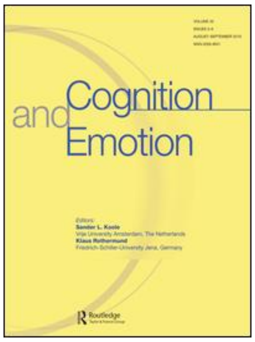

Cognition and Emotion

ISSN: 0269-9931 (Print) 1464-0600 (Online) Journal homepage: https://www.tandfonline.com/loi/pcem20

A meta-analysis of the relationship between emotion recognition ability and intelligence

Katja Schlegel, Tristan Palese, Marianne Schmid Mast, Thomas H. Rammsayer, Judith A. Hall & Nora A. Murphy

To cite this article: Katja Schlegel, Tristan Palese, Marianne Schmid Mast, Thomas H. Rammsayer, Judith A. Hall & Nora A. Murphy (2019): A meta-analysis of the relationship between emotion recognition ability and intelligence, Cognition and Emotion, DOI: 10.1080/02699931.2019.1632801

To link to this article: https://doi.org/10.1080/02699931.2019.1632801

Published online: 21 Jun 2019.

COGNITION AND EMOTION
https://doi.org/10.1080/02699931.2019.1632801

Routledge
Taylor & Francis Group

A meta-analysis of the relationship between emotion recognition ability and intelligence

Katja Schlegel[a], Tristan Palese[b], Marianne Schmid Mast[b], Thomas H. Rammsayer[a], Judith A. Hall[c] and Nora A. Murphy[d]

[a]Institute for Psychology, University of Bern, Bern, Switzerland; [b]Faculty of Business and Economics, University of Lausanne, Lausanne, Switzerland; [c]Department of Psychology, Northeastern University, Boston, MA, USA; [d]Department of Psychology, Loyola Marymount University, Los Angeles, CA, USA

ABSTRACT

The ability to recognise others' emotions from nonverbal cues (emotion recognition ability, ERA) is measured with performance-based tests and has many positive correlates. Although researchers have long proposed that ERA is related to general mental ability or intelligence, a comprehensive analysis of this relationship is lacking. For instance, it remains unknown whether the magnitude of the association varies by intelligence type, ERA test features, as well as demographic variables. The present meta-analysis examined the relationship between ERA and intelligence based on 471 effect sizes from 133 samples and found a significant mean effect size (controlled for nesting within samples) of $r = .19$. Different intelligence types (crystallized, fluid, spatial, memory, information processing speed and efficiency) yielded similar effect sizes, whereas academic achievement measures (e.g. SAT scores) were unrelated to ERA. Effect sizes were higher for ERA tests that simultaneously present facial, vocal, and bodily cues (as compared to tests using static pictures) and for tests with higher reliability and more emotions. Results were unaffected by most study and sample characteristics, but effect size increased with higher mean age of the sample. These findings establish ERA as sensory-cognitive ability that is distinct from, yet related to, intelligence.

ARTICLE HISTORY
Received 15 April 2018
Revised 9 May 2019
Accepted 12 June 2019

KEYWORDS
Emotion recognition ability; intelligence; meta-analysis; emotional intelligence; interpersonal accuracy

A meta-analysis of the relationship between emotion recognition ability and intelligence

The perception of other people's emotions from nonverbal cues is a fundamental component of interpersonal communication (McArthur & Baron, 1983). Individual differences in the ability to accurately detect and label emotions from nonverbal channels including the face, voice, and body are referred to as emotion recognition ability (ERA) and are typically measured with performance-based tests (for an overview, see Bänziger, 2016). In these tests, participants are presented with pictures or recordings of nonverbal expressions (including, for example, facial muscle movements, tone of voice or prosody, postures, or gestures) and report which emotion they think is being shown (for a brief description of some frequently used tests, see Appendix A).

Being good at recognising emotions is typically conceptualised as an adaptive skill in both children and adults, as it should help inferring interaction partners' intentions, anticipating their behaviours, and adapting one's own behaviour in order to achieve interpersonal goals (Halberstadt, Denham, & Dunsmore, 2001; Hall, Andrzejewski, & Yopchick, 2009; Hampson, van Anders, & Mullin, 2006). A large body of research has shown that ERA is related to various positive outcomes such as relationship quality, social adjustment, mental health, and academic and workplace performance (e.g. Elfenbein, Foo, White, Tan, & Aik, 2007; Elfenbein, Marsh, & Ambady, 2002; Hall et al., 2009). For example, medical students with higher accuracy in understanding patients' affect showed more patient-centred behaviour (Hall et al., 2015), and

CONTACT Katja Schlegel katja.schlegel@psy.unibe.ch

female physicians with high ERA showed more behavioural adaptability, i.e. they adapted their nonverbal behaviour to patients' individual needs and generally had more positive consultation outcomes (Carrard, Schmid Mast, Jaunin-Stalder, Junod Perron, & Sommer, 2018). Conversely, lower ERA is linked to psychological disorders such as schizophrenia (Kohler, Walker, Martin, Healey, & Moberg, 2010) as well as maladaptive traits such as trait anger, anxiety, and alexithymia (Schlegel, Fontaine, & Scherer, 2017).

Some of the correlates of ERA are also consistently predicted by psychometric intelligence or general mental ability, notably workplace performance (Schmidt & Hunter, 2004), academic success (Poropat, 2009; Richardson, Abraham, & Bond, 2012), and income (Strenze, 2007). Therefore, two questions arise: First, are ERA and psychometric intelligence related? And second, if they are related, does ERA predict other variables above and beyond intelligence or can the correlates of ERA be explained through the link between ERA and intelligence?

With respect to the first question, which is addressed in the present article, researchers have long proposed that the two constructs are likely to be connected. For example, Davitz et al. (1964) argued that fluid cognitive ability is required to perceive and integrate the numerous and subtle nonverbal characteristics of emotional expressions, and crystallized ability is required to interpret their meaning and name them with the appropriate emotion word. Crystallized ability may entail more general components (such as vocabulary knowledge and verbal fluency) as well as domain-specific knowledge, such as knowledge about nonverbal cues (Rosip & Hall, 2004).

ERA tasks should also draw on intelligence as they are maximum performance tests with correct and incorrect answers (Côté, 2014). Accordingly, some early studies found low-to-medium correlations between ERA and intelligence (e.g. Kanner, 1931). Sometimes the ERA – intelligence link was also assessed as part of the construct validation process of ERA tests, with low to moderate positive correlations (e.g. Profile of Nonverbal Sensitivity, or PONS, Rosenthal, Hall, DiMatteo, Rogers, & Archer, 1979; Geneva Emotion Recognition Test, or GERT, Schlegel, Fontaine, et al., 2017). These correlations were interpreted as providing evidence that ERA is sufficiently distinct from intelligence or test-taking ability to be a separate construct while still being an ability as opposed to a personality trait.

A more recent line of research suggests that ERA also draws on abilities other than intelligence, for example sensory discrimination (Schlegel, Witmer, & Rammsayer, 2017). Sensory discrimination refers to the basic processing of information in visual, auditory, olfactory, and other sensory modalities, specifically to the ability to make fine distinctions among stimuli such as tones differing in pitch or visual stimuli differing in their duration. Individuals with higher sensory discrimination tend to perform better on ERA tasks (Castro & Boone, 2015; Schlegel, Witmer, et al., 2017). Other basic information processing abilities correlated with higher ERA are better face identity recognition and face identity memory (Hildebrandt, Sommer, Schacht, & Wilhelm, 2015; Palermo, O'Connor, Davis, Irons, & McKone, 2013). Hildebrandt and colleagues (2015) concluded, based on a series of laboratory experiments with a large battery of emotion and face perception tasks, that a large portion of the variance in facial ERA (at least for static displays of basic emotions) can be explained through intelligence and face identity processing. Overall, ERA appears to involve cognitive or intellectual as well as sensory skills and might therefore be conceptualised as a sensory-cognitive ability that is related to, but distinct from, intelligence.

However, the relationship between ERA and intelligence might vary depending on how ERA is measured and what component of intelligence is considered. A meta-analysis found that ERA is a heterogeneous construct, with accuracy in different cue channels (i.e. face, voice, and body) showing only low to medium intercorrelations (Schlegel, Boone, & Hall, 2017). That is, a person that is good at recognising emotions from the face is not necessarily good at recognising emotions from the voice. Psychometric intelligence also encompasses different subcomponents. For example, in an attempt to integrate classic factor analytic models (e.g. Cattell, 1971) and more contemporary perspectives on intelligence such as the Cattell–Horn-Carroll model (e.g. Schneider & McGrew, 2012), Mackintosh (2011) proposed five major constituents of human intelligence: speed and efficiency of information processing (Gs), verbal or crystallized ability (Gc), visual-spatial ability (Gv), non-verbal reasoning or fluid ability (Gf), and memory. Given that ERA and intelligence are both multi-faceted constructs, their association might vary depending on which cue channel (face, voice, or body, or a combination of these) and which intelligence component is being assessed.

Past research on the relationship between ERA and intelligence

Two meta-analyses on the ERA – intelligence link have been published: The first meta-analysis focused mainly on children and adolescents and found a significant correlation of $r = .20$ in 19 samples (Halberstadt & Hall, 1980). The second meta-analysis examined the link between intelligence and interpersonal sensitivity, which was defined to include ERA but also tests of accuracy in judging deception, personality, and other states and traits (Murphy & Hall, 2011). The 11 samples from eight published sources that specifically correlated ERA with intelligence had a significant mean effect size of $r = .22$. However, this meta-analysis only included published studies conducted in the USA until 2006.

Since 2006, research on ERA has considerably expanded. Several new tests to measure ERA were published (e.g. Multimodal Emotion Recognition Test, or MERT, Bänziger, Grandjean, & Scherer, 2009; GERT, Schlegel, Grandjean, & Scherer, 2014). These tests closed important gaps in ERA testing by covering a larger range of emotions and presenting emotional expressions in different sensory modalities. Many newer studies were inspired by the popularity of the emotional intelligence (EI) construct (e.g. MacCann, Joseph, Newman, & Roberts, 2014). In the most widely accepted approach to date, EI refers to the ability to perceive and use emotional information to guide one's goal-directed thinking and behaviour (Mayer, Caruso, & Salovey, 2016). In the ability EI model, ERA is one of the fundamental components and precedes the ability to understand and manage one's own and others' emotions. MacCann and colleagues (2014) proposed that ability EI represents an additional second-stratum factor of intelligence of similar standing as fluid intelligence within the Cattell–Horn-Carroll model of intelligence (e.g. Schneider & McGrew, 2012). In addition, in their theoretical analysis, Schneider, Mayer, and Newman (2016) posited that ERA should be related to perceptual abilities as well as to crystallized ability (Gc), fluid ability (Gf), working memory, and processing speed (Gs). Studies in the ability EI field generally found small to moderate positive correlations with intelligence (Austin, 2004, 2005; Davies, Stankov, & Roberts, 1998; MacCann, Roberts, Matthews, & Zeidner, 2004, 2014). However, results were inconsistent for different intelligence types and all studies measured only facial ERA without including other sensory modalities.

Objectives of the present meta-analysis

Previous research implies a positive association between ERA and intelligence because ERA is conceptualised and measured as an ability with performance-based tests and because it is theoretically embedded in the EI construct. However, the magnitude of this association remains unclear, as well as how its strength changes depending on a) which intelligence component is examined, b) how ERA is measured (in particular, which cue channels are included), and c) what kind of population is assessed. These questions are relevant to the ERA field in several ways.

First, an analysis of different intelligence components would advance understanding of the potential mechanisms underlying high emotion recognition performance. If for example crystallized intelligence correlated most highly with ERA, this might suggest that ERA is an acquired skill relying mostly on explicit knowledge, and that it can be trained (e.g. Blanch-Hartigan, Andrzejewski, & Hill, 2016; Rosip & Hall, 2004).

Second, assessing the magnitude of the ERA – intelligence association has implications for research into the predictive validity of ERA. If the association is substantial, the relationship between ERA and a given outcome might be explained partially or entirely through participants' general level of intelligence, which should be acknowledged and investigated in future studies.

And, third, the present meta-analysis contributes to the ongoing debate about the legitimacy of EI as a type of intelligence. While previous research in EI mostly relied on facial ERA as measured with static pictures, the scope of the present analysis is much bigger and encompasses a large variety of measures and studies from almost a decade of research, allowing for more reliable conclusions regarding the link between ERA as a basic EI component and intelligence.

