AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychological Bulletin

© 2024 The Author(s)
ISSN: 0033-2909

2024, Vol. 150, No. 9, 1094–1117
https://doi.org/10.1037/bul0000442

Age-Related Changes in Emotion Recognition Across Childhood: A Meta-Analytic Review

Christopher Riddell[1, 2], Milica Nikolić[3], Elise Dusseldorp[4], and Mariska E. Kret[1, 2]
[1] Cognitive Psychology Unit, Institute of Psychology, Leiden University
[2] Leiden Institute for Brain and Cognition, Leiden University
[3] Research Institute of Child Development and Education, University of Amsterdam
[4] Methodology and Statistics Unit, Institute of Psychology, Leiden University

Children's ability to accurately recognize the external emotional signals produced by those around them represents a milestone in their socioemotional development and is associated with a number of important psychosocial outcomes. A plethora of individual studies have examined when, and in which order, children acquire emotion knowledge over the course of their development. Yet, very few attempts have been made to summarize this body of work quantitatively. To address this, the present meta-analysis examined the age-related trajectories of emotion recognition across childhood and the extent to which typically developing children's recognition of external emotional cues (in the face, voice, and body) is influenced by a host of participant-, task-, and stimulus-related factors. We analyzed children's emotion recognition overall (independent of specific emotion categories) and for specific basic emotions. In total, $k = 129$ individual studies, investigating a total of $N = 31{,}101$ 2–12-year-old children's emotion recognition abilities were included in our analyses. Children's recognition accuracy across all emotion categories was significantly above chance and improved with age in the same manner for all emotions. Emotion recognition accuracy was also moderated by region of study and task type. The order in which children became proficient at identifying specific emotions was consistent with previous qualitative reviews: Happiness was the easiest emotion to recognize, and disgust and fear were the most difficult to recognize across age. Task- and stimulus-related moderator variables also influenced specific emotion categories in different ways. We contextualize these results with regard to children's socioemotional development more broadly, and we discuss how our findings can be used to guide researchers and practitioners interested in children's social skills.

Public Significance Statement

The accurate recognition of others' emotions is a fundamental social skill, relevant for navigating the social world from early childhood. The present meta-analysis demonstrates that children's emotion recognition accuracy increases across childhood in the same manner for different emotions (i.e., happiness, sadness, fear, anger, surprise, and disgust). In addition, children's emotion recognition accuracy was found to depend on tasks that are used to test these abilities. These findings suggest that child development researchers and practitioners should carefully consider age as well as the tests they use when assessing emotion recognition in childhood.

Keywords: emotion recognition, emotion knowledge, basic emotions, socioemotional development

Supplemental materials: https://doi.org/10.1037/bul0000442.supp

Blair T. Johnson served as action editor.
Christopher Riddell https://orcid.org/0000-0002-0605-5526
Christopher Riddell and Mariska E. Kret's research on emotion recognition was supported by the European Research Council (Starting Grant 804582) awarded to Mariska E. Kret. Milica Nikolić's research on emotional development was supported by an Nederlandse Organisatie voor Wetenschappelijk Onderzoek VENI Grant (Veni.201G.017). This research was funded in whole, or in part, by the Netherlands Organisation for Scientific Research (NWO; Grant Veni.201G.017). The authors thank Janneke Staaks for assisting them with their literature search. The authors declare no conflicts of interest. For the purpose of open access, the author has applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission. All analytic data, analysis code, and research materials (including our coding scheme) are available at https://osf.io/68g7z/?view_only=4b5d5b8dec65463fbbe099ee89e58b37.

This work is licensed under a Creative Commons Attribution-Non Commercial-No Derivatives 4.0 International License (CC BY-NC-ND 4.0; https://creativecommons.org/licenses/by-nc-nd/4.0). This license permits copying and redistributing the work in any medium or format for noncommercial use provided the original authors and source are credited and a link to the license is included in attribution. No derivative works are permitted under this license.

Christopher Riddell played a lead role in formal analysis, visualization, and writing–original draft and an equal role in conceptualization, investigation, and methodology. Milica Nikolić played an equal role in conceptualization, methodology, validation, and writing–review and editing. Elise Dusseldorp played an equal role in formal analysis, methodology, and writing–review and editing. Mariska E. Kret played a lead role in funding acquisition and a supporting role in methodology and writing–review and editing.

Correspondence concerning this article should be addressed to Christopher Riddell, Cognitive Psychology Unit, Institute of Psychology, Leiden University, Wassenaarseweg 52, 2333 AK, Leiden, the Netherlands. Email: c.d.riddell@fsw.leidenuniv.nl

The ability to accurately recognize and interpret the emotional signals produced by other human beings is a vital social skill that is gradually honed throughout childhood. Typically, studies in the field refer to this ability as *emotion recognition*—the ability to infer the internal emotional state of an individual based on a series of external cues (e.g., a configuration of facial muscle movements, body language, or vocal prosody)—and differentiate this state from a list of other emotion categories (Ruba & Pollak, 2020). Many researchers have adopted vastly different approaches and paradigms to examine how children develop these skills, but only a few attempts have been made to analyze what is now a wealth of literature in a systematic way (Herba & Phillips, 2004; Ruba & Pollak, 2020; Widen, 2013). The current article attempts to address this problem, providing the first comprehensive meta-analysis that estimates the extent to which age, task-related properties, and their interaction, influence typically developing children's emotion recognition abilities.

The large size of the literature on children's emotion recognition competencies may be ascribed at least partly to the fact that emotion recognition skills are related to a variety of psychosocial outcomes. Children who are more proficient in recognizing the emotions of others perform better in school (Izard et al., 2001; Torres et al., 2015; Widen, 2019), have higher self-confidence (Widen, 2019), and increased social adjustment (Goodfellow & Nowicki, 2009). Conversely, emotion recognition deficits have been related to a number of disorders present in childhood, including attention deficit hyperactivity disorder (Da Fonseca et al., 2009), autism (Lozier et al., 2014), and psychopathic traits (Dawel et al., 2012; for review, see; Collin et al., 2013). As such, emotion recognition abilities have been argued to represent an important life skill that is relevant even in early childhood (Herba & Phillips, 2004).

Children become increasingly better at recognizing external emotional cues across childhood (Herba & Phillips, 2004). Still, given the large volume of studies on children's emotion recognition abilities, a wide variety of tasks and stimuli have been used, making it difficult to delineate how this process occurs. In other words, the extent to which task- and stimulus-related features, that, along with age, may influence these developmental trajectories is currently unclear. In this meta-analysis, we attempt to address this problem. We consider all conceptualizations and paradigms of emotion recognition in children aged 2 to 12 that use stimuli of discrete basic emotions (happiness, sadness, fear, anger, surprise, disgust). In the following sections, we will summarize the current state of the field, reviewing the myriad of studies that have examined children's emotion recognition competencies. Furthermore, we highlight some of the methodological factors that, along with age, may influence children's responses in emotion recognition paradigms. We then perform two separate meta-analyses: one that tracks the development of emotion recognition across age more generally (independent of emotion categories), and another that considers the trajectories of specific basic emotions separately.

An important caveat, here, is that much of the work in this field implicitly endorses the "basic emotion theory." According to this view, a set of basic emotions—happiness, sadness, anger, fear, disgust, and surprise—are argued to all have an evolutionarily rooted, culturally universal facial expression (Ekman, 1970). In the literature, these categories of expression are typically composed of rigid and highly intense combinations of external facial cues that are argued to represent a direct "readout" of one's internal emotional state. For example, the facial muscles *zygomaticus major* and *orbicularis oculi* contract in unison (resulting in a smile and the raising of the cheeks), and this indicates that one feels the emotion happiness. This view has received criticism (Barrett et al., 2019; Cowen et al., 2019; Durán & Fernández-Dols, 2021; Russell & Barrett, 1999). For example, the constructionist approach to emotions (e.g., Barrett, 2017; Barrett et al., 2019) suggests that expressions of these emotions may not be as culturally universal as has been previously claimed and there appears to exist substantial variation in how individuals display certain emotions, even in identical contexts.

According to this constructionist approach, which sees cultural learning and the language acquisition as crucial in emotional development, children learn to differentiate between discrete emotion categories within their culture as emotion words and concepts are acquired throughout infancy and childhood (Shablack & Lindquist, 2019). Thus, infants are not born able to discriminate between discrete emotion categories but start with experiencing and perceiving general core affective states, which they start differentiating and refining into specific emotion categories with the development of the conceptual knowledge about emotions (Hoemann et al., 2019; Shablack & Lindquist, 2019). Accordingly, children start recognizing happiness for pleasant feelings and sadness (or sometimes anger) for unpleasant feelings at around the age of two. At around this time, they also acquire these emotion words, and, with age, they expand their vocabularies to include fear, surprise, and finally disgust (Widen, 2013; Widen & Russell, 2008a). Contrastingly, although the basic emotion theory postulates that humans are born prepared to experience and differentiate between discrete emotions, these theories also assume that emotion differentiation may increase with age as maturation and social determinants may influence this ability throughout childhood (Izard, 1971, 1977). Therefore, according to both theories, emotion recognition should increase with age in childhood, but both emphasize different mechanisms.

Although we structure our meta-analysis within the framework of basic emotion theory, we do not necessarily endorse it. We are not attempting to solve this theoretical debate (for more detailed discussions, see; Adolphs et al., 2019; Barrett, 2014; Cowen et al., 2019); rather, our goal is to attempt to analyze developmental emotion recognition trajectories within the context of the existing literature, which largely conforms to this view. In addition, given the pertinent psychosocial outcomes associated with poor emotion recognition skills in childhood discussed above (Izard et al., 2001; Torres et al., 2015; Widen, 2019), such an analysis remains needed and useful.

Emotion Recognition Across Age

A number of studies, in a body of literature that now spans roughly half a century, have investigated the precise developmental trajectory in which infants and children learn to differentiate particular emotion categories (for reviews, see: Gross & Ballif, 1991; Herba & Phillips, 2004; Ruba & Pollak, 2020; Widen, 2013). Few have attempted to summarize this wealth of studies in any systematic way. In one such attempt, Herba and Phillips' (2004) review the behavioral and neuroimaging studies concerning children's facial emotion recognition abilities. They find a typical pattern across studies investigating facial expressions, in which children begin by being able to recognize expressions of happiness earliest in childhood, followed by expressions of sadness and anger, then by fear and surprise, ending in disgust. The authors highlight that methodological inconsistencies between studies make it difficult to draw systematic conclusions. This problem has become no less apparent in recent years, as more and more studies have attempted to clarify how children make sense of such external emotional cues (Ruba & Pollak, 2020).

Widen (2013) examined the specific age-related trajectories associated with children's freely produced labels of facial expressions. The author argues that infants first interpret emotional cues as valence-based categories (e.g., pleasant or unpleasant) and gradually develop their abilities in order to deduce discrete emotion categories across childhood. This process has been termed the "broad-to-differentiated" hypothesis (Widen, 2013; Widen & Russell, 2008a). In free-label tasks (where children are asked to produce their own labels for categories of emotional stimuli), a similar pattern to that suggested by Herba and Phillips (2004) was found. Happy smiles were the first facial expression of emotion children consistently identified (from around 24 months of age). As children hone their expressive language abilities and gain experience in the social world, they add more emotion terms to their arsenal. From around 3 years of age, they correctly produce labels for angry scowls and sad cries. It is not until much later during middle-childhood (from around 9 years of age) that children correctly identify disgust nose scrunches in free-labeling tasks (Widen, 2013) and these are still often confused with angry scowls (Widen & Russell, 2010b).

Methodological Features of Emotion Recognition Paradigms

Whereas it is known that the recognition of discrete facial emotion categories (particularly in free-label type tasks) follows a broadly consistent trajectory, it is unclear how much such age-related trajectories are robustly found independently of the methodology used. For example, some individual studies find inconsistent recognition trajectories. Contrary to claims made by Widen (2013), Montirosso et al. (2010), in their study with dynamic expressions, found that angry and sad faces were the least accurately recognized at low intensity in their sample of 4–18-year-old children, with recognition rates lower than that of disgust and fear faces. Moreover, a forced-choice study by Garcia and Tully (2020) found that recognition rates of photographs displaying happiness and sadness facial expressions were highly similar when these emotions were shown at low intensity, with a difference only becoming apparent with high intensity expressions. Finally, Gagnon et al. (2010) found that 5- and 6-year-olds were highly accurate in discriminating fear and disgust facial expressions from other emotion categories when they were asked to match them based on perceptual features (rather than based on specific emotion terms). These inconsistencies highlight the potential interaction between age and certain methodological properties and how it may influence children's responding in emotion recognition tasks.

Other task- and stimulus-related features may also influence emotion recognition. For example, in early investigations, the majority of works in the field of emotion recognition focused on children's ability to identify static, highly posed, and highly intense emotional expressions of adult faces. For both practical and theoretical reasons, studies tended to make use of the same stimulus databases (e.g., Pictures of Facial Affect; Ekman, 1976). In recent years, more studies have begun to make use of more naturalistic (nonposed; e.g., Dawel et al., 2015), subtle (e.g., Gao & Maurer, 2009; Montirosso et al., 2010; Rosenqvist et al., 2014), and dynamic (e.g., Kessels et al., 2014; Montirosso et al., 2010; Richoz et al., 2018; Widen & Russell, 2015) stimuli. In addition, researchers have begun to examine other expression modalities, such as body posture (e.g., de Gelder et al., 2010; Mondloch et al., 2013) and vocalizations/emotional speech (e.g., Chronaki et al., 2015; Gil et al., 2016; Sauter et al., 2013). Finally, more recent studies have used stimuli derived from children and adolescents (e.g., Theurel et al., 2016), rather than only those of adults. All of these many stimuli types may be more representative of the emotional contexts that children are more likely to encounter in daily life and, thus, more representative of the trajectory in which these abilities develop (Vieillard & Guidetti, 2009).

These stimulus-type considerations are not trivial. In their meta-analysis of older and younger adults' emotion recognition competencies, Gonçalves et al. (2018) found that stimulus features (monochrome vs. color, virtual vs. natural, static vs. dynamic) moderated disgust recognition in older adults. Similarly, in a meta-analysis by Hayes et al. (2020), also examining recognition accuracy differences between older and younger adults, the authors found strong age-related effects on emotion recognition—but these were also moderated by stimulus features. Here, older adults performed significantly better than younger adults in recognizing disgust in studies using Ekman's (1976) Pictures of Facial Affect, but this result did not hold when studies used alternative stimulus sets (with older adults demonstrating inferior recognition than younger adults). As such, at least in adults, different stimulus properties appear to interact with age to produce measurable differences in emotion recognition competencies.

There have been no similar meta-analyses with a specific focus on stimulus-related effects on emotion recognition in the child development literature. Nevertheless, some individual studies point to some key differences in children's recognition competencies between particular stimulus types. For example, Richoz et al. (2018) examined the recognition of static and dynamic facial expressions across the lifespan, testing a group in the range of 5–96-years-old. For the majority of the emotions included in the study, the authors found a dynamic over static advantage that differed in magnitude across age. This result demonstrates that stimulus feature considerations may interact with age—potentially providing children at some ages with a "toe-hold" to better recognition accuracy. Yet, this result is inconsistent with two other studies (Nelson et al., 2013; Widen & Russell, 2015), that found that children were no better at identifying dynamic expressions than they were at identifying static expressions (albeit in a free-labeling paradigm).

In addition to considerations of dynamic versus static stimuli, stimulus intensity also appears to be important when considering how emotion recognition develops. As for dynamic stimuli, some discrete emotion categories may be better recognized at lower intensities than others. Indeed, emotions in real life are seldom experienced at the full, and often exaggerated, intensity used in many studies (Kret, 2015; Matsumoto & Hwang, 2014). Children encounter a variety of subtler emotional cues in their social interactions, and it makes sense that some of these less intense signals may be more proficiently recognized than others. This notion is reflected in the literature. For example, Gao and Maurer (2009) examined children's sensitivity to facial expressions of happiness, sadness, and fear. To manipulate intensity, the authors used computer software to morph posed happy, sad, and fearful facial expressions with a neutral expression—resulting in a series of images ranging from 5% to 100% emotional intensity. The authors found that recognition thresholds differed markedly across emotion categories. Five-year-old children's recognition abilities of happiness expressions were similar to those of adults, even in cases where intensity was low. In the case of fear, rather, children's recognition competencies were not adultlike until the age of 10. Thus, emotion category, intensity, and age may interact to produce very different recognition accuracy outcomes.

As mentioned, while the majority of works in this field have focused exclusively on children's recognition of emotional signals produced on the face, far less work has been conducted on the recognition of bodily and vocal signals of emotion. In one study of point-light displays of bodily expressions of emotion, Pollux et al. (2016) found a pattern of recognition accuracy broadly comparable to studies of facial expressions. In this study, they found that children were better at identifying point-light bodily displays of anger, happiness, fear, and sadness compared to those of disgust and surprise. In another study, Witkower et al. (2021) examined children's recognition of bodily expressions of three negative emotions: anger, fear, and sadness. The authors found that bodily expressions of sadness were the first to be recognized above chance (from approximately 3-years-old), followed by bodily expressions of fear (from approximately 4–5-years-old), and then by expressions of anger (from approximately 6–8-years-old).

In terms of studies of vocal emotion, Nelson and Russell (2011a) found that children's ability to identify emotional information in vocal information was less than their ability to identify facial and bodily expressions. In their sample of 3–5-year-old children, the authors found that children were most accurate at identifying sadness and anger intonation in voice clips, and least accurate in identifying happiness and fear. In further work, Sauter et al. (2013) examined 5–10-year-old children's ability to recognize emotional information contained in nonverbal vocalizations (i.e., grunts, laughs, and sighs) and emotionally inflected speech (i.e., sentences expressing positive, neutral, and negative states). The authors used stimuli from a variety of emotion categories (beyond that of simply basic emotions). Children were above chance in recognizing the emotions in both types of vocal signals, and they became more proficient with age. In terms of the basic emotions included in the study, children were the least accurate in identifying sad nonverbal vocalizations and they were most proficient in identifying disgust vocalizations. Contrastingly, expressions of sadness, anger, and surprise were easily identified in inflected speech, whereas disgust was not.

Finally, in addition to the aforementioned stimulus considerations, the age of the individuals in the stimulus material may also influence recognition accuracy. In studies of adults, individuals appear to be more proficient at processing the faces of in-group compared to out-group members more generally (Elfenbein & Ambady, 2002), which may extend to recognition advantages to those similar to the perceiver (e.g., adults better at recognizing the emotions of other adults). Results of studies comparing age-congruency effects are, however, unclear. One study that compared children's recognition of other children's, young adults', and older adults' body expressions from point-light displays found that children who reported more exposure to older adults were better at recognizing older actors' happiness and anger body expressions (Pollux et al., 2016). However, the authors found no own-age bias in the categorization of body expressions overall. This result is supported by work by Griffiths et al. (2015), who similarly, in a sample of 5–12-year-old children, found that children were no better at recognizing the emotions of other children than they were of adult expressions.

