Emotion Recognition Ability, Metacognition, and Metaemotion: A Multimodal Online-Assessment of Swedish Adults

Supervisor: Petri Laukka
MASTER THESIS, 30 ECTS, 2018

STOCKHOLMS UNIVERSITET
PSYKOLOGISKA INSTITUTIONEN

Alexandra Israelsson
Data obtained in laboratory settings is a valid but resource-demanding approach. Moreover, although aspects of both metacognition and metaemotion have been proposed to be important for socioemotional functioning, such associations have rarely been studied previously. This study aimed to examine the feasibility of a multimodal online-assessment of emotion recognition ability, and to investigate its associations with metacognition and metaemotion. The sample consisted of 106 students from three Swedish universities. The online-survey included a multimodal emotion recognition test (ERAM) with added trial-by-trial confidence judgments (to measure metacognition) and questionnaires related to metaemotion. Online-data showed great consistency with previous data collected in lab. Well-calibrated adults had higher emotion recognition accuracy than under-confident adults. Higher levels of negative metaemotions were associated with higher emotion recognition accuracy. In conclusion, online-assessments of emotional abilities may be a useful approach. Further research is required to understand relationships between metacognition, metaemotion, and emotion recognition ability more fully.

Emotion recognition ability is an essential facet of social interactions, functioning, and maintenance of interpersonal relationships (Bänziger, Grandjean & Scherer, 2009). Emotion recognition involves the ability to interpret bodily, facial and vocal emotional cues accurately, which is a dynamic and multimodal process; however, the majority of researchers have investigated individual differences in emotion recognition ability with a limited set of emotions and unimodal instruments (Bänziger et al., 2009; Schlegel, Fontaine & Scherer, 2017; Schlegel & Scherer, 2016). Although researchers have adopted a more comprehensive approach, current knowledge about why some individuals have better emotion recognition ability than others within the general population is still limited. Additionally, even though online-surveys nowadays is a commonly used method, online-distributed multimodal emotion recognition tasks have not yet become a predominating method despite their potential. Therefore, this study aims to examine if a multimodal online-assessment is a feasible method to investigate emotion recognition ability within the general population, as well as to investigate associations between adults’ multimodal emotion recognition ability and measures of metacognition and metaemotion.

Current Findings of Individual Differences in Emotion Recognition
Traditionally, research on individual differences in emotion recognition ability has centered on deficiencies and psychopathology, wherein the general population has served as a healthy control. Consequently, current knowledge about what psychological factors
might underlie variability within the general population is still limited. However, research has reported individual differences with respect to gender (Thompson & Voyer, 2014; Wright, Riedel, Sechrest, Lane & Smith, 2018), age (Ruffman, Henry, Livingstone & Phillips, 2008), and between cultures (Dailey et al., 2010; Elfenbein & Ambady, 2002). Moreover, researchers also propose emotional intelligence (EI) as a fundamental aspect of emotion recognition ability, which involves the ability to perceive emotions, use emotions to facilitate cognition, understand emotions, and manage emotions (Mayer, Salovey & Caruso, 2008). However, research has shown that emotion recognition ability and EI measurements correlate relatively low (Bänziger et al., 2009). Furthermore, interest in psychological factors related to socioemotional functioning such as metacognition has recently emerged as well.

**Metacognition: General Assumptions and Ideas**
Flavell (1979) initiated the notion of metacognition in the 1970s, which has since become a well-established field of educational and cognitive psychology. Metacognition as an umbrella term refers to the psychological processes and operations engaged in the executive, monitoring, understanding, and modification of cognition (Flavell, 1979; Wells & Cartwright-Hatton, 2004). Whereas cognitive functions are central to the acquisition of knowledge, metacognition constitutes the capability to evaluate, develop, and monitor cognition (Gourgey, 1998). Moreover, metacognitive judgments and feelings refer to the aware processes of an individual’s monitoring and control of metamemory (Nelson & Narens, 1990). Metacognitive judgments constitute fundamental aspects of an individual’s monitoring of knowledge, which are both crucial for knowing what we know and what we do not know (Koriat & Goldsmith, 1996; Koriat & Levy-Sadot, 1999).

**Metacognitive Judgments and Dynamics of Meta-Accuracy**
Metacognitive judgments involve the aware monitoring and regulatory processes of knowledge, which may be either prospective, current or retrospective relative to the task. However, these types of judgments vary from not only a time perspective but also in metacognitive operations (Kelly & Metcalfe, 2011). In research on metacognition, absolute and relative meta-accuracy involves different metacognitive aspects and phenomena. Absolute meta-accuracy pertains to how well an individual’s judgment and actual performance corresponds, which may vary from perfect calibration to over- and under-confidence (Keren, 1991). Relative meta-accuracy refers to the correlation between confidence and correctness at an item level, which may be either related or unrelated to the actual performance (Nelson, 1984). Furthermore, although relative and absolute measures represent different aspects of metacognitive accuracy, these measures nevertheless provide vital information when used together. For instance, an individual’s confidence may be highly correlated relative to the performance, and in an absolute sense be over- or under-confident (Juslin, Olsson & Winman, 1996; Schraw, 2009).

Research has mainly investigated individuals’ meta-accuracy relative to learning and academic performance (Zimmerman, 2008). However, other fields have also investigated metacognitive judgments. For instance, research on social cognition has investigated confidence judgments regarding interpersonal accuracy such as deception detection (e.g., DePaulo, Charlton, Cooper, Lindsay & Muhlenbruck, 1997), and memory accuracy in eyewitness identification (Brewer & Wells, 2006). Research on social inferences has reported that individuals’ meta-accuracy tend to be relatively poor regarding accuracy of social judgments (Todorov, Olivola, Dotsch & Mende-Siedlecki, 2014). However, biased
metacognitive judgments may arise due to sampling error, which may lead to task-related disruptions such as the hard-easy effect and scale-end effect (Juslin & Olsson, 1997; Juslin, Winman & Olsson, 2000; Yang, Sun & Shanks, 2018). Researchers have also discussed whether judgments are domain-general or domain-specific processes (Gutierrez, Schraw, Kuch & Richmond, 2016).

Metacognition in Relation to Social and Emotional Functioning
Research has related metacognition to individual differences in socioemotional functioning and maintenance, and to the development of prosocial and emotional competencies. Researchers have also related social dysfunction in autism spectrum disorder (ASD) to individual differences in metacognition (Zalla, Miele, Leboyer & Metcalfe, 2015). Attention to emotional processes has emerged as well. For example, research has reported that metacognitive knowledge organizes an individual’s acquisition of information for emotional and motivational factors relevant for self-regulatory means (Boekaerts, 2011). Meta-analysis has also indicated that metacognitive dysfunctions are highly prevalent among persons diagnosed with anxiety disorders (Sun, Zhu & So, 2017). Studies also report metacognition as a critical aspect of social and emotional skills (Kelly & Metcalfe, 2011). However, although theory-of-mind (ToM) also involves vital aspects of an individual’s social and emotional functioning, and is closely related to metacognition, it is nevertheless essential to distinguish these constructs. Specifically, ToM refers to the capability to understand that others have mental states, internal beliefs, intentions, desires, and feelings (Wellman, Cross & Watson, 2001). Whereas metacognition refers to the awareness, and reflectiveness of knowledge and internal beliefs about oneself and others (Flavell, 1979; Kuhn, 2000).

Attention to emotion recognition ability has become an emerging research area as well. For instance, Kelly and Metcalfe (2011) argue that there is an essential difference between the ability to identify what emotion another person is showing correctly, and the ability to evaluate that one does not know what emotion another person is showing correctly. Moreover, studies have indicated that persons with ASD and Asperger’s Syndrome have impaired social skills, as well as a reduced ability to recognize others’ emotions through facial expressions. However, researchers argue that impaired emotion recognition ability might be due to insufficient means to evaluate one’s accuracy, rather than an inability to recognize emotions (Sawyer, Williamson & Young, 2014). Research has reported that persons with ASD have insufficient regulation and application of strategies relevant for metacognitive self-monitoring, which is vital for making accurate judgments (Wojcik, Waterman, Lestié, Moulin & Souchay, 2014). Sawyer et al. (2014) reported that adults with Asperger’s also have insufficient self-monitoring regarding filtering inaccurate from accurate responses, even though they discriminated and labeled the emotions adequately. Nevertheless, Kelly and Metcalfe (2011) reported that meta-accuracy was above chance among healthy adults and correlated with emotion recognition ability. However, research on metacognition and emotion recognition ability within the general population is limited.

