Diff), Model 1: GATB with three indicators (figural, numerical, verbal): $\chi^2(225) = 283.69$ ($p = .005$), RMSEA = .04 ($p = .930$), CFI = .92; [c]Model 2: including three separate factors for figural, numerical, and verbal instead of GMA: $\chi^2(221) = 264.41$ ($p = .024$), RMSEA = .03 ($p = .980$), CFI = .94; *$p < .05$; **$p < .01$.

Table 6. Descriptive statistics and correlations of personality, FEMT, and DANVA-F in Study 2, Samples 1 and 3 (construct validation)

	M	SD	1	2	3	4	5	6	7
1. FEMT	57.01	20.49	.75	.61**	−.07	.15*	.12	.20**	.21**
2. DANVA-F	75.90	10.90	.28**	.76[a]	−.02	.11*	.09	.11*	.09
3. Neuroticism	2.51	0.66	−.04	.00	.87[e]	−.52**	.05	−.19**	−.34**
4. Extraversion	3.45	0.52	.08*	.08*	−.44**	.79[e]	.06	.35**	.31**
5. Openness	3.46	0.51	.06	.06	.03	.04	.71[e]	.13*	−.08
6. Agreeableness	3.77	0.45	.11**	.07	−.15**	.27**	.09*	.73[e]	.21**
7. Conscientiousness	3.88	0.48	.12**	.05	−.29**	.26**	−.06	.16**	.79

Note. N = 654. FEMT = Face-Based Emotion Matching Test; DANVA-F = Diagnostic Analysis of Nonverbal Accuracy – Faces. Correlations of the manifest variables are presented below the diagonal. Cronbach's αs in the diagonal. Correlations of the 2nd order factor of the FEMT model are presented above the diagonal; [a]ω; Goodness of fit: $\chi^2(383) = 529.85$ ($p = .000$), RMSEA = .02 ($p = 1.00$), CFI = .96; *$p < .05$; **$p < .01$.

scores were calculated for the branches of using, understanding, and managing emotions and the MSCEIT total score (see Table 4 for the descriptive statistics of the MSCEIT).

Cognitive Intelligence

In S2, we measured general cognitive intelligence with the German version (Schmale & Schmidtke, 2008) of the GATB (Weiss, 1972). As proposed by the authors, we used the three subtests visual-perceptual ability (Scale 2), numerical reasoning (Scale 6), and verbal intelligence (Scale 7) to capture the three components of cognitive intelligence as well as overall general mental ability. Test-retest reliabilities for the three subtests provided by the authors ranged from r_{tt} = .89 to .90 (see Table 5 for the descriptive statistics of the GATB).

NEO-Five Factor Inventory (NEO-FFI)

We used the German version (Borkenau & Ostendorf, 1993) of the NEO-FFI in S1 and S3 to assess the Big Five personality dimensions. The NEO-FFI contains 60 self-report items that assess the five dimensions, namely agreeableness, conscientiousness, extraversion, neuroticism, and openness with 12 items each. The items are

answered on a 5-point Likert scale ranging from 1 = *does not apply at all* to 5 = *applies completely* (see Table 6 for the descriptive statistics of the NEO-FFI).

Task Performance

Coworkers rated targets' task performance using the German version (Blickle, Kramer, et al., 2011) of the scale developed and validated by Ferris et al. (2001). A sample item is "This person finds resourceful and creative solutions to complex technical problems."

Social Astuteness

To assess targets' social astuteness, the validated German translation (5 items; Blickle, Ferris, et al., 2011) of the Political Skill Inventory (PSI; Ferris et al., 2005) was used for other-assessment. A sample item is "This person is particularly good at sensing the motivations and hidden agendas of others."