Some other works have found age-congruency effects. For example, in a study using an adult sample that compared young, middle-aged, and older women, Malatesta et al. (1987) found an own-age recognition bias, with recognition accuracy varying depending on the age difference between the viewer and the stimulus. Young adults, for example, made more errors in identifying the facial expressions of older adults compared to those of younger adults. Moreover, when comparing typically developing adolescents to those with autism spectrum disorder, Hauschild et al. (2020) found evidence of an own-age emotion recognition bias regardless of autism status. The extent to which these results hold in the context of typically developing children's emotion recognition abilities remains an unanswered question.

In addition to an over-reliance on the use of static and adult faces to pry recognition competencies, the types of paradigms used across studies also appears to be quite homogenous. For example, the majority of tasks employ a forced-choice method, asking children to select a stimulus that matches an experimenter-defined label (e.g., "pick the face that looks happy"; Nelson & Russell, 2016), or select an experimenter-defined label from a list that best describes a stimulus (e.g., "does this person feel happy, sad, or angry?"; Garcia & Tully, 2020; Verpaalen et al., 2019; Woody et al., 2019). We term these "forced-choice" tasks. Another common task asks children to match emotion categories on the basis of perceptual features without a written or verbal label (e.g., "which two people feel the same?"; Gao & Maurer, 2009; LoBue et al., 2018). We term these "match-to-sample" tasks. More uncommon are studies that ask children to sort stimuli into (experimenter or self-defined) emotion categories (e.g., "put the faces that feel the same way in the same pile"; Matthews et al., 2022), or to provide their own emotion labels (e.g., "how does this person feel?"; Nelson & Russell, 2011a; Wang et al., 2014; Widen & Russell, 2015). We term these "pile-sorting" and "free-label" tasks, respectively.

These differences in task type may be highly relevant for the measurement of the trajectories in which emotion recognition abilities develop. For example, Matthews et al. (2022) argue that free-labeling tasks may underestimate children's competencies, as they rely on the size of children's own emotional vocabularies in order for them to be able to actually produce a label. Despite not knowing the specific verbal label for said emotion, children may still understand its features (e.g., its antecedents, appraisals, and experience) and be able to use this information to make accurate decisions (e.g., display caution in the face of an adult scowling and clenching his or her fists). On the other hand, work by Nelson and Russell (2016) found that children can use a process of elimination strategy to infer correct emotion categories in forced-choice tasks, which may actually result in an overestimation of children's competencies. Yet, as with properties of stimulus-type, it is currently unknown how consistent such effects are across paradigm types. Furthermore, it is known that task variation within individual studies can lead to different recognition accuracies for the same emotional expression. For example, Vicari et al. (2000) found that children's recognition accuracy scores for disgust facial expressions were 40% lower in a free-labeling compared to a perceptual matching task. Therefore, it is also important to investigate whether different task types influence emotion recognition in the same way overall, as well as for specific emotion categories.

The Present Study

Ultimately, despite the wealth of research on this topic, the only meta-analysis on typically developing children's emotion recognition competencies was performed by McClure (2000). Nevertheless, the focus of the McClure study was specifically on gender differences in recognition accuracy and therefore provides only limited insights into the developmental patterns of emotion recognition abilities across childhood, as well as the different experimental factors that may influence children's emotion recognition accuracy. In addition, as mentioned, other review articles have attempted to summarize this literature, but are limited by their qualitative nature (Gross & Ballif, 1991; Herba & Phillips, 2004) and exclusive focus on free-label tasks (Widen, 2013). Given the aforementioned between-study heterogeneity in stimuli and tasks, as well as the lack of understanding of the developmental patterns of emotion recognition throughout childhood, a logical step forward is to attempt to consolidate the literature, in an integrative way, on the basis of more modern works. To this end, we use a quantitative meta-analytic approach to chart the age-related changes in emotion recognition across childhood, and the factors that may influence children's emotion recognition accuracy across age. More specifically, this meta-analysis aims to track the trajectory of typically developing children's emotion recognition competencies, from toddlerhood to the beginning of adolescence (2-to-12-years-old).

As discussed above, to aid in the interpretation of the findings, we perform two types of analyses with two separate coding schemes. First, in what we term the *overall* analysis, we examine children's emotion recognition competencies independent of specific emotion categories. Second, in what we term the *emotion-specific* analysis, we examine children's recognition of individual emotion categories. As most of the previous studies investigated basic emotions, we made the decision to focus exclusively on these studies that include basic emotions (i.e., happiness, sadness, fear, surprise, disgust, anger). Basic emotions also constitute the majority of past works on emotion recognition in children.

In addition, we investigate whether patterns in both overall and emotion-specific analyses are moderated by participant-, stimulus-, and task-related considerations. We examine the effects of participant-related characteristics, including age, sex, and region of study. Moreover, we assess the impact of various stimulus-related features (e.g., stimulus dynamicity, age, color, intensity, etc.) and task-related features (e.g., free-label, forced-choice, perceptual matching, pile-sorting). All of these potential moderators may independently influence recognition competencies in general, and in specific ways for different emotion categories. Finally, all of these moderators may interact with age. Such an analysis is needed, considering that this field has grown dramatically over recent years, but the extent to which developmental trends, design considerations (or their combination) may moderate recognition abilities is largely unclear.

Method

Literature Search

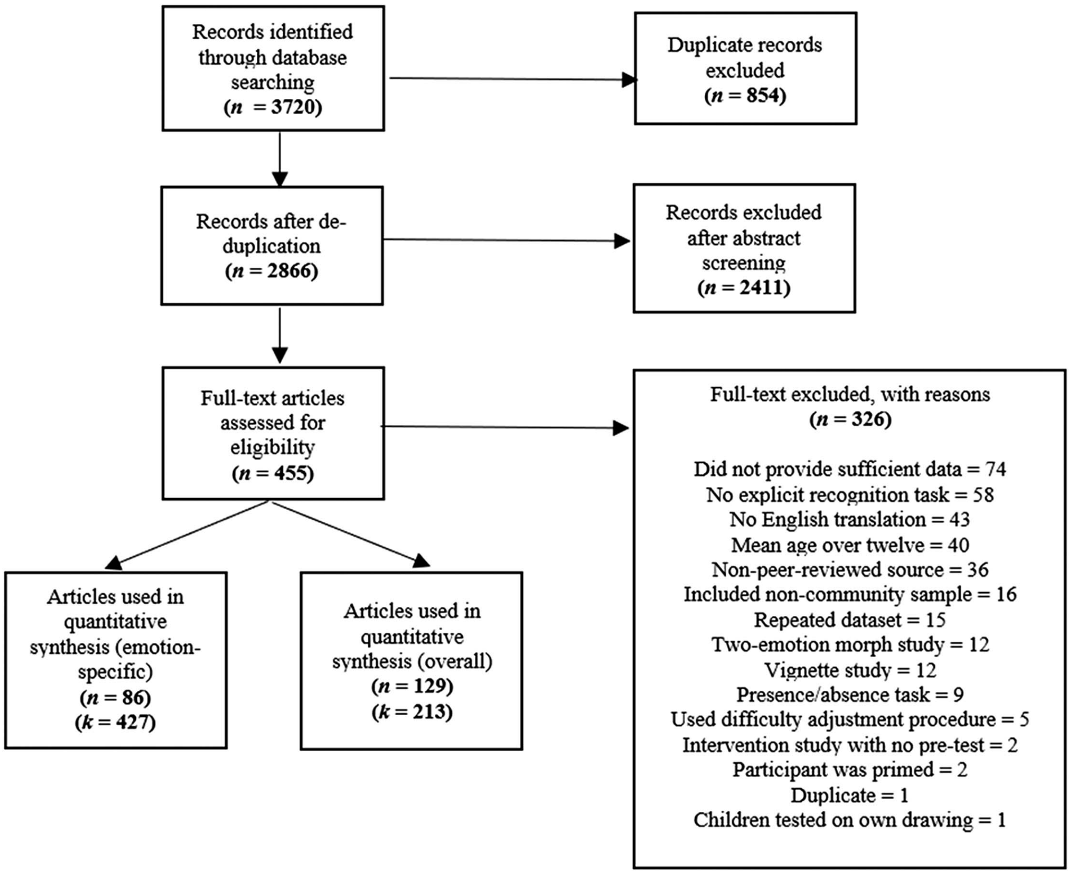
Studies that are potentially relevant for our research aims are many, as such we used APA PsycInfo, Web of Science Core Collection,[1] and Linguistics and Language Behavior Abstracts databases to search for works related to emotion recognition in typically developing childhood samples. For example, our APA PsycInfo search was constructed as follows:

> Emotion recognition/OR facial affect recognition/OR (((recogni* OR perception OR perceive* OR process* OR label* OR identif*) ADJ3 (emotion* OR expression*)) OR affect recognition). (preschool age 2 5 yrs OR school age 6 12 yrs).ag. OR (toddler* OR preschool* OR child* OR kid OR kids OR teen* OR girl* OR boy*). (((faces OR face OR facial) ADJ3 (emot* OR expressi*)) OR photograph OR photographs OR body posture OR body express* OR body movement* OR body language OR emot* stimul* OR prosod* OR vocali*tion OR sorting task OR match* task* OR match to sample OR choice from array OR free response).ti,ab,id.

The search was conducted in January 2021 and, after removing duplicates, resulted in a total of 2,866 abstracts. In line with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines for meta-analysis and systematic review (Page et al., 2021), the first author screened these abstracts to identify studies that were clearly ineligible (e.g., because they included participants of the wrong age, were animal studies, or included the incorrect emotion categories). The second author also screened 574 (20%) of these original 2,866 abstracts to ensure reliability at this stage. Reliability was excellent (Cohen's κ = 0.88). This initial abstract screening left a total of 455 articles to be assessed at the full-text stage. The first and second authors worked independently and screened the full text of the remaining list of studies for inclusion, achieving excellent reliability in the selection of articles (Cohen's κ = 0.86). Any potential disagreements were discussed until both researchers reached consensus. This full-text screening led to a total of 252 studies that could potentially be included in the analyses. These works were then coded for all demographic, task-related, and stimulus-related information (i.e., the moderators), as well as the relevant effect size(s) by the first author. The first and second authors double-coded 15% of the 252 included articles to ensure reliability of the coded moderators and effect sizes, which was good (Cohen's κ > 0.70 for all moderators). Any coding disagreements were discussed until consensus was reached between both authors. In addition to this, the first author also contacted the authors of any works in which relevant information was missing or incomplete. Authors were contacted and sent a reminder if the first author received no response after 2 weeks. Studies in which the corresponding author did not reply after a second reminder email were removed from the inclusion list. In total, there were $N = 129$ studies included in the overall analysis, and a total of $N = 83$ studies included in the emotion-specific analysis (more studies reported overall scores without scores for specific emotions, discussed in detail below). The flow of reports into the meta-analysis, which highlights the phases of our literature search, is displayed in Figure 1.

[1] Web of Science Core Collection (Web of Science Core Collection Editions: Science Citation Index Expanded (SCI-EXPANDED), 1975—present, Social Sciences Citation Index (SSCI), 1975—present, Arts & Humanities Citation Index (A&HCI), 1975—present, Emerging Sources Citation Index (ESCI), 2015—present).

Figure 1
Flow of Reports Into the Meta-Analysis



Eligibility Criteria

In order to be included in our analyses, an individual study needed to (a) examine emotion recognition competencies in 2–12-year-old children; (b) test the recognition of specific emotion categories (i.e., not only valence, arousal, or intensity judgments); (c) use external cues (visual or auditory) to test competency (this criterion excluded those studies that used vignettes exclusively, as children may use other external cues embedded in such stories to make recognition judgments); (d) include basic emotions (e.g., happiness, sadness, anger, fear, disgust, surprise)—studies were also still considered eligible if they included other emotion categories, but we did not include these in the calculations of effect sizes in both overall and emotion-specific analyses. Studies that included stimuli involving morphs of two different emotion categories (e.g., happy and sad morphed together) were excluded, as there is no “correct” alternative in this case. Studies that morphed neutral faces with another emotion category (creating a subtler version of said face) were still included. Additionally, studies were eligible only if they (e) used a sample with only typically developing children (as indicated above, we used this criterion to constrain our search); (f) examined at least two different emotion categories, e.g., happiness and sadness, but not happiness and neutral, as one could argue that these tasks represent an emotion detection (presence/absence) task, rather than one that explicitly taps into recognition competencies; (g) did not use any per-participant stair-case or difficulty adjustment procedures (as accuracy scores are therefore not comparable across participants); (h) were written in English, or available with an English translation; and (i) had the relevant statistics required for the calculation of our effect sizes (see calculation of effect sizes) available in-text or provided to us upon our request.

Coding of Studies

As previously mentioned, we coded two separate data sets for the overall and emotion-specific analyses, respectively. We chose to create two data sets, rather than simply average the effect sizes in the emotion-specific analyses to derive an overall score, as authors often reported accuracies collapsed across emotion categories and not for the individual categories themselves. Despite contacting all authors who did this to ask for emotion category accuracies, a number of authors did not reply to our requests (see above). Therefore, rather than simply removing these studies from our analyses, we included them in the overall analysis data set (but not in the emotion-specific data set). As such, our analysis of children’s emotion recognition skills overall included more studies and had more power. In addition, this means that the overall analysis is not simply an averaged version of the emotion-specific analysis. Rather, it is a separate data set derived to answer a different research question.

As a general rule, individual studies could only contribute one observation (effect size) with one stimulus-type to each data set (the two exceptions of this rule concerned age-group and intensity, explained below). Our effect size of interest was participants' proportion of correctly identified emotions across stimuli (i.e., proportion correct), which we later corrected to account for chance guessing (see below). To constrain the large number of paradigm types, in the case that an individual article contained multiple studies, multiple experimental conditions (e.g., a free-label and a forced-choice task; Vicari et al., 2000), or multiple stimulus types (e.g., static and dynamic; Richoz et al., 2018), we included and always coded the first result reported in-text in the results section of the article. In the case of intervention (e.g., Riquelme & Montero, 2013) or longitudinal (e.g., Cecilione et al., 2017) studies, we coded the pretest (or time one) scores for the calculation of the effect size. As previous works found clear age-related effects on emotion recognition (Widen, 2013), for studies that included multiple age groups (e.g., Richoz et al., 2018), multiple effect sizes were calculated, but only if that individual study reported recognition accuracy means for each age group separately. If this was not the case, then we coded the mean age of the sample for said study. Similarly, for studies that included stimuli manipulated by software to achieve multiple levels of emotional intensity (e.g., through the use of some sort of mask/degradation technique; Chronaki et al., 2015), multiple effect sizes were calculated as long as the study reported accuracies at each intensity.

In addition to the proportion correct score itself, for each study, we collected the relevant demographic data, including mean age, age range, region of study, percentage of girls in the sample, and total number of participants. In addition, we included multiple stimulus- and task-related sources of information, including: stimulus format (photo, video, vocalization), stimulus-type (posed or naturalistic; human or nonhuman [cartoon, robot, emoticon, point-light display]; adult or child; monochrome or color; face, body, both face and body), intensity (full intensity, not full intensity), method of intensity degradation (posed as low-intense by actor, morphed with neutral, masked), number of trials, stimulus presentation time, image set (derived from a database or author-own), stimulus validation status (validated, unvalidated, unknown), number of response alternatives (e.g., number of choice options a participant was given), number of basic emotions included in the study overall, and task type (free-label, match-to-sample, pile-sorting [with label], pile-sorting [no label], forced-choice).

Calculation of Effect Sizes

The majority of included studies used multiple-choice type questioning, and the number of response alternatives offered to children differed between studies. As such, it was necessary to first correct all raw proportion correct scores (P) for chance performance due to guessing alone. For example, a child's proportion correct in a task that includes six equally likely response alternatives per trial (without correction) is not directly comparable to a task that includes two equally likely response alternatives per trial. To account for this, we used Rosenthal and Rubin's (1989, p.337) corrected proportion index to adjust all raw proportion correct scores P. The corrected proportion index is a modification of the standard proportion index (proposed by the same authors), as this standard index is less suitable for small sample sizes (i.e., number of trials in a task) and may lead to problems of nonnormality (in the preliminary analyses of our data using the standard proportion index we encountered strong ceiling effects). We therefore adopted the authors' adjusted proportion index by first correcting the observed proportion correct scores for small sample size (trials) and adding half a failure and half a success to scores, as follows:

$$\tilde{\pi} = \frac{P + \left(\frac{1}{2n}\right)}{1 + \frac{1}{n}}, \tag{1}$$

where P equals the proportion correct, and n refers to the number of trials. Then, in a second step, this adjusted $\tilde{\pi}$ score is converted to a logit scale to address issues of nonnormality:

$$\text{logit}(\pi) = \text{logit}(\tilde{\pi}) - \text{logit}\left(\frac{1}{c}\right), \tag{2}$$

where c refers to the number of response alternatives in each trial. This effect size measure, logit(π), was calculated for each study and used as an outcome measure in our meta-analysis. Accordingly, the standard error of logit(π) was calculate following formula (3) in Rosenthal and Rubin (1989, pp. 334). Note that logit(π) is equal to the log of the odds and varies theoretically between $-\infty$ (when $\tilde{\pi} = 0$) and $+\infty$ (when $\tilde{\pi} = 1$). The effect size logit(π) provides a readily interpretable effect size estimate in which the proportion correct is scaled to account for differences in response alternatives between studies. Thus, in the previous example, the (adjusted) proportion index scales the accuracy in the task with six response alternatives such that it is equivalent to a task in which there had been only two.

Importantly, "free-label" and "pile-sorting" tasks present a problem for the calculation of this effect size. In these tasks, a child is not constrained by a set of labels and can therefore give a potentially infinite number of responses. One solution to this problem is to set the number of response alternatives in such tasks to an (arbitrarily) higher number. Doing so accounts for the marked difficulty in responding correctly in such tasks as compared to multiple-choice tasks and, additionally, would allow their direct comparison. However, in preliminary analyses, this led to a systematic overestimation of free-label and pile-sorting task accuracy across studies because of the exponential rate in which π increases with single unit increases in c (see Table 1 Rosenthal & Rubin, pp. 333). Because free-label paradigms represent a significant portion of the literature, we thought it nonetheless important to include them in our analyses. As such, we performed all analyses on these tasks separately. Importantly, we therefore did not correct for guessing in such tasks, and simply use the logit of the proportion correct scores corrected for small sample size (given in Equation 1 above) as our effect size in these analyses.