The present meta-analysis includes all published and unpublished studies we could locate in which at least one ERA test and one intelligence test (measures of academic achievement were also included, see below) were administered together to the same sample of adult participants. In addition to establishing the overall correlation between ERA and intelligence, we also examined moderating variables in order to answer the questions raised above. *Type of intelligence* and *cue channels* assessed in the ERA test were included to examine which specific mental abilities are linked to the perception of emotions,

depending on different visual and auditory modalities. Relatedly, we aimed to test whether this association is moderated by the complexity of how ERA is measured, assuming that more complex or comprehensive ERA tests might correlate more strongly with intelligence. Variables related to complexity were the *number of measured cue channels, presentation mode* (static versus dynamic stimuli), *number of emotions*, and *number of targets in the ERA test*. Given that ERA tests tend to vary a lot in terms of internal consistency (Schlegel, Boone, et al., 2017) and internal consistency can affect correlations with other variables, we included *Cronbach's alpha* of ERA as well as of intelligence tests, when available, as moderator variables. As Cronbach's alpha of the ERA test was available only for about half of the effect sizes, we also coded *number of items* in the ERA test (which was available for all effect sizes), assuming that longer tests tend to be more reliable, but also more complex (e.g. including more channels, emotions, or actors). *Sample characteristics*, including demographics such as gender, age, and country of data collection were included as moderators because ERA has been shown to differ by gender (e.g. Hall, Gunnery, & Horgan, 2016) and age (Isaacowitz, Vicaria, & Murry, 2016); furthermore, mean level differences across countries and languages have been found for some ERA tests (e.g. Schlegel, Fontaine, et al., 2017). Other moderators were *study characteristics* such as publication year, whether data collection was done in the lab or online, and whether the sample was tested as part of a clinical study or not were assessed in an exploratory fashion. Characteristics related to the assessment of *publication bias* were also coded.

Method

Inclusion and exclusion criteria

Studies meeting the following criteria were included: (1) Reported in English; (2) Participants at least 18 years old on average; (3) Participant sample size at least 10; (4) Participants from a typically-developing, non-clinical population (healthy control participants from studies with a clinical focus were also eligible); (5) one or more correlations between an ERA test and an intelligence measure were given.

ERA tests had to meet the following criteria: (1) Participants were presented with portrayals of human individuals expressing emotions (such as joy, sadness, disgust) or affective states (such as expressing motherly love) nonverbally through the face, voice, or body. Vocal stimuli were typically content-free; that is, targets uttered a standard sentence without meaning or content was electronically filtered to make the words unintelligible. Tests with meaningful verbal (linguistic) information in the voice were also considered. (2) Participants judged which emotion or affective state was being expressed either in a multiple choice ("Select the emotion word that best describes what the target expressed"), dimensional rating ("To what extent does this picture express sadness?"), or open response format ("Describe what the target person is feeling"). Self- or other-ratings of ERA were not included. (3) The emotion or affect judgments were compared against a criterion to provide an overall assessment of participants' ERA. Criteria included the target's intention (e.g. emotion the target was instructed to portray), targets' self-report (e.g. emotion the target reported to have felt), and consensus (e.g. proportion of experts that chose each option). ERA scores could be computed as either the sum or average of the scores across items or as the correlation between participants' emotion ratings and the criterion ratings, across stimuli. Subtest scores for different nonverbal cue channels were used if provided (e.g. separate scores for MERT audio, pictures, video, and audio-video subtests); in this case the total score of the test (e.g. MERT total score) was not included to avoid repeated inclusion of the same data. (4) All participants in a study had to complete the same items and to be unacquainted with the target individuals in the test.

Intelligence tests had to assess one of the following mental abilities according to Mackintosh (2011): *Crystallized ability*, Gc (including tests of vocabulary, verbal and reading comprehension, speed of lexical access, verbal fluency, reading or word span, numerical ability, mental arithmetic, and general knowledge); nonverbal *reasoning or fluid ability*, Gf (e.g. matrix completion); *visual-spatial ability*, Gv (e.g. visual imagery and mental rotation tasks, including tasks assessing closure); *speed and efficiency of information processing*, Gs (e.g. speeded detection of symbols), and *short- and long-term memory (including working memory).* Measures related to *academic achievement* such as Graduate Record Exam (GRE) or SAT scores and end-of-year grades were also included because intelligence is an important predictor of these measures (e.g. Poropat, 2009) and they were also part of Murphy and Hall's meta-analysis (2011). The scores on these measures could have been self-reported by the

participant. Finally, measures of *full scale IQ from standard intelligence test batteries* (e.g. Wechsler Adult Intelligence Scale) were included. Tests were not included if their primary use is to screen for cognitive impairments or to measure premorbid intelligence levels, e.g. the National Adult Reading Test (NART).

Literature search and information sources

The literature search was conducted in three ways. First, we used the authors' own published and unpublished studies, but also relevant studies included in the Murphy and Hall (2011) meta-analysis. Second, we conducted a systematic literature search on Google Scholar using all combinations of intelligence- and ERA-related terms. We opted for Google Scholar because it includes all articles that would be included in more specific databases (e.g. PsycINFO). The intelligence-related terms were intelligence, IQ, mental ability, cognitive ability, aptitude, faculty, capability, giftedness, information processing, attention, working memory, executive control, executive functioning, associative learning, memory, problem solving skills, vocabulary test, scholastic aptitude test, grade point average, graduate record exam, SAT, GPA, GRE, LSAT, and MCAT. The ERA-related terms were emotion recognition ability, interpersonal sensitivity, nonverbal communication, emotion perception, affect perception, emotion detection, affect detection, nonverbal sensitivity, empathic accuracy, interpersonal accuracy, emotion recognition accuracy, Profile of Nonverbal Sensitivity (PONS), Diagnostic Analysis of Nonverbal Accuracy (DANVA), Brief Affect Recognition Test (BART), MiniPONS, Japanese and Caucasian Brief Affect Recognition Test (JACBART), Emotion Recognition Index (ERI), Micro-Expression Training Tool (METT), Communication of Affect Receiving Ability Test (CARAT), MERT, GERT, and Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT).

When the search term combination yielded too many results, we refined the research with additional operators to exclude non-relevant studies. For example, the combination "Cognitive ability" AND "Emotion Recognition Ability" yielded 540 results, but we saw that many of the papers examined children or diseases such as autism and schizophrenia. We therefore refined the search with the additional operators "- schizophrenia – disease – children – autism" and obtained 70 results.

Abstracts and, when necessary, full texts were reviewed to see whether at least one ERA test and one intelligence measure had been administered to the same sample of participants. In the example above, this resulted in 6 studies that then were further reviewed for eligibility using the inclusion criteria defined below. We did this for all the search term combinations and then eliminated duplicate studies. In case of doubt about the study eligibility, one of the other authors was asked and a common decision was taken. In eligible studies, we looked at whether the correlations between the ERA and the intelligence measures were reported. If this was not the case, we requested these correlations directly from the authors. We did not keep track of the numbers of full texts screened at each stage of the literature search because there were so many overlaps between search terms, but we systematically used the same procedure on all the search results obtained.

Third, emails were sent to researchers known to study ERA to request published or unpublished results, and announcements were posted on the listservs of the Society for Personality and Social Psychology, the International Society for Research on Emotions, and Researchgate. Figure 1 illustrates how many sources and effect sizes were identified using each of the three research methods. The final database of studies was once more independently reviewed by two authors to ensure that eligibility criteria were met.

Final database and descriptive statistics

The final database consisted of 133 independent samples (studies) from 106 sources (e.g. articles) with a total of 471 effect sizes, all of which were Pearson correlations. If effect sizes were available for subgroups within one study (e.g. men vs. women), these subgroups were considered as separate samples. Table 1 shows the descriptive statistics of the 133 samples, and Table 2 contains a stem-and-leaf plot of all effect sizes. In total, there were 84 unique ERA measures and 133 unique intelligence measures (subparts and different versions of the same test were counted as separate measures). Appendix A contains the most frequent ERA and intelligence measures in the meta-analysis.

Coding of test, source, and sample characteristics

All moderator variables are presented in Table 3. Additional explanations for selected variables are provided below.

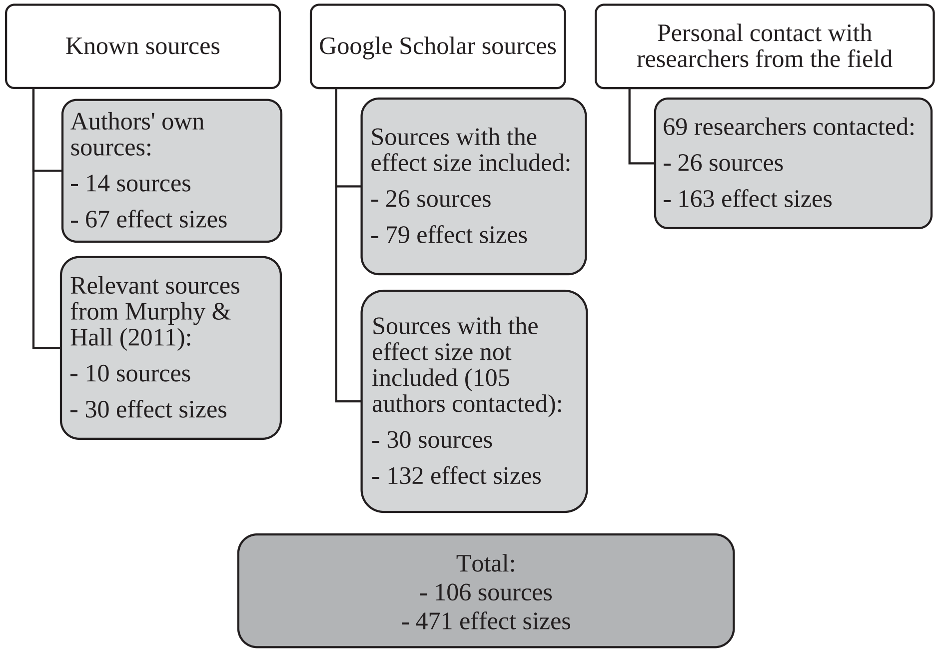

Figure 1. Summary of the literature search.

Table 1. Descriptive Statistics for 133 Independent Samples of Participants.

Measure	Statistic
Year	*Md* = 2012, *M* = 2008.95 (*SD* = 12.34), range = 1931–2017
Publication type	Journal article or book chapter = 53%, unpublished = 28%[a], published plus = 19%[b]
Total number of participants	16,094
Number of participants per sample	*Md* = 72, *M* = 120.44 (*SD* = 142.58), range 10–764
Age of sample	*Md* = 23.50, *M* = 31.80 (*SD* = 15.93), range = 18.26–76.11
Females in sample	*Md* = 57%, *M* = 57% (*SD* = 22%), range = 0–100%
Ethnicity of sample	>60% Caucasian = 26%, other = 14%, unknown = 60%
Location of data collection	In the lab = 65%, online = 10%, in work or class setting = 7%, unknown or lab/ online combined = 18%
Study focus	Comparison of clinical and non-clinical samples = 17%[c], other = 83%
Study language	English = 71%, other = 29%
Country of data collection	USA = 45%, UK = 9%, Australia = 9%, other = 37%

Note: Mean age of sample was unknown for 22 samples, percentage of females was unknown for 12 samples, and ethnicity was unknown for 80 samples. [a] unpublished data, theses, and dissertations. [b] studies were published but desired effect sizes were not published; author supplied these. [c] only effect sizes obtained from nonclinical control groups were included.

Table 2. Stem and Leaf Display of 471 Effect Sizes (Correlations between Emotion Recognition Ability and Intelligence Tests).

+.6	233457
+.5	0012456889
+.4	0000001111111222333444444444445677778899
+.3	00000000000111111112222222333444444444456666777788888888888888899999
+.2	000000000111111111112223333333333344444444444445555555556666666667777777778888888888888899999
+.1	000000000000111111222222222223333333333334444444455555555556666666667777777788888888888888899999999
+.0	0000001111111122222233333444444444555555556666666666667777777788888889999999999
−.0	9988877765543333322222222222221111111111
−.1	988776665555444333332222221110
−.2	8866554322
−.3	53
−.4	32
−.5	50
−.6	10

Table 3. Test and Sample Characteristics as Moderators.