“Thoughts and Emotions about Emotions”: The Inquiry of Metaemotion
Gottman and colleagues introduced metaemotion in the 1990s in their research on parent-child interactions. Gottman, Katz, and Hooven (1996) theorize metaemotion as an organized set of internal concepts and emotions about emotions. Like metacognition, which refers to the executive functions of cognition (Flavell, 1979), metaemotion refers to the executive functions of emotions, wherein the emotion itself figures as the object
(Gottman et al., 1996). Furthermore, Gottman et al. (1996) also postulate that whenever we experience emotions, we also experience emotions about having felt that emotion. However, unlike primary emotions, which is an appraisal-driven process elicited by a given event (Scherer, 2005), metaemotion involves the elicitation of a given thought or emotion about an emotion (Gottman et al., 1996). Gottman and colleagues (1996) conceptualize an individual’s thoughts and feelings about emotions as a profound approach to emotions:

For some people, emotions are a welcome and enriching part of their lives; they believe, in a fundamental way, that it is OK to have feelings. However, for other people, emotions are to be avoided and minimized; the world of negative emotions is seen as dangerous (p. 245).

**Metaemotion as a Multifaceted Construct: Underlying Factors and Related Concepts**

Researchers have argued for the necessity of further investigations of metaemotion to uncover the underlying dynamics of the construct. For instance, Mendonça (2013) argues that “though the existence of metaemotions is undeniable, their impact has not been totally explored and literature upon the subject is still scarce” (p. 390). Norman and Furnes (2016) also emphasize the importance of exploring this concept further and propose metaemotion as a multifaceted construct (i.e., metaemotional knowledge, experiences, and strategies). Moreover, Norman and Furnes (2016) define metaemotional knowledge as an organized set of declarative and general knowledge about one’s own and others’ emotions, as well as behavioral and situational factors. However, although EI also refers to vital aspects of an individual’s emotional understanding, metaemotional knowledge involves the awareness of, as well as internal beliefs and general knowledge about, one’s own and others’ emotions (Norman & Furnes, 2016).

Norman and Furnes (2016) conceptualize metaemotional experiences as an individual’s experiential feelings that accompany one’s own emotions or thoughts about emotions. According to this notion, one related construct is emotional clarity, which pertains to the emotional awareness of the cause and distinctiveness of emotions (Boden & Berenbaum, 2011; Salovey, Mayer, Goldman, Turvey & Palfai, 1995). Another related concept to note is the facet difficulty in identifying feelings of the Toronto Alexithymia Scale, which refers to the appraisal-driven process involved in discriminating experiential feelings from bodily sensations (Bagby, Parker & Taylor, 1994). However, whereas these constructs include appraisals of primary emotions, metaemotional experiences involve the phenomenological aspects of past, present and future emotions, which may occur at various levels of subjective awareness and control (Norman & Furnes, 2016).

Norman and Furnes (2016) propose metaemotional strategies as the control and regulatory mechanisms of metaemotion. An evident construct related to this notion is emotional self-regulation. For instance, reappraisal involves the reappraising processes of situations and emotional states to regulate emotions (John & Gross, 2004). The concept of mood repair also relates to essential aspects of metaemotional strategies. Specifically, Salovey et al. (1995) argue that regulation of mood involves an individual’s attitudes and perceptions of one’s own emotional experiences. However, although individuals’ perception of emotions involves cognitive appraisals of emotional experiences, it does not necessarily mean that it is devoted to primary cognitions exclusively. Accordingly, these processes might nevertheless involve vital aspects of an individual’s acquisition and
application of emotional information, which in turn organize metaemotional strategies (Norman & Furnes, 2016).

Metaemotion in Relation to Emotional Functioning and Interpersonal Processes

According to the framework by Gottman et al. (1996), metaemotion organizes parents’ perception of their own and their child’s emotions. Specifically, they postulate that parents who perceive one’s own emotions as essential also perceive their child’s emotions as important, whereas parents who perceive one’s own emotions as secondary also perceive their child’s emotions as unimportant. Research has reported that parents with dismissing-metaemotion show dismissive attitudes about their child’s emotions, and parents with coaching-metaemotion show supportive attitudes about their child’s emotions (Katz, Maliken & Stettler, 2012). Morey and Gentzler (2017) reported that parents with coaching-metaemotion rated greater responsiveness to negative emotions and accuracy when labeling children’s facial expressions. Furthermore, attention to metaemotion has recently emerged in educational settings as well. Research has shown that teachers with beliefs about emotions as important tend to pay attention to and regulate children’s emotions more efficiently (Ciucci, Baroncelli, Toselli & Denham, 2018).

Despite Gottman and colleagues (1996) definition of metaemotion as an individual’s thoughts and feelings about one’s own and others’ emotions, researchers have not systematically investigated adults’ metaemotion within the general population. Consequently, knowledge about the underlying mechanisms and dynamic interplay between an individual’s perception of another person’s emotions relative to one’s internal concepts and emotions remains unknown. However, other domains have related adults’ internal beliefs and perceptions to emotional functioning, interpersonal processes, as well as emotion recognition ability. For instance, research has linked adults with insecure attachment to attentional biases and distorted emotion recognition ability (Farley, Niedenthal, Marks, Brumbaugh & Vicary, 2006). Furthermore, attachment researchers argue that persons with avoidant attachment might be more attentive to emotional cues in order to down-regulate emotions more efficiently. For instance, Maier et al. (2005) reported that adults with dismissing attachment tendencies show higher vigilance to social and emotional stimuli. Nevertheless, researchers also theorize that adults with anxious attachment have maladaptive emotional processing of emotional cues due to insufficient emotion regulation (Dewitt & Houwer, 2008).

The Present Study

Regardless of the magnitude of research about emotion recognition ability, knowledge about the underlying psychological factors within the general population is still limited. Moreover, although researchers on metacognition and metaemotion have related these concepts to vital aspects of socioemotional processes, knowledge about the general population in that regard is still limited. Additionally, the majority of researchers in emotion recognition conduct their experiments in laboratory settings, even though online-surveys may be an efficient approach. Therefore, the primary aims of this study are: a) to examine if a multimodal online-assessment is a feasible method to investigate emotion recognition ability within the general population, and b) to investigate associations between multimodal emotion recognition ability and facets of metacognition and metaemotion.
Data obtained in laboratory settings are a valid but resource-demanding procedure. However, if online assessments yield valid and reliable results, this approach could facilitate both the investigation of emotion recognition ability with lesser resources and to obtain comprehensive samples more efficiently. Therefore, the research question for the first aim is:

**Q1:** Does an online-distributed multimodal emotion recognition task yield valid and reliable results that are consistent with prior data collected in laboratory settings?

Research has linked metacognitive functions to important aspects of socioemotional functioning; however, research on metacognition and emotion recognition ability is still limited. Although researchers have proposed metaemotion as essential for emotional functioning, research has not yet investigated how the notion of metaemotion relates to adults’ emotion recognition ability within the general population. For that reason, the investigation of how adults’ metacognition and metaemotion is associated with emotion recognition ability is of both empirical and conceptual importance. Therefore, the research questions for the second aim are:

**Q2:** Do adults know when they are accurate at identifying others’ emotions through multiple modalities?

**Q3:** Does adults’ meta-accuracy in an absolute sense correlate with multimodal emotion recognition ability?

**Q4:** Do metaemotional facets correlate with adults’ multimodal emotion recognition ability?

### Method

**Participants**

The sample consisted of students from three universities in Middle Sweden. Only students who understood Swedish were qualified to participate. In the present study, 106 of 135 distributed surveys were completed (completion rate 79%). Ninety-six participants were students who participated in the study for course credit. The participants consisted of 82 females and 24 males and were between 20 to 53 years old (\(M = 26, SD = 6\)).