Finally, we note that in some studies authors reported some already corrected form of accuracy (e.g., Wagner's unbiased hit rate; d' indices from signal detection theory; or *Pr* from two-high threshold theory). Such indices are preferable, as variance in recognition accuracy can be explained by both chance guessing as well as response bias (Wagner, 1993). That said, the calculation of such effect sizes requires a complete confusion matrix (indexing children's correct *and* error judgments for each emotion category), which only few studies reported in full. It was therefore impractical

Table 1
Descriptive Statistics for Variables in the Overall Analysis

Variable	*k*	*M* (*SD*)	*Mdn*	Mode	Range
Sample Size	210	148.10 (567.21)	48.00	24 (*k* = 14)	8–6,506
Age (years)	213	7.11 (2.47)	7.08	9.50 (*k* = 8)	2.02–12.00
Percent female	190	51.78 (11.12)	50.00	50 (*k* = 46)	0–100
Trials	213	28.92 (34.66)	16.00	12 (*k* = 32)	3–264
Presentation time (ms)	33	4,803 (4769.60)	3,000	2000 (*k* = 8)	1,000–15,000
Intensity (reduced vs. full)	213	78% full		Full (*k* = 168)	
Color (monochrome vs. color)	147	52% monochrome		Monochrome (*k* = 76)	
Humanness (human vs. nonhuman)	213	92% human		Human (*k* = 196)	
Posed (posed vs. naturalistic)	211	99% posed		Posed (*k* = 209)	
Face versus body versus both	174	93% face		Face (*k* = 162)	
Child versus adult versus both	196	69% adult		Adult (*k* = 136)	
Stimulus format (photo vs. video vs. vocalization)	211	76% photo		Photo (*k* = 160)	
Country	213	30% U.S.		U.S. (*k* = 63)	
Stimulus set	212	21% author-own		Author-own (*k* = 45)	
Validated (yes vs. no)	178	85% validated		Validated (*k* = 151)	

to use these indices in our analyses, and we opted for the alternative logit(π) effect size highlighted above.

Statistical Analysis

Overall Analysis

In the overall analysis, we aimed to track children's emotion recognition generally, independent of specific emotion categories. Because some studies could contribute multiple effect sizes logit(π) for separate age and/or intensity categories to our analyses, the assumption of independence between effect sizes necessary for traditional meta-analysis was not met. As such, we used a multilevel approach, which accounts for the dependency in effect sizes within and between individual studies using *rma.mv* function in the metafor package (Viechtbauer, 2010) in R Studio, Version 1.2.5 (R Core Team, 2020). We fitted a three-level random effects model, using logit(π) as an outcome measure, and considering the following levels of variance: sampling variance, given by the standard error of logit(π) per study, effect size variance (referring to the variance in effect sizes within individual studies), and study variance (referring to the variance between studies). We assumed a t-distribution, and restricted maximum-likelihood estimation was used to derive the standard deviation of the true effect sizes (as recommended by Assink & Wibbelink, 2016). The overall analysis resulted in a pooled estimate of logit(π) across studies. In the results, we computed the inverse of the logit (with the use of the *invlogit* function in the LaplacesDemon package in R; Hall et al., 2021) to convert estimates on the logit scale back to proportions (*P*) for ease of interpretation and comparison to previous works.

Emotion-Specific Analysis

In the emotion-specific analysis, we aimed to track children's recognition of specific emotion categories. As an individual study could contribute an effect size for each (basic) emotion included in the study, the assumption of independence between effect sizes needed for traditional meta-analysis was similarly not met. As such, we also used a multilevel approach for our emotion-specific analyses in the same way as described above. One caveat here is that we could not include effect sizes for separate age categories, intensities, *and* emotions, because proportion correct scores (*P*) were rarely given per emotion, per age group, and per intensity in included studies. As such, in the emotion-specific analyses, if a study included multiple age categories (or intensities), we used the average age of the participants within our accepted age range (or intensity) in our calculations of individual emotion category scores, and hence calculated our effect sizes logit(π) on the basis of the proportion scores *P* averaged across all of these.

Moderator Analysis

We used a hypothesis-driven strategy to examine the effects of our moderator variables on children's recognition abilities. We performed several metaregression analyses using the calculated effect sizes, logit(π) as an outcome. Given the potential for model overspecification in the emotion-specific analysis (e.g., a model including the interaction between six emotion categories and six countries requires the estimation of 32 coefficients), we adopted different approaches to moderation analysis for the overall and emotion-specific analyses. Following the approach of Hox (2010), in the overall analysis, we tested the effect of each moderating variable in univariate models (inspecting the omnibus *F*-test for a significant moderator effect). After this, we pooled significant moderator variables in one multiple metaregression to inspect their unique contribution. This way, we could ensure that the significant effect of a given moderator was robust and did not disappear when considering the effects of other significant univariate moderators.

For the emotion-specific analysis we took a two-step approach. First, we aimed to ascertain the relative "importance" of a moderator. To do this, we examined the differences in fit between main effects models (including emotion category and the addition of a given moderator) compared to interaction models (including emotion category and its interaction with the same moderator), using likelihood ratio tests. Importantly, as models using restricted maximum-likelihood estimation cannot be compared using likelihood ratio tests (Hox, 2010, pp.15), we used maximum-likelihood estimation for these models. A significant difference between the full (interaction) and reduced (main effects) model on a given moderator indicates the difference in accuracy between emotions is influenced by said moderator, and therefore may be important for moderation analysis for an individual emotion category. In a second step, we tested the effects of only those moderators that produced a significant likelihood ratio test in a series of models for each emotion category separately. Because each study could only contribute one emotion type, we used simpler two-level meta-analytic models using the *rma* function in the metafor package (Viechtbauer, 2010). In cases of model nonconvergence, we adjusted the Fisher scoring algorithm step length to 0.5 and increased the max number of iterations to 10,000.

Because the outcome measure for all our moderation analyses was logit(π), we took the exponent of the model coefficients in order to interpret significant effects of the moderators at the odds ratio (*OR*) level. In this way, we follow the standard interpretation of the effects of predictor variables in a logistic regression framework. Odds ratios that are greater than 1 indicate that the chance of a correct response increases as the predictor increases. Odds ratios that are less than 1 indicate that the chance of a correct response decreases as the predictor increases.

Another caveat is that for clarity and ease of communication, we refer to these predictor variables as "moderators," although wish to stress that these are not moderators in the traditional meta-analytic sense (as our effect sizes are calculated on the basis of a single group, rather than standardized mean difference or correlation scores between groups).

Publication Bias

The results of meta-analyses may be compromised by publication biases of various forms (Nakagawa et al., 2022). In particular, they may be biased because studies that report small(er) effect sizes and nonsignificant results have a lower chance of publication (Thornton & Lee, 2000). To assess this "file-drawer" effect, we used a two-method approach. First, we inspected the funnel plot of both overall and emotion-specific data sets, for both multiple-choice and free-label tasks separately. In addition, we tested for the effect of small sample size (i.e., trials) using a method suitable for multilevel analyses (Nakagawa et al., 2022). The authors propose to use the square root of the effective sample size (i.e., number of trials) in each study and include this term as a moderator in a multilevel meta-analysis. A significant negative effect of this moderator indicates, for our meta-analysis, that studies with a small number of trials bear undue influence on the recognition accuracy estimate.

Transparency and Openness

We adhered to the Meta-Analysis Reporting Standards guidelines for meta-analytic reporting (Appelbaum et al., 2018). All analytic data, analysis code, and research materials are available at https://osf.io/68g7z/?view_only=4b5d5b8dec65463fbbe099ee89e58b37 (Riddell et al., 2024). This review project was not preregistered.

Results

We begin our results by reporting the study characteristics of the overall emotion recognition data set both in-text and in Table 1. Importantly, we do not describe study characteristics for the emotion-specific data set, as all studies in the emotion-specific data set are also coded in the overall data set. Then, we proceed with moderation analysis by examining how age, as well as task-, participant-, and stimulus-related features impact emotion recognition overall, as well as for specific emotion categories. We conduct separate analyses for free-label tasks, given the aforementioned issues in adjusting for response options in these paradigms. Finally, we discuss publication bias.

Overall Emotion Recognition Accuracy

In the overall analysis, we examine the developmental trajectory of children's emotion recognition competencies independent of specific emotion categories. We identified a total of 129 unique studies and 213 effect sizes (k) from a total of 31,101 participants. The 213 calculated effect sizes, logit(π), ranged from −1.31 to 4.02, equivalent to accuracy scores between 21.21 and 98.23%.

Study Characteristics and Pooled Model

From the total 213 effect sizes derived for our overall analyses the majority (k = 160) included static photographs. Videos (k = 29) were the next most common in our sample, followed by vocalizations (k = 22). Stimuli of adults (k = 136) were much more common than children (k = 51), and full intensity stimuli (k = 161) were more popular than reduced intensity stimuli (k = 52). The mean age of participants was 85.19 months (7.11 years) (SD_{age} = 29.78, range = 31–144 months), and studies included an average of 51.78% of female participants (SD_{female} = 11.12, range = 0–100). Children from the United States (k = 63) and Canada (k = 28) were the most represented, followed by the United Kingdom (k = 23), Japan (k = 10), and Australia (k = 8). To allow for moderator analysis, we recoded individual countries to represent six broader geographic regions: Australia/New Zealand (k = 10); Europe (k = 81); South America (k = 3); North America (k = 91); Asia (k = 19); and other (k = 9). A wide variety of stimulus sets were included, but author-own stimulus sets were the most popular (k = 45). In terms of established and validated stimulus databases, the majority of effect sizes (k = 30) came from studies using some iteration of the Ekman and Friesen (1976) *Pictures of Facial Affect*, followed by some iteration of the *Diagnostic Analysis of Nonverbal Accuracy* (k = 17), NimStim (k = 14), and *Radboud Faces Database* (RaFD; k = 8). Image sets with k = 3 or more observations were included in moderator analyses, and image sets with less than this amount were designated as "other." Additionally, authors who used their own stimulus sets were designated as "author-own." Unfortunately, this resulted in a large number of observations in the "other" and "author-own" categories, and we therefore excluded these categories from the stimulus set moderation analyses. A copy of the summary table of studies for which effect sizes could be calculated is available in the Supplemental Materials. In addition, estimated summary statistics for each level of each moderator variable included in our analyses are available in Supplemental Table S1.

The estimate of the pooled logit(π) across studies (including all stimulus, task, and participant characteristics, but not including free-label tasks) indicated that accuracy was significantly different from chance, logit(π) = 2.01, $t(172) = 28.83$, $p < .0001$, 95% CI [1.88, 2.15]. This is equal to a pooled estimated accuracy (P) of 88%. Pooled heterogeneity (across all three levels of the model) was also significant, $Q(172) = 437.61$, $p < .0001$. Approximately 33% of the variance could be attributed to sampling variance (Level 1), less than 1% of the variance could be attributed to within-study variance (Level 2), and approximately 66% of the variance could be attributed to between-study variance (Level 3).

Moderator Analysis: Participant-, Task-, and Stimulus-Related Effects

In our first moderator analysis, we aimed to test participant-related effects on overall emotion recognition. Four potential moderators were tested in separate models: region of study, mean-centered age (in years), number of participants, and the percentage of female participants in the sample. Omnibus tests revealed significant effects of both region of study and age on emotion recognition accuracy; $F(5, 167) = 2.38$, $p = .041$, $F(1, 171) = 38.76$, $p < .0001$, respectively. Yet, we found no evidence of a female advantage, nor an effect of number of participants in the sample; $F(1, 150) = 1.36$, $p = .246$, $F(1, 171) = 1.04$, $p = .310$, respectively.

In addition to participant-related features, we were also interested in examining how stimulus and task-related features influence emotion recognition competencies. To this end, we derived individual models for: image set, image intensity (no intensity adjustment vs. intensity adjustment), naturalness (posed vs. naturalistic), humanness (human vs. nonhuman), physical modality (body vs. face vs. both), age (child vs. adult vs. both), stimulus format (video vs. photo vs. vocalization), task type (match-to-sample, forced-choice, other), intensity reduction method (mask/blur, morph, posed), color (monochrome vs. color), validation (validated vs. not validated), presentation time, and task type (forced-choice, match-to-sample). Omnibus tests on individual models revealed no effect of image set, stimulus naturalness, intensity, intensity reduction method, humanness, physical modality, stimulus format, age of actor, presentation time, and color on children's recognition rates (all *p*s $> .05$). However, results found a significant moderating effect of task type, $F(1, 169) = 6.82$, $p = .010$.

Finally, given the theoretical importance of examining age-related effects on specific moderators, we derived a series of models to test the interaction between age and all other moderators. There was a significant interaction effect between age and stimulus validation, demonstrating that children's accuracy on validated stimulus sets increased more with increased age compared to unvalidated stimulus sets, $OR = 1.11$, 95% CI [1.02, 1.21], $\text{logit}(\pi) = 0.11$, $t(143) = 2.38$, $p = .019$. This indicates that, with a 1-year increase in age, the odds of a correct response in validated sets are 1.11 times the odds of a correct response in unvalidated stimulus sets. In addition, there was a significant positive interaction effect between age and presentation time, although this effect was so small it prevented its proper interpretation, $OR = 1.00$, 95% CI [0.00, 0.00], $\text{logit}(\pi) = 0.00$, $t(25) = 2.13$, $p = .043$. Among the remaining moderators, there were no significant interactions with age (all *p*s $>$ .05), indicating that the influence of these remaining moderators on emotion recognition accuracy was similar across age.

Multiple Metaregression Model

Finally, a multiple metaregression analysis was then conducted in order to examine whether the results of individual models held when controlling for the effect of other significant univariate moderators. To this end, we derived a model with effect sizes as the dependent variable and included all significant moderators in the above univariate models (mean participant age, region of study, and task type) as predictors. The results of the multiple regression model are reported in Table 2. The omnibus test of moderators was significant, $F(7, 163) = 7.48$, $p < .0001$, indicating that the model (with all moderators pooled together) explained a significant amount of variance in the effect sizes. In addition, in line with the individual models discussed above, and after controlling for potentially confounding moderators, mean age (in years) was a significant predictor of recognition accuracy, $OR = 1.11$, 95% CI [1.07, 1.15], $\text{logit}(\pi) = 0.10$, $t(163) = 5.53$, $p < .001$. This indicates that with a 1-year increase in age, the odds of a correct response increase with a factor of 1.11. In addition, the effect of task type also remained significant, as the odds of a correct response in match-to-sample tasks decreased with a factor of 0.74 compared to forced-choice tasks, $OR = 0.74$, 95% CI [0.56, 0.99], $\text{logit}(\pi) = -0.30$, $t(163) = -2.04$, $p = .044$. Finally, the model revealed that the odds of a correct response decreased by a factor of 0.69 for North American samples compared to European samples, $OR = 0.69$, 95% CI [0.52, 0.91], $\text{logit}(\pi) = -0.37$, $t(163) = -2.60$, $p = .010$. After changing the reference category in the model to account for all possible contrasts between other countries, no differences between other regions were established. Ultimately, the results of this analysis indicated that the effects of participant age, region of study, and task type remained significant when controlling for the effects of other significant univariate moderators.

Overall Effects in Free-Label Tasks

As discussed above, free-label and pile-sorting tasks presented a problem for the calculation of our effect sizes and needed to be analyzed separately to multiple-choice tasks. No pile-sorting tasks included the appropriate information to satisfy our inclusion criteria and were therefore not included in these analyses. From the 129 total studies in the overall data set, we derived a total of 21 unique free-label studies and $k = 40$ effect sizes. The estimate of the pooled

Table 2
Multiple Metaregression Model for Multiple-Choice Tasks

Variable	*OR*	95% CI	logit(π)
Intercept	9.66***	[7.78, 12.00]	2.27[a]
Age (years)	1.11***	[1.07, 1.15]	0.10
Region			
Europe	1.00[b]		
Australia/New Zealand	0.99	[0.55, 1.77]	−0.01
South America	0.64	[0.31, 1.31]	−0.45
North America	0.69*	[0.52, 0.91]	−0.37
Asia	0.86	[0.51, 1.42]	−0.16
Other	0.89	[0.55, 1.42]	−0.12
Task type			
Forced-choice	1.00[b]		
Match-to-sample	0.74*	[0.56, 0.99]	−0.30
Other	1.66	[0.52, 5.28]	0.50

Note. An *OR* greater than 1 (smaller than 1, respectively) indicates that as the predictor increases the odds of a correct response increase (decrease, respectively). *OR* = odds ratio; CI = confidence interval.
[a] Estimated accuracy = 90.6%. [b] Intercept represents estimate for reference categories Europe and forced-choice at mean age.
* $p < .05$. *** $p < .001$.

logit(π) indicated that emotion recognition was better than chance, logit(π) = 0.41, $t(39) = 2.95$, $p = .005$, 95% CI [0.13, 0.70]. This is equivalent to a recognition accuracy (*P*) of approximately 60%, lower than that of multiple-choice tasks indicated in the previous section. Examination of the variance components also revealed lower levels of heterogeneity. Approximately 86% of the total variance was attributable to sampling variance (Level 1), less than 1% at Level 2 (within studies), and 13% at Level 3 (between studies).

Examination of omnibus tests in univariate models revealed significant effects of age and stimulus humanness; $F(1, 38) = 12.49$, $p = .001$, $F(1, 38) = 4.77$, $p = .035$, respectively. No other moderators influenced recognition accuracy (all *p*s > .05). After pooling mean age and stimulus humanness in one multiple metaregression model, the effect of stimulus humanness became nonsignificant, *OR* = 0.49, 95% CI [0.18, 1.34], logit(π) = −0.71, $t(37) = -1.44$, $p = .159$. Age remained a significant moderator of recognition accuracy in free-label tasks, *OR* = 1.17, 95% CI [1.06, 1.30], logit(π) = 0.16, $t(37) = 3.06$, $p = .004$. This indicates that a 1-year increase in age predicts a 1.17 factor increase in the odds of a correct response in free-label paradigms. The results of the metaregression are presented in Table 3. Ultimately, the results of this analysis suggest that age is significantly related to children's recognition accuracy on free-label tasks. When controlling for the effects of age, however, stimulus humanness was not found to be related to children's recognition accuracy on free-label tasks.

Emotion-Specific Recognition Accuracy

To recap, in the emotion-specific analysis, we attempted to examine children's emotion recognition competencies for specific emotion categories separately. In addition, we aimed to test how the recognition of these specific emotions may be influenced by our list of moderating variables. As many studies reported overall scores (collapsed across all emotion categories), but not scores for individual emotions, this analysis contained fewer unique studies, but more effect sizes than the overall analysis. We identified a total of 102 unique studies and 427 effect sizes (*k*).

First, we modeled recognition accuracy for specific emotion categories by deriving a model which included emotion category as a moderator (six levels: happiness, sadness, fear, surprise, disgust, and anger). The omnibus test revealed that emotion category was a highly significant moderator of emotion recognition accuracy, $F(5, 327) = 8.42$, $p < .0001$. Importantly, individual emotion categories explained a significant amount of variance, and there was no residual heterogeneity after inclusion of emotion as a moderator, $Q(327) = 326.05$, $p = .504$. The model estimates are presented in Table 4. For interpretability, summary statistics, including proportion correct scores (with sample size adjustment, but no adjustment for the number of response alternatives) for each emotion category, are reported in Table 5.