	Overall test of significance	Pearson correlation	*k*
Level 1 predictors (test characteristics)			
Cue channels of ERA test	$F(6, 353.46) = 5.31$, $p < .001$		
body only		.03 ($p = 0.506$)[a]	20
voice only		.18 ($p < 0.001$)	60
eyes only		.19 ($p < 0.001$)	40
face only		.16 ($p < 0.001$)[b]	252
face and body		.16 ($p = 0.008$)	12
face and voice		.13 ($p = 0.109$)	8
face, voice, and body		.29 ($p < 0.001$)[a, b]	79
Type of intelligence measured in intelligence test	$F(6, 395.30) = 7.89$, $p < .001$		
full scale intelligence tests		.21 ($p < 0.001$)	37
Gc (crystallized)		.22 ($p < 0.001$)	150
Gf (fluid)		.20 ($p < 0.001$)	109
Gv (spatial)		.12 ($p = 0.026$)	15
Gs (speed/ efficiency of information processing)		.22 ($p < 0.001$)	17
memory (long, short, & working)		.21 ($p < 0.001$)	80
achievement tests		.01 ($p = 0.620$)[c]	63
Reliability (Cronbach's α) of ERA test	$F(1, 193.24) = 3.57$, $p = .060$ Intercept = .19 (*SE* = .02) Estimate = .03 (*SE* = .02)		217
Reliability (Cronbach's α) of intelligence test	$F(1, 51.61) = 3.00$, $p = .089$ Intercept = .17 (*SE* = .03) Estimate = .03 (*SE* = .02)		56
Number of items in ERA test	$F(1, 466.50) = 6.14$, $p = .014$ Intercept = .19 (*SE* = .01) Estimate = .02 (*SE* = .01)		469
Standard vs. non-standard ERA test	$F(2, 211.36) = 0.21$, $p = .808$		
Standard		.19 ($p < 0.001$)	294
Modification of standard		.14 ($p = .042$)	11
Nonstandard		.19 ($p < 0.001$)	166
Standard vs. non-standard intelligence test	$F(2, 420.43) = 1.81$, $p = .164$		
Standard		.19 ($p < .001$)	433
Modification of standard		.33 ($p < .001$)	16
Nonstandard		.20 ($p < .001$)	8
ERA stimulus presentation mode	$F(2, 441.05) = 4.31$, $p = .014$		
Static (i.e. pictures)		.17 ($p < .001$) [a]	283
Dynamic (audio and/or video recordings)		.20 ($p < .001$)	165
Static and dynamic		.29 ($p < .001$) [a]	23
ERA stimulus creation mode	$F(1, 220.51) = 0.62$, $p = .432$		
Posed (deliberately enacted behavior for stimulus creation)		.19 ($p < .001$)	440
Spontaneous (relatively unconstrained behavior, e.g. "getting acquainted" conversation)		.14 ($p = .036$)	13
ERA response format	$F(2, 218.19) = 8.57$, $p < .001$		
multiple choice		.21 ($p < .001$) [a]	366
rating scale		.11 ($p < .001$) [a]	101
open response		.04 ($p = .720$)	4
Number of different target individuals in ERA test	$F(1, 355.37) = 3.07$, $p = .081$ Intercept = .18 (*SE* = .02) Estimate = .02 (*SE* = .01)		368
Number of emotions in ERA test	$F(1, 167.27) = 8.71$, $p = .004$ Intercept = .18 (*SE* = .01) Estimate = .04 (*SE* = .01)		395
Level 2 predictors (source and sample characteristics)			
Publication status	$F(2, 100.88) = 0.13$, $p = .848$		
Published (journal article or book chapter)		.19 ($p < .001$)	183
Unpublished (including dissertation, thesis)		.17 ($p < .001$)	215
Published plus (unreported effect size from published study, supplied by study author)		.19 ($p < .001$)	73
Study language	$F(1, 118.92) = 0.02$, $p = .899$		
English		.19 ($p < .001$)	366
Other		.19 ($p < .001$)	105
Country of data collection	$F(3, 105.59) = 1.02$, $p = .389$		
USA		.21 ($p < .001$)	265

(*Continued*)

Table 3. Continued.

	Overall test of significance	Pearson correlation	*k*
UK		.14 ($p = .006$)	28
Australia		.16 ($p < .001$)	53
Other		.18 ($p < .001$)	125
Ethnic composition	$F(1, 42.18) = 4.50$, $p = .040$		
>60% Caucasian		.19 ($p < .001$)	153
other		.28 ($p < .001$)	49
Study focus	$F(2, 111.18) = 6.62$, $p = .011$		
Comparison of clinical and non-clinical samples		.26 ($p < .001$)	85
Other		.17 ($p < .001$)	386
Location of data collection	$F(3, 206.71) = 1.95$, $p = .123$		
In the lab		.17 ($p < .001$)	373
Online		.25 ($p < .001$)	30
In work or class setting		.15 ($p = .038$)	9
Unknown or lab/ online combined		.24 ($p < .001$)	59
Publication year	$F(1, 104.76) = 0.40$, $p = .529$ Intercept = .19 ($SE = .01$) Estimate = .01 ($SE = .01$)		471
Sample size	$F(1, 102.75) = 0.65$, $p = .422$ Intercept = .19 ($SE = .01$) Estimate = −.01 ($SE = .01$)		471
Mean age in sample	$F(1, 90.23) = 7.97$, $p = .006$ Intercept = .19 ($SE = .01$) Estimate = .04 ($SE = .01$)		386
Percentage of female participants	$F(1, 76.92) = 0.10$, $p = .747$ Intercept = .19 ($SE = .01$) Estimate = .01 ($SE = .02$)		435

Note: *P*-values for mean correlations refer to comparisons against zero. When k does not add up to 471, the respective information was unavailable for the remaining effect sizes.
[a,b]Categories differ significantly from each other ($p < .05$).
[c]Category differs significantly from all other categories except Gv (spatial intelligence) ($p < .05$).

ERA test cue channels

For each test, the presence or absence of each of the following cue channels was coded: *Face, voice* (tone of voice or prosody as conveyed by content-masked speech or a standard sentence used across all items such as "I am going out of the room now"), *body* (arms, legs, and/or torso, but not head), *linguistic* (intelligible free speech containing potentially meaningful content to infer the emotion), and *eyes* (coded only when stimuli consisted of only the eyes). Thus, a test could consist of a single channel or any combination of these channels. These codes were applied to the test as a whole. That is, if an effect size was based on an ERA test consisting of one subtest showing only the face and one subtest only showing the body, both the face and the body would be coded as present for this test. Ten different cue channel configurations were present among the ERA tests. Given that few measures contained linguistic information, the linguistic and vocal tone channels were combined, leading to seven final cue channel configurations; *body only*, *voice only*, *eyes only*, *face only*, *face and body*, *face and voice*, and *face, voice, and body*.

ERA test number of items

The number of items in the ERA test was coded to assess whether more comprehensive and therefore longer tests may correlate more strongly with intelligence. Number of test items was not coded for intelligence tests because the wide range of tasks made this variable less meaningful for the evaluation of reliability.

Standard test

A *standard* ERA or intelligence test was defined as a named test that was used repeatedly in the literature and/or for which the developer had published at least one validity article. Variants, shortened versions, and subparts of such tests were coded as *modifications of standard* tests. *Non-standard* tests were, by contrast, typically developed for a particular study, were less likely to have psychometric reporting, and were not known to be used by other investigators much, if at all.

Number of emotions

In multiple-choice tests, this was the number of emotion or affect categories that were presented as response options after each item (e.g. in the GERT

there are 14 emotions to choose from). This number corresponds to the total number of different emotions portrayed in the tests (e.g. the GERT contains videos for each of the 14 emotions), with the exception of the PONS test. In this test, a total of 20 affective states are portrayed, but for each item only two out of the 20 states are presented as response options to choose from (e.g. item 1: "admiring nature" versus "helping a customer"; item 2: "leaving on a trip" versus "ordering food in a restaurant"). This variable was thus not coded for the PONS. For tests using a rating scale format, this variable was the number of emotions in the test that participants were asked to rate (e.g. in the MSCEIT participants rated a total of seven emotions). For tests using an open response format, number of emotions was the number of discrete emotion categories in the coding scheme used to score the responses.

Coder reliability

Sample characteristics (e.g. year) and effect sizes were directly retrieved from the respective publication or email communication with the study author and did not require coding. Test characteristics were coded by the first or second author (KS or TP) using templates for standard tests in order to avoid mistakes. In addition, the first and second author performed extensive cross-checking of the database on a study-by-study and test-by-test basis; any disagreement was resolved by the two authors.

Statistical analysis

Given that one study could have correlated multiple ERA and intelligence tests with each other, studies could provide more than one effect size. The number of effect sizes originating from the same participant sample ranged from 1 to 35. To account for the nesting of effect sizes within samples, the data were analysed using multilevel modelling (MLM) with sample ID as the random effects nesting variable. MLM has been previously proposed and used for meta-analysis when number of effect sizes varies considerably between studies (e.g. Hedges, Tipton, & Johnson, 2010; Hox & de Leeuw, 2003; Konstantopoulos, 2011; Schlegel, Boone, et al., 2017). This approach takes into account within and between study sampling error as well as the possibility that effect sizes within one study are more similar than effect sizes across studies. As there was considerable heterogeneity in the variables of interest (for example, a wide range of intelligence types and measures, a lot of variation in sample mean age), it was assumed that the "true" effect size would vary from study to study due to these variables. Also, because the often used random effects model that weights by sample size is not fully random at the level of studies due to the weighting (Borenstein, Hedges, Higgins, & Rothstein, 2010), we chose to use the unweighted random effects model that weights each study (and its moderator variables) equally in order to maximise generalisation to future studies that have different measures or sample characteristics (Goh, Hall, & Rosenthal, 2016; Hall & Rosenthal, 2018). Unweighted random effects models have long been used in meta-analysis (e.g. Frattaroli, 2006; White, 1982; Zuckerman, Silberman, & Hall, 2013; see Hall & Rosenthal, 2018, for a review).

The dependent variable in all analyses was the Pearson correlation between an ERA and an intelligence test transformed into Fisher's z for normalisation. Effect sizes were transformed back to the r-metric for all data presentations. To assess the overall association between ERA and intelligence, an unconditional means model with effect size as the dependent variable and no predictors was computed. For comparison, the average effect size without nesting was also computed. To assess the moderating influence of test and sample characteristics on effect size, each potential moderator was separately added as a fixed effect to the unconditional means model, yielding a separate analysis for each moderator. Level 1 moderators included test characteristics that varied on the effect size level, such as test reliability or ERA test response format. Level 2 moderators were characteristics of the sample such as sample size or publication year. Continuous level 1 and level 2 moderators (e.g. reliability or year) were standardised prior to the analysis.

Results

Overall correlation between emotion recognition ability and intelligence tests

Correlations between tests ranged from $r = -.61$ to $r = .67$ with a mean correlation of $r = .16$ and a median of .17 (both uncontrolled for nesting). The average correlation between tests accounting for nesting as indicated by the intercept in the unconditional means model was $r = .19$ ($SE = .01$; $p < .001$). As indicated by

a Wald test, the estimated variance of the intercept was significant, suggesting that effect sizes were heterogeneous and might be affected by moderator variables ($Z = 2.99$, $p = .003$).

Level 1 moderators (test characteristics)

Cue channels in the ERA test

In order to evaluate how the cue channel(s) in ERA measures influence effect sizes, we ran a multilevel model with the cue channel configuration of the ERA test as a Level 1 predictor (see Table 3). Effect sizes were significantly higher than zero for all configurations except *body only* and *face and voice*. However, only few effect sizes were available for each of these configurations ($k = 20$ and $k = 8$, respectively). The largest average effect size ($r = .29$, $p < .001$) was found for ERA tests that included *face, voice, and body*. To assess which configurations significantly differed from each other, we conducted pairwise comparisons with Bonferroni-corrected *p*-values (*p*-values were multiplied by the number of comparisons) as part of the multilevel model described above. Results showed that ERA tests including *face, voice, and body* were significantly more strongly correlated with intelligence than ERA tests including the *body only* or the *face only* ($p < .05$). It should be noted, however, that the number of effect sizes varied greatly across configurations, which may have limited statistical power of the analysis. We further explored to what extent each of the cue channels (face, voice, body, eyes, linguistic information) influenced effect sizes by running, for each cue channel, one multilevel model with a dummy-coded variable (1 = the ERA test included this modality; 0 = the ERA test did not include this modality) as a predictor. Each of these five analyses compared all effect sizes based on an ERA test including the modality against all effect sizes in which the ERA test did not include this modality. Results showed that effect sizes based on an ERA test that included the voice were significantly higher than effect sizes in which the ERA test did not include the voice ($p < .001$). In addition, effect sizes based on an ERA test that included the body were significantly higher than effect sizes in which the ERA test did not include the body ($p < .001$). For the other cue channels, effect sizes did not differ depending on whether the channel was measured in the ERA test or not. Taken together, these findings suggest that correlations between ERA and intelligence become stronger the more cue channels are measured in the ERA test. In order to test whether these results might be due to a confound of channels assessed and test length as well as test reliability (i.e. tests with more channels might be longer and more reliable, and hence yield higher correlations with intelligence), we additionally compared tests assessing more than one channel in different subtests (e.g. ERI) and tests assessing multiple channels simultaneously (e.g. using videos with voice, as in the GERT). The latter category of tests did not significantly differ in test length compared to tests assessing single channels ($M = 38$ items vs. $M = 36$ items), but nevertheless yielded higher correlations with intelligence ($r = .25$ vs. $r = .16$). In addition, we repeated the moderator analysis for cue channels (Table 3, first section) adding ERA test reliability as a covariate to the model (271 effect sizes). Reliability was not significant in this model ($p = .543$) and the effect size estimates for the different channels were very similar to the ones reported in Table 3 for all 471 effect sizes.