Adaptive Performance

Coworkers rated targets' adaptive performance using the corresponding five items of the scale developed and validated by Blickle, Kramer, et al. (2011). A sample is "This

Table 7. Descriptive statistics, correlations, and regression analyses in Study 2, Sample 3 (criterion validation)

	M	SD	1	2	3	4	5	6	7
1. FEMT	55.90	21.51	.78	.48**	−.14	−.28**	.09	.22*	.19*
2. DANVA-F	75.75	10.58	.22**	.70[a]	−.11	−.29**	.11	.10	.16
3. Sex	1.47	.50	−.10	−.08		−.01	−.14*	−.12	−.11
4. Age	43.48	10.69	−.20**	−.18**	.00		.07	−.01	−.01
5. Task Performance	3.80	0.74	.05	.08	−.13*	.06	.88[e]	.36**	.94**
6. Social Astuteness	4.91	0.91	.13*	.03	−.10	−.01	.27**	.78[e]	.34**
7. Adaptive Performance	3.75	0.71	.11*	.09	−.10	−.01	.83**	.27**	.85

Regression model	Social astuteness		Adaptive performance		Task performance	
	Manifest	Latent	Manifest	Latent	Manifest	Latent
Predictor	β	β	β	β	β	β
Sex	−.09	−.08	−.09	−.08	−.08	−.07
Age	.02	.06	.02	.05	.05	.08
DANVA-F	.01	−.01	.07	.06	.07	.10
FEMT	.12*	.22[†]	.09	.16	.03	.05
ΔR^2	.01*		.01		.01	

Note. N = 310. FEMT = Face-Based Emotion Matching Test; DANVA-F = Diagnostic Analysis of Nonverbal Accuracy – Faces. Upper part: Correlations of the manifest variables are presented below the diagonal. Cronbach's αs in the diagonal. Correlations of the 2nd order factor of the FEMT model are presented above the diagonal. 0 = female, 1 = male; [a]ω; Goodness of fit: $\chi^2(610)$ = 763.71 (p = .001), RMSEA = .03 (p = 1.00), CFI = .93. Lower part: manifest and latent regression analyses. [†]p < .05 (one-tailed), *p < .05 (two-tailed); **p < .01.

person successfully handles emergencies, interruptions, and losses at work."

Additionally, we asked target participants to report their gender and age.

Data Analysis

In confirmatory factor analyses, we tested the goodness of fit indices of all measures (Table 3). We report the percentage of correct answers on the FEMT and its correlations with the other manifest test scores. We also assessed the relations between the FEMT and the other variables at the construct level, beyond measurement error, by using SEM (Muthén & Muthén, 2012); more specifically, we assessed the relations between the FEMT second-order factor and the other constructs using SEM (Muthén & Muthén, 2012). To reduce statistical model complexity, we used parceled construct indicators for the DANVA and NEO-FFI items (odd- vs. even-numbered items).

Results and Discussion

As expected (see Table 4), the convergent relations of the FEMT with accurately linguistically labeling emotions, that is, ρ(FEMT, DANVA-F) = .89 (p < .01), ρ(FEMT, PEI-Faces) = .53 (p < .01) were higher than the discriminant relation of the FEMT with rating smileys with reference to pictures of art and landscapes, that is, ρ(FEMT, PEI-Pictures) = .28 (p < .01): ρ = .89 versus ρ = .28, N = 344, z = 15.31, p < .01; ρ = .53 versus ρ = .28, N = 344, z = 5.75, p < .01. These findings provide support for the FEMT's convergent and discriminant validity.

Table 8. Descriptive statistics and correlations of the variables in Study 2, Sample 4 (Test of interethnic fit)

	M	SD	1	2	3
1. FEMT	53.02	24.23	.83		
2. KDEF	58.71	18.01	.74**	.82	
3. JACFEE	64.54	19.53	.71**	.78**	.88

Note. N = 383. FEMT = Face-Based Emotion Matching Test; KDEF = Karolinska Directed Emotional Faces; JACFEE = Japanese and Caucasian Facial Expressions of Emotion. Cronbach's αs in the diagonal; **p < .01.

As expected (see Table 5), the FEMT (ρ = .29, p < .01) and the DANVA-F (ρ = .37, p < .01) were positively related to cognitive intelligence and the FEMT to conscientiousness (see Table 6), (ρ = .21, p < .01). These findings further support the convergent validity of the FEMT, because it has the same pattern of relations with cognitive intelligence and conscientiousness as established measures of emotion recognition ability in faces.

Furthermore (Table 7), as expected, the FEMT correlated with manifest and latent measures of social astuteness (ρ = .22, p < .05) and adaptive performance (ρ = .19, p < .05). Additionally, the FEMT even showed incremental validity in predicting coworker ratings of targets' social astuteness (ρ = .22, p < .05, one-tailed) above and beyond the DANVA2-F, sex, and age, despite being substantially correlated with the DANVA2-F. These findings support the FEMT's criterion validity in the work context.