Table 3
Multiple Metaregression for Free-Label Tasks

Variable	*OR*	95% CI	logit(π)
Intercept	1.60***	[1.25, 2.04]	0.47[a]
Age (years)	1.17**	[1.06, 1.30]	0.16
Stimulus humanness			
Human	1.00[b]		
Nonhuman	0.49	[0.18, 1.34]	−0.71

Note. An *OR* greater than 1 (smaller than 1, respectively) indicates that as the predictor increases the odds of a correct response increase (decrease, respectively). *OR* = odds ratio; CI = confidence interval.
[a] Estimated accuracy = 61.5%. [b] Intercept represents estimate for reference category human at mean age.
** $p < .01$. *** $p < .001$.

Happiness was the best recognized emotion across all studies, $M_{\text{logit}(\pi)} = 2.43$; $M_{\text{accuracy}} = 91.90\%$, followed by anger, $M_{\text{logit}(\pi)} = 2.04$; $M_{\text{accuracy}} = 88.49\%$, surprise, $M_{\text{logit}(\pi)} = 1.83$; $M_{\text{accuracy}} = 86.18\%$, and sadness, $M_{\text{logit}(\pi)} = 1.82$; $M_{\text{accuracy}} = 86.06\%$. Children were the least accurate at identifying disgust, $M_{\text{logit}(\pi)} = 1.79$; $M_{\text{accuracy}} = 85.69\%$, and fear, $M_{\text{logit}(\pi)} = 1.53$; 82.20% stimuli. To identify whether these differences between emotion categories were significant, we ran models changing the reference category with each iteration, accounting for all possible pairwise comparisons. We found that recognition for anger ($p = .006$), disgust ($p = .002$), fear ($p < .0001$), sadness ($p < .0001$), and surprise ($p = .001$) stimuli was significantly lower than for happiness. Accuracy for fearful stimuli was significantly worse than both anger ($p < .0001$) and sadness ($p = .034$). No significant accuracy differences were found for all other emotion category combinations.

Moderator Analysis: Participant-, Task-, and Stimulus-Related Effects

To examine any potential interaction effects between specific emotion categories and our list of moderator variables, a series of moderator analyses were performed. Although the inclusion of specific emotion categories as a moderator left little heterogeneity in our model (see above), we decided to nonetheless proceed with our moderation analyses given the theoretical importance of examining how various methodological choices impact children's emotion recognition accuracy.

Moderator Selection. We began by running likelihood ratio tests between main effects (emotion category + moderator) and interaction (Emotion Category × Moderator) models to determine the relative importance of each moderator. The results of the log-likelihood test revealed that full models (including interactions with emotion category) were significantly different from main effects models for the moderators: stimulus format, $\chi^2(10) = 26.09$, $p = .004$; stimulus color, $\chi^2(9) = 18.03$, $p = .003$; and stimulus validation, $\chi^2(9) = 19.34$, $p = .002$. The likelihood ratio tests with all other potential moderating variables (most notably, including age) were nonsignificant (all *p*s > .05). Figure 2 displays the recognition of each emotion category as a trajectory across age, and visual inspection of the slopes appears to support a lack of interaction between emotion category and age. Based on these tests, we selected stimulus format (photo vs. video vs. vocalization), color (monochrome vs. color), and stimulus validation (validated vs. not validated) as potential moderators for individual emotion category analyses.

We found no significant interaction between age and emotion category. Yet, because tracking the age-related trajectories of emotion recognition was a specific aim of this meta-analysis, we estimated the recognition accuracy of each emotion category in multiple-choice tasks at two age points: 4-years-old (representing early childhood) and 12-years-old (representing late childhood). These estimates are presented in Table 6. For example, at 4-years-

Table 4
Emotion Recognition Accuracy Model Estimates for Specific Emotion Categories in Multiple-Choice Tasks

Variable	k	OR	95% CI	logit(π)
Intercept		11.40	[9.14, 14.21]***	2.43
Emotion				
Happiness	80	1.00[a]		
Anger	68	0.67	[0.51, 0.89]**	−0.39
Surprise	31	0.55	[0.38, 0.79]**	−0.61
Sadness	70	0.54	[0.42, 0.71]***	−0.61
Disgust	23	0.53	[0.35, 0.79]**	−0.64
Fear	61	0.41	[0.31, 0.54]***	−0.90

Note. An OR greater than 1 (smaller than 1, respectively) indicates that as the predictor increases the odds of a correct response increase (decrease, respectively). OR = odds ratio; CI = confidence interval.
[a] Intercept represents estimate for reference category happiness.
** $p < .01$. *** $p < .001$.

old, happiness recognition accuracy was estimated to be 88.20%, while at 12-years-old, this accuracy estimate increased to 94.97%.

Moderation Analysis With Individual Emotion Categories. An overview of the results of the moderation analysis for each emotion category is displayed in Tables 7 and 8. Note that in the cases where the level of a moderator had less than three observations, this level was excluded from the analysis. Despite demonstrating an increase in model fit in the selection stage, stimulus format had no significant influence on any of the individual emotion categories (all ps > .05). Similarly, stimulus color had no effect on recognition accuracy for: anger, $Q(1) = 0.91$, $p = .400$; disgust, $Q(1) = 0.06$, $p = .810$; happiness, $Q(1) = 0.00$, $p = .947$; sadness, $Q(1) = 0.46$, $p = .497$; and surprise, $Q(1) = 0.26$, $p = .612$. Conversely, the effect of stimulus color on the recognition of fearful stimuli was significant in the omnibus test, $Q(1) = 22.14$, $p < .0001$, with the metaregression revealing that odds of a correct response for fearful stimuli presented in black and white decreased by a factor of 0.38, $OR = 0.38$, 95% CI [0.26, 0.57], logit(π) = −0.95, $p < .0001$. Finally, stimulus validation had no effect on recognition accuracy for: anger, $Q(1) = 0.36$, $p = .546$; happiness, $Q(1) = 0.88$, $p = .350$; sadness, $Q(1) = 0.72$, $p = .395$; and surprise stimuli, $Q(1) = 0.15$, $p = .701$. However, the omnibus test revealed a significant effect of stimulus validation on the recognition of fearful stimuli, $Q(1) = 11.73$, $p = .001$, and the metaregression indicated that the odds of a correct response for unvalidated fearful stimuli decreasing by a factor of 0.44 compared to validated stimuli, $OR = 0.44$, 95% CI [0.28, 0.71], logit(π) = −0.82, $p = .001$. Given that there were only two studies which used unvalidated disgust stimuli, we were unable to test for the effect of stimulus validation for the emotion disgust.

Table 5
Summary Statistics for Each Emotion Category in Multiple-Choice Tasks

Emotion	k	$M_{proportion}$	$SD_{proportion}$
Happiness	80	0.78	0.13
Anger	68	0.67	0.15
Sadness	70	0.66	0.14
Surprise	31	0.62	0.13
Disgust	23	0.59	0.17
Fear	61	0.60	0.15

Note. Mean proportions are adjusted for small sample size (trials) according to Equation 1 of the article. Scores not adjusted for number of response alternatives.

Overall, the results of the moderation analysis suggest that stimulus format did not influence the recognition rates of any specific emotion categories. Moreover, stimulus color and validation influenced the recognition of fearful stimuli significantly, an effect that was not seen in other emotion categories.

Emotion-Specific Effects in Free-Label Tasks

Like in the overall analysis, we analyzed free-label tasks separately to multiple-choice tasks in the emotion-specific analysis. We identified a total of 19 unique free-label studies and $k = 94$ effect sizes. The estimate of the pooled logit(π) indicated that emotion recognition accuracy was no better than chance, logit(π) = 0.20, $t(93) = 1.60$, $p = .112$, 95% CI [−0.05, 0.45]. This is equivalent to a recognition accuracy of approximately 55%, lower than that of the task types tested in the previous section. Yet, examination of the variance components revealed that, unlike in the overall analysis, there was zero between- *and* within-study heterogeneity for free-label tasks in the emotion-specific data set. This made analysis of any (potential) moderating effects on emotion recognition in free-label studies unsuitable and we, therefore, forewent these analyses. Nonetheless, given the theoretical importance of examining the developmental trajectory of children's emotion recognition abilities, we plot the recognition accuracy for each emotion category in free-label tasks across age in Figure 3.

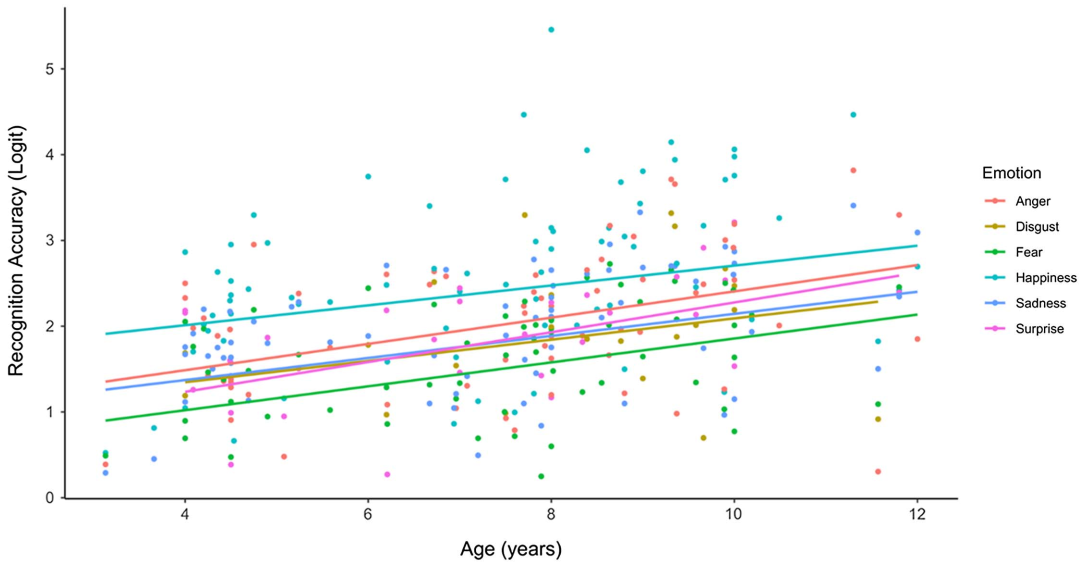
Publication Bias

To examine any potential "file-drawer" effects, we used a two-step method: first by inspecting the funnel plots for each data set, and then by quantifying publication bias with statistics appropriate for three-level meta-analytic models. Inspection of the funnel plots in both the overall and emotion-specific analyses for multiple-choice-type tasks revealed no significant deviations from symmetry, indicating no serious publication bias (Supplemental Figures S1 and S2). As per Nakagawa et al. (2022), we assessed publication bias statistically by including the inverse square root of the number of trials as a moderator in each of our models. A significant negative effect of this moderator indicates, for our meta-analysis, that studies with a small number of trials bear undue influence on the recognition accuracy estimate. No significant effects were found for multiple-choice type tasks in both overall and emotion-specific data sets; $t(171) = -1.64$, $p = .102$, $t(331) = -1.21$, $p = .227$, respectively. As such, we can be confident about the robustness of our estimates in these analyses.

For free-label tasks in the overall analysis, although visual inspection of the funnel plot (Supplemental Figure S3) revealed no major deviations from symmetry, examination of the inverse square root model indicated a significant negative effect of trial number, logit(π) = −4.99, $t(38) = -3.07$, $p = .004$. As such, we can be less confident about the robustness of our estimates in these analyses, as it appears that studies with low trial contribute significantly to the model estimates. For free-label tasks in the emotion-specific analysis, both visual inspection of the funnel plot (Supplemental Figure S4) and examination of the inverse square root model revealed no reason to suggest small-study effects, $t(92) = 0.74$, $p = .463$. Finally, to assess whether the interpretation of our results changed with the

Figure 2

Recognition Accuracy in Multiple-Choice Tasks, Expressed as Estimated Logit(π), as a Function of Age and Emotion Category



Note. If estimated logit(π) is greater than 0, proportion correct is greater than 0.5. See the online article for the color version of this figure.

inclusion of unpublished literature, we performed an additional analysis in which we included this unpublished literature next to published studies (see Supplemental Material 5). The interpretation of our results remained unchanged when including such literature, suggesting that our results are not subject to "file-drawer" effects.

Discussion

This meta-analysis sought to summarize the findings on the developmental trajectories of emotion recognition abilities from early childhood to adolescence, and to examine how such trajectories may be modulated by a host of methodological factors. We focused on more general recognition competencies (independent of specific emotion categories), as well as the recognition of discrete basic emotions. We found that children's emotion recognition accuracy improved across age, but that there were key differences in the order in which emotion categories were recognized with proficiency. In addition, a variety of participant-, stimulus-, and task-related features influenced children's recognition accuracy both overall and for specific emotions.

Table 6

Model-Estimated Recognition Accuracy for 4- and 12-Year-Old Children Across Emotion Categories Based on Age × Emotion Category Model

	4-year-old		12-year-old	
Emotion	logit(π)	Accuracy	logit(π)	Accuracy
Happiness	2.01	88.20	2.94	94.97
Anger	1.49	81.56	2.71	93.77
Sadness	1.37	79.77	2.40	91.69
Surprise	1.24	77.47	2.62	93.23
Disgust	1.35	79.35	2.34	91.21
Fear	1.02	73.51	2.14	89.43

Note. Estimated only from studies that include multiple-choice tasks.

Our analyses answer two important questions. First, on a more conceptual level, we can delineate how emotion recognition abilities improve with age. A number of single studies have investigated this process, but are limited to the extent that they examine small age ranges, use single tasks, and stimulus types (among others). Here, we could use a large sample, and consider a much wider range of experimental paradigms which ultimately provides a more comprehensive account of this conceptual question. Second, we examined how different methodological considerations (e.g., participant-, task-, and stimulus-related features) modulate the findings on children's developing emotion recognition abilities. Similarly, this question is difficult to address with single studies, and this meta-analysis, therefore, represents a crucial first step in analyzing this literature in a quantitative manner. This is necessary given the large proliferation of works on emotion recognition in childhood in recent decades and its importance for children's social lives.

Main Findings

Overall Analysis

Our main findings in the overall analysis were that, as expected, children's emotion recognition competencies appeared to increase across age. This is in line with two major theories regarding the nature of emotions—basic emotion and constructivist theories—which both assume that emotion recognition abilities increase with age, albeit as a result of different mechanisms (Barrett, 2017; Ekman, 1992; Hoemann et al., 2019; Izard, 1971). Emotion

Table 7

Univariate Models for the Effect of Important Moderator Variables on the Recognition of Emotions Anger, Disgust, and Fear in Multiple-Choice Tasks

	Anger				Disgust				Fear			
Moderator	*k*	*OR*	95% CI	logit(π)	*k*	*OR*	95% CI	logit(π)	*k*	*OR*	95% CI	logit(π)
Format												
Intercept		7.00	**[5.46, 8.97]**	1.95		6.28	**[3.97, 9.92]**	1.84		5.04	**[4.00, 6.35]**	1.62
Photo	51	1.00[a]			17	1.00[a]			48	1.00[a]		
Video	10	1.24	[−0.72, 2.13]	0.21	5	1.17	[0.66, 2.09]	0.16	8	0.61	[0.36, 1.03]	−0.50
Vocal	5	1.64	[0.81, 3.34]	0.50	1[b]				3	1.43	[0.61, 3.38]	0.36
Color												
Intercept		7.38	**[5.45, 10.00]**	2.00		7.45	**[4.29, 12.97]**	2.01		6.23	**[4.93, 7.87]**	1.83
Yes	29	1.00[c]			11	1.00[c]			27	1.00[c]		
No	16	1.28	[0.77, 2.11]	0.25	8	0.92	[0.48, 1.76]	−0.08	15	0.38	**[0.26, 0.57]**	−0.95
Validated												
Intercept		7.72	**[5.99, 9.95]**	2.04						5.68	**[4.54, 7.11]**	1.74
Yes	49	1.00[d]			17				44	1.00[d]		
No	6	0.81	[0.41, 1.61]	−0.21	2[b]				7	0.44	**[0.28, 0.71]**	−0.82

Note. Significant terms are indicated in bold. *k* = number of observations in the level of moderator. An *OR* greater than 1 (smaller than 1, respectively) indicates that as the predictor increases the odds of a correct response increase (decrease, respectively). *OR* = odd's ratio; CI = confidence interval.
[a] Intercept represents estimate for reference category Photo. [b] Levels with less than three observations are excluded from analyses. [c] Intercept represents estimate for reference category color ("Yes"). [d] Intercept represents estimate for reference category validated ("Yes").

recognition accuracy in multiple-choice type tasks (collapsed across all stimuli formats) was, on average, above chance (88%), suggesting generalized recognition competencies, even in early childhood. This result broadly reflects previous works which examined children's emotion recognition using multiple-choice tasks, although our estimated accuracy was slightly higher than some individual studies (e.g., Chronaki et al., 2015; LoBue et al., 2018; Verpaalen et al., 2019). Although the nature of our analyses prevented their direct comparison (multiple-choice type tasks could be corrected for chance guessing, whereas free-label tasks could not), children's accuracy in free-label tasks (collapsed across all stimulus formats) was lower than multiple-choice type tasks (60%), owing to the likely higher levels of difficulty associated with tasks of this type. This difference supports a previous study by Vicari et al. (2000), which found that free-label tasks were generally more difficult than multiple-choice tasks, although the magnitude of the difference in our analyses appears smaller.