Type of intelligence measured

In order to evaluate how intelligence type influenced effect sizes, we ran a multilevel model with intelligence type as a Level 1 predictor (Table 3). Effect sizes were significantly larger than zero for all types of intelligence, but was virtually zero for achievement tests ($r = .01$, $p = .62$). All other effect sizes ranged from $r = .20$ to $r = .22$, with the exception of Gv which showed a somewhat lower correlation with ERA ($r = .12$, $p = .03$). The pairwise comparisons revealed that achievement measures had a significantly lower effect size than all intelligence tests ($p < .05$) except Gv. None of the other pairwise comparisons was significant. It should again be noted that the number of effect sizes was unequal across the categories, which may have restricted statistical power.

Combination of cue channels and intelligence type

Table 4 shows the effect size estimates and frequencies for each of the 35 cue channel – intelligence type combinations. Only seven of the combinations were represented with 20 or more effect sizes and 14 possible combinations have, to our knowledge, not been tested before. The biggest and most robust effect sizes were found for the combination of Gc with tests measuring emotion recognition from the *eyes* only ($r = .25$; $p < .001$) and with tests measuring the *face, voice, and body* ($r = .32$; $p < .001$),

Table 4. Average Effect Size (Pearson Correlation) for Each Modality – Intelligence Type Combination.

	full scale IQ	Gc (crystallized)	Gf (fluid)	Gv (spatial)	Gs (speed)	achievement	memory
body only	.25 (2)	.17 (2)	.36**[a] (2)	–	–	–.15* (14)	–
voice only	.14 (4)	.26***[a] (15)	.21***[a] (16)	–	–	.03 (20)	.26**[a] (5)
eyes only	.16 (5)	.25***[abc] (23)	.04(6)	.01 (2)	.27 (1)	–	.01 (3)
face only	.19***[a] (17)	.18***[a] (86)	.17***[a] (52)	.12* (13)	.12 (8)	.06 (38)	.18***[a] (38)
face and body	.41* (1)	.19 (2)	.11 (7)	–	–	–	.25 (2)
face and voice	–	.09 (3)	.17 (3)	–	.12 (2)	–	–
face, voice, and body	.28***[a] (8)	.32***[abc] (19)	.28***[abc] (23)	–	.34***[a] (6)	.09 (8)	.26***[a] (15)

Note: Numbers of effect sizes for each combination are displayed in parentheses. Effect sizes were estimated using MLM. The asterisks refer to the significance levels compared to zero, *$p < .05$. **$p < .01$. ***$p < .001$.
[a]these effect sizes are significantly higher than the achievement/ body only combination effect size.
[b]these effect sizes are significantly higher than the achievement/ face only effect size.
[c]these effect sizes are significantly higher than the achievement/ voice only effect size.

as well as Gf with tests measuring the *face, voice, and body* ($r = .28$, $p < .001$).

Test reliability and number of items

Reliabilities (internal consistency coefficients, expressed with Cronbach's alpha) for the ERA tests were available for 217 (46%) of all effect sizes. For the intelligence tests, Cronbach's alpha was only available for 56 (12%) of all effect sizes. The average reliability for ERA tests was $\alpha = .62$ ($SD = .19$) and $\alpha = .76$ ($SD = .12$) for intelligence tests. As shown in Table 3, reliability of both types of measures was marginally positively related to effect sizes ($p = .06$ for ERA tests' reliability and $p = .09$ for intelligence tests' reliability). For ERA tests, the intercept of .19 indicates that the correlation for ERA tests of average reliability (i.e. $\alpha = .62$) is estimated to be $r = .19$. The estimate of .03 means that a one standard deviation change in α will lead to an increase or decrease in the effect size of .03. For example, the correlation of an ERA test with an α of one standard deviation above the mean (i.e. $.62 + .19 = .81$) would increase by .03 from $r = .19$ to $r = .22$.

Number of items in the ERA test was available for 469 effect sizes ($M = 43.83$; $SD = 40.34$) and was positively related to effect size as well. Collectively, these findings demonstrate that more reliable tests and longer tests tend to yield somewhat larger effect sizes.

Other level 1 moderators

ERA stimulus presentation mode

ERA tests using both static and dynamic stimuli showed significantly higher correlations with intelligence ($r = .29$) than tests using only static stimuli (i.e. pictures). However, the *static and dynamic* category contained few effect sizes ($k = 23$), most of which came from one single study using the PONS test. As the PONS test is much longer than most other tests, we added test length as a covariate to the model. In this model, stimulus presentation mode no longer significantly affected the correlation between ERA and intelligence.

ERA response format

Multiple-choice ERA tests yielded higher correlations with intelligence than ERA tests using dimensional rating scales ($r = .22$ versus $r = .11$). Note that most of the effect sizes in the latter category were based on the Faces subtest of the MSCEIT which has been repeatedly criticised in the literature for its psychometric properties and the use of consensus scoring (Fiori et al., 2014).

ERA number of targets

The number of target individuals in the ERA test ($M = 9.11$, $SD = 11.34$) had a marginally significant ($p = .08$) positive effect on the correlation with intelligence.

ERA number of emotions

The number of emotions portrayed in the ERA measure and presented as response options ($M = 7.22$, $SD = 3.59$) was positively correlated with intelligence ($p = .004$).

Whether the ERA test or the intelligence test was standard or not, and the stimulus creation mode of the ERA test, did not affect the ERA-intelligence relationship.

Level 2 moderators (source and sample characteristics)

Publication status, publication year, study language and country, location of data collection, sample size,

and percentage of female participants did not affect the correlation between ERA and intelligence (see Table 3).

Ethnic composition

Correlations between ERA and intelligence were significantly higher in samples with a proportion of non-Caucasian participants above 40% than in samples with more than 60% Caucasian participants ($r = .29$ versus $r = .19$). All of the samples with a known proportion of non-Caucasian participants above 40% were from the USA or Japan. As more detailed sample characteristics were not available, further research is required to explain this finding.

Study focus

Effect sizes were significantly higher among healthy individuals that served as control subjects in clinical studies than among healthy individuals from studies outside the clinical context ($r = .27$ versus $r = .17$). It can be assumed that clinical samples typically have higher variability in demographic characteristics, ERA, and intelligence than participants in non-clinical studies with university students. As healthy controls in such studies are usually matched on these characteristics to the clinical sample, their variability in ERA and intelligence might also have been higher.

Sample mean age

The mean age of the sample was significantly positively related to effect size. The estimated correlation for samples with an average age of 30.2 years (which was the mean age across the 471 effect sizes) was $r = .19$. Samples with an average age of one standard deviation above the mean (30.2 + 15.3 = 45.5 years) had an estimated correlation of $r = .23$ (.19 + .03). In order to explore this association further, Table 5 shows the correlations between ERA and intelligence split by age groups, intelligence types, and ERA cue channels. The highest increase in the correlation with higher age was observed for Gc (from $r = .18$ to $r = .28$), and slightly smaller increases were found for full scale IQ (from $r = .20$ to $r = .28$) and Gf (from $r = .18$ to $r = .25$). With respect to cue channels, the increase in effect size with age was most pronounced in ERA tests measuring the face only (from $r = .13$ to $r = .26$), but few effect sizes were available for older participants in other channels.

Table 5. Effect sizes by Age Group, Intelligence Type, and Cue Channel.

	mean age	
	under 35	over 35
By intelligence type		
Full scale IQ	.20***[a] (26)	.28**[a] (6)
Gc (crystallized)	.18***[a] (92)	.28***[a] (45)
Gf (fluid)	.18***[a] (85)	.25***[a] (14)
Gv (spatial)	.15* (11)	−.12 (2)
Gs (speed)	.29** (4)	.28***[a] (8)
Achievement	−.03 (55)	–
Memory	.18***[a] (22)	.23***[a] (16)
By cue channel		
body only	.02 (14)	–
voice only	.16*** (44)	.30** (5)
eyes only	.16** (15)	.13* (16)
face only	.13***[c] (158)	.26*** (38)
face and body	.13* (11)	.43*[c] (1)
face and voice	.04 (3)	.20 (4)
face, voice, and body	.26***[bc] (50)	.31***[bc] (27)

Note: The cutoff of age 35 was set based on frequency distribution of age, such that about 20% of effect sizes were in the older category. Total number of correlations was 386 (for the remaining effect sizes, mean age was unknown). Numbers of correlations for each combination are displayed in parentheses. The correlations were adjusted for nesting within samples using MLM. The asterisks refer to the significance levels compared to zero, *$p < .05$. **$p < .01$. ***$p < .001$.
[a]Effect size significantly larger than for achievement tests under 35 years ($p < .05$, Bonferroni-adjusted for multiple comparisons).
[b]Effect size significantly larger than for body only under 35 years ($p < .05$, Bonferroni-adjusted for multiple comparisons).
[c]Effect size significantly larger than for face only under 35 years ($p < .05$, Bonferroni-adjusted for multiple comparisons).

Publication bias and selective reporting

Nearly half of all effect sizes were obtained from unpublished studies or upon request from authors of published studies who had not included the relevant results in their publication (see Table 1). Effect sizes of published studies were not larger than effect sizes of unpublished studies (see Table 3), speaking against a publication bias. As recommended by Hox and de Leeuw (2003), we also examined whether larger positive effect sizes were predominantly found in studies with smaller samples by including sample size as a moderator in the unconditional means model. This analysis follows the same rationale as visual inspections of funnel plots in traditional meta-analyses which plot effect sizes versus sample sizes. It is assumed that if no publication bias is present, this plot will have the shape of a funnel because studies with smaller sample size will produce more variable results. Publication bias might be present when visual inspection of the funnel plot reveals an asymmetry in that large positive effects are found predominantly in smaller studies. The estimate of this analysis was non-significant (see Table 3), speaking

against this potential bias. It should be noted that this approach assumes that publication bias supresses small effect sizes, but other approaches examining publication bias based on the distribution of *p*-values of published studies have become increasingly popular (Schimmack & Brunner, 2017; Simonsohn, Nelson, & Simmons, 2014). However, to date methods to examine such bias are unavailable for nested data.

Discussion

The goal of the present meta-analysis was to provide a comprehensive overview of the association between different types of tests of emotional recognition accuracy (ERA) and basic facets of intelligence or general mental ability. The last meta-analysis on the topic included 11 effect sizes for ERA tests and found an effect size of $r = .22$ (Murphy & Hall, 2011), but covered only published studies conducted in the US until 2006. The present study included Murphy and Hall's 11 effect sizes as well as unpublished results, studies conducted outside the US, and studies published until 2017, yielding 471 effect sizes. In light of the current replicability debate in psychological science (e.g. Open Science Collaboration, 2015), a special effort was made to identify effect sizes that were either unpublished or not directly reported in published studies. Given that almost half of the present database consisted of such effect sizes and analyses of publication status and sample size suggested that publication bias did not affect results, the present findings can be interpreted with a reasonable degree of confidence. In addition, the included studies differed widely in their primary study goals, encompassing the assessment of predictive validity or age differences, the comparison of clinical and typically developing samples, or test development and validation. In most studies, ERA and intelligence measures were not administered for the express purpose of exploring the link between them, which further decreases the likelihood of publication bias in relation to whether an expected correlation was found or not. As a limitation of the present approach, it should be noted that although the literature search strategy was systematic, we did not document the exact numbers of manuscripts found for each search term combination, the numbers of overlapping results, and the number of excluded manuscripts at every stage of the search process as it is recommended in some recent guidelines (e.g. Atkinson, Koenka, Sanchez, Moshontz, & Cooper, 2015).