Finally (Table 8), the FEMT correlated with both the Caucasian face stimuli (r = .74, p < .01) and the Japanese face stimuli (r = .71, p < .01). The difference between these correlations was nonsignificant. This indicates that

individuals with a high FEMT score were able to consistently match expressions for the same emotion and distinguish different emotions in faces with both the same and different ethnic origins.

General Discussion

The purpose of the current research was to develop and validate an objective and nonlinguistic measure of ERA-O (Sauter et al., 2011) that is distinct from but related to traditional linguistic measures of emotion recognition ability and can be applied in vocational and organizational settings. Across five studies we found that the nonlinguistic ERA-O dimension can be assessed with the FEMT – an objective test not a self-report measure. We also found that scores on the ERA-O are related to coworker assessed social astuteness and adaptive performance in the workplace.

We do not suggest that the nonlinguistic perception of emotions is more important than combined perception and labeling measures. Instead, we offer researchers a new assessment tool to distinguish between the nonlinguistic sensitivity to an emotional expression and the ability to correctly report it (Sauter et al., 2011). With this measure, researchers can develop a more comprehensive understanding of the perception of emotions in others and their social consequences and can use it in research on emotional labor, leadership behavior, and effectiveness in conflict behavior, negotiations, sales performance, and service behavior. Furthermore, although this test was developed and validated with workers and workplace-specific criteria, there is nothing precluding its use as a general nonlinguistic emotion recognition test.

In the validation study, we used the DANVA2-F or the MSCEIT test, or both. Both tools comprise Caucasian and non-Caucasian face stimuli. In addition, in Study 2 we directly addressed the interethnic fit of our test items and found empirical support for it. Our findings thus support the assumption that nonlinguistic emotion recognition operates independently from ethnic-specific items. In sum, we not only contribute to the literature by adding a new theoretical perspective and a comprehensive validation of a respective assessment tool, but we also suggest a promising assessment approach (see Figure 2) that can be applied not only to facial stimuli from different ethnic backgrounds but also to postural and auditory emotional stimuli.

Although social astuteness and adaptive performance are important constructs for successful social behavior, their relations with objective emotion recognition ability have rarely been empirically investigated. We followed the call by Jundt et al. (2015) to address the role of emotions in predicting adaptive performance. Our research also answers calls by forensic and clinical scholars for improved

psychological measurement of emotion recognition ability (Cigna et al., 2017). The large sample size in combination with established scales and the hetero-source, hetero-method approach reduces the probability of biased parameter estimates (Podsakoff et al., 2012).

In line with previous literature (Côté, 2014), we suggest that some emotional states can be read from a person's face by those who have high ERA-O. Our test was designed to assess this ability. However, information about the situation in which an emotion is expressed can also play an important role in assessing the emotional state of other persons. Future research should address this because expressive and emotional situation knowledge is often also important to accurately read others' emotions and correctly anticipate their consequences.

The FEMT can assist in the process of hiring potentially successful personnel. Organizations might apply the FEMT when hiring employees for roles where emotion recognition ability may enhance performance, such as jobs that demand complex social perception in enterprising and social work environments. Supplemented by other sources of information, such as cognitive intelligence tests and interviews, the FEMT can make the recruitment and selection process more efficient. In addition, the FEMT could also be used with current employees to evaluate their ongoing level of functioning and well-being (Schlegel et al., 2021). This could help indicate organizational positions for which emotion recognition ability is more or less important.

Our study is not without limitations: First, the empirical research in this study is limited in its focus to one particular modality of emotion expressions (pictures of faces). The extent to which the effects that we found apply to other modalities (e.g., videos of gestures or recordings of voices) deserves to be studied in future research. Second, in all studies, the recruited participants were recruited through students fulfilling course requirements. In future research, additional recruiting strategies should be applied.