**Material**

*Multimodal emotion recognition ability.* To assess adults’ emotion recognition, an online version of the Emotion Recognition Assessment in Multiple modalities (ERAM) task was applied (Laukka et al., 2015; see also Holding et al., 2017; Hovey et al., 2018). ERAM is a multimodal emotion recognition task with forced choice responses. The test aims at measuring individuals’ ability to identify emotions through visual, auditory, and multimodal modalities. The video clips are from the Geneva Multimodal Emotion Portrayals (GEMEP) corpus and consist of actors pronouncing pseudo-sentences while portraying a specific emotion (Bänziger, Mortillaro & Scherer, 2012). ERAM consists of 72 video clips and 12 emotions: (sadness = sorg; anger = ilska; despair = förtvivlan; anxiety = ängslan; relief = lättnad; joy = glädje; interest = intresse; pride = stolthet; fear = rädsla; disgust = avsmak; irritation = irritation; and pleasure = välbehag). The test included ten different actors; each video clip was a frontal view of the actor’s upper torso and face, which displayed bodily, facial, and vocal emotional cues. The duration of the clips ranged from 1 to 5 seconds and was presented in three conditions: 24 video-only,
24 audio-only, and 24 video-audio. The coding of the scores was dichotomous: 1 represented correct responses and 0 incorrect responses. The total score for each modality ranged from 0 to 24.

**Metacognition.** In the present study, adults’ metacognition was assessed using retrospective trial-by-trial confidence judgments. Retrospective confidence judgments are a valid approach to assess individuals’ metacognitive monitoring (Nelson, 1984). Participants were asked (after each video clip) to indicate how certain they were that they had identified the correct emotion using a slider. The sliders were anchored at 0 % and 100 % at each end, and had fixed intervals of ten units (i.e., 0 %, 10 %, 20 %, 30 %, etc.). Both participants’ absolute and relative meta-accuracy was calculated to determine to what extent their level of confidence corresponded and correlated with their actual performance. The calculation of participants’ absolute meta-accuracy was achieved by subtracting the mean proportion of correct answers with the mean proportion of confidence, and compiled into O/U-indexes. The O/U-indexes range from -1 to +1. Negative values denote under-confidence, which means that participants on average are less confident relative to the proportion correct. Positive values denote over-confidence, which means that participants are on average more confident relative to the proportion correct. A value of zero denotes perfect calibration, which represents that participants’ confidence is equivalent to the proportion correct. The calculations of participants’ relative meta-accuracy were done by correlating the mean proportion correct with mean proportion of confidence.

**Metaemotion.** The present study used three instruments involving aspects of metaemotion: the twenty-item Toronto Alexithymia Scale (TAS-20), the Trait Meta-Mood Scale (TMMS), and the Meta-Emotion Scale (MES). Due to the absence of a published Swedish version of the TMMS and MES, the author has translated these scales to Swedish with guidance from the thesis supervisor (the translated scales are available in Appendix A and B).

The TAS-20 measures alexithymia and emotional understanding (Bagby et al., 1994). However, researchers propose TAS-20 also to involve metaemotional aspects (Norman & Furnes, 2016). The present study used the validated Swedish version by Simonsson-Sarnecki et al. (2000). The scale consists of 20 items and is a self-reported measurement with three subscales. Items are rated using a 5-point Likert scale (1 = strongly disagree, 2 = disagree, 3 = neither, 4 = agree, and 5 = strongly agree). The difficulty identifying feelings subscale (DIF) measures the ability to recognize one’s feelings and include seven items (e.g., Jag har känslor som jag inte riktigt kan identifiera). The difficulty describing feelings subscale (DDF) measures the ability to communicate emotions and includes five items (e.g., Det är svårt för mig att hitta rätt ord för mina känslor). The externally-oriented thinking subscale (EOT) measures cognitive style regarding one’s preferences at attending to external rather than internal factors and includes eight items (e.g., Jag föredrar att prata med folk om deras vardagsaktiviteter, snarare än om deras känslor). Items 4, 5, 10, 18, and 19 were reversed, and scores were calculated separately for each dimension and not as a global index. Lower scores on the subscales indicate higher levels of metaemotion, while higher scores represent lower levels of metaemotion. In the present study, the internal consistency of the subscales was relatively good: DIF; α = .80, DDF; α = .85, and EOT; α = .63.
The TMMS measures the capability to regulate, perceive, and discriminate emotions, and is a commonly used instrument in assessments of EI. The present study used the short version, which consists of 30 items in total (Salovey et al., 1995). The scale is a validated self-reported instrument with three subscales. Items are rated using a 5-point Likert scale (1 = strongly disagree, 2 = disagree, 3 = neither, 4 = agree, and 5 = strongly agree). The attention to feelings subscale (TMMS-A) measures an individual’s preferences to think and notice emotions and consists of 13 items (2, 3, 4, 7, 10, 12, 17, 18, 21, 23, 24, 27, and 29, see Appendix A). The clarity of feelings subscale (TMMS-C) measures the ability to understand one’s own emotional experiences, and consists of 11 items (5, 6, 11, 14, 15, 16, 20, 22, 25, 28, and 30). The mood repair subscale (TMMS-R) measures an individual’s efforts to maintain pleasant moods or repair unpleasant moods and consists of six items (1, 8, 9, 13, 19, and 26). The following items were reversed during coding: 2, 3, 4, 5, 9, 11, 14, 16, 17, 18, 19, 22, 23, 27, and 29. High scores on the subscales represent higher levels of metaemotion, and low scores on the subscales represent lower levels of metaemotion. The Swedish version showed a relatively good internal consistency: TMMS-A, α = .84; TMMS-C, α = .87; and TMMS-R, α = .79.

The MES measures metaemotional experiences and consists of 28 items in total (Mitmansgruber, Beck, Höfer & Schüßler, 2009). The scale is a validated self-reported instrument involving positive and negative metaemotions. Items are rated using a 6-point Likert scale (1 = “is not at all true for me” to 6 = “is completely true for me”). The participant was instructed to rate the items according to their actual experiences and not as they think that they should react. The positive dimension measures positive valence metaemotions and includes meta-compassionate care with seven items (2, 4, 7, 9, 11, 22, and 24, see Appendix B) and meta-interest with five items (13, 16, 18, 23, and 28). The negative dimension measures negative valence metaemotions and include meta-anger with four items (3, 5, 15, and 27); meta-contempt/Shame with five items (12, 14, 17, 20, and 21); meta-tough control with five items (1, 6, 8, 19, and 26); and meta-suppression with two items (10 and 25). Item 20 was reversed during the coding of scale scores. The calculations of the scores were dimensional, and organized one positive index (P-MES) and one negative index (N-MES). High scores on the positive index represent higher levels of positive metaemotions, and low scores lower levels of positive metaemotions. Likewise, high scores on the negative index represent higher levels of negative metaemotions, and low scores lower levels of negative metaemotions. The Swedish version showed good internal consistency: P-MES, α = .80; and N-MES, α = .91.

Procedure
Permission was granted from the course coordinators at the universities to recruit participants’ email addresses during lecture breaks, as well as via an electronically posted call of interest at university student websites. The participants were clarified that the obtained contact information was for the distribution of the survey only, and did not constitute the basis for their consent. Google forms was used to collect participants’ call of interest online, wherein they entered their email addresses without seeing other individuals’ contact information. The online-posted call of interest was downloaded and deleted from the server when the data collection was ended.

The survey was distributed using the email invitation function in Qualtrics.com. The participants received an individual invitation with an anonymized link. The invitation included the aim of the study, technical requirements (i.e., only performing the test on a
computer), and recommendations such as recommended browsers and use of headphones. The participants were required to give an informed consent online by either selecting “Yes, I consent” to continue or “No, I do not consent” to withdraw and not participate. If a student did not consent, an online message was displayed, which thanked the student for shown interest and then terminated the survey. The “prevent ballot box stuffing” function in Qualtrics.com was used to prevent students from participating more than once. The survey was designed so that the participant was unable to go back to check their answers, and used the “request answer” function to notify in cases of missed answers, thereby prevent participants to not answering all items by mistake.

The survey encompassed 81 questions, 75 video clips, and 75 confidence judgments (including three example trials) which took approximately 35 to 40 minutes to complete. The survey consisted of four blocks. The first block was demographics wherein participants entered their year of birth, gender and whether they were a university student or not. The participant was terminated from the survey if they were not a university student or under 18 years of age. The second block included three different questionnaires and constituted 78 items in total. The third block was an example trial of the emotion recognition task and consisted of three practice video clips (video-only, audio-only, and video-audio) and three retrospective trial-by-trial confidence judgments. The final block included the ERAM emotion recognition task with 72 video clips and 72 confidence judgments. All instructions were in Swedish (see Appendix C for more information). Researchers at the University of Ghent programmed the online version of ERAM.