ERA as a sensory-cognitive ability

Overall, the small to moderate correlation between ERA and intelligence of $r = .19$ is comparable in magnitude to the results of the previous meta-analyses on this same topic (Halberstadt & Hall, 1980; Murphy & Hall, 2011). It is also very similar to the meta-analytic correlation found between ERA and other tests measuring accurate person perception (Schlegel, Boone, et al., 2017). Overall, this finding is in line with the conceptualisation of ERA as a sensory-cognitive ability that partly draws on individual differences in intelligence, but also on general sensory perception, and more specific cognitive processes like face identity recognition (e.g. Hildebrandt et al., 2015). With ERA being a central component of ability EI, the present finding is also compatible with the conceptualisation of ability EI as a second-stratum factor of intelligence (MacCann et al., 2014) or a broad intelligence (Schneider et al., 2016).

Moderator analyses provided insight into how the magnitude of the ERA – intelligence link varied depending on ERA test features and intelligence types. The finding that more complex tests are more strongly related to intelligence suggests that these tests might rely more strongly on perceptual abilities than less complex measures. For example, tests using videos and voice recordings might draw more strongly on higher sensory discrimination skills than tests using static pictures. Sensory discrimination subsumes a wide range of specific abilities such as pitch, loudness, brightness, or visual duration discrimination (e.g. Rammsayer & Troche, 2012). Better sensory discrimination might facilitate the perception and interpretation of brief and subtle auditory and visual elements, especially in dynamic and multimodal emotional expressions. Indeed, Castro and Boone (2015) found a stronger association between ERA and sensitivity to rhythm and visual features of shapes when ERA was measured with dynamic than static stimuli. In line with this argument, Schlegel, Witmer, et al. (2017) found that better visual duration discrimination predicted higher scores on the GERT (a complex test with multimodal stimuli and many emotions) above and beyond Gf. These findings are also consistent with Schneider et al.'s (2016) theoretical account, suggesting that perceptual abilities should play an important role in ERA.

The correlations with ERA were remarkably similar for most intelligence components, excluding academic achievement measures. It could be that each component contributes uniquely to better ERA, or that all components relate to ERA through their loading on a broad general intelligence ("g") factor. From a neuroscience perspective, the different stages of the emotion recognition process draw on multiple intelligence components (e.g. models by Adolphs, 2002; and Dricu & Frühholz, 2016). At the first stage in these models, sensory information is extracted and encoded, requiring perceptual abilities. At the second stage, this information is integrated with existing representations of emotional displays and associated semantic knowledge, drawing on fluid abilities, crystallized ability, and long-term memory. Finally, as these representations are accessed and an emotion label for the stimulus is chosen, the choice options need to be maintained in the working memory. It should be noted that crystallized ability within the framework of the present meta-analysis consisted of general facets such as verbal fluency, vocabulary, and numeric abilities, and did not include domain-specific knowledge. As demonstrated by Rosip and Hall (2004) or Schlegel and Scherer (2018), domain-specific knowledge about emotion concepts and nonverbal cues is much more strongly correlated with ERA than broad crystallized intelligence in the present meta-analysis.

It also seems plausible to assume that better ERA is explained by a higher broad mental capacity rather than single specific abilities, but more studies assessing a range of abilities within the same sample are needed to answer this question. Such studies could estimate the g-saturation of each ability or task and disentangle the contribution of general and specific cognitive abilities to ERA. Elementary information processing speed could be one broad mental capacity underlying ERA, intelligence, and sensory discrimination, which would explain the modest correlation between ERA and intelligence found here (e.g. Acton & Schroeder, 2001). The present correlation is comparable in magnitude to associations between sensory discrimination and intelligence (e.g. Acton & Schroeder, 2001; Schlegel, Witmer, et al., 2017), and between mental processing speed and intelligence (Sheppard & Vernon, 2008). ERA might thus be considered as part of the range of mental abilities that are associated with, but distinct from, intelligence.

Whereas ERA was related to all intelligence facets, it was unrelated to measures of academic achievement such as GRE and SAT scores or grades. Similarly, in Murphy and Hall's (2011) meta-analysis, these measures showed lower correlations with interpersonal sensitivity than intelligence tests. The present finding might indicate that achievement test performance depends on factors that are unrelated to ERA, such as conscientiousness (Poropat, 2009). Another reason for the zero correlation might be that some studies took place years after participants passed the achievement test. Participants' abilities might have changed over time and some might not have accurately remembered or reported their score, as these were typically self-reported. In contrast to these results, in children ERA has been more consistently related to academic achievement (Elfenbein et al., 2002; Halberstadt & Hall, 1980; Nowicki & Duke, 1994). This might be explained by the fact that achievement in children is typically rated by their teachers through end-of-year grades, whereas achievement tests for adults are standardised and do not involve personal interaction. Halberstadt and Hall (1980) found that nonverbally sensitive children were perceived as more intelligent by their teachers even when their actual cognitive ability was controlled for, suggesting a halo effect of ERA that might contribute to the positive ERA – achievement link in this age group.

ERA and intelligence as predictors of psychosocial variables

In light of the non-trivial correlation between ERA and intelligence, an interesting question is whether links between ERA and other variables such as workplace performance, health, social effectiveness, or personality traits are unique to ERA or whether they can be explained partly or entirely through its link with intelligence. Researchers examining relationships between ERA and other variables might therefore consider adding a measure of intelligence as a control variable to tease apart the unique and shared contributions of ERA and intelligence. This might be especially useful when ERA is measured with complex tests including several modalities and many emotions, as correlations between ERA and intelligence are higher for these tests. Given that the correlations were similar for all forms of intelligence in this analysis, a test measuring any of these may be used. However, if future studies find that correlations with ERA increase as the g-saturation of an intelligence test gets stronger, researchers should preferably choose a test with high g-saturation.

Controlling for intelligence has not been routinely done in research on the predictive power of ERA, but can have important theoretical implications. For example, in Olderbak, Mokros, Nitschke, Habermeyer, and Wilhelm's (2018) study, ERA no longer predicted psychopathy among prisoners when intelligence was controlled for, suggesting that psychopaths' deficits in emotional processing are not specifically emotional, but rather due to a more general intellectual deficit. In contrast, Schlegel, Mehu, van Peer, and Scherer (2018) showed that ERA rather than general intelligence is necessary to explain better negotiation performance. Similarly, Palese and Schmid Mast (2017) found that female leaders with higher ERA displayed higher behavioural adaptability even when intelligence was controlled for, i.e. they better adjusted the extent of participative behaviour to the preferred leadership style of their subordinates. It might be that ERA and intelligence correlate more strongly in the lower ability range (as might have been the case in the prisoner sample of Olderbak et al., 2018) than in the higher ability range (as might have been the case in the student samples of Schlegel et al., 2018, and Palese & Schmid Mast, 2017; see also Rosenthal et al., 1979, and Legree, Mullins, LaPort, & Roberts, 2016).

Sample and test characteristics as moderators

With respect to sample characteristics, correlations between ERA and intelligence were higher when the sample was older, when the sample was more diverse in ethnicity, and when samples were healthy controls in studies with a clinical focus. A likely reason for this finding is that these samples are more heterogeneous in their levels of ERA and intelligence than college student samples, possibly also due to a higher diversity in background variables such as education level which might affect both ERA and intelligence. Although not examined here, previous research also suggests that these two abilities are more strongly related in clinical samples (Rosenthal et al., 1979), which might also be explained by third variables affecting ERA and intelligence such as attention or test-taking abilities. In older adults, the stronger association might also be explained by the cognitive de-differentiation hypothesis that posits that during aging, different abilities become more dependent on similar executive or organising resources such as sensory functioning and thus become more highly correlated (Cabeza, 2002; Li & Lindenberger, 1999). Therefore, intelligence might especially be considered as a control variable in ERA studies of non-student samples.

The analysis of test characteristics revealed that overall, reliability of ERA tests tended to be modest with a mean Cronbach's alpha of .62, in line with psychometric shortcomings noted by others (see Schlegel, Boone, et al., 2017). Moderator analyses showed that more reliable ERA tests, tests with more emotions, and tests with more items in general yielded higher correlations with intelligence, suggesting that the mean effect of $r = .19$ might underestimate the true relationship between the two constructs. Another test feature affecting the results was response format, with multiple choice yielding higher effect sizes than rating scales like in the MSCEIT Faces test. As Legree and colleagues (2014) noted, this difference might be due to response tendencies associated with rating scales, for instance tendency towards the mean or towards extreme values. Considering these results, future test developers can expect moderately high correlations with intelligence if their test meets common criteria for good internal consistency, uses a multiple choice format, includes a large number of emotions, and is not too short.

Recommendations for methodological improvements in ERA measurement

A large number of effect sizes (35%) was based on non-standard ERA tests that did not undergo regular test development phases such as testing out a first item pool with a development sample, selecting good items based on item difficulty and item discrimination, and assessing factor structure (e.g. DeVellis, 2016). Further, internal consistency was only reported for 23% of the non-standard ERA tests as opposed to 60% for standard tests. Although internal consistency did not differ between standard and non-standard tests (when it was reported), we recommend that future studies should report basic test and item properties such as internal consistency, factor structure, and item difficulty and discrimination if they build a custom test in order to ensure reliable measurement of ERA. However, even for standard tests information regarding internal consistency, factor structure, and item difficulties and discrimination is missing in some of the original publications, suggesting that the field lacks the strong psychometric tradition of other areas such as intelligence testing. We therefore urge future ERA test developers to follow all steps

outlined in recent test development guidelines (e.g. DeVellis, 2016), and to consider using Item Response Theory (IRT) as the psychometric framework. IRT is especially useful in the ERA field because it offers models adapted to the binary (i.e. correct/ incorrect) item format often used in ERA tests, which is not the case for standard Classical Test Theory (CTT) methods. For example, Cronbach's alpha (a traditional CTT reliability index) tends to underestimate the internal consistency of tests with binary items, whereas the IRT framework provides specific reliability indices for such tests (for more details, see Boone & Schlegel, 2016). Rigorously following test development guidelines and using IRT to select the best items may help increasing the reliability of new tests (e.g. see Schlegel et al., 2014).

Gaps in the literature and conclusion

The present analysis points to several gaps in the literature. For example, only few studies to date examined visual-spatial ability and speed and efficiency of information processing as correlates of ERA. However, as noted earlier, elementary information processing might be the most basic mental ability linking intelligence and ERA. Statistically, this imbalance in available effect sizes across some of the moderator variables (particularly, intelligence types and ERA cue channel configurations), may have limited statistical power to detect differences between the different categories. In addition, few studies examined adults beyond their mid-twenties. More studies with more diverse samples in terms of age will advance our understanding of the mechanisms underlying high performance in emotion recognition across the lifespan, and might inform the development of interventions to improve emotion recognition and other emotional competencies.

Taken together, the present meta-analysis provides some evidence for a substantial link between ERA and intelligence in adults, considering that the effect size of $r = .19$ might be on the lower end of the true relationship due to psychometric limitations and limited complexity of some of the included tests. It establishes ERA as sensory-cognitive ability amidst other mental abilities (including intellectual and perceptual abilities) that are distinct from each other yet may share an elementary cognitive basis. We would like to conclude with the observation that more than half (302 out of 471) of the effect sizes in this meta-analysis came from studies conducted or published in 2012 or later, possibly highlighting researchers' increasing awareness to consider ERA alongside other cognitive abilities. We appreciate this progress in integrating the ERA and intelligence fields and believe that it offers exciting avenues for future research. For example, an open question is whether higher ERA develops as a consequence of higher intelligence, i.e. whether intelligence is a prerequisite for ERA, whether both abilities develop independently, or whether other variables affect the development of both abilities. As Halberstadt et al. (2001) suggested, the development of emotional competence is likely a complex interplay between maturation and opportunities provided by socialisation experiences. Future research might therefore also look into shared and unique family, educational, social and other predictors of both constructs across the lifespan to better understand the developmental trajectories.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Thomas H. Rammsayer http://orcid.org/0000-0002-9140-067X

References

Acton, G. S., & Schroeder, D. H. (2001). Sensory discrimination as related to general intelligence. *Intelligence*, *29*(3), 263–271.

*Adler, N., Nadler, B., Eviatar, Z., & Shamay-Tsoory, S. G. (2010). The relationship between theory of mind and autobiographical memory in high-functioning autism and Asperger syndrome. *Psychiatry Research*, *178*(1), 214–216.

Adolphs, R. (2002). Neural systems for recognizing emotion. *Current Opinion in Neurobiology*, *12*(2), 169–177.

Army Individual Test Battery. (1944). *Manual of directions and scoring*. Washington, DC: War Department, Adjunct General's Office.

Atkinson, K. M., Koenka, A. C., Sanchez, C. E., Moshontz, H., & Cooper, H. (2015). Reporting standards for literature searches and report inclusion criteria: Making research syntheses more transparent and easy to replicate. *Research Synthesis Methods*, *6*, 87–95.