References

Bechtoldt, M. N., Rohrmann, S., de Pater, I. E., & Beersma, B. (2011). The primacy of perceiving: Emotion recognition buffers negative effects of emotional labor. *The Journal of Applied Psychology, 96*(5), 1087–1094. https://doi.org/10.1037/a0023683

Blickle, G., Ferris, G. R., Munyon, T. P., Momm, T., Zettler, I., Schneider, P. B., & Buckley, M. R. (2011). A multi-source, multi-study investigation of job performance prediction by political skill. *Applied Psychology, 60*(3), 449–474. https://doi.org/10.1111/j.1464-0597.2011.00443.x

Blickle, G., Kranefeld, I., Wihler, A., Kückelhaus, B. P., & Menges, J. I. (2021). *Supplementary materials for https://doi.org/10.1027/1015-5759/a000656.* https://osf.io/sxtr3/?view_only=c55a8948a97b43e589e149a225eadb36

Blickle, G., Kramer, J., Schneider, P. B., Meurs, J. A., Ferris, G. R., Mierke, J., Witzki, A. H., & Momm, T. D. (2011). Role of political skill in job performance prediction beyond general mental ability and personality in cross-sectional and predictive studies. *Journal of Applied Social Psychology, 41*(2), 488–514. https://doi.org/10.1111/j.1559-1816.2010.00723.x

Borkenau, P., & Ostendorf, F. (1993). *NEO-Fünf-Faktoren-Inventar (NEO-FFI) nach Costa und McCrae* [NEO-Five Factor Inventory (NEO-FFI) according to Costa and McCrae]. Hogrefe.

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 & Emotion, 14*(2), 193–211. https://doi.org/10.1080/026999300378932

Bryan, V. M., & Mayer, J. D. (2020). A meta-analysis of the correlations among broad intelligences: Understanding their relations. *Intelligence, 81*, Article 101469. https://doi.org/10.1016/j.intell.2020.101469

Cigna, M.-H., Guay, J.-P., & Renaud, P. (2017). Psychopathic traits and their relation to facial affect recognition. *Personality and Individual Differences, 117*, 210–215. https://doi.org/10.1016/j.paid.2017.06.014

Côté, S. (2014). Emotional intelligence in organizations. *Annual Review of Organizational Psychology and Organizational Behavior, 1*, 459–488. https://doi.org/10.1146/annurev-orgpsych-031413-091233

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods, 1*(1), 16–29. https://doi.org/10.1037/1082-989X.1.1.16

Elfenbein, H. A., & Ambady, N. (2002). Predicting workplace outcomes from the ability to eavesdrop on feelings. *Journal of Applied Psychology, 87*(5), 963–971. https://doi.org/10.1037/0021-9010.87.5.963

Elfenbein, H. A., Foo, M. D., White, J., Tan, H. H., & Aik, V. C. (2007). Reading your counterpart: The benefit of emotion recognition accuracy for effectiveness in negotiation. *Journal of Nonverbal Behavior, 31*, 205–223. https://doi.org/10.1007/s10919-007-0033-7

Elfenbein, H. A., Marsh, A., & Ambady, N. (2002). Emotional intelligence and the recognition of emotion from facial expressions. In L. F. Barrett & P. Salovey (Eds.), *Emotions and social behavior. The wisdom in feeling: Psychological processes in emotional intelligence* (pp. 37–59). Guilford Press.

Enticott, P. G., Johnston, P. J., Herring, S. E., Hoy, K. E., & Fitzgerald, P. B. (2008). Mirror neuron activation is associated with facial emotion processing. *Neuropsychologia, 46*(11), 2851–2854. https://doi.org/10.1016/j.neuropsychologia.2008.04.022

Ferris, G. R., Treadway, D. C., Kolodinsky, R. W., Hochwarter, W. A., Kacmar, C. J., Douglas, C., & Frink, D. D. (2005). Development and validation of the Political Skill Inventory. *Journal of Management, 31*(1), 126–152. https://doi.org/10.1177/0149206304271386

Ferris, G. R., Witt, L. A., & Hochwarter, W. A. (2001). Interaction of social skill and general mental ability on job performance and salary. *Journal of Applied Psychology, 86*(6), 1075–1082. https://doi.org/10.1037/0021-9010.86.6.1075

Fiori, M. (2009). A new look at emotional intelligence: A dual-process framework. *Personality and Social Psychology Review, 13*(1), 21–44. https://doi.org/10.1177/1088868308326909