Data Analysis
The data were analyzed using SPSS version 25.0. Extreme values were excluded using the outlier-labeling rule (Hoalgin & Iglewicz, 1987). In the evaluation of the online version of ERAM, data were compared with baseline data obtained from a sample of 600 persons measured in lab settings (Laukka et al., 2015). The “psychometric” package in R-studio was used in analyzing the item discrimination (Fletcher, 2015).

Result

Descriptive Statistics and Validity of the Multimodal Emotion Recognition Task
In the evaluation of the online version of ERAM, mean accuracy for overall performance and each modality was calculated and compared with the baseline data. Figure 1a shows mean accuracy for overall performance at easy and difficult items in the online-distributed and prior data. In the online-distributed data, mean accuracy for overall performance was 68 % (SD = 13 %) for easy items and 45 % (SD = 12 %) for difficult items. The mean difference for easy and difficult items was significant, $M = 23 \%, SD = 10 \%, t(104) = 24.06, p < .001; CI: 95 \% [.21, .24]$. In prior data, mean accuracy for overall performance was 66 % (SD = 12 %) for easy items and 46 % (SD = 11 %) for difficult items. This indicates that the online-distributed emotion recognition test and prior data is consistent. Figure 1b shows the mean accuracy for overall performance in the online-distributed and prior data. The mean accuracy was 56 % (SD = 10 %) in prior data and 58 % (SD = 10 %)
in the online-distributed data. The result suggests that the online-distributed data is consistent with prior data obtained in lab settings.

Figure 1. (a) Mean accuracy in percentages for overall performance with error-bars [CI: 95 %]. Prior data: valid $N = 593$, and online-data: valid $N = 105$. (b) Mean accuracy for overall performance in percentages with error-bars [CI: 95 %] for prior data: valid $N = 593$, and online-data: valid $N = 105$.

Figure 2 shows mean accuracy across emotion recognition modalities for the online-distributed and prior data. In the visual modality, the mean accuracy was 54 % ($SD = 13$ %) in prior data and 55 % ($SD = 14$ %) in the online-distributed data. In the auditory modality, the mean accuracy was 48 % ($SD = 12$ %) in prior data and 50 % ($SD = 14$ %) in the online-distributed data. In the multimodal modality, the mean accuracy was 66 % ($SD = 13$ %) in prior data and 68 % ($SD = 12$ %) in the online-distributed data. These results suggest that the online-distributed data is consistent with prior data conducted in traditional settings across modalities.

Figure 2. Mean accuracy in percentages for each emotion recognition modality with error-bars [CI: 95 %]. Prior data: valid $N = 593$, Online-data: video-only; valid $N = 106$, audio-only; valid $N = 106$, video-audio; valid $N = 105$. 
Mean differences were analyzed using a within-subject ANOVA for repeated measures. The one-way ANOVA shown a significant difference between modalities, $F(2, 103) = 135.59, \ p < .001, \ \eta^2 = .73$. Pairwise multiple comparisons (Bonferroni corrected t-tests) showed that mean accuracy was greater for video-audio relative to video-only (mean difference $M = 12 \ %, \ SE = .01, \ p < .001, \ CI: 95 \ % [.10, .15]$) and audio-only (mean difference $M = 17 \ %, \ SE = .01, \ p < .001, \ CI: 95 \ % [.14, .20]$). This result indicates that emotion recognition accuracy was greatest in the multimodal condition. The Bonferroni test showed that mean accuracy was greater for video-only relative to audio-only (mean difference $M = 5 \ %, \ SE = .02, \ p = .010, \ CI: 95 \ % [.01, .80]$). This result suggests that emotion recognition accuracy was slightly greater in the visual modality relative to the auditory modality.

Emotion recognition accuracy was also analyzed separately for easy and difficult items across modalities for the online-distributed data (see Figure 3). For video-only, the mean accuracy for easy items was 65 % ($SD = 17 \ %$) and for difficult items 42 % ($SD = 15 \ %$). The mean difference for easy and difficult items was significant, $M = 23 \ %, \ SD = 17 \ %, \ t(105) = 13.86$. For audio-only, the mean accuracy for easy items was 60 % ($SD = 16 \ %$) and for difficult items 41 % ($SD = 17 \ %$). The mean difference in the proportion for easy and difficult items was significant, $M = 19 \ %, \ SD = 17 \ %, \ t(105) = 11.77$. For video-audio, the mean accuracy for easy items was 80 % ($SD = 14 \ %$) and for difficult items 55 % ($SD = 15 \ %$). The mean difference in accuracy for easy and difficult items was significant, $M = 25 \ %, \ SD = 16 \ %, \ t(104) = 15.82$.

The results showed no significant difference in mean accuracy between participants that completed the questionnaires before ERAM ($M = 59 \ %, \ SD = 10 \ %$) and after ERAM ($M = 57 \ %, \ SD = 11 \ %$), $t(103) = .78, \ p = .436, \ CI: 95 \ % [-.02, .05]$. Furthermore, demographics were analyzed using Pearson correlations, wherein the encoding was 1 for females and 0 for males. The results showed no gender-related ($r = .17, \ p = .091$) or age-related ($r = -.07, \ p = .432$) differences in emotion recognition accuracy.
Reliability of the Online-Distributed Multimodal Emotion Recognition Task
An item discrimination analysis was used to evaluate the reliability of the discrimination between low- and high performing participants, and indexes for easy and difficult items were computed separately for each modality. The index for the video-only items ranged from .02 to .60, which indicates that the overall discrimination between high- and low performing participants is relatively good. The item discrimination index for easy video-only items ranged from .02 to .60, whereas the index for difficult items ranged from .20 to .57. This indicates that the overall discrimination between high- and low performing participants was relatively good for difficult video-only items, but relatively low for easy items. The index for the audio-only items ranged from -.08 to .60, which indicates that the overall discrimination between high- and low performing participants was relatively low and in some cases negative. Accordingly, negative item discrimination indicates that low performing participants tend to perform better than high performing participants do for those specific items. However, in the analysis of easy audio-only items, the index ranged from 0 to .54 and for difficult items from .20 to .57. This indicates that the discrimination between high- and low performing participants was relatively low for easy audio-only items, but relatively good for difficult items. The index for the video-audio items ranged from .02 to .54, which indicates that the overall discrimination between high- and low performing participants is relatively low. The item discrimination index for easy video-audio items ranged from .06 to .66 whereas the index for difficult items ranged from .20 to .49. This indicates that discrimination between high- and low performing participants was relatively low for easy video-audio items, but relatively good for difficult items.

Metacognition and Multimodal Emotion Recognition Ability
Relative meta-accuracy was analyzed using Goodman-Kruskal gamma correlation. The values range from -1 to +1. Positive values represent agreement; negative values represent inversion, whereas zero represents no association. Mean values ranged from 0 to 100 % and ranked with an interval of ten units (i.e., 0 %, 10 %, 20 %, 30 %, etc.). The gamma correlation between relative meta-accuracy and emotion recognition accuracy was unrelated across modalities; video-only: \( G(106) = .03, p = .803 \); audio-only: \( G(106) = -.01, p = .887 \); and video-audio: \( G(105) = -.04, p = .688 \). The results suggest that participants did not know when they identified an emotion accurately.

The O/U-indexes refer to the absolute difference of the mean proportion correct and mean proportion confidence, with values ranging from -1 to +1. Negative values indicate under-confidence, zero perfect calibration, and positive values over-confidence. The O/U-index for the total performance was negative and significantly different from zero, \( M = -.07, SD = .17; t(104) = -3.98, p < .001, CI: 95 \% [-.10, -.34] \). The result suggests that participants were slightly under-confident relative to the proportion correct. The O/U-indexes for each modality were negative and significantly different from zero, video-only \( (M = -.09, SD = .20); t(105) = -56.55, p < .001, CI: 95 \% [-1.13, -1.05] \), audio-only \( (M = -.08, SD = .21); t(105) = -52.47, p < .001, CI: 95 \% [-1.19, -1.04] \), and video-audio \( (M = -.05, SD = .19); t(105) = -56.76, p < .001, CI: 95 \% [-1.05, -1.09] \). The O/U-indexes for each modality indicate that participants were slightly under-confident relative to the proportion correct. The one-way ANOVA for repeated measures showed a
main effect between modalities, $F(2, 103) = 4.45, p = .014, \eta^2 = .08$. The Bonferroni post-hoc test showed that the O/U-index for video-only was lower relative to the O/U-index for video-audio, $M = .04, SE = .02, p = .030, CI: 95\% [-.08, .00]$. The mean differences were not significant between audio-only and video-only or between audio-only and video-audio.