Some authors have argued that the accuracy difference found between free-label and multiple-choice tasks is due to the fact that free-label paradigms rely on children's ability to produce emotion labels (Matthews et al., 2022). Children may, therefore, be unable to respond "correctly" to a given stimulus, simply because they are unable to

Table 8

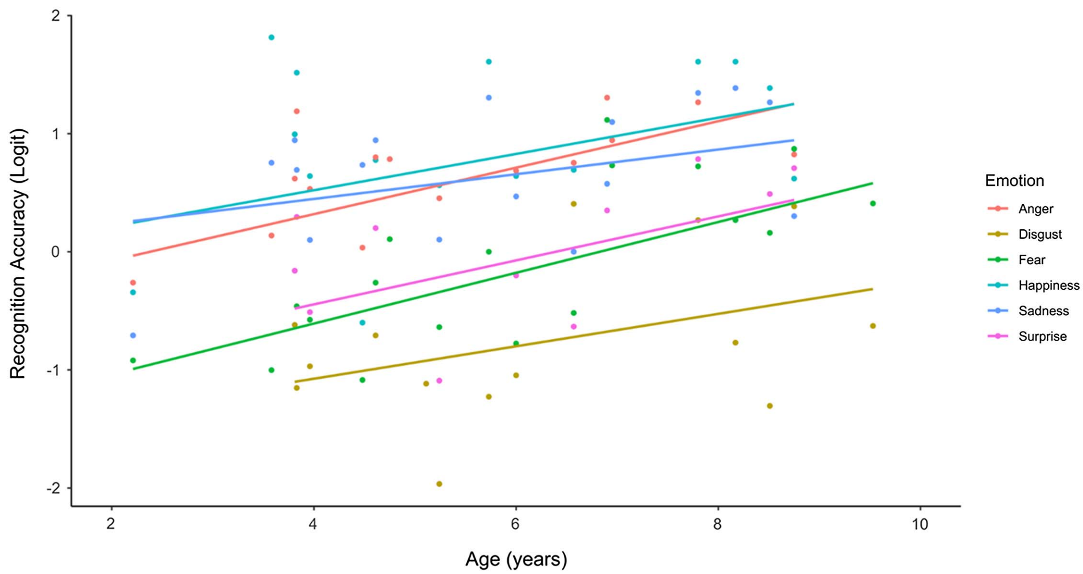
Univariate Models for the Effect of Important Moderator Variables on the Recognition of Emotions Happiness, Sadness, and Surprise in Multiple-Choice Tasks

	Happiness				Sadness				Surprise			
Moderator	*k*	*OR*	95% CI	logit(π)	*k*	*OR*	95% CI	logit(π)	*k*	*OR*	95% CI	logit(π)
Format												
Intercept		12.24	**[9.21, 16.26]**	2.50		5.84	**[4.70, 7.26]**	1.76		5.32	**[3.74, 7.57]**	1.67
Photo	61	1.00[a]			55	1.00[a]			24	1.00[a]		
Video	11	1.41	[0.68, 2.92]	0.35	7	1.10	[0.65, 1.87]	0.10	5	1.59	[0.97, 2.63]	0.47
Vocal	6	0.48	[0.22, 1.03]	−0.74	6	1.08	[0.52, 2.21]	0.07	1[b]			
Color												
Intercept		13.58	**[9.15, 20.15]**	2.61		6.48	**[4.86, 8.63]**	1.87		7.40	**[4.11, 13.34]**	2.00
Yes	35	1.00[c]			31	1.00[c]			10	1.00[c]		
No	20	0.98	[0.51, 1.86]	−0.02	15	0.84	[0.51, 1.38]	−0.17	12	0.83	[0.40, 1.71]	−0.19
Validated												
Intercept		11.32	**[8.46, 15.14]**	2.43		5.78	**[4.65, 7.20]**	1.76		6.56	**[4.07, 10.57]**	1.88
Yes	58	1.00[d]			50	1.00[d]			18	1.00[d]		
No	8	1.52	[0.63, 3.68]	0.42	7	1.28	[0.72, 2.26]	0.25	6	0.87	[0.42, 1.80]	−0,14

Note. Significant terms are indicated in bold. *k* = number of observations in the level of moderator. An *OR* greater than 1 (smaller than 1, respectively) indicates that as the predictor increases the odds of a correct response increase (decrease, respectively). *OR* = odd's ratio; CI = confidence interval.
[a] Intercept represents estimate for reference category Photo. [b] Levels with less than three observations are excluded from analyses. [c] Intercept represents estimate for reference category color ("Yes"). [d] Intercept represents estimate for reference category validated ("Yes").

Figure 3

Recognition Accuracy in Free-Label Tasks, Expressed as Estimated Logit(π), as a Function of Age and Emotion Category



Note. If estimated logit(π) is greater than 0, proportion correct is greater than 0.5. See the online article for the color version of this figure.

produce an emotion label, despite having conceptual knowledge of several key factors associated with said emotion (e.g., its antecedents, consequences, and typically associated behavioral response). It has therefore been argued that free-label tasks may underestimate children's "true" recognition competencies (Matthews et al., 2022). Given the differences in accuracy between these two task types found in our analyses, this assessment may have merit, although we could not test it statistically. On the contrary, Nelson et al. (2018) found evidence in their work that children can utilize a process of elimination strategy to infer the correct emotion labels in multiple-choice tasks with experimenter-defined labels. Therefore, the opposing assertion may also have merit—children may be able to respond "correctly" in multiple-choice type tasks without having any conceptual knowledge of a target emotion, but rather by simply knowing that a given choice in array is *not* equivalent to an already acquired emotion label. Therefore, children's "true" emotion recognition abilities may lie somewhere between these two extreme points, which may be an underestimation or overestimation, respectively.

Focusing on the differences within the multiple-choice type tasks themselves, our analyses revealed that children performed better in forced-choice as compared to match-to-sample paradigms. This result supports recent work on the importance of language as a toe-hold to emotion recognition, emphasized in constructivist theories of emotion (Barrett, 2017; Lindquist et al., 2015; Price et al., 2022). In one study, Price et al. (2022) examined 2- and 3-year-old children's performance in a task in which children either heard emotion labels before sorting faces (e.g., "sad faces go in this pile") or heard no labels (e.g., "faces like this go in this pile"). These two conditions are equivalent to forced-choice and match-to-sample task types we coded in our analyses. Ultimately, the authors found that labels improved recognition accuracy for emotion categories typically learned later in childhood (e.g., disgust, surprise, fear), but not for those typically learned earlier in childhood (e.g., happy, sad, angry). In our analyses, we found a forced-choice advantage overall, but this advantage was not reflected in accuracy differences for specific emotion categories (discussed in detail below). This result appears to support Barrett's (2017) suggestion that a focus on emotion words may be critical if we are to understand how children learn emotion concepts in the face of large intraindividual variation.

Our overall analyses also pointed to variation across certain geographical regions (Table 2). The results of our metaregression revealed significant accuracy differences between North American and European samples, with European children scoring significantly higher than their North American counterparts. This difference was not due to the differences in mean age between North American and European samples, as the effect of study region remained significant in the metaregression model when age was included as a covariate. No other differences between geographical regions could be found. Despite this, previous works have suggested that certain cultural factors may influence children and adults' response accuracy in emotion recognition tasks (Camras et al., 2006; Collins & Nowicki, 2001; Elfenbein & Ambady, 2002; Rosenqvist et al., 2017). In addition, constructivist theories of emotion emphasize the important role of culture in acquiring certain emotion knowledge (Shablack & Lindquist, 2019). A study by Rosenqvist et al. (2017) found significant variation in children's performance on the Affect Recognition subscale of the Developmental Neuropsychological Assessment-II between three countries (United States, Italy, Finland). The authors indicate that these differences may echo variation in political, educational, and cultural factors between these countries, and these differences may have been captured in our data. For example, it is possible that children who start formal schooling earlier may have more exposure to emotion words and, therefore, perform better in tasks that require the use of these words.

Critically, we found no differences in emotion recognition accuracy between boys and girls. This result stands in contrast to a previous meta-analysis (McClure, 2000), which found a small (but significant) female advantage in emotion recognition abilities, beginning from infancy and continuing into adolescence. The most likely explanation for this discrepancy lies in the models themselves. In our analyses, we used the percentage of female participants in the sample as a moderator in our estimation of recognition accuracy. In McClure's (2000) work, the author modeled difference scores between boys and girls directly, using this standardized difference as an effect size, providing a likely more sensitive measure of difference. Inspection of the distribution of our percentage female scores revealed that the majority of studies included around 50% females, with extreme values rare. This lack of between-study variance may be a contributing factor to this null effect and should be examined more carefully in future works.

Finally, of key theoretical interest were any potential interactions between age and our other moderating variables. That said, we found no evidence for any participant-, task-, or stimulus-feature interactions with age, except for stimulus validation status and presentation time. Thus, emotion recognition accuracy is not differentially influenced by task- and stimulus-related features (barring stimulus validation and presentation time) for different ages. In the case of stimulus validation status, emotion recognition accuracy increased more across age for validated stimuli compared to those that were not validated. This result could be because validated stimuli are independently rated by a (typically) larger cohort and, therefore, certain stimuli are more likely to reflect "consensus" on how a certain stimulus best be labeled. Additionally, such stimuli are likely to be more rigorous and of greater methodological quality, as they are more likely to control for extraneous low-level features that may influence emotion recognition abilities. As children age, their judgments may be more likely to conform to such consensus and this may explain this interaction. In terms of stimulus presentation time, while it did interact significantly with age, the magnitude of the effect was so small that its interpretation was impossible. As such, we can say very little about the interaction between these two variables and whether this produces meaningful differences in emotion recognition accuracy.

Emotion-Specific Analyses

Our emotion-specific analyses also revealed a number of key findings. In terms of specific emotion categories, estimated accuracy scores suggest a pattern that is broadly typical of that found in previous individual studies and reviews of children's understanding of emotions (e.g., Herba & Phillips, 2004; Ruba & Pollak, 2020; Widen, 2013), although we found some noteworthy differences. When considering all stimulus and task types, happiness was the most easily recognized emotion category, followed by anger, surprise, sadness, disgust, and fear. Age did not interact with any emotion category to produce differences in recognition accuracy. This finding has an important implication, namely, that recognition of all basic emotions appears to improve across development in a similar fashion—such that, for example, basic emotions that are the worst recognized at 3-years-old, are also the worst recognized at 10-years-old (see Figure 2). Happiness, for example, was by far the best recognized emotion across age, supporting previous findings which indicate that happiness stimuli (including facial, body expressions, and vocalizations) are consistently recognized and labeled even in early childhood (e.g., Amorim et al., 2021; Grosbras et al., 2018; Herba & Phillips, 2004; Widen, 2013; Widen & Russell, 2008a). In addition, this result supports a previous meta-analysis on an adult sample which found that happiness facial expressions were the most easily recognized among all other basic emotions (Elfenbein & Ambady, 2002).

An important note here is that happiness is the only positively valanced emotion among the list of Ekman's (1992) basic emotions (Sauter, 2010). As it has been argued that emotions are first recognized based on valence (positive/negative) in young childhood before each specific emotion can be recognized (broad-to-differentiated hypothesis; Widen, 2017), this may be the reason why happiness was recognized so consistently well. Such a pop out effect is likely not the only explanation, given the high levels of accuracy found for happiness stimuli in free-labeling tasks found in our analyses. Another explanation is that happiness may simply be better recognized because of its ubiquitousness from birth—children likely encounter expressions of happiness more than any other emotion in early childhood (Kotsoni et al., 2001), and this familiarity could also contribute to its high recognition rates.

Disgust and fear stimuli were consistently the worst recognized across all ages and stimulus types. This result is interesting considering evidence for an attentional bias toward fear-inducing stimuli during infancy (Grossmann & Jessen, 2017). Yet, the low recognition of disgust stimuli supports previous studies which found that, at least in the facial expression domain (which accounted for the largest proportion of studies), children easily confuse disgust nose scrunches with anger scowls (Widen & Russell, 2010b). In addition, a previous review found that facial expressions of both disgust and fear were the last to be acquired emotion categories, with children reaching chance recognition accuracy only by 9 years of age (Widen, 2013).

It is important to note that the recognition accuracy of disgust stimuli was only significantly lower than happy stimuli. This finding is generally in line with the constructionist perspective that disgust should have the lowest recognition rate across emotion categories overall, as the word "disgust" is learned only in later childhood (Price et al., 2022; Shablack & Lindquist, 2019). However, the fact that the recognition accuracy of disgust was not significantly lower than the other emotion categories, such as sad and angry, across all ages, speaks against constructivist theories of emotion that highlight the crucial role of language in emotion recognition development (Barrett et al., 2007; Hoemann et al., 2019). In particular, if language plays such a critical role in emotion differentiation, we would expect to see worse recognition of disgust stimuli compared to other emotion categories, especially at an early age (when children have typically not yet learned this word). Instead, we found no interaction between age and any emotion category, suggesting that, as discussed, the recognition for all emotions follows the same developmental trajectory (no matter when the words for emotions are learned in the course of development). This finding may support basic emotion theories, which, unlike constructivist theories, do not assume sudden jumps in emotion recognition when emotion words are learned. Rather, these theories suggest that children are born ready to differentiate between all emotions and become more proficient in emotion recognition across age due to maturation and social determinants (Ekman, 1999; Izard, 1994). This notion, should, nonetheless, be further clarified in future works, as we note that some emotion categories had only a small number of observations at certain ages, potentially limiting our ability to detect significant interaction effects.

When considering differences between specific stimulus formats (vocalizations, photos, and videos) across emotion categories this pattern was similar—although the nature of our analyses prevented examination of the interaction between certain emotion types and stimulus formats statistically. In particular, we did not find any differences in accuracy for the various stimulus formats within specific emotion categories. This pattern of results is unexpected, given previous works which found clear differences in recognition accuracy between static images and video stimuli of emotional expressions in both childhood and adulthood (particularly at low stimulus intensities; Richoz et al., 2018). Some authors have argued that video stimuli may bolster the recognition of certain emotions as their dynamicity is more in keeping with the kinds of information children encounter in their daily lives (Vieillard & Guidetti, 2009). Moreover, dynamic stimuli convey additionally important visual information to children that may assist in categorization (e.g., a stimulus' temporal properties or intensity; Richoz et al., 2018). Rather, our results seem to be more in keeping with an alternative set of studies which find no significant advantages for dynamic over static stimuli (Nelson & Russell, 2011b; Widen & Russell, 2015). It is perhaps possible that multiple stimulus features (e.g., stimulus format and intensity) interact to produce these between-study differences, although we did not examine these interactions specifically. We wish to stress that because a number of these pairwise comparisons included only a small number of observations (particularly for video stimuli and vocalizations), the results of these within-emotion category analyses of stimulus format should be treated cautiously. The overwhelming majority of effect sizes came from static image stimuli. Therefore, it will be important for future works to examine whether such stimulus format difference exist across other format types and, in particular, whether these differences are similar across emotion categories.

Despite finding no clear differences between stimulus format types, we found that some other task- and stimulus-related features moderated children's recognition accuracy for specific emotions. This pattern was particularly the case for fearful stimuli, which appeared to be the most influenced by our included moderators, beyond those of other emotion categories (Table 7). In particular, fearful stimuli presented in color were recognized significantly better than those presented in monochrome. We could identify only one study which looked specifically at the influence of color on the perception of facial emotions in adults (Silver & Bilker, 2015). Here, the authors found a generalized recognition advantage (in terms of both accuracy and response latency) for emotions presented in color, although they only included sad, happy, and neutral facial expressions in their analyses. Interestingly, we only found such an advantage for fearful stimuli, which may suggest that this advantage may not extend to children. Finally, unvalidated fearful stimuli were recognized significantly worse than those that were validated. Again, validated stimuli are rated independently by a (typically) larger cohort of individuals and also tend to be controlled for extraneous low-level features that may influence participants' responding. As such, it holds that an advantage for validated over unvalidated stimuli was found. We can only speculate as to why we found this relationship for fearful stimuli only. In particular, because fearful facial expressions are encountered much less frequently in daily life than other basic emotions (Calvo et al., 2014), unvalidated fearful stimuli may be more sensitive to misattribution errors than validated stimuli. Indeed, evidence suggests that children often confuse fearful facial expressions with those of surprise (Roy-Charland et al., 2015). This problem may be exacerbated in unvalidated stimulus sets where low-level perceptual features and potentially confusing facial movements are likely less controlled.

Limitations and Future Directions

Despite its comprehensiveness, there are still several important noteworthy limitations in our analysis. First, our analysis was constrained by its exclusive focus on basic emotions. Therefore, our results cannot be generalized beyond basic emotions. It will be an important avenue of future research to examine whether our pattern of results holds for other emotion categories. For example, some scholars (e.g., Izard, 2007; Lewis et al., 1989) argue that self-conscious emotions are more cognitively complex than basic emotions, and it is important to examine whether certain task- or stimulus-related features aid the recognition of these emotions above and beyond those included in the present analyses.

Moreover, in a related issue, the majority of our studies focused on children from the Global North. This problem was particularly marked given that we had only included manuscripts written in English (or with an appropriate English translation) in the synthesis of our evidence. Given the importance of considering cross-cultural perspectives in studies of emotion (Keltner et al., 2022; Matsumoto & Wilson, 2022), the plethora of studies from the Global North limit the generalizability of our findings to other cultural contexts and represent an important gap in the literature. Future studies should focus on investigating emotion recognition capacities in non-Western, educated, industrialized, rich, and democratic societies to better understand whether current findings hold.

Another critical limitation of our analyses is that we only consider children's response accuracy and did not consider any other biases or misattributions. To consider response bias, we would have needed response confusion matrices for every coded study, which presented an issue of feasibility (such matrices are seldom included in published works). Given this, our analyses are limited in that they only present one side of the story—namely, children's correct judgments. We echo calls from other scholars (e.g., Chronaki et al., 2015; Widen, 2013) who emphasize that a focus on children's misattributions (i.e., incorrect judgments) is critical if we want to fully understand how emotion recognition competencies develop across ontogeny. Finally, as discussed above, a further limitation of our analysis method was that we were unable to appropriately combine the effect sizes of free-label and multiple-choice tasks in the one analysis. As such, we were unable to compare these popular task types directly, limiting the conclusions we can draw regarding the impact of task type on children's emotion recognition abilities.

Despite these limitations, our data highlight some further important considerations for researchers interested in emotion recognition in childhood. First, given the clear differences in accuracy found between the various task types analyzed, researchers are advised to think clearly about low-level task considerations when designing emotion recognition paradigms for children. For instance, match-to-sample type tasks may index more general perceptual discrimination abilities, whereas free-label paradigms are heavily reliant on children's lexical knowledge and ability to produce labels ad hoc

(Matthews et al., 2022). The effects of stimulus-related features, however, were less clear, and showed clear differences depending on the emotion category under examination. In terms of practical recommendations, it may be useful for researchers and clinicians to ensure that they include more than one style of task in their assessments of emotion recognition. Critically, other potentially moderating variables (e.g., emotions used, number of response alternatives) should be kept consistent across tasks. For comparing children's response across multiple-choice tasks with differing response alternatives, we suggest using the proportion index measure highlighted above (Rosenthal & Rubin, 1989). Finally, given that children's accuracy improved significantly more across age in tasks using validated stimuli, we suggest using only validated stimuli in assessments of emotion recognition.

The results of the present work also provide some novel opportunities for future experimental paradigms. We noticed very few studies that used naturalistic (nonposed, spontaneous) stimuli to examine emotion recognition. Previous work in adult samples has found that the recognition of naturalistic emotional expressions follows a different pattern to that of the posed stimuli typically included in emotion recognition paradigms (Krumhuber et al., 2021; Motley & Camden, 1988). Posed stimuli, which are typically included in the most popular stimulus sets, often display highly exaggerated and stereotyped features, which may not be in keeping with the expressions individuals make in real life (Naab & Russell, 2007). This point also stands for stimuli derived from non-Western individuals, which were generally poorly represented in our included studies. We suggest that researchers interested in children's emotion recognition test children's abilities cross culturally, using stimuli derived in naturalistic settings and from various cultural groups. In addition to overcoming some of the limitations in previous works, such studies may also be useful in contributing to the current theoretical discussions regarding the nature of emotions (Adolphs et al., 2019; Barrett, 2017; Cowen et al., 2019).