Austin, E. J. (2004). An investigation of the relationship between trait emotional intelligence and emotional task performance. *Personality and Individual Differences*, *36*(8), 1855–1864.

*Austin, E. J. (2005). Emotional intelligence and emotional information processing. *Personality and Individual Differences*, *39*(2), 403–414.

*Austin, E. J. (2010). Measurement of ability emotional intelligence: Results for two new tests. *British Journal of Psychology*, *101*(3), 563–578.

*Bailey, P. E., & Henry, J. D. (2008). Growing less empathic with age: Disinhibition of the self-perspective. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *63*(4), P219–P226.

Bänziger, T. (2016). Accuracy of judging emotions. In J. A. Hall, M. Schmid Mast, & T. V. West (Eds.), *The social Psychology of Perceiving others accurately* (1 edition., pp. 23–51). Cambridge, United Kingdom: Cambridge University Press.

Bänziger, T., Grandjean, D., & Scherer, K. R. (2009). Emotion recognition from expressions in face, voice, and body: The multimodal emotion recognition test (MERT). *Emotion*, *9*(5), 691–704.

Bänziger, T., Scherer, K. R., Hall, J. A., & Rosenthal, R. (2011). Introducing the MiniPONS: A short multichannel version of the profile of nonverbal sensitivity (PONS). *Journal of Nonverbal Behavior*, *35*, 189–204.

*Barchard, K. A. (2001). *Emotional and social intelligence: Examining its place in the nomological network* (Doctoral dissertation). Retrieved from https://open.library.ubc.ca/cIRcle/collections/ubctheses/831/items/1.0090848

*Barchard, K. A. (2003). Does emotional intelligence assist in the prediction of academic success? *Educational and Psychological Measurement*, *63*(5), 840–858.

Baron-Cohen, S., Jolliffe, T., Mortimore, C., & Robertson, M. (1997). Another advanced test of theory of mind: Evidence from very high functioning adults with autism or Asperger syndrome. *Journal of Child psychology and Psychiatry*, *38*(7), 813–822.

Baum, K. M., & Nowicki, S. (1998). Perception of emotion: Measuring decoding accuracy of adult prosodic cues varying in intensity. *Journal of Nonverbal Behavior*, *22*(2), 89–107.

*Bedwell, S. E. (2010). *Beyond static representations: development of a multi-channel, dynamic assessment of emotion perception* (Doctoral dissertation). Retrieved from https://www.ideals.illinois.edu/handle/2142/16704

*Bernieri, F. J., Brown, J. A., & Reyna, N. (2016). Unpublished data raw data.

Blanch-Hartigan, D., Andrzejewski, S. A., & Hill, K. M. (2016). Training people to be interpersonally accurate. In J. A. Hall, M. Schmid Mast, & T. V. West (Eds.), *The social Psychology of Perceiving others accurately* (pp. 253–269). Cambridge University Press.

*Bommer, W. H., Pesta, B. J., & Storrud-Barnes, S. F. (2011). Nonverbal emotion recognition and performance: Differences matter differently. *Journal of Managerial Psychology*, *26*(1), 28–41.

Boone, R. T., & Schlegel, K. (2016). Is there a general skill in perceiving others accurately? In J. A. Hall, M. Schmid Mast, & T. V. West (Eds.), *The social Psychology of Perceiving others accurately* (pp. 379–403). Cambridge: Cambridge University Press.

Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, *1*(2), 97–111.

Byron, K., Terranova, S., & Nowicki, S. (2007). Nonverbal emotion recognition and salespersons: Linking ability to perceived and actual success. *Journal of Applied Social Psychology*, *37*(11), 2600–2619.

Cabeza, R. (2002). Hemispheric asymmetry reduction in older adults: The HAROLD model. *Psychology and Aging*, *17*(1), 85–100.

*Camargo, M. A. (2007). *Hypothesized fitness indicators and mating success* (Doctoral dissertation). Retrieved from https://dspace.sunyconnect.suny.edu/handle/1951/42568

Carrard, V., Schmid Mast, M., Jaunin-Stalder, N., Junod Perron, N., & Sommer, J. (2018). Patient-Centeredness as Physician Behavioral adaptability to patient Preferences. *Health Communication*, *33*(5), 593–600.

*Carrard, V., & Schmid Mast, M. (2015). Unpublished raw data.

*Carroll, J. M., & Yung, C. K. (2006). Sex and discipline differences in empathising, systemising and autistic symptomatology: Evidence from a student population. *Journal of Autism and Developmental Disorders*, *36*(7), 949–957.

*Carter, J. D., & Hall, J. A. (2008). Individual differences in the accuracy of detecting social covariations: Ecological sensitivity. *Journal of Research in Personality*, *42*(2), 439–455.

Castro, V. L., & Boone, R. T. (2015). Sensitivity to spatiotemporal percepts predicts the perception of emotion. *Journal of Nonverbal Behavior*, *39*(3), 215–240.

*Castro, V. L., & Isaacowitz, D. M. (2017). Unpublished raw data.

Cattell, R. B. (1950). *Culture fair intelligence test: A measure of "g"*. Savoy, IL: Institute for Personality and Ability Testing.

Cattell, R. B. (1971). Abilities: Their structure, growth, and action.

Côté, S. (2014). Emotional intelligence in Organizations. *Annual Review of Organizational Psychology and Organizational Behavior*, *1*(1), 459–488.

*Davies, M., Stankov, L., & Roberts, R. D. (1998). Emotional intelligence: In search of an elusive construct. *Journal of Personality and Social Psychology*, *75*(4), 989–1015.

*Davitz, J. R., Beldoch, M., Blau, S., Dimitrovsky, L., Levitt, E., & Kempner Levy, P. (1964). Personality, perceptual, and cognitive correlates of emotional sensitivity. In J. R. Davitz (Ed.), *The communication of emotional meaning* (pp. 57–68). New York: McGraw-Hill.

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (2000). *California verbal learning test-second edition (CVLT-II)*. San Antonio, TX: Psychological Corporation.

DeVellis, R. F. (2016). *Scale development: Theory and applications* (Vol. 26). Sage publications.

*Doucet, L., Shao, B., Wang, L., & Oldham, G. R. (2016). I know how you feel, but it does not always help: Integrating emotion recognition, agreeableness, and cognitive ability in a compensatory model of service performance. *Journal of Service Management*, *27*(3), 320–338.

Dricu, M., & Frühholz, S. (2016). Perceiving emotional expressions in others: Activation likelihood estimation meta-analyses of explicit evaluation, passive perception and incidental perception of emotions. *Neuroscience & Biobehavioral Reviews*, *71*, 810–828.

*Dziobek, I., Fleck, S., Kalbe, E., Rogers, K., Hassenstab, J., Brand, M., … Convit, A. (2006). Introducing MASC: A movie for the assessment of social cognition. *Journal of Autism and Developmental Disorders*, *36*(5), 623–636.

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*(4), 205–223.

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

*Elfenbein, H. A., Jang, D., Sharma, S., & Sanchez-Burks, J. (2017). Validating emotional attention regulation as a component of emotional intelligence: A Stroop approach to individual differences in tuning in to and out of nonverbal cues. *Emotion*, *17* (2), 348–358.

*Farrelly, D., & Austin, E. J. (2007). Ability EI as an intelligence? Associations of the MSCEIT with performance on emotion processing and social tasks and with cognitive ability. *Cognition and Emotion*, *21*(5), 1043–1063.

*Ferguson, F. J., & Austin, E. J. (2010). Associations of trait and ability emotional intelligence with performance on Theory of Mind tasks in an adult sample. *Personality and Individual Differences*, *49*(5), 414–418.

Fiori, M. (2015). Emotional intelligence compensates for low IQ and boosts low emotionality individuals in a self-presentation task. *Personality and Individual Differences*, *81*, 169–173.

Fiori, M., Antonietti, J.-P., Mikolajczak, M., Luminet, O., Hansenne, M., & Rossier, J. (2014). What is the ability emotional intelligence test (MSCEIT) good for? An evaluation using item response theory. *PloS One*, *9*(6), e98827.

*Fiori, M., & Antonakis, J. (2011). The ability model of emotional intelligence: Searching for valid measures. *Personality and Individual Differences*, *50*(3), 329–334.

*Fiori, M., & Antonakis, J. (2012). Selective attention to emotional stimuli: What IQ and openness do, and emotional intelligence does not. *Intelligence*, *40*(3), 245–254.

*Franklin Jr, R. G., & Zebrowitz, L. A. (2017). Age differences In emotion recognition: Task Demands Or perceptual Dedifferentiation? *Experimental Aging Research*, *43*(5), 453–466.

Frattaroli, J. (2006). Experimental disclosure and its moderators: A meta-analysis. *Psychological Bulletin*, *132*, 823–865.

*Frommeyer, N., & Klonek, F. (2015). Unpublished data.

*Fujiwara, K. (2016). Unpublished raw data.

Goh, J. X., Hall, J. A., & Rosenthal, R. (2016). Mini meta-analysis of your own studies: Some arguments on why and a primer on how. *Social and Personality Psychology Compass*, *10*(10), 535–549.

*Grainger, S. A., Henry, J. D., Phillips, L. H., Vanman, E. J., & Allen, R. (2015). Age deficits in facial affect recognition: The influence of dynamic cues. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *72*(4), 622–632.

Halberstadt, A. G., Denham, S. A., & Dunsmore, J. C. (2001). Affective social competence. *Social Development*, *10*(1), 79–119.

Halberstadt, A. G., & Hall, J. A. (1980). Who's getting the message? Children's nonverbal skill and their evaluation by teachers. *Developmental Psychology*, *16*(6), 564–573.

*Halberstadt, J., Ruffman, T., Murray, J., Taumoepeau, M., & Ryan, M. (2011). Emotion perception explains age-related differences in the perception of social gaffes. *Psychology and Aging*, *26*(1), 133–136.

Hall, J. A., Andrzejewski, S. A., & Yopchick, J. E. (2009). Psychosocial correlates of interpersonal sensitivity: A meta-analysis. *Journal of Nonverbal Behavior*, *33*(3), 149–180.

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

Hall, J. A., & Rosenthal, R. (2018). Choosing between random effects models in meta-analysis: Units of analysis and the generalizability of obtained results. *Social and Personality Psychology Compass*, 12, e12417.

Hall, J. A., Ship, A. N., Ruben, M. A., Curtin, E. M., Roter, D. L., Clever, S. L., … Pounds, K. (2015). Clinically relevant correlates of accurate perception of patients' Thoughts and Feelings. *Health Communication*, *30*(5), 423–429.

Hampson, E., van Anders, S. M., & Mullin, L. I. (2006). A female advantage in the recognition of emotional facial expressions: Test of an evolutionary hypothesis. *Evolution and Human Behavior*, *27*(6), 401–416.

*Hassenstab, J., Dziobek, I., Rogers, K., Wolf, O. T., & Convit, A. (2007). Knowing what others know, feeling what others feel: A controlled study of empathy in psychotherapists. *The Journal of Nervous and Mental Disease*, *195*(4), 277–281.

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, *1*, 39–65.

*Henry, J. D., Phillips, L. H., Beatty, W. W., McDonald, S., Longley, W. A., Joscelyne, A., & Rendell, P. G. (2009). Evidence for deficits in facial affect recognition and theory of mind in multiple sclerosis. *Journal of the International Neuropsychological Society*, *15*(2), 277–285.

Hildebrandt, A., Sommer, W., Schacht, A., & Wilhelm, O. (2015). Perceiving and remembering emotional facial expressions—A basic facet of emotional intelligence. *Intelligence*, *50*, 52–67.

*Holding, B. C., Laukka, P., Fischer, H., Bänziger, T., Axelsson, J., & Sundelin, T. (2017). Multimodal emotion recognition Is Resilient to Insufficient Sleep: Results from cross-Sectional and Experimental studies. *Sleep*, *40*, zsx145.

Hox, J. J., & de Leeuw, E. D. (2003). Multilevel models for meta-analysis. In S. P. Reise & N. Duan (Eds.), *Multilevel modeling: Methodological Advances, Issues, and Applications* (pp. 90–111). Mahwah, NJ, US: Lawrence Erlbaum Associates.

Isaacowitz, D. M., Vicaria, I. M., & Murry, M. W. (2016). A lifespan developmental perspective on interpersonal accuracy. In J. A. Hall, M. Schmid Mast, & T. V. West (Eds.), *The social Psychology of Perceiving others accurately* (pp. 206–229). Cambridge, United Kingdom: Cambridge University Press.