In the analysis of adults’ absolute meta-accuracy relative to multimodal emotion recognition ability, the O/U-indexes and emotion recognition accuracy were analyzed using Pearson correlations. The Pearson correlation showed a positive relationship between the O/U-index and overall emotion recognition accuracy ($r = .59, p < .001$). The result indicates that calibrated participants had greater emotion recognition accuracy than under-confident participants. The correlations for each modality showed a positive relationship between the O/U-indexes and emotion recognition accuracy. For video-only, there was a positive and relatively high correlation between the O/U-index and emotion recognition ability ($r = .68, p < .001$). Similar correlations were observed between the O/U-index and audio-only emotion recognition ($r = .65, p < .001$) and video-audio emotion recognition ($r = .62, p < .001$). These results indicate that more well-calibrated participants show greater emotion recognition accuracy compared to more under-confident participants across modalities.

Metaemotion and Multimodal Emotion Recognition Ability
Metaemotion and multimodal emotion recognition were analyzed using Pearson correlations to examine the direction and strength of the associations. Table 1 shows results from the correlation analyses and descriptive statistics. Only one of the metaemotion measurements correlated significantly with emotion recognition ability. A positive correlation was observed between N-MES and total emotion recognition ability ($r = .27, p = .007$), which suggests that adults’ with higher levels of negative metaemotions have greater emotion recognition ability than adults’ with lower levels of negative metaemotions. Similar correlations were also observed for each presentation modality: video-only ($r = .25, p = .009$), audio-only ($r = .27, p = .005$), and video-audio ($r = .21, p = .031$).

The correlation analyses further showed moderate associations between the conditions in the emotion recognition task, which indicates that emotion recognition from video-only, audio-only, and video-audio stimuli constitute related but still independent measures of adults’ overall emotion recognition ability.

Finally, the correlations between the various metaemotion instruments indicated that the TAS-20 subscales correlated moderately with each other. The result shows a positive correlation between DIF and DDF, which indicates that adults’ with greater difficulties in identifying feelings also have greater difficulties in describing feelings. In addition, the DDF and EOT correlated positively, which suggests that adults’ with greater difficulties in describing feelings, also have greater externally-oriented thinking. The TMMS subscales also correlated moderately with each other. A positive correlation between TMMS-A and TMMS-C was observed, which suggests that adults’ with greater attention to feelings may also have greater emotional clarity. Additionally, the TMMS-C and TMMS-R correlated positively, which suggests that adults’ with greater emotional clarity also have greater emotional repair.
Table 1
Pearson Correlation between Metaemotion Instruments and Multimodal Emotion Recognition Accuracy

<table>
<thead>
<tr>
<th>Variables</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. DIF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>16.96</td>
<td>5.39</td>
</tr>
<tr>
<td>2. DDF</td>
<td>.62***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12.87</td>
<td>4.75</td>
</tr>
<tr>
<td>3. EOT</td>
<td>.17</td>
<td>.30**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>16.92</td>
<td>4.34</td>
</tr>
<tr>
<td>4. TMMS-A</td>
<td>-.22*</td>
<td>-.33***</td>
<td>-.63***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>50.48</td>
<td>7.22</td>
</tr>
<tr>
<td>5. TMMS-C</td>
<td>-.82***</td>
<td>-.67***</td>
<td>-.21*</td>
<td>.29**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>36.75</td>
<td>8.00</td>
</tr>
<tr>
<td>6. TMMS-R</td>
<td>-.51***</td>
<td>-.38***</td>
<td>.13</td>
<td>.08</td>
<td>.47***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>19.70</td>
<td>4.86</td>
</tr>
<tr>
<td>7. P-MESa</td>
<td>-.26**</td>
<td>-.18</td>
<td>-.38***</td>
<td>.34***</td>
<td>.34***</td>
<td>.43***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>44.97</td>
<td>8.80</td>
</tr>
<tr>
<td>8. N-MESb</td>
<td>.51***</td>
<td>.41***</td>
<td>.09</td>
<td>.03</td>
<td>-.53***</td>
<td>-.45***</td>
<td>-.12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>60.10</td>
<td>5.64</td>
</tr>
<tr>
<td>9. Video</td>
<td>.06</td>
<td>.10</td>
<td>-.09</td>
<td>.09</td>
<td>-.06</td>
<td>-.05</td>
<td>-.04</td>
<td>.25**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>55%</td>
<td>14%</td>
</tr>
<tr>
<td>10. Audio</td>
<td>.07</td>
<td>.10</td>
<td>-.12</td>
<td>.11</td>
<td>-.06</td>
<td>-.06</td>
<td>.10</td>
<td>.27**</td>
<td>.37***</td>
<td></td>
<td></td>
<td></td>
<td>50%</td>
<td>14%</td>
</tr>
<tr>
<td>11. Video-Audioc</td>
<td>.11</td>
<td>.06</td>
<td>.02</td>
<td>.02</td>
<td>-.11</td>
<td>-.13</td>
<td>-.17</td>
<td>.21*</td>
<td>.57***</td>
<td>.52***</td>
<td></td>
<td></td>
<td>68%</td>
<td>12%</td>
</tr>
<tr>
<td>12. Totald</td>
<td>.13</td>
<td>.12</td>
<td>-.06</td>
<td>.06</td>
<td>-.12</td>
<td>-.12</td>
<td>-.12</td>
<td>.27**</td>
<td>.78***</td>
<td>.76***</td>
<td>.86***</td>
<td></td>
<td>58%</td>
<td>10%</td>
</tr>
</tbody>
</table>


a One pairwise excluded as missing value, *N* = 105.
b One pairwise excluded as missing value, *N* = 105.
c One excluded as outlier, *N* = 105.
d One excluded as outlier, *N* = 105.
Discussion

The primary aims of the study were to examine if a multimodal online-assessment is a feasible method to investigate emotion recognition ability within the general population, as well as to investigate correlations between multimodal emotion recognition ability and measures of metacognition and metaemotion. The evaluation of the multimodal emotion recognition task showed good validity and reliability. In comparison with the baseline data (Laukka et al., 2015), results indicate good consistency between the online version of ERAM and previous data collected in lab settings. The analysis of metacognition showed a positive correlation between the O/U-indexes and emotion recognition ability for overall performance and across modalities. This suggests that more calibrated adults had greater emotion recognition ability than adults who were more under-confident. The results showed no associations between relative meta-accuracy and emotion recognition ability, which indicates that adults did not know when they identified an emotion accurately. In the analysis of metaemotion, only one of the instruments correlated with emotion recognition accuracy, which suggests that the other metaemotional facets may be unrelated to adults’ multimodal emotion recognition ability. Nevertheless, negative metaemotions and multimodal emotion recognition ability correlated positively, both in overall performance and across modalities. These results suggest that adults with higher levels of negative metaemotions have greater emotion recognition ability than adults with lower levels of negative metaemotions.

**Evaluation of the Multimodal Online-Assessment**

The online-assessment shows good consistency with the baseline data (Laukka et al., 2015), which suggests that adults’ multimodal emotion recognition ability is assessable online. The item analysis showed feasible reliability and an expected response pattern for easy and difficult items. This result indicates that the participants followed and understood the instructions in the survey. Specifically, the instructions included definitions of the emotions, as well as guidelines such as using headphones and only performing the test when they would not be disturbed, which are vital for completing the task as intended and for minimizing errors. Taken together, because the online-distributed and prior data yielded highly similar results, it is also important to emphasize that participants complied with the guidelines well without the presence of an experimenter. Furthermore, the item discrimination also shows relatively good discrimination between high- and low performing participants, which is vital for determining whether the variability in accuracy is due to the actual task or other factors. Therefore, because the item discrimination indexes are feasible, it indicates that the variability originates from the variation in participants’ emotion recognition ability and not potential disruptions in the instructions or systematic errors in the data.