Summary and Conclusions

This meta-analysis provides some important insights into the various participant-, task-, and stimulus-related features that influence children's responding in emotion recognition paradigms. We found that the methodological quality of studies did not influence our results. Effect sizes were not typically related to potential markers of study quality—including sample size and stimulus validation status (except for the emotion fear). Our results reiterate and strengthen the conclusions of previous qualitative reviews, which suggest general increases in emotion recognition capacities across age, but with key differences between specific emotion categories at certain ages (Herba & Phillips, 2004; Ruba & Pollak, 2020; Widen, 2013). In addition, our results support the arguments of Hayes et al. (2020), who highlight that it is not only the specific emotions, but that a host of other methodological factors may lead to significant variation in responding in recognition tasks. This variation may be particularly marked for children, who are only beginning to acquire knowledge of certain emotions and have more limited experience interacting in complex social environments.

To return to our initial aim, the results of this meta-analysis highlight the complexities associated with delineating exactly how, and in what order, children's knowledge of external emotion cues develops across the lifespan. The responses that children give in paradigms that aim to test these patterns are likely to depend on a complex interaction between stimulus, task, and participant. These interactions should be carefully considered by those interested in how children's emotion recognition abilities develop.

References

References marked with an asterisk indicate studies included in the meta-analysis.

Adolphs, R., Mlodinow, L., & Barrett, L. F. (2019). What is an emotion? *Current Biology*, *29*(20), R1060–R1064. https://doi.org/10.1016/j.cub.2019.09.008

*Agnoli, S., Mancini, G., Pozzoli, T., Baldaro, B., Russo, P. M., & Surcinelli, P. (2012). The interaction between emotional intelligence and cognitive ability in predicting scholastic performance in school-aged children. *Personality and Individual Differences*, *53*(5), 660–665. https://doi.org/10.1016/j.paid.2012.05.020

*Alonso-Alberca, N., Vergara, A. I., Fernandez-Berrocal, P., Johnson, S. R., & Izard, C. E. (2012). The adaptation and validation of the Emotion Matching Task for preschool children in Spain. *International Journal of Behavioral Development*, *36*(6), 489–494. https://doi.org/10.1177/0165025412462154

*Amici, F., Waterman, J., Kellermann, C. M., Karimullah, K., & Bräuer, J. (2019). The ability to recognize dog emotions depends on the cultural milieu in which we grow up. *Scientific Reports*, *9*(1), Article 16414. https://doi.org/10.1038/s41598-019-52938-4

Amorim, M., Anikin, A., Mendes, A. J., Lima, C. F., Kotz, S. A., & Pinheiro, A. P. (2021). Changes in vocal emotion recognition across the life span. *Emotion*, *21*(2), 315–325. https://doi.org/10.1037/emo0000692

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist*, *73*(1), 3–25. https://doi.org/10.1037/amp0000191

*Arterberry, M. E., Perry, E. T., Price, C. M., & Steimel, S. A. (2020). Emotional understanding predicts facial recognition in 3-to 5-year-old children. *European Journal of Developmental Psychology*, *17*(2), 293–306. https://doi.org/10.1080/17405629.2019.1589445

Assink, M., & Wibbelink, C. J. (2016). Fitting three-level meta-analytic models in R: A step-by-step tutorial. *The Quantitative Methods for Psychology*, *12*(3), 154–174. https://www.tqmp.org/RegularArticles/vol12-3/p154/

*Balas, B., Auen, A., Saville, A., & Schmidt, J. (2018). Body emotion recognition disproportionately depends on vertical orientations during childhood. *International Journal of Behavioral Development*, *42*(2), 278–283. https://doi.org/10.1177/0165025417690267

*Balconi, M., Vanutelli, M. E., & Finocchiaro, R. (2014). Multilevel analysis of facial expressions of emotion and script: Self-report (arousal and valence) and psychophysiological correlates. *Behavioral and Brain Functions*, *10*(1), Article 32. https://doi.org/10.1186/1744-9081-10-32

Barrett, L. F. (2014). The conceptual act theory: A précis. *Emotion Review*, *6*(4), 292–297. https://doi.org/10.1177/1754073914534479

Barrett, L. F. (2017). *How emotions are made: The secret life of the brain*. Houghton Mifflin Harcourt.

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements. *Psychological Science in the Public Interest*, *20*(1), 1–68. https://doi.org/10.1177/1529100619832930

Barrett, L. F., Lindquist, K. A., & Gendron, M. (2007). Language as context for the perception of emotion. *Trends in Cognitive Sciences*, *11*(8), 327–332. https://doi.org/10.1016/j.tics.2007.06.003

*Barth, J. M., & Archibald, A. (2003). The relation between emotion production behavior and preschool social behavior: In the eye of the beholder. *Social Development*, *12*(1), 67–90. https://doi.org/10.1111/1467-9507.00222

*Battaglia, M., Zanoni, A., Ogliari, A., Crevani, F., Falzone, L., Bertoletti, E., & Di Serio, C. (2010). Identification of gradually changing emotional expressions in schoolchildren: The influence of the type of stimuli and of specific symptoms of anxiety. *Cognition and Emotion*, *24*(6), 1070–1079. https://doi.org/10.1080/02699930903112761

*Beck, A., Canamero, L., Hiolle, A., Damiano, L., Cosi, P., Tesser, F., & Sommavilla, G. (2013). Interpretation of emotional body language displayed by a humanoid robot: A case study with children. *International Journal of Social Robotics*, *5*(3), 325–334. https://doi.org/10.1007/s12369-013-0193-z

*Bedford, R., Wagner, N. J., Rehder, P. D., Propper, C., Willoughby, M. T., & Mills-Koonce, R. W. (2017). The role of infants' mother-directed gaze, maternal sensitivity, and emotion recognition in childhood callous unemotional behaviours. *European Child & Adolescent Psychiatry*, *26*(8), 947–956. https://doi.org/10.1007/s00787-017-0967-1

*Berzenski, S. R., & Yates, T. M. (2017). The differential influences of parenting and child narrative coherence on the development of emotion recognition. *Developmental Psychology*, *53*(10), 1912–1923. https://doi.org/10.1037/dev0000395

*Bijsterbosch, G., Mobach, L., Verpaalen, I. A. M., Bijlstra, G., Hudson, J. L., Rinck, M., & Klein, A. M. (2020). Validation of the child models of the Radboud Faces Database by children. *International Journal of Behavioral Development*, *45*(2), 146–152. https://doi.org/10.1177/0165025420935631

*Birmingham, E., Meixner, T., Iarocci, G., Kanan, C., Smilek, D., & Tanaka, J. W. (2013). The moving window technique: A window into developmental changes in attention during facial emotion recognition. *Child Development*, *84*(4), 1407–1424. https://doi.org/10.1111/cdev.12039

*Bisson, J. B. (2019). It's written all over their faces: Preschoolers' emotion understanding. *Social Development*, *28*(1), 74–89. https://doi.org/10.1111/sode.12322

*Broerena, S., Murisa, P., Bouwmeestera, S., Field, A. P., & Voerman, J. S. (2011). Processing biases for emotional faces in 4- to 12-year-old non-clinical children: An exploratory study of developmental patterns and relationships with social anxiety and behavioral inhibition. *Journal of Experimental Psychopathology*, *2*(4), 454–474. https://doi.org/10.5127/jep.016611

*Brosgole, L., & Weisman, J. (1995). Mood recognition across the ages. *International Journal of Neuroscience*, *82*(3–4), 169–189. https://doi.org/10.3109/00207459508999800

*Bullock, M., & Russell, J. A. (1985). Further evidence in preschoolers' interpretation of facial expressions. *International Journal of Behavioral Development*, *8*(1), 15–38. https://doi.org/10.1177/016502548500800103

Calvo, M. G., Gutiérrez-García, A., Fernández-Martín, A., & Nummenmaa, L. (2014). Recognition of facial expressions of emotion is related to their frequency in everyday life. *Journal of Nonverbal Behavior*, *38*(4), 549–567. https://doi.org/10.1007/s10919-014-0191-3

Camras, L. A., Perlman, S. B., Fries, A. B. W., & Pollak, S. D. (2006). Post-institutionalized Chinese and Eastern European children: Heterogeneity in the development of emotion understanding. *International Journal of Behavioral Development*, *30*(3), 193–199. https://doi.org/10.1177/0165025406063608

*Castro, V. L., Cooke, A. N., Halberstadt, A. G., & Garrett-Peters, P. (2018). Bidirectional linkages between emotion recognition and problem behaviors in elementary school children. *Journal of Nonverbal Behavior*, *42*(2), 155–178. https://doi.org/10.1007/s10919-017-0269-9

Cecilione, J. L., Rappaport, L. M., Verhulst, B., Carney, D. M., Blair, R. J. R., Brotman, M. A., Leibenluft, E., Pine, D. S., Roberson-Nay, R., & Hettema, J. M. (2017). Test–retest reliability of the facial expression labeling task. *Psychological Assessment*, *29*(12), 1537–1542. https://doi.org/10.1037/pas0000439

*Champoux-Larsson, M.-F., Dylman, A. S., Ornkloo, H., & Esteves, F. (2019). Identification of facial expressions of emotion by 4-year-old children from different linguistic environments. *The International Journal of Bilingualism*, *23*(5), 1208–1219. https://doi.org/10.1177/1367006918781069

*Chronaki, G., Broyd, S., Garner, M., Hadwin, J. A., Thompson, M. J. J., & Sonuga-Barke, E. J. S. (2012). Isolating N400 as neural marker of vocal anger processing in 6–11-year old children. *Developmental Cognitive Neuroscience*, *2*(2), 268–276. https://doi.org/10.1016/j.dcn.2011.11.007

*Chronaki, G., Broyd, S. J., Garner, M., Benikos, N., Thompson, M. J. J., Sonuga-Barke, E. J. S., & Hadwin, J. A. (2018). The moderating effect of self-reported state and trait anxiety on the late positive potential to emotional faces in 6–11-year-old children. *Frontiers in Psychology*, *9*, Article 125. https://doi.org/10.3389/fpsyg.2018.00125

Chronaki, G., Hadwin, J. A., Garner, M., Maurage, P., & Sonuga-Barke, E. J. S. (2015). The development of emotion recognition from facial expressions and non-linguistic vocalizations during childhood. *British Journal of Developmental Psychology*, *33*(2), 218–236. https://doi.org/10.1111/bjdp.12075

*Cohen, I., Looije, R., & Neerincx, M. A. (2014). Child's perception of robot's emotions: Effects of platform, context and experience. *International Journal of Social Robotics*, *6*(4), 507–518. https://doi.org/10.1007/s12369-014-0230-6

*Coleman, J. R. I., Lester, K. J., Keers, R., Munafò, M. R., Breen, G., & Eley, T. C. (2017). Genome-wide association study of facial emotion recognition in children and association with polygenic risk for mental health disorders. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics*, *174*(7), 701–711. https://doi.org/10.1002/ajmg.b.32558

*Colle, L., & Del Giudice, M. (2011). Patterns of attachment and emotional competence in middle childhood. *Social Development*, *20*(1), 51–72. https://doi.org/10.1111/j.1467-9507.2010.00576.x

Collin, L., Bindra, J., Raju, M., Gillberg, C., & Minnis, H. (2013). Facial emotion recognition in child psychiatry: A systematic review. *Research in Developmental Disabilities*, *34*(5), 1505–1520. https://doi.org/10.1016/j.ridd.2013.01.008

*Collins, M., & Nowicki, S., Jr. (2001). African American children's ability to identify emotion in facial expressions and tones of voice of European Americans. *The Journal of Genetic Psychology*, *162*(3), 334–346. https://doi.org/10.1080/00221320109597487

Cowen, A., Sauter, D., Tracy, J. L., & Keltner, D. (2019). Mapping the passions: Toward a high-dimensional taxonomy of emotional experience and expression. *Psychological Science in the Public Interest*, *20*(1), 69–90. https://doi.org/10.1177/1529100619850176

*Crivelli, C., Jarillo, S., Russell, J. A., & Fernández-Dols, J.-M. (2016). Reading emotions from faces in two indigenous societies. *Journal of Experimental Psychology: General*, *145*(7), 830–843. https://doi.org/10.1037/xge0000172

Da Fonseca, D., Seguier, V., Santos, A., Poinso, F., & Deruelle, C. (2009). Emotion understanding in children with ADHD. *Child Psychiatry and Human Development*, *40*(1), 111–121. https://doi.org/10.1007/s10578-008-0114-9

Dawel, A., O'Kearney, R., McKone, E., & Palermo, R. (2012). Not just fear and sadness: Meta-analytic evidence of pervasive emotion recognition deficits for facial and vocal expressions in psychopathy. *Neuroscience and Biobehavioral Reviews*, *36*(10), 2288–2304. https://doi.org/10.1016/j.neubiorev.2012.08.006

*Dawel, A., Palermo, R., O'Kearney, R., & McKone, E. (2015). Children can discriminate the authenticity of happy but not sad or fearful facial expressions, and use an immature intensity-only strategy. *Frontiers in Psychology*, *6*, Article 462. https://doi.org/10.3389/fpsyg.2015.00462

de Gelder, B., Van den Stock, J., Meeren, H. K. M., Sinke, C. B. A., Kret, M. E., & Tamietto, M. (2010). Standing up for the body. Recent progress in uncovering the networks involved in the perception of bodies and bodily expressions. *Neuroscience and Biobehavioral Reviews*, *34*(4), 513–527. https://doi.org/10.1016/j.neubiorev.2009.10.008

*DiLalla, L. F., & John, S. G. (2020). A genetically informed examination of the relations between inaccurate emotion expression and recognition and experiencing peer victimization. *Social Development*, *29*(2), 391–410. https://doi.org/10.1111/sode.12410

*Doherty, C. P., Fitzsimons, M., Asenbauer, B., & Staunton, H. (1999). Discrimination of prosody and music by normal children. *European Journal of Neurology*, *6*(2), 221–226. https://doi.org/10.1111/j.1468-1331.1999.tb00016.x

*Doi, H., Kato, A., Hashimoto, A., & Masataka, N. (2008). Role of biological-motion information in recognition of facial expressions by young children. *Perception*, *37*(9), 1399–1411. https://doi.org/10.1068/p5673

*Dunn, E. C., Crawford, K. M., Soare, T. W., Button, K. S., Raffeld, M. R., Smith, A. D. A. C., Penton-Voak, I. S., & Munafò, M. R. (2018). Exposure to childhood adversity and deficits in emotion recognition: Results from a large, population-based sample. *Journal of Child Psychology and Psychiatry*, *59*(8), 845–854. https://doi.org/10.1111/jcpp.12881

*Dunsmore, J. C., Her, P., Halberstadt, A. G., & Perez-Rivera, M. B. (2009). Parents' beliefs about emotions and children's recognition of parents' emotions. *Journal of Nonverbal Behavior*, *33*(2), 121–140. https://doi.org/10.1007/s10919-008-0066-6

Durán, J. I., & Fernández-Dols, J.-M. (2021). Do emotions result in their predicted facial expressions? A meta-analysis of studies on the co-occurrence of expression and emotion. *Emotion*, *21*(7), 1550–1569. https://doi.org/10.1037/emo0001015

*Edwards, R., Manstead, A. S., & MacDonald, C. J. (1984). The relationship between children's sociometric status and ability to recognize facial expressions of emotion. *European Journal of Social Psychology*, *14*(2), 235–238. https://doi.org/10.1002/ejsp.2420140212

Ekman, P. (1970). Universal facial expressions of emotion. *California Mental Health Research Digest*, *8*(4), 151–158. https://psycnet.apa.org/record/1972-06605-001

Ekman, P. (1976). *Pictures of facial affect*. Consulting Psychologists Press.

Ekman, P. (1992). Are there basic emotions? *Psychological Review*, *99*(3), 550–553. https://doi.org/10.1037/0033-295X.99.3.550

Ekman, P. (1999). Basic emotions. In T. Dalgleish & M. Power (Eds.), *Handbook of cognition and emotion* (pp. 45–60). Wiley. https://doi.org/10.1002/0470013494.ch3

Ekman, P., & Friesen, W. V. (1976). *Pictures of facial affect*. Consulting Psychologists Press.