*Ivan, L., & Duduciuc, A. (2011). Social skills, nonverbal sensitivity and academic success. The key role of centrality in student networks for higher grades achievement. *Review of Research and Social Intervention*, *33*, 151–166.

*Jang, D., & Elfenbein, H. A. (2015). Unpublished raw data.

Jastak, J. F., & Jastak, S. R. (1978). *WRAT: Wide range achievement test*. Jastak Associates.

*Kanner, L. (1931). Judging emotions from facial expression. *Psychological Monographs*, *41*(3), i–91.

*Karpouzian, T. M., Alden, E. C., Reilly, J. L., & Smith, M. J. (2016). High functioning individuals with schizophrenia have preserved social perception but not mentalizing abilities. *Schizophrenia Research*, *171*(1), 137–139.

Kaufman, A. S., & Kaufman, N. L. (1990). *Kaufman brief intelligence test*. Circle Pines, MN: American Guidance Service.

Kaufman, A. S., & Kaufman, N. L. (2014). *Kaufman adolescent and adult intelligence test*. Encyclopedia of Special Education. American Cancer Society.

Kohler, C. G., Walker, J. B., Martin, E. A., Healey, K. M., & Moberg, P. J. (2010). Facial emotion perception in schizophrenia: A meta-analytic review. *Schizophrenia Bulletin*, *36*(5), 1009–1019.

*Kong, D. T. (2016). Ostracism perception as a multiplicative function of trait self-esteem, mindfulness, and facial emotion recognition ability. *Personality and Individual Differences*, *93*, 68–73.

Konstantopoulos, S. (2011). Fixed effects and variance components estimation in three-level meta-analysis. *Research Synthesis Methods*, *2*(1), 61–76.

*Kunzmann, U., Wieck, C., & Dietzel, C. (2018). Empathic accuracy: Age differences from adolescence into middle adulthood. *Cognition and Emotion*, *0*(0), 1–14.

*Lawrence, K., Campbell, R., Swettenham, J., Terstegge, J., Akers, R., Coleman, M., & Skuse, D. (2003). Interpreting gaze in Turner syndrome: Impaired sensitivity to intention and emotion, but preservation of social cueing. *Neuropsychologia*, *41*(8), 894–905.

Legree, P. J., Mullins, H. M., LaPort, K. A., & Roberts, R. D. (2016). SLODR-house rules: EI tests less g loaded in higher ability groups. *Intelligence*, *59*, 32–38.

Legree, P. J., Psotka, J., Robbins, J., Roberts, R. D., Putka, D. J., & Mullins, H. M. (2014). Profile similarity metrics as an alternate framework to score rating-based tests: MSCEIT reanalyses. *Intelligence*, *47*, 159–174.

*Lewis, G. J., Lefevre, C. E., & Young, A. W. (2016). Functional architecture of visual emotion recognition ability: A latent variable approach. *Journal of Experimental Psychology: General*, *145*(5), 589–602.

Li, S.-C., & Lindenberger, U. (1999). Cross-level unification: A computational exploration of the link between deterioration of neurotransmitter systems and dedifferentiation of cognitive abilities in old age. In *Cognitive neuroscience of memory* (pp. 103–146). Hogrefe & Huber.

*Libbrecht, N., & Lievens, F. (2012). Validity evidence for the situational judgment test paradigm in emotional intelligence measurement. *International Journal of Psychology*, *47*(6), 438–447.

*Lima, C. F., Alves, T., Scott, S. K., & Castro, S. L. (2014). In the ear of the beholder: How age shapes emotion processing in nonverbal vocalizations. *Emotion*, *14*(1), 145–160.

*Lippens, L. (2015). *Short form of the Wilde Intelligenztest: Psychometric qualities and utility of a 12-minute intelligence test for personnel selection* (Master's thesis). Retrieved from https://lib.ugent.be/fulltxt/RUG01/002/274/635/RUG01-002274635_2016_0001_AC.pdf

*Lopes, P. N., Salovey, P., & Straus, R. (2003). Emotional intelligence, personality, and the perceived quality of social relationships. *Personality and Individual Differences*, *35*(3), 641–658.

*Lyusin, D., & Ovsyannikova, V. (2016). Measuring two aspects of emotion recognition ability: Accuracy vs. Sensitivity. *Learning and Individual Differences*, *52*, 129–136.

*MacCann, C., Joseph, D. L., Newman, D. A., & Roberts, R. D. (2014). Emotional intelligence is a second-stratum factor of intelligence: Evidence from hierarchical and bifactor models. *Emotion*, *14*(2), 358–374.

*MacCann, C., Pearce, N., & Roberts, R. (2011). Emotional intelligence as assessed by situational judgment and emotion recognition tests: Building the nomological net. *Psychological Topics*, *20*(3), 393–412.

*MacCann, C., Roberts, R. D., Matthews, G., & Zeidner, M. (2004). Consensus scoring and empirical option weighting of performance-based emotional intelligence (EI) tests. *Personality and Individual Differences*, *36*(3), 645–662.

Mackintosh, N. J. (2011). *IQ and human intelligence*. Oxford, UK: Oxford University Press.

*Mar, R. A., Oatley, K., Hirsh, J., dela Paz, J., & Peterson, J. B. (2006). Bookworm versus nerds: Exposure to fiction versus non-fiction, divergent associations with personality, ability, and achievement. *Journal of Research in Personality*, *40*, 694–712.

*Mason, A., & Tang, R. (2010). Unpublished raw data.

Matsumoto, D., LeRoux, J., Wilson-Cohn, C., Raroque, J., Kooken, K., Ekman, P., … Goh, A. (2000). A New test to measure emotion recognition ability: Matsumoto and Ekman's Japanese and Caucasian brief affect recognition test (JACBART). *Journal of Nonverbal Behavior*, *24*(3), 179–209.

Mayer, J. D., Caruso, D. R., & Salovey, P. (2016). The ability model of emotional intelligence: Principles and updates. *Emotion Review*, *8*(4), 290–300.

Mayer, J. D., Salovey, P., Caruso, D. R., & Sitarenios, G. (2003). Measuring emotional intelligence with the MSCEIT V2.0. *Emotion*, *3*(1), 97–105.

McArthur, L. Z., & Baron, R. M. (1983). Toward an ecological theory of social perception. *Psychological Review*, *90*(3), 215–238.

*Mestre, J. M. (2016). Unpublished raw data.

*Mestre, J. M., Larrán, C., Herrero, J., Guil, R., & de la Torre, G. G. (2015). PERVALE-S: A new cognitive task to assess deaf people's ability to perceive basic and social emotions. *Frontiers in Psychology*, *6*(1148).

*Miller, A. B., & Lenzenweger, M. F. (2012). Schizotypy, social cognition, and interpersonal sensitivity. *Personality Disorders: Theory, Research, and Treatment*, *3*(4), 379–392.

*Murphy, N. A. (2015). Unpublished raw data.

Murphy, N. A., & Hall, J. A. (2011). Intelligence and interpersonal sensitivity: A meta-analysis. *Intelligence*, *39*(1), 54–63.

*Murry, M. W. E., & Isaacowitz, D. M. (2017). Age differences in emotion perception: The effects of the social environment. *International Journal of Behavioral Development*, *41*(5), 597–604.

Nowicki, S., & Duke, M. P. (1994). Individual differences in the nonverbal communication of affect: The Diagnostic analysis of nonverbal accuracy scale. *Journal of Nonverbal Behavior*, *18*(1), 9–35.

Olderbak, S. G., Mokros, A., Nitschke, J., Habermeyer, E., & Wilhelm, O. (2018). Psychopathic men: Deficits in general mental ability, not emotion perception. *Journal of Abnormal Psychology*, *127*(3), 294–304.

*Olderbak, S., Wilhelm, O., Olaru, G., Geiger, M., Brenneman, M. W., & Roberts, R. D. (2015). A psychometric analysis of the reading the mind in the eyes test: Toward a brief form for research and applied settings. *Frontiers in Psychology*, *6*, 1503.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*, aac4716.

Otis, A. S. (1954). *Otis quick-scoring mental ability tests*. Harcourt, Brace and World.

*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), e68126.

Palese, T., & Schmid Mast, M. (2017). Interpersonal accuracy is related to leadership behavioral adaptability in women. *Manuscript under review*.

Peterson, E., & Miller, S. (2012). The eyes test as a measure of individual differences: How much of the variance reflects verbal IQ? *Frontiers in Psychology*, *3*(220).

*Phillips, L. H., MacLean, R. D., & Allen, R. (2002). Age and the understanding of emotions: Neuropsychological and sociocognitive perspectives. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *57*(6), P526–P530.

*Pickett, C. L., Gardner, W. L., & Knowles, M. (2004). Getting a cue: The need to belong and enhanced sensitivity to social cues. *Personality and Social Psychology Bulletin*, *30*(9), 1095–1107.

*Pinkham, A. E., Penn, D. L., Green, M. F., & Harvey, P. D. (2015). Social cognition psychometric evaluation: Results of the initial psychometric study. *Schizophrenia Bulletin*, *42*(2), 494–504.

Pitterman, H., & Nowicki, S. (2004). A test of the ability to identify emotion in human standing and sitting postures: The diagnostic analysis of nonverbal accuracy-2 posture test (DANVA2-POS). *Genetic, Social, and General Psychology Monographs*, *130*(2), 146–162.

*Plesa-Skwerer, D., Faja, S., Schofield, C., Verbalis, A., & Tager-Flusberg, H. (2006). Perceiving facial and vocal expressions of emotion in individuals with Williams syndrome. *American Journal on Mental Retardation*, *111*(1), 15–26.

*Plesa Skwerer, D., Verbalis, A., Schofield, C., Faja, S., & Tager-Flusberg, H. (2006). Social-perceptual abilities in adolescents and adults with Williams syndrome. *Cognitive Neuropsychology*, *23*(2), 338–349.

*Pool, L. D., & Qualter, P. (2012). Improving emotional intelligence and emotional self-efficacy through a teaching intervention for university students. *Learning and Individual Differences*, *22*(3), 306–312.

Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin*, *135*(2), 322–338.

Rammsayer, T. H., & Troche, S. J. (2012). On sex-related differences in auditory and visual sensory functioning. *Archives of Sexual Behavior*, *41*(3), 583–590.

Raven, J. (1981). *Manual for Raven's progressive matrices and vocabulary scales*. Oxford: Oxford Psychologists Press.

Raven, J., Raven, J. C., & Court, J. H. (1998). *Raven manual: Section 4, advanced progressive matrices*. Oxford: Oxford Psychologists Press.

*Realo, A., Allik, J., Nolvak, A., Valk, R., Ruus, T., Schmidt, M., & Eilola, T. (2003). Mind-reading ability: Beliefs and performance. *Journal of Research in Personality*, *37*(5), 420–445.

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin*, *138*(2), 353–387.

*Roberts, R. D., Schulze, R., O'brien, K., MacCann, C., Reid, J., & Maul, A. (2006). Exploring the validity of the Mayer-Salovey-Caruso emotional intelligence test (MSCEIT) with established emotions measures. *Emotion*, *6*(4), 663–669.

*Rode, J. C., Mooney, C. H., Arthaud-Day, M. L., Near, J. P., Rubin, R. S., Baldwin, T. T., & Bommer, W. H. (2008). An examination of the structural, discriminant, nomological, and incremental predictive validity of the MSCEIT© V2. 0. *Intelligence*, *36*(4), 350–366.

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

Rosip, J. C., & Hall, J. A. (2004). Knowledge of nonverbal cues, gender, and nonverbal decoding accuracy. *Journal of Nonverbal Behavior*, *28*(4), 267–286.

*Ryan, M., Murray, J., & Ruffman, T. (2010). Aging and the perception of emotion: Processing vocal expressions alone and with faces. *Experimental Aging Research*, *36*(1), 1–22.

*Sawyer, A. C., Williamson, P., & Young, R. L. (2012). Can gaze avoidance explain why individuals with Asperger's syndrome can't recognise emotions from facial expressions? *Journal of Autism and Developmental Disorders*, *42*(4), 606–618.

Scherer, K. R., & Scherer, U. (2011). Assessing the ability to recognize facial and vocal expressions of emotion: Construction and validation of the emotion recognition Index. *Journal of Nonverbal Behavior*, *35*(4), 305–326.

Schimmack, U., & Brunner, J. (2017, November 17). *Z-Curve*. doi:10.31219/osf.io/wr93f

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.

Schlegel, K., Fontaine, J. R. J., & Scherer, K. R. (2017). The nomological network of emotion recognition ability: Evidence from the Geneva emotion recognition test. *European Journal of Psychological Assessment*. doi:10.1027/1015-5759/a000396

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.