In sum, the results indicated that the online-assessment was highly consistent with the previous data (Laukka et al., 2015; see also Holding et al., 2017; Hovey et al., 2018), although the sample size in the present study was relatively restricted. Therefore, the results strongly suggest that online assessment of adults’ multimodal emotion recognition ability is a reliable and valid approach for future researchers to consider.
The Relationship between Metacognition and Multimodal Emotion Recognition

The present study found no associations between adults’ relative meta-accuracy and emotion recognition accuracy, regardless of modality. These results suggest that participants did not know when they identified an emotion accurately, or at least showed a lack of consistency in their level of confidence relative to their average performance. However, the present study did not analyze relative meta-accuracy together with item difficulty and emotion category. Therefore, future research should focus on the dynamics of task parameters such as item difficulty and emotion categories in the investigation of adults’ relative meta-accuracy and multimodal emotion recognition ability.

The study found that participants were slightly under-confident regardless of modality. In the correlation analysis of the O/U-indexes and emotion recognition ability, the results showed a relatively high and positive correlation for each modality, as well as for overall performance. This suggests that more calibrated adults have greater emotion recognition ability relative to adults that are more under-confident in the visual, auditory, and multimodal modality. These results are consistent with assumptions in metacognition (Keren, 1991). However, because research regarding metacognition and emotion recognition ability is limited, this conclusion should be treated as tentative only. Furthermore, in the multiple comparisons of the O/U-indexes, results showed that participants were slightly more under-confident in the visual modality relative to the multimodal modality, which suggests that adults’ confidence judgments might be less biased in the multimodal modality than in the visual modality. From this notion, the lower under-confidence in the multimodal modality might reflect that participants’ ratings were more consistent with their actual performance in an absolute sense. However, further research is necessary to reveal what mechanisms might underlie these tendencies.

Variability of Relative Meta-Accuracy between Multimodal and Unimodal Assessments

The present study found no relationship between adults’ relative meta-accuracy and emotion recognition ability, which is inconsistent with the results Kelly and Metcalfe (2011) reported in their study of healthy adults. However, it is important to note that despite the fact that both studies used the same methods to measure metacognition, they included different emotions, stimuli, and modalities – which are essential factors to consider. From this perspective, because the ERAM emotion recognition task has both easy and difficult items, as well as 12 emotions with varying intensity and valence, these factors might have affected the consistency of the judgments to some extent. For instance, previous research has reported that item difficulty is a factor that might influence individuals’ meta-accuracy (Juslin et al., 2000). Nevertheless, the inconsistent findings might also be due to different approaches to the measurement of adults’ emotion recognition ability. For instance, the present study measured 12 emotions, whereas in the study by Kelly and Metcalfe (2011) there were only five emotions assessed (including only one positive emotion). Accordingly, Bänziger et al. (2009) emphasized that a limited set of emotions and diametrical arousal dimensions might introduce the risk of measuring individuals’ discrimination ability rather than emotion recognition ability. In addition, because the current study did not analyze individual emotion categories, valence or intensity ratings, the extent to which these factors might have affected the consistency of the ratings remains unknown. Consequently, future research should focus on analyzing these components in relation to adults’ relative meta-accuracy to gain a more comprehensive knowledge.
Another aspect to consider in this context is the differences in the stimuli used to assess adults’ emotion recognition ability. Accordingly, Bänziger et al. (2012) emphasize that using still photographs of faces to measure dynamic processes might be problematic. From this perspective, the inconsistency between the present findings and results reported in Kelly and Metcalfe (2011) might relate to differences between using unimodal and multimodal assessments. Specifically, Kelly and Metcalfe (2011) assessed participants’ emotion recognition ability with still photographs of faces, whereas the present study used a multimodal emotion recognition task. However, researchers in metacognition have discussed whether self-monitoring is domain-general or if it consists of domain-specific processes (Gutierrez et al., 2016). From this perspective, if metacognitive monitoring were a domain-general process, individuals’ meta-accuracy would be relatively consistent even in low knowledge situations. Conversely, if metacognitive monitoring were a domain-specific process, individuals’ meta-accuracy would be relatively inconsistent between high- and low knowledge situations. Nonetheless, further research is required to evaluate if the current findings are robust. Therefore, future research should focus on using multimodal assessments to uncover the dynamics of adults’ relative meta-accuracy and emotion recognition ability.

**Dynamics of Under-Confidence in Research of Metacognition**

The current study found that adults tend to be slightly under-confident across modalities, and that well-calibrated adults had greater emotion recognition ability than adults who are more under-confident. Nonetheless, because research on metacognition and emotion recognition ability is a newly emerged area, knowledge of potential factors that might underlie adults’ under-confidence in emotion recognition is limited. In other research areas on metacognition, researchers have discussed the occurrence of biased judgments in various ways. For instance, Juslin et al. (2000) noted that biased confidence judgments might arise due to sampling error of items rather than cognitive biases. Furthermore, researchers have also discussed biases in terms of anchoring effects, wherein under-confidence may emerge when participants use low anchors (Yang et al., 2018). However, because the O/U-indexes in the present study were relatively low for overall performance and across modalities, an anchoring effect or task-related disruptions would most likely produce a highly skewed rather than a slightly skewed distribution. In conclusion, it is premature to conclude whether adults’ under-confidence in emotion recognition is due to task-related disruptions or cognitive biases, and further investigations are necessary.

Another aspect that might have contributed to adults’ under-confidence in emotion recognition is the utilization of non-diagnostic cues when estimating their performance. Accordingly, Koriat (1997) emphasize that individuals tend to rate higher on answers that come to mind more readily relative to those that come with less ease, even if they would be incorrect. Regarding the present findings, participants might have used processing difficulty as an indicator for incorrect responses, which may have facilitated biased ratings in emotion recognition. However, because the current study did not analyze the O/U-indexes for easy and difficult items, as well as emotion categories separately, it nevertheless remains unknown to which extent these factors might have contributed to this effect. For instance, if the accuracy probabilities vary between easy and difficult items, as well as emotion categories, it becomes problematic to analyze them together, rather than analyzing them separately. Therefore, future research should focus on
conducting a more systematic analysis of these components relative to adults’ meta-accuracy to evaluate the dynamics of multimodal emotion recognition ability.

Lastly, from a more speculative perspective, the observed under-confidence might originate from dissociations between adults’ evaluation and actual performance as a function of insufficient monitoring of one’s emotion recognition ability. Specifically, emotion recognition as a fundamental skill might consist of automatic processes that operate non-consciously, which in turn may affect individuals’ calibration to some extent. For instance, the perceptual decoding of emotional cues might operate on more subtle levels of cognitive control, and hence influence individuals’ self-monitoring and thereby meta-accuracy in emotion recognition. From this perspective, adults’ under-confidence might arise due to a lacking awareness, and reflectiveness of one’s ability to recognize others’ emotions. However, it is hard to speculate whether the observed under-confidence in emotion recognition is due to insufficient monitoring of fundamental skills and further research is required to explore this notion.

The Relationship between Metaemotional Facets and Multimodal Emotion Recognition
In the present study, only one of the metaemotion instruments correlated with adults’ multimodal emotion recognition ability. Nevertheless, the results showed a cohesive direction between the different subscales that is in line with previous research. For instance, subscales intended to measure adaptive aspects of metaemotion such as positive metaemotions, emotional clarity, repair, and attention, correlated negatively with negative metaemotions, difficulty in identifying and describing feelings, and externally-oriented thinking. However, it is premature to propose that the instruments do not have any moderating or mediating effects on each other. Furthermore, regarding the instruments used in this study, it is essential to note that they are not devoted to measuring metaemotion explicitly. Consequently, whether or to what extent different metaemotional facets relate to adults’ multimodal emotion recognition ability remains unknown until psychometric tests developed to measure the complexity of metaemotion are available.

Negative metaemotions and emotion recognition ability correlated positively, regardless of modality. This indicates that adults who have higher levels of negative metaemotions tend to perform better at recognizing others’ emotions relative to adults with lower levels of negative metaemotions. Furthermore, the relationship between the level of negative metaemotions and emotion recognition accuracy is highest in the auditory modality relative to the visual and multimodal modality. Taken together, the overall direction between negative metaemotions and emotion recognition accuracy appear to be relatively consistent across modalities.