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin*, *128*(2), 203–235. https://doi.org/10.1037/0033-2909.128.2.203

*Felleman, E. S., Carlson, C. R., Barden, R. C., Rosenberg, L., & Masters, J. C. (1983). Children's and adults' recognition of spontaneous and posed emotional expressions in young children. *Developmental Psychology*, *19*(3), 405–413. https://doi.org/10.1037/0012-1649.19.3.405

*Fernandez-Sanchez, M., Gimenez-Dasi, M., & Quintanilla, L. (2014). Toddlers' understanding of basic emotions: Identification, labeling and causality. *Infancia y Aprendizaje*, *37*(3), 569–601. https://doi.org/10.1080/02103702.2014.965463

*Fong, F. T. K., Mondloch, C. J., & Nelson, N. L. (2020). Interactive situations reveal more about children's emotional knowledge. *Journal of Experimental Child Psychology*, *198*, Article 104879. https://doi.org/10.1016/j.jecp.2020.104879

*Forslund, T., Kenward, B., Granqvist, P., Gredebäck, G., & Brocki, K. C. (2017). Diminished ability to identify facial emotional expressions in children with disorganized attachment representations. *Developmental Science*, *20*(6), Article e12465. https://doi.org/10.1111/desc.12465

Gagnon, M., Gosselin, P., Hudon-ven der Buhs, I., Larocque, K., & Milliard, K. (2010). Children's recognition and discrimination of fear and disgust facial expressions. *Journal of Nonverbal Behavior*, *34*(1), 27–42. https://doi.org/10.1007/s10919-009-0076-z

*Gao, X., & Maurer, D. (2009). Influence of intensity on children's sensitivity to happy, sad, and fearful facial expressions. *Journal of Experimental Child Psychology*, *102*(4), 503–521. https://doi.org/10.1016/j.jecp.2008.11.002

*Garcia, S. E., & Tully, E. C. (2020). Children's recognition of happy, sad, and angry facial expressions across emotive intensities. *Journal of Experimental Child Psychology*, *197*, Article 104881. https://doi.org/10.1016/j.jecp.2020.104881

*Gil, S., Hattouti, J., & Laval, V. (2016). How children use emotional prosody: Crossmodal emotional integration? *Developmental Psychology*, *52*(7), 1064–1072. https://doi.org/10.1037/dev0000121

*Glanville, D. N., & Nowicki, S. (2002). Facial expression recognition and social competence among African American elementary school children: An examination of ethnic differences. *Journal of Black Psychology*, *28*(4), 318–329. https://doi.org/10.1177/009579802237540

Gonçalves, A. R., Fernandes, C., Pasion, R., Ferreira-Santos, F., Barbosa, F., & Marques-Teixeira, J. (2018). Effects of age on the identification of emotions in facial expressions: A meta-analysis. *PeerJ*, *6*, Article e5278. https://doi.org/10.7717/peerj.5278

Goodfellow, S., & Nowicki, S., Jr. (2009). Social adjustment, academic adjustment, and the ability to identify emotion in facial expressions of 7-year-old children. *The Journal of Genetic Psychology*, *170*(3), 234–243. https://doi.org/10.1080/00221320903218281

*Gosselin, P., & Simard, J. (1999). Children's knowledge of facial expressions of emotions: Distinguishing fear and surprise. *The Journal of Genetic Psychology*, *160*(2), 181–193. https://doi.org/10.1080/00221329909595391

Griffiths, S., Penton-Voak, I. S., Jarrold, C., & Munafò, M. R. (2015). No own-age advantage in children's recognition of emotion on prototypical faces of different ages. *PLOS ONE*, *10*(5), Article e0125256. https://doi.org/10.1371/journal.pone.0125256

*Grosbras, M.-H., Ross, P. D., & Belin, P. (2018). Categorical emotion recognition from voice improves during childhood and adolescence. *Scientific Reports*, *8*(1), Article 14791. https://doi.org/10.1038/s41598-018-32868-3

Gross, A. L., & Ballif, B. (1991). Children's understanding of emotion from facial expressions and situations: A review. *Developmental Review*, *11*(4), 368–398. https://doi.org/10.1016/0273-2297(91)90019-K

Grossmann, T., & Jessen, S. (2017). When in infancy does the "fear bias" develop? *Journal of Experimental Child Psychology*, *153*, 149–154. https://doi.org/10.1016/j.jecp.2016.06.018

*Guarnera, M., Magnano, P., Pellerone, M., Cascio, M. I., Squatrito, V., & Buccheri, S. L. (2018). Facial expressions and the ability to recognize emotions from the eyes or mouth: A comparison among old adults, young adults, and children. *The Journal of Genetic Psychology*, *179*(5), 297–310. https://doi.org/10.1080/00221325.2018.1509200

Hall, B., Hall, M., Statisticat, L. L. C., Brown, E., Hermanson, R., Charpentier, E., Heck, D., Laurent, S., Gronau, Q. F., & Singmann, H. (2021). *LaplacesDemon: Complete environment for bayesian inference* (Version 16.1.6). https://CRAN.R-project.org/package=LaplacesDemon

*Harrigan, J. A. (1984). The effects of task order on children's identification of facial expressions. *Motivation and Emotion*, *8*(2), 157–169. https://doi.org/10.1007/BF00993071

Hauschild, K. M., Felsman, P., Keifer, C. M., & Lerner, M. D. (2020). Evidence of an own-age bias in facial emotion recognition for adolescents with and without autism spectrum disorder. *Frontiers in Psychiatry*, *11*, Article 428. https://doi.org/10.3389/fpsyt.2020.00428

Hayes, G. S., McLennan, S. N., Henry, J. D., Phillips, L. H., Terrett, G., Rendell, P. G., Pelly, R. M., & Labuschagne, I. (2020). Task characteristics influence facial emotion recognition age-effects: A meta-analytic review. *Psychology and Aging*, *35*(2), 295–315. https://doi.org/10.1037/pag0000441

Herba, C., & Phillips, M. (2004). Annotation: Development of facial expression recognition from childhood to adolescence: Behavioural and neurological perspectives. *Journal of Child Psychology and Psychiatry*, *45*(7), 1185–1198. https://doi.org/10.1111/j.1469-7610.2004.00316.x

*Herba, C. M., Benson, P., Landau, S., Russell, T., Goodwin, C., Lemche, E., Santosh, P., & Phillips, M. (2008). Impact of familiarity upon children's developing facial expression recognition. *Journal of Child

Psychology and Psychiatry, *49*(2), 201–210. https://doi.org/10.1111/j.1469-7610.2007.01835.x

Hoemann, K., Xu, F., & Barrett, L. F. (2019). Emotion words, emotion concepts, and emotional development in children: A constructionist hypothesis. *Developmental Psychology*, *55*(9), 1830–1849. https://doi.org/10.1037/dev0000686

*Horning, S. M., Cornwell, R. E., & Davis, H. P. (2012). The recognition of facial expressions: An investigation of the influence of age and cognition. *Neuropsychology, Development, and Cognition. Section B, Aging, Neuropsychology and Cognition*, *19*(6), 657–676. https://doi.org/10.1080/13825585.2011.645011

Hox, J. J. (2010). *Multilevel analysis: Techniques and applications*. Routledge.

*Hwa-Froelich, D. A., & Matsuo, H. (2019). Pragmatic language performance of children adopted internationally. *American Journal of Speech-Language Pathology*, *28*(2), 501–514. https://doi.org/10.1044/2018_AJSLP-18-0075

*Hwa-Froelich, D. A., Matsuo, H., & Becker, J. C. (2014). Emotion identification from facial expressions in children adopted internationally. *American Journal of Speech-Language Pathology*, *23*(4), 641–654. https://doi.org/10.1044/2014_AJSLP-14-0009

Izard, C. E., Fine, S., Schultz, D., Mostow, A., Ackerman, B., & Youngstrom, E. (2001). Emotion knowledge as a predictor of social behavior and academic competence in children at risk. *Psychological Science*, *12*(1), 18–23. https://doi.org/10.1111/1467-9280.00304

Izard, C. E. (1971). *The face of emotion*. Appleton-Century-Crofts.

Izard, C. E. (1977). *Human emotions* (1977th ed.). Plenum Press.

Izard, C. E. (1994). Innate and universal facial expressions: Evidence from developmental and cross-cultural research. *Psychological Bulletin*, *115*(2), 288–299. https://doi.org/10.1037/0033-2909.115.2.288

Izard, C. E. (2007). Basic emotions, natural kinds, emotion schemas, and a new paradigm. *Perspectives on Psychological Science*, *2*(3), 260–280. https://doi.org/10.1111/j.1745-6916.2007.00044.x

Keltner, D., Sauter, D., Tracy, J. L., Wetchler, E., & Cowen, A. S. (2022). How emotions, relationships, and culture constitute each other: Advances in social functionalist theory. *Cognition and Emotion*, *36*(3), 388–401. https://doi.org/10.1080/02699931.2022.2047009

*Kessels, R. P. C., Montagne, B., Hendriks, A. W., Perrett, D. I., & de Haan, E. H. F. (2014). Assessment of perception of morphed facial expressions using the Emotion Recognition Task: Normative data from healthy participants aged 8–75. *Journal of Neuropsychology*, *8*(1), 75–93. https://doi.org/10.1111/jnp.12009

*Kestenbaum, R. (1992). Feeling happy versus feeling good—The processing of discrete and global categories of emotional expressions by children and adults. *Developmental Psychology*, *28*(6), 1132–1142. https://doi.org/10.1037/0012-1649.28.6.1132

Kotsoni, E., de Haan, M., & Johnson, M. H. (2001). Categorical perception of facial expressions by 7-month-old infants. *Perception*, *30*(9), 1115–1125. https://doi.org/10.1068/p3155

Kret, M. E. (2015). Emotional expressions beyond facial muscle actions. A call for studying autonomic signals and their impact on social perception. *Frontiers in Psychology*, *6*, Article 711. https://doi.org/10.3389/fpsyg.2015.00711

Krumhuber, E. G., Küster, D., Namba, S., Shah, D., & Calvo, M. G. (2021). Emotion recognition from posed and spontaneous dynamic expressions: Human observers versus machine analysis. *Emotion*, *21*(2), 447–451. https://doi.org/10.1037/emo0000712

*Kujawa, A., Dougherty, L., Durbin, C. E., Laptook, R., Torpey, D., & Klein, D. N. (2014). Emotion recognition in preschool children: Associations with maternal depression and early parenting. *Development and Psychopathology*, *26*(1), 159–170. https://doi.org/10.1017/S0954579413000928

*Lawrence, K., Campbell, R., & Skuse, D. (2015). Age, gender, and puberty influence the development of facial emotion recognition. *Frontiers in Psychology*, *6*, Article 761. https://doi.org/10.3389/fpsyg.2015.00761

*Lázaro, E., Amayra, I., López-Paz, J. F., Martínez, O., Pérez Alvarez, M., Berrocoso, S., Al-Rashaida, M., García, M., Luna, P., Pérez-Núñez, P., Rodriguez, A. A., Fernández, P., Parada Fernández, P., & Oliva-Macías, M. (2020). Using a virtual serious game (Deusto-e-motion1.0) to assess the theory of mind in primary school children: Observational descriptive study. *JMIR Serious Games*, *8*(2), Article e12971. https://doi.org/10.2196/12971

*Leime, J. L., Neto, J. R., Alves, S. M., & Torro-Alves, N. (2013). Recognition of facial expressions in children, young adults and elderly people. *Estudos de Psicologia*, *30*(2), 161–167. https://doi.org/10.1590/S0103-166X2013000200002

*Leitzke, B. T., & Pollak, S. D. (2016). Developmental changes in the primacy of facial cues for emotion recognition. *Developmental Psychology*, *52*(4), 572–581. https://doi.org/10.1037/a0040067

Lewis, M., Sullivan, M. W., Stanger, C., & Weiss, M. (1989). Self development and self-conscious emotions. *Child Development*, *60*(1), 146–156. https://doi.org/10.2307/1131080

Lindquist, K. A., MacCormack, J. K., & Shablack, H. (2015). The role of language in emotion: Predictions from psychological constructionism. *Frontiers in Psychology*, *6*, Article 444. https://doi.org/10.3389/fpsyg.2015.00444

*LoBue, V., Baker, L., & Thrasher, C. (2018). Through the eyes of a child: Preschoolers' identification of emotional expressions from the child affective facial expression (CAFE) set. *Cognition and Emotion*, *32*(5), 1122–1130. https://doi.org/10.1080/02699931.2017.1365046

Lozier, L. M., Vanmeter, J. W., & Marsh, A. A. (2014). Impairments in facial affect recognition associated with autism spectrum disorders: A meta-analysis. *Development and Psychopathology*, *26*(4), 933–945. https://doi.org/10.1017/S0954579414000479

*Lydon, D. M., & Nixon, E. (2014). Facial mimicry interference affects reaction time but not accuracy on an emotion recognition task in children, adolescents and adults. *The Irish Journal of Psychology*, *35*(2–3), 106–120. https://doi.org/10.1080/03033910.2014.982145

*Maassarani, R., Gosselin, P., Montembeault, P., & Gagnon, M. (2014). French-speaking children's freely produced labels for facial expressions. *Frontiers in Psychology*, *5*, Article 555. https://doi.org/10.3389/fpsyg.2014.00555

*MacDonald, P. M., Kirkpatrick, S. W., & Sullivan, L. A. (1996). Schematic drawings of facial expressions for emotion recognition and interpretation by preschool-aged children. *Genetic, Social, and General Psychology Monographs*, *122*(4), 373–388. https://europepmc.org/article/med/8976595

Malatesta, C. Z., Izard, C. E., Culver, C., & Nicolich, M. (1987). Emotion communication skills in young, middle-aged, and older women. *Psychology and Aging*, *2*(2), 193–203. https://doi.org/10.1037/0882-7974.2.2.193

*Mancini, G., Agnoli, S., Baldaro, B., Bitti, P. E. R., & Surcinelli, P. (2013). Facial expressions of emotions: Recognition accuracy and affective reactions during late childhood. *The Journal of Psychology*, *147*(6), 599–617. https://doi.org/10.1080/00223980.2012.727891

*Mardo, E., Avidan, G., & Hadad, B.-S. (2018). Adults' markers of face processing are present at age 6 and are interconnected along development. *Perception*, *47*(10–11), 1002–1028. https://doi.org/10.1177/0301006618794943

*Markham, R., & Adams, K. (1992). The effect of type of task on children's identification of facial expressions. *Journal of Nonverbal Behavior*, *16*(1), 21–39. https://doi.org/10.1007/BF00986877

Matsumoto, D., & Hwang, H. C. (2014). Judgments of subtle facial expressions of emotion. *Emotion*, *14*(2), 349–357. https://doi.org/10.1037/a0035237

Matsumoto, D., & Wilson, M. (2022). A half-century assessment of the study of culture and emotion. *Journal of Cross-Cultural Psychology*, *53*(7–8), 917–934. https://doi.org/10.1177/00220221221084236

Matthews, C. M., Thierry, S. M., & Mondloch, C. J. (2022). Recognizing, discriminating, and labeling emotional expressions in a free-sorting task: A developmental story. *Emotion*, *22*(5), 945–953. https://doi.org/10.1037/emo0000851

McClure, E. B. (2000). A meta-analytic review of sex differences in facial expression processing and their development in infants, children, and adolescents. *Psychological Bulletin*, *126*(3), 424–453. https://doi.org/10.1037/0033-2909.126.3.424

*McClure, E. B., & Nowicki, S., Jr. (2001). Associations between social anxiety and nonverbal processing skill in preadolescent boys and girls. *Journal of Nonverbal Behavior*, *25*(1), 3–19. https://doi.org/10.1023/A:1006753006870

*Mondloch, C. J., Gerada, A., Proietti, V., & Nelson, N. L. (2019). The influence of subtle facial expressions on children's first impressions of trustworthiness and dominance is not adult-like. *Journal of Experimental Child Psychology*, *180*, 19–38. https://doi.org/10.1016/j.jecp.2018.12.002

*Mondloch, C. J., Horner, M., & Mian, J. (2013). Wide eyes and drooping arms: Adult-like congruency effects emerge early in the development of sensitivity to emotional faces and body postures. *Journal of Experimental Child Psychology*, *114*(2), 203–216. https://doi.org/10.1016/j.jecp.2012.06.003

Montirosso, R., Peverelli, M., Frigerio, E., Crespi, M., & Borgatti, R. (2010). The development of dynamic facial expression recognition at different intensities in 4- to 18-year-olds. *Social Development*, *19*(1), 71–92. https://doi.org/10.1111/j.1467-9507.2008.00527.x

Motley, M. T., & Camden, C. T. (1988). Facial expression of emotion: A comparison of posed expressions versus spontaneous expressions in an interpersonal communication setting. *Western Journal of Speech Communication*, *52*(1), 1–22. https://doi.org/10.1080/10570318809389622

*Muñoz, L. C. (2009). Callous-unemotional traits are related to combined deficits in recognizing afraid faces and body poses. *Journal of the American Academy of Child & Adolescent Psychiatry*, *48*(5), 554–562. https://doi.org/10.1097/CHI.0b013e31819c2419

Naab, P. J., & Russell, J. A. (2007). Judgments of emotion from spontaneous facial expressions of New Guineans. *Emotion*, *7*(4), 736–744. https://doi.org/10.1037/1528-3542.7.4.736

Nakagawa, S., Lagisz, M., Jennions, M. D., Koricheva, J., Noble, D. W. A., Parker, T. H., Sánchez-Tójar, A., Yang, Y., & O'Dea, R. E. (2022). Methods for testing publication bias in ecological and evolutionary meta-analyses. *Methods in Ecology and Evolution*, *13*(1), 4–21. https://doi.org/10.1111/2041-210X.13724

*Naruse, S., Hashimoto, T., Mori, K., Tsuda, Y., Takahara, M., & Kagami, S. (2013). Developmental changes in facial expression recognition in Japanese school-age children. *The Journal of Medical Investigation*, *60*(1–2), 114–120. https://doi.org/10.2152/jmi.60.114

*Nelson, N. L., Hudspeth, K., & Russell, J. A. (2013). A story superiority effect for disgust, fear, embarrassment, and pride. *British Journal of Developmental Psychology*, *31*(3), 334–348. https://doi.org/10.1111/bjdp.12011

*Nelson, N. L., & Mondloch, C. J. (2018). Children's visual attention to emotional expressions varies with stimulus movement. *Journal of Experimental Child Psychology*, *172*, 13–24. https://doi.org/10.1016/j.jecp.2018.03.001

*Nelson, N. L., & Mondloch, C. J. (2019). Children's perception of emotions in the context of live interactions: Eye movements and emotion judgements. *Behavioural Processes*, *164*, 193–200. https://doi.org/10.1016/j.beproc.2019.05.006

*Nelson, N. L., Nowicki, E., Diemer, M. C., Sangster, K., Cheng, C., & Russell, J. A. (2018). Children can create a new emotion category through a process of elimination. *Cognitive Development*, *47*, 117–123. https://doi.org/10.1016/j.cogdev.2018.04.006

*Nelson, N. L., & Russell, J. A. (2011a). Preschoolers' use of dynamic facial, bodily, and vocal cues to emotion. *Journal of Experimental Child Psychology*, *110*(1), 52–61. https://doi.org/10.1016/j.jecp.2011.03.014

*Nelson, N. L., & Russell, J. A. (2011b). Putting motion in emotion: Do dynamic presentations increase preschooler's recognition of emotion? *Cognitive Development*, *26*(3), 248–259. https://doi.org/10.1016/j.cogdev.2011.06.001

Nelson, N. L., & Russell, J. A. (2016). A facial expression of pax: Assessing children's "recognition" of emotion from faces. *Journal of Experimental Child Psychology*, *141*, 49–64. https://doi.org/10.1016/j.jecp.2015.07.016

*Nowicki, S., Bliwise, N., & Joinson, C. (2019). The association of children's locus of control orientation and emotion recognition abilities at 8 years of age and teachers' ratings of their personal and social difficulties at 10 years. *Journal of Nonverbal Behavior*, *43*(3), 381–396. https://doi.org/10.1007/s10919-019-00304-3

*Oleszkiewicz, A., Frackowiak, T., Sorokowska, A., & Sorokowski, P. (2017). Children can accurately recognize facial emotions from emoticons. *Computers in Human Behavior*, *76*, 372–377. https://doi.org/10.1016/j.chb.2017.07.040

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *The BMJ*, *372*, Article n71. https://doi.org/10.1136/bmj.n71

*Parker, A. E., Mathis, E. T., & Kupersmidt, J. B. (2013). How is this child feeling? Preschool-aged children's ability to recognize emotion in faces and body poses. *Early Education and Development*, *24*(2), 188–211. https://doi.org/10.1080/10409289.2012.657536

*Petitclerc, A., Henry, J., Feng, B., Poliakova, N., Brendgen, M., Dionne, G., Vitaro, F., Tremblay, R. E., & Boivin, M. (2019). Genetic correlation between child callous-unemotional behaviors and fear recognition deficit: Evidence for a neurocognitive endophenotype. *Journal of Abnormal Child Psychology*, *47*(9), 1483–1493. https://doi.org/10.1007/s10802-019-00529-2