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

Schlegel, K., & Scherer, K. R. (2018). The nomological network of emotion knowledge and emotion understanding in adults: Evidence from two new performance-based tests. *Cognition and Emotion*, *32*, 1514–1530.

*Schlegel, K., & Hall, J. A. (2016). Unpublished raw data.

*Schlegel, K., Mehu, M., van Peer, J. M., & Scherer, K. R. (2018). Sense and sensibility: The role of cognitive and emotional intelligence in negotiation. *Journal of Research in Personality*, *74*, 6–15.

*Schlegel, K., & Mortillaro, M. (2018). The Geneva Emotional Competence Test (GECo): An ability measure of workplace emotional intelligence. *Journal of Applied Psychology*, *104*(4), 559–580.

*Schlegel, K., Vicaria, I. M., Isaacowitz, D. M., & Hall, J. A. (2017). Effectiveness of a short audiovisual emotion recognition training program in adults. *Motivation and Emotion*, *41*(5), 646–660.

*Schlegel, K., Witmer, J. S., & Rammsayer, T. H. (2017). Intelligence and sensory sensitivity as predictors of emotion recognition ability. *Journal of Intelligence*, *5*(4), 35.

*Schmid Mast, M. (2015). Unpublished raw data.

Schmidt, F. L., & Hunter, J. (2004). General mental ability in the world of work: Occupational attainment and job performance. *Journal of Personality and Social Psychology*, *86*(1), 162–173.

Schneider, W. J., Mayer, J. D., & Newman, D. A. (2016). Integrating Hot and Cool Intelligences: Thinking Broadly about broad abilities. *Journal of Intelligence*, *4*(1), 1.

Schneider, W. J., & McGrew, K. S. (2012). The Cattell-Horn-Carroll model of intelligence. In D. P. Flanagan & P. L. Harrison (Eds.),

Contemporary intellectual assessment: Theories, tests, and issues (3rd ed., pp. 99–144). New York, NY: Guilford.

Sheppard, L. D., & Vernon, P. A. (2008). Intelligence and speed of information-processing: A review of 50 years of research. *Personality and Individual Differences*, *44*(3), 535–551.

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). p-curve and effect size: Correcting for publication bias using only significant results. *Perspectives on Psychological Science*, *9*(6), 666–681.

*Simpson, C., Pinkham, A. E., Kelsven, S., & Sasson, N. J. (2013). Emotion recognition abilities across stimulus modalities in schizophrenia and the role of visual attention. *Schizophrenia Research*, *151*(1), 102–106.

*Slessor, G., Phillips, L. H., & Bull, R. (2007). Exploring the specificity of age-related differences in theory of mind tasks. *Psychology and Aging*, *22*(3), 639–643.

*Smith, R., Alkozei, A., & Killgore, W. D. S. (2016). Contributions of self-report and performance-based individual differences measures of social cognitive ability to large-scale neural network functioning. *Brain Imaging and Behavior*, *11*(3), 685–697.

Stankov, L. (1997). *The Gf/Gc quickie test battery* (Unpublished test battery from the School of Psychology). University of Sydney, Australia.

*Sternberg, R. J., & Smith, C. (1985). Social intelligence and decoding skills in nonverbal communication. *Social Cognition*, *3*(2), 168–192.

Strenze, T. (2007). Intelligence and socioeconomic success: A meta-analytic review of longitudinal research. *Intelligence*, *35*(5), 401–426.

*Suzuki, A., & Akiyama, H. (2013). Cognitive aging explains age-related differences in face-based recognition of basic emotions except for anger and disgust. *Aging, Neuropsychology, and Cognition*, *20*(3), 253–270.

*Suzuki, A., Hoshino, T., Shigemasu, K., & Kawamura, M. (2007). Decline or improvement?: Age-related differences in facial expression recognition. *Biological Psychology*, *74*(1), 75–84.

*Szily, E., & Kéri, S. (2009). Anomalous subjective experience and psychosis risk in young depressed patients. *Psychopathology*, *42*(4), 229–235.

Thiébaut, E., & Bidan-Fortier, C. (2003). *Batterie de tests d'aptitudes cognitives NV5-R – Manuel*. Paris: Hogrefe.

*Trimmer, C. G., & Cuddy, L. L. (2008). Emotional intelligence, not music training, predicts recognition of emotional speech prosody. *Emotion*, *8*(6), 838–849.

*Tso, I. F., Grove, T. B., & Taylor, S. F. (2010). Emotional experience predicts social adjustment independent of neurocognition and social cognition in schizophrenia. *Schizophrenia Research*, *122*(1), 156–163.

*Turkstra, L. S. (2008). Conversation-based assessment of social cognition in adults with traumatic brain injury. *Brain Injury*, *22*(5), 397–409.

*Turkstra, L. S. (2016). Unpublished raw data.

*Valadez-Sierra, M. D., Borges del Rosal, M. A., Ruvalcaba Romero, N., Villegas, K., & Lorenzo, M. (2013). Emotional intelligence and its relationship with gender, academic performance and intellectual abilities of Undergraduates. *Electronic Journal of Research in Educational Psychology*, *11*(2), 395–412.

*Vicaria, I. M. (2016). Unpublished raw data.

*Walter, F., Cole, M. S., van der Vegt, G. S., Rubin, R. S., & Bommer, W. H. (2012). Emotion recognition and emergent leadership: Unraveling mediating mechanisms and boundary conditions. *The Leadership Quarterly*, *23*(5), 977–991.

*Walther, S., Stegmayer, K., Sulzbacher, J., Vanbellingen, T., Müri, R., Strik, W., & Bohlhalter, S. (2015). Nonverbal social communication and gesture control in schizophrenia. *Schizophrenia Bulletin*, *41*(2), 338–345.

*Webb, C. A., Schwab, Z. J., Weber, M., DelDonno, S., Kipman, M., Weiner, M. R., & Killgore, W. D. (2013). Convergent and divergent validity of integrative versus mixed model measures of emotional intelligence. *Intelligence*, *41*(3), 149–156.

Wechsler, D. (1955). *Wechsler adult intelligence scale*. New York: Psychological Corporation.

*Weisgerber, C. A. (1956). Accuracy in judging emotional expressions as related to college entrance test scores. *The Journal of Social Psychology*, *44*(2), 233–239.

Weiss, R. H. (2006). *Grundintelligenztest Skala 2 (CFT 20-R) mit Wortschatztest (WS) und Zahlenfolgentest (ZF) - Revision*. Göttingen: Hogrefe.

*Westbrook, M. (1974). Judgement of emotion: Attention versus accuracy. *British Journal of Social and Clinical Psychology*, *13*, 383–390.

White, K. R. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin*, *91*, 461–481.

*Wieck, C., & Kunzmann, U. (2015). Age differences in empathy: Multidirectional and context-dependent. *Psychology and Aging*, *30*(2), 407–419.

*Wolf, R. C., Pujara, M., Baskaya, M. K., & Koenigs, M. (2016). Emotion recognition deficits associated with ventromedial prefrontal cortex lesions are improved by gaze manipulation. *Cortex*, *82*, 255–262.

Wonderlic, E. F. (1961). *Wonderlic personnel test manual (No. 1)*. EF Wonderlic & Associates.

Zachary, R. (1986). *Shipley institute of living scale, revised manual*. Los Angeles: Western Psychological Services.

*Zentner, M., Beermann, U., & Strauss, H. (2016). Unpublished raw data.

Zuckerman, M., Silberman, J., & Hall, J. A. (2013). The relation between intelligence and religiosity: A meta-analysis and some proposed explanations. *Personality and Social Psychology Review*, *17*, 325–354.

*Zuckerman, M., Lipets, M. S., Koivumaki, J. H., & Rosenthal, R. (1975). Encoding and decoding nonverbal cues of emotion. *Journal of Personality and Social Psychology*, *32*(6), 1068–1076.

References marked with an asterisk provided data used in the meta-analysis.

Appendix A

ERA and intelligence tests included in the meta-analysis. Frequency includes all test versions or forms, as well as subtests of a particular test. Tests with a frequency of 10 or higher are displayed.

Test	Number of effect sizes	Brief description of task	Source
ERA tests			
Diagnostic Analysis of Nonverbal Accuracy (DANVA)	80	Participants view pictures of faces (Faces subtest) or listen to voice recordings of a standard sentence (Voices subtest) or view pictures of full body postures with the face blacked out (Postures subtest), and choose, for each item, which out of four emotions was being expressed	Nowicki and Duke (1994); Baum and Nowicki (1998); Pitterman and Nowicki (2004)
Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT) Faces Subtest	75	Participants view pictures of faces and rate, for each picture, to what extent each of five emotions is expressed in the picture	Mayer, Salovey, Caruso, and Sitarenios (2003)
Reading Mind in the Eyes Test (RMET)	39	Participants view pictures of eye regions and choose, for each picture, which out of four affective words is expressed by the eyes	Baron-Cohen, Jolliffe, Mortimore, and Robertson (1997)
Profile of Nonverbal Sensitivity (PONS) and MiniPONS	35	Participants view brief recordings of one actor that include her voice (random spliced to mask content), face, or body, or a combination of these. After each clip, participants choose which out of two affective situations was expressed by the actor	Rosenthal et al. (1979); Bänziger, Scherer, Hall, and Rosenthal (2011)
Geneva Emotion Recognition Test (GERT)	32	Participants view brief video recordings (upper body and face visible) with sound in which actors express 14 emotions while saying a standard sentence; after each video they choose which emotion was expressed	Schlegel et al. (2014); Schlegel and Scherer (2016)
Multimodal Emotion Recognition Test (MERT)	14	Video clips (upper body and face visible) of actors expressing 10 emotions while saying a standard sentence are presented in four modalities: still picture, video without sound, video with sound, audio only; after each item, participants choose which emotion was expressed	Bänziger et al. (2009)
Emotion Recognition Index (ERI)	13	Participants view pictures of faces (Faces subtest) or listen to voice recordings of a standard sentence (Voices subtest) and choose, for each item, which out of five emotions was expressed	Scherer and Scherer (2011)
other ERA measures	183		
Intelligence and academic achievement tests			
Wechsler Adult Intelligence Scale (WAIS)	63	Full scale intelligence battery measuring verbal comprehension, working memory, perceptual organisation, and processing speed	Wechsler (1955)
Scholastic Aptitude Test (SAT)	40	College admission test including subtests for reading, writing, language, and mathematics	–
Raven's Progressive Matrices (standard and advanced)	40	Fluid intelligence test consisting of visual geometric designs with a missing piece; participants choose which out of a range of missing pieces completes the design	Raven (1981); Raven, Raven, and Court (1998)
Gf/ Gc Quickie Test Battery	39	Test battery containing subtests measuring fluid intelligence (e.g. matrix completion, completing letter series) and crystallized intelligence (e.g. vocabulary and general knowledge tests)	Stankov (1997)
NV5-R	37	Full scale intelligence battery including subtests measuring fluid (e.g. completing number series) and crystallized intelligence (e.g. vocabulary)	Thiébaut and Bidan-Fortier (2003)
Culture Fair Intelligence Test (CFT/ CFIT)	21	Fluid intelligence test requiring participants to infer complex relationships between elements of figures	Cattell (1950); Weiss (2006)

(*Continued*)

Continued.

Test	Number of effect sizes	Brief description of task	Source
Shipley Vocabulary Test	18	Crystallized intelligence test in which participants are asked to identify which out of six words has the same meaning as a given target word	Zachary (1986)
Grade Point Average GPA (all levels)	18	Average value of the accumulated final grades earned in courses over time	–
California Verbal Learning Test (CVLT)	15	Memory test measuring immediate, short- and long-term recall	Delis, Kramer, Kaplan, and Ober (2000)
Wonderlic Personnel Test	12	Full scale intelligence test including question types such as analogies, word definitions, analysis of geometric figures, and arithmetic	Wonderlic (1961)
Wide Range Achievement Test (WRAT)	11	Crystallized intelligence test assessing reading comprehension, spelling, and arithmetic computation	Jastak and Jastak (1978)
Kaufman Brief Intelligence Test (KBIT) and Kaufman Adolescent and Adult Intelligence Test (KAIT)	11	Full scale intelligence battery including subtests measuring sequential reasoning, long-term memory, word knowledge etc.	Kaufman and Kaufman (1990, 2014)
Otis Quick-Scoring Mental Ability Test	10	Full scale intelligence test including question types such as word comparisons and analysis of geometric figures, and arithmetic	Otis (1954)
Trail Making Test	10	Measures speed and efficiency of information processing by asking participants to connect a set of dots as quickly and accurately as possible	Army Individual Test Battery (1944)
other intelligence measures	126		