Although the present study found only a relatively low correlation between adults’ negative metaemotions and emotion recognition ability, an evident aspect to consider in this context is the relatively low correlations generally observed between self-reports and ability-based measures (Bänziger et al., 2009; Schlegel et al., 2017). Associations between ability measures and self-reports vary across studies as well. For instance, Schlegel et al. (2017) reported a negative correlation between multimodal emotion recognition ability and TAS-20 in a large community sample, whereas in the current study none of the TAS-20 subscales correlated with adults’ multimodal emotion recognition ability. From this perspective, factors that might contribute to this variation may nevertheless relate to the discrepancy between an individual’s actual ability and the
subjective self-assessment of one’s emotional competencies. In addition, because the study did not analyze individual emotion categories, it is hard to speculate if the direction or strength of associations with metaemotion varies as a function of individual emotions. Consequently, further research is required to explore this notion, as well as to conclude to what extent these findings are robust among adults within the general population.

**Negative Metaemotions in Relation to Multimodal Emotion Recognition Ability**

The current study found that adults with higher levels of negative metaemotions have greater emotion recognition ability than adults with lower levels of negative metaemotions, regardless of modality. According to the framework proposed by Gottman et al. (1996), parents’ introspection and perception about their own emotions constitute the basis for how they perceive their child’s emotions. Additionally, research has linked adults’ negative beliefs to dismissing emotional perception of children’s emotions, as well as the development of dysfunctional emotional understanding among children (Ciucci et al., 2018; Katz et al., 2012; Morey & Gentzler, 2017). However, these findings pertain to the dynamics of adults’ metaemotion and possible influences in children’s emotional socialization and regulation, and do not explicitly state that adults with negative thoughts and emotions about emotions have a reduced or impaired emotion recognition ability. Therefore, a positive correlation between negative metaemotions and multimodal emotion recognition ability may be consistent with findings from other domains. For instance, attachment researchers theorize that individuals with a dismissing attachment state of mind might detect emotional cues more rapidly to avoid emotional experiences (Maier et al., 2005). From this perspective, adults with higher levels of negative metaemotions might be more attentive to emotional cues, which could facilitate their emotion recognition ability in the visual, auditory, and multimodal modality to some extent.

Another relevant aspect to discuss in relation to the present findings is negative metaemotions as metaemotional strategies. Accordingly, Norman and Furnes (2016) propose metaemotion as a multifaceted construct underpinned by metaemotional knowledge, experiences, and strategies, wherein the dynamic interplay constitutes vital aspects of individuals emotional functioning. In addition, Mitmansgruber et al. (2009) proposed negative and positive metaemotions as two dimensions of metaemotional experiences that constitute regulatory means. From this perspective, adults’ negative metaemotions might involve metaemotional strategies essential for emotion recognition ability. For instance, if metaemotional experiences constitute metaemotional strategies, negative metaemotions might organize underlying mechanisms of strategies aimed at recognizing and avoiding others’ emotions, which nevertheless become relevant for avoiding emotions in interpersonal relationships and social situations. However, further research is required to uncover the underlying structure and dynamics of adults’ metaemotion within the general population.

In the present study, neither emotion categories, valence, nor intensity was analyzed relative to adults’ negative metaemotions. Therefore, the extent to which these factors might relate to the relationship between adults’ negative metaemotions and multimodal emotion recognition ability thus remains unknown. However, previous research has linked psychiatric disorders as well as attachment patterns to variability in recognition of specific emotion categories (Farley et al., 2006). Although research has associated individual differences in emotion categories, valence, intensity, and modality with
psychological biases due to various dysfunctions, variability nevertheless also occurs within the general population. Consequently, future research should focus on investigating adults’ negative metaemotions and emotion recognition ability with multimodal assessments, as well as to conduct systematic analyses in order to obtain a more comprehensive knowledge.

**Limitations**

Overall, the validity and reliability of the findings are good. For instance, influences of potential errors that might invalidate or interfere with the data were reduced using different survey-tools in Qualtrics.com. The online-survey was designed so that responses could not be changed afterward, which is essential to minimize biased responses on the recognition task, questionnaires, as well as on the retrospective confidence judgments. Because this study used a student sample and that the sample size was relatively restricted, it is important to note that further research is necessary to determine if the findings are robust. However, the present study used ERAM to investigate adults’ emotion recognition ability, which in previous research has shown both good validity and reliability (Laukka et al., 2015; see also Holding et al., 2017; Hovey et al., 2018). One additional limitation of the present study that is essential to note is that the emotion recognition task was not analyzed using unbiased hit rates to validate the scores (see Wagner, 1993, for a critical review). Nevertheless, the overall pattern of responses showed no indication of systematic errors, which indicates that a further analysis to validate the scores is not necessary.

Regarding the metaemotion instruments, in the current study were three different self-reported measures used, which previous research has reported to be valid and reliable. However, an important limitation to address concerns the TMMS and MES. Specifically, the author translated the TMMS and MES to Swedish and did not test the instruments before using them. However, the Cronbach’s alphas were relatively good, which to some extent indicates that the translations are reliable. Furthermore, because the majority of the self-reported instruments used in the study were not originally or explicitly aimed to measure metaemotion, this could affect to what extent the scales actually reflect metaemotional facets. Accordingly, although specific items of the scales might pertain to metaemotional aspects, the subscales as a whole aim to measure other psychological features. Therefore, whether or to what extent metaemotional facets and multimodal emotion recognition ability are related remains inconclusive until psychometric tests devoted to measuring the complexity of metaemotion are available.

A final aspect to consider is the notion of metaemotion as a future area of research on emotion recognition ability within the general population. Accordingly, the necessity of developing metaemotion as a construct is evident; however, it is arguable whether this aspect should be treated as a unique contribution or not, and if it should be prioritized in research on emotion recognition. However, although the study only found a relatively low correlation, it is premature to conclude that metaemotion is unrelated to adults’ emotion recognition ability. Therefore, to conclude whether or to what extent adults’ metaemotion constitute fundamental aspects of emotion recognition ability or not, further research is required.

**Conclusions and Future Directions**

Online-assessment of adults’ multimodal emotion recognition ability within the general population appears to be a valid and reliable approach. For that reason, it is essential to
emphasize the significance and opportunity to investigate multimodal emotion recognition ability in relation to psychological factors more efficiently. Furthermore, the current study found that adults were slightly under-confident relative to the proportion accuracy on the emotion recognition task in an absolute sense, regardless of modality. In addition, the correlation analysis showed a positive and relatively strong relationship between adults’ level of confidence and emotion recognition ability across modalities, which indicates that more calibrated adults have greater emotion recognition ability than adults that are more under-confident in the visual, auditory, and multimodal modality. Conclusively, these results are essential findings both empirically and conceptually regarding the field of metacognition and emotion recognition. However, further research is necessary to uncover the underlying mechanisms of adults’ under-confidence in emotion recognition, especially by using multimodal assessments. Regarding metaemotion, although only one of the facets correlated with adults’ multimodal emotion recognition ability, the current study is nonetheless the first to explore this notion. In addition, further research is needed to uncover the relationships between an individual’s internal concepts and perceptions of one’s own and others’ emotions to conclude if this psychological construct constitutes an underlying mechanism of adults’ emotion recognition ability or not.