Fiori, M., & Vesely-Maillefer, A. K. (2018). Emotional intelligence as an ability: Theory, challenges, and new directions. In K. Keefer, J. Parker, & D. Saklofske (Eds.), *Emotional Intelligence in Education* (pp. 23–47). Springer. https://doi.org/10.1007/978-3-319-90633-1_2

Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change, 114*, 254–280. https://doi.org/10.1016/j.techfore.2016.08.019

Griffin, M. A., Neal, A., & Parker, S. K. (2007). A new model of work role performance: Positive behavior in uncertain and interdependent contexts. *Academy of Management Journal, 50*(2), 327–347. https://doi.org/10.5465/amj.2007.24634438

Herzmann, G., Danthiir, V., Schacht, A., Sommer, W., & Wilhelm, O. (2008). Toward a comprehensive test battery for face cognition: Assessment of the tasks. *Behavior Research Methods, 40*(3), 840–857. https://doi.org/10.3758/BRM.40.3.840

International Organization for Migration. (2019). *World Migration Report 2020*. https://publications.iom.int/system/files/pdf/wmr_2020.pdf

Joseph, D. L., & Newman, D. A. (2010). Emotional intelligence: An integrative meta-analysis and cascading model. *Journal of Applied Psychology, 95*(1), 54–78. https://doi.org/10.1037/a0017286

Jundt, D. K., Shoss, M. K., & Huang, J. L. (2015). Individual adaptive performance in organizations: A review. *Journal of Organizational Behavior, 36*(1), 53–71. https://doi.org/10.1002/job.1955

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition and Emotion, 24*(8), 1377–1388. https://doi.org/10.1080/02699930903485076

Lundqvist, D., Flykt, A., & Öhman, A. (1998). *The Karolinska Directed Emotional Faces – KDEF* [CD ROM]. Department of Clinical Neuroscience, Psychology section, Karolinska Institutet.

Matsumoto, D., & Ekman, P. (1988). *Japanese and Caucasian Facial Expressions of Emotion (JACFEE)* [Slides]. Intercultural and Emotion Research Laboratory, Department of Psychology, San Francisco State University.

Mayer, J. D., Salovey, P., & Caruso, D. R. (2002). *Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT) user's manual*. MHS.

Momm, T. D., Blickle, G., & Liu, Y. (2010). Political skill and emotional cue learning. *Personality and Individual Differences, 49*, 396–401. https://doi.org/10.1016/j.paid.2010. 04.006

Momm, T., Blickle, G., Liu, Y., Wihler, A., Kholin, M., & Menges, J. I. (2015). It pays to have an eye for emotions: Emotion recognition ability indirectly predicts annual income. *Journal of Organizational Behavior, 36*, 147–163. https://doi.org/10.1002/job.1975

Muthén, L., & Muthén, B. (2012). *Mplus user's guide* (7th ed.). Muthén & Muthén.

Nowicki, S., & Duke, M. P. (2001). Nonverbal receptivity: The Diagnostic Analysis of Nonverbal Accuracy (DANVA). In J. A. Hall & F. J. Bernieri (Eds.), *Interpersonal sensitivity: Theory and measurement* (pp. 183–198). Erlbaum.

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 recognise vocal emotions and facial identity. *PLoS One, 8*(6), Article e68126. https://doi.org/10.1371/journal.pone.0068126

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology, 63*, 539–569. https://doi.org/10.1146/annurev-psych-120710-100452

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Applying bifactor statistical indices in the evaluation of psychological measures. *Journal of Personality Assessment, 98*(3), 223–237. https://doi.org/10.1080/00223891.2015.1089249

Rubin, R. S., Munz, D. C., & Bommer, W. H. (2005). Leading from within: The effects of emotion recognition and personality on transformational leadership behavior. *Academy of Management Journal, 48*(5), 845–858. https://doi.org/10.2307/20159701

Sauter, D. A., LeGuen, O., & Haun, D. B. M. (2011). Categorical perception of emotional facial expressions does not require lexical categories. *Emotion, 11*(6), 1479–1483. https://doi.org/10.1037/a0025336

Scherer, K. R., & Ellgring, H. (2007). Are facial expressions of emotion produced by categorical affect programs or dynamically driven by appraisal? *Emotion, 7*(1), 113–130. https://doi.org/10.1037/1528-3542.7.1.113

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online, 8*, 23–74.