Pollux, P. M. J., Hermens, F., & Willmott, A. P. (2016). Age-congruency and contact effects in body expression recognition from point-light displays (PLD). *PeerJ*, *4*, Article e2796. https://doi.org/10.7717/peerj.2796

Price, G. F., Ogren, M., & Sandhofer, C. M. (2022). Sorting out emotions: How labels influence emotion categorization. *Developmental Psychology*, *58*(9), 1665–1675. https://doi.org/10.1037/dev0001391

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

*Rehder, P. D., Mills-Koonce, W. R., Willoughby, M. T., Garrett-Peters, P., Wagner, N. J., & the Family Life Project Key Investigators. (2017). Emotion recognition deficits among children with conduct problems and callous-unemotional behaviors. *Early Childhood Research Quarterly*, *41*, 174–183. https://doi.org/10.1016/j.ecresq.2017.07.007

*Reichenbach, L., & Masters, J. C. (1983). Children's use of expressive and contextual cues in judgments of emotion. *Child Development*, *54*(4), 993–1004. https://doi.org/10.2307/1129903

*Richoz, A.-R., Lao, J., Pascalis, O., & Caldara, R. (2018). Tracking the recognition of static and dynamic facial expressions of emotion across the life span. *Journal of Vision*, *18*(9), Article 5. https://doi.org/10.1167/18.9.5

Riddell, C., Nikolic, M., Dusseldorp, E., & Kret, M. (2024, January 25). *Age-related changes in emotion recognition across childhood: A meta-analytic review*. https://osf.io/68g7z/?view_only=4b5d5b8dec65463fbbe099ee89e58b37

*Riquelme, E., & Montero, I. (2013). Improving emotional competence through mediated reading: Short term effects of a children's literature program. *Mind, Culture, and Activity*, *20*(3), 226–239. https://doi.org/10.1080/10749039.2013.781185

*Riquelme-Mella, E., & Montero Garcia-Celay, I. (2016). Long-term effects of a mediated reading programme on the development of emotional competencies. *Cultura Y Educacion*, *28*(3), 435–467. https://doi.org/10.1080/11356405.2016.1196900

*Rojahn, J., Singh, N. N., Singh, S. D., Baker, J. A., Lawrence, M. A., & Davis, C. M. (2002). Concurrent validity studies of the Facial Discrimination Task. *Journal of Child and Family Studies*, *11*(2), 203–215. https://doi.org/10.1023/A:1015129726221

*Rosenqvist, J., Lahti-Nuuttila, P., Laasonen, M., & Korkman, M. (2014). Preschoolers' recognition of emotional expressions: Relationships with other neurocognitive capacities. *Child Neuropsychology*, *20*(3), 281–302. https://doi.org/10.1080/09297049.2013.778235

Rosenqvist, J., Lahti-Nuuttila, P., Urgesi, C., Holdnack, J., Kemp, S. L., & Laasonen, M. (2017). Neurocognitive Functions in 3- to 15-Year-Old Children: An International Comparison. *Journal of the International Neuropsychological Society*, *23*(4), 367–380. https://doi.org/10.1017/S1355617716001193

Rosenthal, R., & Rubin, D. B. (1989). Effect size estimation for one-sample multiple-choice-type data: Design, analysis, and meta-analysis. *Psychological Bulletin*, *106*(2), 332–337. https://doi.org/10.1037/0033-2909.106.2.332

*Ross, P. D., Polson, L., & Grosbras, M.-H. (2012). Developmental changes in emotion recognition from full-light and point-light displays of body movement. *PLOS ONE*, *7*(9), Article e44815. https://doi.org/10.1371/journal.pone.0044815

*Roy-Charland, A., Perron, M., Young, C., Boulard, J., & Chamberland, J. A. (2015). The confusion of fear and surprise: A developmental study of the Perceptual-Attentional Limitation Hypothesis using eye movements. *The Journal of Genetic Psychology*, *176*(5), 281–298. https://doi.org/10.1080/00221325.2015.1066301

Ruba, A. L., & Pollak, S. D. (2020). The development of emotion reasoning in infancy and early childhood. *Annual Review of Developmental Psychology*, *2*(1), 503–531. https://doi.org/10.1146/annurev-devpsych-060320-102556

Russell, J. A., & Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology*, *76*(5), 805–819. https://doi.org/10.1037/0022-3514.76.5.805

*Salmon, K., Evans, I. M., Moskowitz, S., Grouden, M., Parkes, F., & Miller, E. (2013). The components of young children's emotion knowledge: Which are enhanced by adult emotion talk? *Social Development*, *22*(1), 94–110. https://doi.org/10.1111/sode.12004

Sauter, D. (2010). More than happy: The need for disentangling positive emotions. *Current Directions in Psychological Science*, *19*(1), 36–40. https://doi.org/10.1177/0963721409359290

*Sauter, D. A., Panattoni, C., & Happé, F. (2013). Children's recognition of emotions from vocal cues. *British Journal of Developmental Psychology*, *31*(1), 97–113. https://doi.org/10.1111/j.2044-835X.2012.02081.x

*Schapira, R., Anger Elfenbein, H., Amichay-Setter, M., Zahn-Waxler, C., & Knafo-Noam, A. (2019). Shared environment effects on children's emotion recognition. *Frontiers in Psychiatry*, *10*, Article 215. https://doi.org/10.3389/fpsyt.2019.00215

*Schermerhorn, A. C. (2019). Associations of child emotion recognition with interparental conflict and shy child temperament traits. *Journal of Social and Personal Relationships*, *36*(4), 1343–1366. https://doi.org/10.1177/0265407518762606

Shablack, H., & Lindquist, K. A. (2019). The role of language in emotional development. In V. LoBue, K. Pérez-Edgar, & K. A. Buss (Eds.), *Handbook of emotional development* (pp. 451–478). Springer. https://doi.org/10.1007/978-3-030-17332-6_18

*Shaw, D. J., Mareček, R., Grosbras, M.-H., Leonard, G., Pike, G. B., & Paus, T. (2016). Co-ordinated structural and functional covariance in the adolescent brain underlies face processing performance. *Social Cognitive and Affective Neuroscience*, *11*(4), 556–568. https://doi.org/10.1093/scan/nsv138

*Shin, N., Krzysik, L., & Vaughn, B. E. (2014). Emotion expressiveness and knowledge in preschool-age children: Age-related changes. *Child Studies in Asia-Pacific Context*, *4*(1), 1–12. https://doi.org/10.5723/csac.2014.4.1.001

Silver, H., & Bilker, W. B. (2015). Colour influences perception of facial emotions but this effect is impaired in healthy ageing and schizophrenia. *Cognitive Neuropsychiatry*, *20*(5), 438–455. https://doi.org/10.1080/13546805.2015.1080157

*Smiecinska, J. (2016). Emotional and linguistic prosody development in Polish children: Three different paths. *Poznán Studies in Contemporary Linguistics*, *52*(3), 519–554. https://doi.org/10.1515/psicl-2016-0020

*Smith, M., & Walden, T. (1998). Developmental trends in emotion understanding among a diverse sample of African-American preschool children. *Journal of Applied Developmental Psychology*, *19*(2), 177–197. https://doi.org/10.1016/S0193-3973(99)80035-5

*Steele, H., Steele, M., & Croft, C. (2008). Early attachment predicts emotion recognition at 6 and 11 years old. *Attachment & Human Development*, *10*(4), 379–393. https://doi.org/10.1080/14616730802461409

*Stetina, B. U., Turner, K., Burger, E., Glenk, L. M., McElheney, J. C., Handlos, U., & Kothgassner, O. D. (2011). Learning emotion recognition from canines? Two for the road. *Journal of Veterinary Behavior*, *6*(2), 108–114. https://doi.org/10.1016/j.jveb.2010.11.004

*Strand, P. S., Cerna, S., & Downs, A. (2008). Shyness and emotion-processing skills in preschoolers: A 6-month longitudinal study. *Infant and Child Development*, *17*(2), 109–120. https://doi.org/10.1002/icd.534

*Strand, P. S., Downs, A., & Barbosa-Leiker, C. (2016). Does facial expression recognition provide a toehold for the development of emotion understanding? *Developmental Psychology*, *52*(8), 1182–1191. https://doi.org/10.1037/dev0000144

*Streubel, B., Gunzenhauser, C., Grosse, G., & Saalbach, H. (2020). Emotion-specific vocabulary and its contribution to emotion understanding in 4- to 9-year-old children. *Journal of Experimental Child Psychology*, *193*, Article 104790. https://doi.org/10.1016/j.jecp.2019.104790

*Szekely, E., Tiemeier, H., Jaddoe, V. W. V., Hofman, A., Verhulst, F. C., & Herba, C. M. (2014). Associations of internalizing and externalizing problems with facial expression recognition in preschoolers: The generation R study. *Social Development*, *23*(3), 611–630. https://doi.org/10.1111/sode.12070

*Theurel, A., Witt, A., Malsert, J., Lejeune, F., Fiorentini, C., Barisnikov, K., & Gentaz, E. (2016). The integration of visual context information in facial emotion recognition in 5- to 15-year-olds. *Journal of Experimental Child Psychology*, *150*, 252–271. https://doi.org/10.1016/j.jecp.2016.06.004

*Thompson, W. F., Schellenberg, E. G., & Husain, G. (2004). Decoding speech prosody: Do music lessons help? *Emotion*, *4*(1), 46–64. https://doi.org/10.1037/1528-3542.4.1.46

Thornton, A., & Lee, P. (2000). Publication bias in meta-analysis: Its causes and consequences. *Journal of Clinical Epidemiology*, *53*(2), 207–216. https://doi.org/10.1016/S0895-4356(99)00161-4

*Tolleson, A. M., Tone, E. B., Schroth, E. A., & Robbins, M. M. (2016). Mother and child facial expression labeling skill relates to mutual responsiveness during emotional conversations. *Journal of Nonverbal Behavior*, *40*(3), 205–219. https://doi.org/10.1007/s10919-016-0232-1

Torres, M. M., Domitrovich, C. E., & Bierman, K. L. (2015). Preschool interpersonal relationships predict kindergarten achievement: Mediated by gains in emotion knowledge. *Journal of Applied Developmental Psychology*, *39*, 44–52. https://doi.org/10.1016/j.appdev.2015.04.008

*Tuminello, E. R., & Davidson, D. (2011). What the face and body reveal: In-group emotion effects and stereotyping of emotion in African American and European American children. *Journal of Experimental Child Psychology*, *110*(2), 258–274. https://doi.org/10.1016/j.jecp.2011.02.016

*Uhls, Y. T., Broome, J., Levi, S., Szczepanski-Beavers, J., & Greenfield, P. (2020). Mobile technologies and their relationship to children's ability to read nonverbal emotional cues: A cross-temporal comparison. *Cyberpsychology, Behavior, and Social Networking*, *23*(7), 465–470. https://doi.org/10.1089/cyber.2019.0174

*Verpaalen, I. A. M., Bijsterbosch, G., Mobach, L., Bijlstra, G., Rinck, M., & Klein, A. M. (2019). Validating the Radboud faces database from a child's perspective. *Cognition and Emotion*, *33*(8), 1531–1547. https://doi.org/10.1080/02699931.2019.1577220

*Vetter, N. C., Drauschke, M., Thieme, J., & Altgassen, M. (2018). Adolescent basic facial emotion recognition is not influenced by puberty or own-age bias. *Frontiers in Psychology*, *9*, Article 956. https://doi.org/10.3389/fpsyg.2018.00956

Vicari, S., Reilly, J. S., Pasqualetti, P., Vizzotto, A., & Caltagirone, C. (2000). Recognition of facial expressions of emotions in school-age children: The intersection of perceptual and semantic categories. *Acta Paediatrica*, *89*(7), 836–845. https://doi.org/10.1080/080352500750043756

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

Vieillard, S., & Guidetti, M. (2009). Children's perception and understanding of (dis)similarities among dynamic bodily/facial expressions of happiness, pleasure, anger, and irritation. *Journal of Experimental Child Psychology*, *102*(1), 78–95. https://doi.org/10.1016/j.jecp.2008.04.005

Wagner, H. (1993). On Measuring performance in category judgement studies of nonverbal behavior. *Journal of Nonverbal Behavior*, *17*(1), 3–28. https://doi.org/10.1007/BF00987006

*Walden, T. A., & Field, T. M. (1982). Discrimination of facial expressions by preschool children. *Child Development*, *53*(5), 1312–1319. https://doi.org/10.2307/1129021

*Wang, Y., Hawk, S. T., Tang, Y., Schlegel, K., & Zou, H. (2019). Characteristics of Emotion Recognition Ability among Primary School Children: Relationships with Peer Status and Friendship Quality. *Child Indicators Research*, *12*(4), 1369–1388. https://doi.org/10.1007/s12187-018-9590-z

*Wang, Y., Zhou, W., Cheng, Y., & Bian, X. (2017). Gaze patterns in auditory-visual perception of emotion by children with hearing aids and hearing children. *Frontiers in Psychology*, *8*, Article 2281. https://doi.org/10.3389/fpsyg.2017.02281

*Wang, Z., Lu, W., Zhang, H., & Surina, A. (2014). Free-labeling facial expressions and emotional situations in children aged 3–7 years: Developmental trajectory and a face inferiority effect. *International Journal of Behavioral Development*, *38*(6), 487–498. https://doi.org/10.1177/0165025414531462

*Watanabe, N., Denham, S. A., Jones, N. M., Kobayashi, T., Bassett, H. H., & Ferrier, D. E. (2019). Working toward cross-cultural adaptation: Preliminary psychometric evaluation of the affect knowledge test in Japanese preschoolers. *Sage Open*, *9*(2), 1–15. https://doi.org/10.1177/2158244019846688

*Watling, D., & Bourne, V. J. (2013). Sex differences in the relationship between children's emotional expression discrimination and their developing hemispheric lateralization. *Developmental Neuropsychology*, *38*(7), 496–506. https://doi.org/10.1080/87565641.2013.826660

*Watling, D., & Damaskinou, N. (2020). Children's facial emotion recognition skills: Longitudinal associations with lateralization for emotion processing. *Child Development*, *91*(2), 366–381. https://doi.org/10.1111/cdev.13188

*Weimer, A. A., Sallquist, J., & Bolnick, R. R. (2012). Young children's emotion comprehension and theory of mind understanding. *Early Education and Development*, *23*(3), 280–301. https://doi.org/10.1080/10409289.2010.517694

*West, M. J., Angwin, A. J., Copland, D. A., Arnott, W. L., & Nelson, N. L. (2020). Cross-modal emotion recognition and autism-like traits in typically developing children. *Journal of Experimental Child Psychology*, *191*, Article 104737. https://doi.org/10.1016/j.jecp.2019.104737

Widen, S. C. (2013). Children's interpretation of facial expressions: The long path from valence-based to specific discrete categories. *Emotion Review*, *5*(1), 72–77. https://doi.org/10.1177/1754073912451492

Widen, S. C. (2017). The development of emotion recognition. In J. M. Fernández-Dols & J. A. Russell (Eds.), *The science of facial expression*, (pp. 297–312). Oxford University Press.

Widen, S. C. (2019). Emotion perception and recognition. In S. Hupp & J. Jewell (Eds.), *The encyclopedia of child and adolescent development* (pp. 1–11). Wiley.

*Widen, S. C., Pochedly, J. T., & Russell, J. A. (2015). The development of emotion concepts: A story superiority effect in older children and adolescents. *Journal of Experimental Child Psychology*, *131*, 186–192. https://doi.org/10.1016/j.jecp.2014.10.009

*Widen, S. C., & Russell, J. A. (2002). Gender and preschoolers' perception of emotion. *Merrill-Palmer Quarterly*, *48*(3), 248–262. https://doi.org/10.1353/mpq.2002.0013

*Widen, S. C., & Russell, J. A. (2003). A closer look at preschoolers' freely produced labels for facial expressions. *Developmental Psychology*, *39*(1), 114–128. https://doi.org/10.1037/0012-1649.39.1.114

*Widen, S. C., & Russell, J. A. (2008a). Children acquire emotion categories gradually. *Cognitive Development*, *23*(2), 291–312. https://doi.org/10.1016/j.cogdev.2008.01.002

*Widen, S. C., & Russell, J. A. (2008b). Children's and adults' understanding of the "disgust face.". *Cognition and Emotion*, *22*(8), 1513–1541. https://doi.org/10.1080/02699930801906744

*Widen, S. C., & Russell, J. A. (2010a). Differentiation in preschooler's categories of emotion. *Emotion*, *10*(5), 651–661. https://doi.org/10.1037/a0019005

*Widen, S. C., & Russell, J. A. (2010b). The "disgust face" conveys anger to children. *Emotion*, *10*(4), 455–466. https://doi.org/10.1037/a0019151

Widen, S. C., & Russell, J. A. (2015). Do dynamic facial expressions convey emotions to children better than do static ones? *Journal of Cognition and Development*, *16*(5), 802–811. https://doi.org/10.1080/15248372.2014.916295

*Wiggers, M., & van Lieshout, C. F. (1985). Development of recognition of emotions: Children's reliance on situational and facial expressive cues. *Developmental Psychology*, *21*(2), 338–349. https://doi.org/10.1037/0012-1649.21.2.338

*Williams, L. M., Mathersul, D., Palmer, D. M., Gur, R. C., Gur, R. E., & Gordon, E. (2009). Explicit identification and implicit recognition of facial emotions: I. Age effects in males and females across 10 decades. *Journal of Clinical and Experimental Neuropsychology*, *31*(3), 257–277. https://doi.org/10.1080/13803390802255635

Witkower, Z., Tracy, J. L., Pun, A., & Baron, A. S. (2021). Can children recognize bodily expressions of emotion? *Journal of Nonverbal Behavior*, *45*(4), 505–518. https://doi.org/10.1007/s10919-021-00368-0

*Woody, M. L., James, K., Foster, C. E., Owens, M., Feurer, C., Kudinova, A. Y., & Gibb, B. E. (2019). Children's sustained attention to emotional facial expressions and their autonomic nervous system reactivity during parent–child interactions. *Biological Psychology*, *142*, 37–44. https://doi.org/10.1016/j.biopsycho.2019.01.005

*Yamamoto, H. W., Kawahara, M., & Tanaka, A. (2020). Audiovisual emotion perception develops differently from audiovisual phoneme perception during childhood. *PLOS ONE*, *15*(6), Article e0234553. https://doi.org/10.1371/journal.pone.0234553

*Zhang, H. (2018). Relations of parent–child interaction to Chinese young children's emotion understanding. *Journal of Pacific Rim Psychology*, *12*, Article e22. https://doi.org/10.1017/prp.2018.11

Received October 19, 2022
Revision received April 15, 2024
Accepted April 16, 2024 ■