References


APPENDIX A

Swedish Translation of the Trait Meta-Mood Scale (TMMS)

1. Jag försöker att tänka bra tankar oavsett hur dåligt jag mår.
2. Människor skulle ha det bättre ifall de kände mindre och tänkte mer.
3. Jag tycker inte att det är vårt att fokusera på ens känslor eller sinnesstämning.
4. Jag bryr mig vanligtvis inte så mycket om vad jag känner.
5. Ibland kan jag inte säga vad jag har för känslor.
6. Jag är sällan förvirrad över hur jag känner.
7. Känslor ger riktning i livet.
8. Trots att jag emellanåt är leden har jag mestadels optimistiska förväntningar.
9. När jag är upprörd upptäcker jag att det goda i livet är illusioner.
12. Det bästa sättet för mig att hantera mina känslor är att uppleva dom till fullo.
13. När jag blir upprörd påminner jag mig själv om allt gott i livet.
15. Jag är ofta medveten om hur jag känner inför en fråga.
17. Man bör aldrig styras av känslor.
19. Trots att jag emellanåt är glad har jag mestadels pessimistiska förväntningar.
20. Jag känner mig bekväm med mina känslor.
26. Oavsett hur dåligt jag mår försöker jag tänka på bra saker.
27. Känslor är en svaghet människor har.
28. Vanligtvis känner jag till mina känslor inför en fråga.
29. Vanligtvis är det slöseri med tid att tänka på ens känslor.
30. Jag vet nästan alltid exakt hur jag känner.
1. När jag är ledsen eller ängslig blir jag krävande mot mig själv.
2. När jag känner mig stressad och upplever negativa känslor, behandlar jag mig själv med medkänsla.
4. Jag försöker att smickra mig själv för att må bättre när jag upplever en stor känslomässig börda.
5. Ibland kan jag bli riktigt arg på mig själv för hur jag reagerade känslomässigt.
6. Gång på gång tvingar jag mig själv att ta mig samman.
7. I stressfyllda situationer försöker jag att behandla mig själv med omtanke, på samma sätt som jag skulle ha gjort mot personer nära mig.
8. När jag ser mina känslor som olämpliga blir jag väldigt strikt med mig själv.
9. När jag är ledset eller ängslig, så gör jag något bra för mig själv så att saker och ting blir lättare.
13. Gång på gång upptäcker jag nya former av upplevelser i mig själv.
14. Gång på gång finns det situationer då jag kritiserar mig själv hårt.
15. Jag tycker ofta att mina känslomässiga reaktioner är felaktiga.
16. Mina tankar och känslor är en oändlig källa till information om mig själv.
17. När jag känner skuld för att jag gjorde ett misstag är jag rätt så oförsonlig med mig själv.
18. Negativa känslor förser mig med intressant information om mig själv.
19. När jag talar med mig själv i mina tankar är jag ofta hård mot mig själv.
20. När jag inte kan leva upp till mina egna förväntningar i det vardagliga livet klankar jag inte ner på mig själv.
22. Jag är snäll mot mig själv när jag känner att mina känslor är betungande.
24. När jag upplever starka negativa känslor, tröstar och uppmuntrar jag mig själv.
27. Gång på gång blir jag irriterad på mina löjliga känslomässiga reaktioner.
28. Gång på gång är mina tankar och känslor fascinerande och viktiga för mig.
APPENDIX C
Swedish Instructions for Example Trials

**Instruktioner**

Detta test har för avsikt att mäta din förmåga att känna igen känslouttryck i människors ansikten och röster.


Efter exemplen, kommer tre delar. (1) 24 inspelningar utan ljud, (2) 24 inspelningar med ljud utan bild, och (3) 24 inspelningar med både ljud och bild. Detta betyder att du i denna del kommer att ge 72 svar (som du väljer från en lista).


Du kommer att behöva ungefär 20 minuter för att slutföra testet. I och med att Du ska besvara alla frågorna löpande (man kan inte ta en paus och fortsätta senare) se till att Du inte kan bli störda.

---

För syftet med denna studie, definieras känsloalternativen enligt följande:

**Stolthet:** Känsla av triumf efter en framgång eller en personlig prestation.

**Ilska:** Extrem missnöje orsakad av någons orättvisa eller fientliga handlingar.

**Glädje:** Känsla framkallad av en fantastisk sak som uppstod oväntat.

**Irritation:** Att uppleva missnöje hos någon ting eller någon medans man fortfarande är lugn.

**Avsmak:** Motvilja orsakad av ett obehagligt objekt eller en miljö.

**Välbehag:** Upplevd känsla av välbefinnande och sensuell glädje.

**Sorg:** Känner sig nere efter förlust av en person, plats eller sak

**Lättnad:** Känsla av lugn i slutet eller upplösningen av en obehaglig, obehaglig eller till och med farlig situation.

**Förtvivlan:** Bekymrad av ett livsproblem utan lösning, tillsammans med en ovillighet att acceptera situationen.

**Intresse:** Att vara fascinerad eller har en uppmärksamhet fångad av en person eller en sak.

**Rädsla:** Att stå inför en överhängande fara som hotar ens fysiska välbefinnande.

**Ångslan:** Rädsla för eller oro för konsekvenserna av en situation som kan vara ogylnsamma för sig själv eller någon nära.
**Block intro 3 exempel**

Detta är det första exemplet.

Klicka på ”fortsätt” för att se en inspelning utan ljud.

Du kan klicka på ”se igen” för att se inspelningen igen.


Om videon inte spelas upp kan Du inte göra testet med de inställningar som Din dator har nu. Var snäll och informera den person som bad dig göra testet, och tala om att video inte spelas upp på din dator.

Vilken känsla visade personen?

Direkt efter Du sett videon, får du dessa känslor-ord att välja mellan.

Din uppgift är att välja det känslor-ord som bäst överensstämmer med den känsla som Du tyckte personen visade.

Vänligen välj ett känslor-ord för att fortsätta till nästa exempel.

Hur säker är du att du identifierat rätt emotion?

Om du är helt säker skall du ange 100 % på skalan.

Om du anger t.ex. 60 % betyder det att du inte är helt säker, utan kommer i idealfallet att ha rätt i just 60 % av fallen (d.v.s. i 6 fall av 10). På samma sätt fungerar de övriga kategorierna.

Om du anger 0 % är du istället helt säker på att du svarat fel.

Detta är det andra exemplet.

Klicka på ”Fortsätt” för att höra en ljudinspelning utan bild.

Du kan klicka på ”lyssna igen” för att höra ljudinspelningen igen. Försök inte förstå vad orden betyder, för orden är påhittade, helt meningslösa, ord.


Om ljudet inte spelas upp kan Du inte göra testet med de inställningar som Din dator har nu. Var snäll och informera den person som bad dig göra testet, och tala om att ljudet inte spelas upp på din dator.

Direkt efter Du hört ljudinspelningen, får du dessa känslor-ord att välja mellan.

Din uppgift är att välja det känslor-ord som bäst överensstämmer med den känsla som Du tyckte personen visade.

Vänligen välj ett känslor-ord för att fortsätta till nästa exempel.

Hur säker är du att du identifierat rätt emotion?

Om du är helt säker skall du ange 100 % på skalan.

Om du anger t.ex. 60 % betyder det att du inte är helt säker, utan kommer i idealfallet att ha rätt i just 60 % av fallen (d.v.s. i 6 fall av 10). På samma sätt fungerar de övriga kategorierna.

Om du anger 0 % är du istället helt säker på att du svarat fel.
Detta är det sista exemplet.

Klicka på "fortsätt" för att se en video med ljud.

Du kan klicka på "se och höra igen" för att se videon igen.

Under testet kommer Du att kunna se på varje inspelning bara en gång. Därför är det viktigt att vara uppmärksam och inte bli störd när Du gör testet.

Om videon inte spelas upp kan Du inte göra testet med de inställningar som Din dator har nu. Var snäll och informera den person som bad dig göra testet, och tala om att videon inte spelas upp på din dator. Vänligen säkerställ att du också hörde ljudet i exemplet.

Direkt efter Du sett videon, får Du dessa känslo-ord att välja mellan.

Din uppgift är att välja det känslo-ord som bäst överensstämmer med den känsla som Du tyckte personen visade.

Vänligen välj ett känslo-ord för att avsluta exemplet.

Hur säker är du att du identifierat rätt emotion?

Om du är helt säker skall du ange 100 % på skalan.

Om du anger t.ex. 60 % betyder det att du inte är helt säker, utan kommer i idealfallet att ha rätt i just 60 % av fallen (d.v.s. i 6 fall av 10). På samma sätt fungerar de övriga kategorierna.

Om du anger 0 % är du istället helt säker på att du svarat fel.

Vänligen försäkra dig om att Du kan göra testet utan att bli störd.

Du behöver 15-20 minuter för att göra hela testet.

Försäkra dig om att ingen stör dig under denna tid (t.ex. tänk på att helt stäng av din mobiltelefon).

Använd gärna hörlurar så att Du hör ordentligt under testet.

Börja testet genom att klicka på "Fortsätt".