Schlegel, K., Boone, R. T., & Hall, J. A. (2017). Individual differences in interpersonal accuracy: A multi-level meta-analysis to assess whether judging other people is one skill or many. *Journal of Nonverbal Behavior, 41*(2), 103–137. https://doi.org/10.1007/s10919-017-0249-0

Schlegel, K., Grandjean, D., & Scherer, K. R. (2014). Introducing the Geneva Emotion Recognition Test: An example of Rasch-based test development. *Psychological Assessment, 26*(2), 666–672. https://doi.org/10.1037/a0035246

Schlegel, K., von Gugelberg, H. M., Makowski, L. M., Gubler, D. A., & Troche, S. J. (2021). Emotion recognition ability as a predictor of well-being during the COVID-19 pandemic. *Social Psychological and Personality Science*,. https://doi.org/10.1177/1948550620982851

Schlegel, K., & Mortillaro, M. (2019). The Geneva Emotional Competence Test (GECo): An ability measure of workplace emotional intelligence. *Journal of Applied Psychology, 104*(4), 559–580. https://doi.org/10.1037/apl0000365

Schlegel, K., Palese, T., Mast, M. S., Rammsayer, T. H., Hall, J. A., & Murphy, N. A. (2020). A meta-analysis of the relationship between emotion recognition ability and intelligence. *Cognition & Emotion, 34*(2), 329–351. https://doi.org/10.1080/02699931.2019.1632801

Schmale, H., & Schmidtke, H. (2008). *Berufseignungstest: BET* [Professional Aptitude Test] (4th ed.). Huber.

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality, 47*(5), 609–612. https://doi.org/10.1016/j.jrp. 2013.05.009

Simmons, J. P., Nelson, L. D., & Simonsohn, U. A. (2012, October 14). *21 Word Solution*. https://ssrn.com/abstract=2160588 or http://dx.doi.org/10.2139/ssrn.2160588

Steinmayr, R., Schütz, A., Hertel, J., & Schröder-Abé, M. (2011). *Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT^{TM}) – German Version of the MSCEIT^{TM}*. Hogrefe.

Svetina, D., Rutkowski, L., & Rutkowski, D. (2020). Multiple-group invariance with categorical outcomes using updated guidelines: An illustration using mplus and the lavaan/semTools packages. *Structural Equation Modeling: A Multidisciplinary Journal, 27*(1), 111–130. https://doi.org/10.1080/10705511.2019.1602776

Weiss, D. (1972). The General Aptitude Test Battery. In O. K. Buros (Ed.), *The seventh mental measurement yearbook* (pp. 1058–1061). Gryphon.

Weiss, H. M., & Cropanzano, R. (1996). Affective events theory: A theoretical discussion of the structure, causes and consequences of affective experiences at work. In B. M. Staw & L. L. Cummings (Eds.), *Research in Organizational Behavior* (Vol. 18, pp. 1–74). Elsevier Science/JAI Press.

Open Science
Open Data: We confirm that there is sufficient information for an independent researcher to reproduce all of the reported results, including codes, see Blickle et al. (2021).
Open Materials: The information needed to reproduce all of the reported methodology is not openly accessible because we are not the copyright holders of the picture stimuli. To request access (info@rafd.nl) to the Radboud Faces Database (RaFD) (http://www.socsci.ru.nl:8180/RaFD2/RaFD), researchers need to be researchers at an accredited university. In order to check that, Radboud University asks you in the request form to provide it with a link to a university webpage mentioning your affiliation. If none such site exists, please send a recent publication with you as an author to info@rafd.nl.
After researchers have admission to use the RaFD the specific material will be made available on request from the paper authors.
Preregistration of Studies and Analysis Plans: Our studies were not preregistered.

ORCID
Gerhard Blickle
https://orcid.org/0000-0002-8443-9015

Gerhard Blickle
Institute of Psychology
Rheinische Friedrich-Wilhelms University of Bonn
Kaiser-Karl-Ring 9
53111 Bonn
Germany
gerhard.blickle@uni-bonn.de