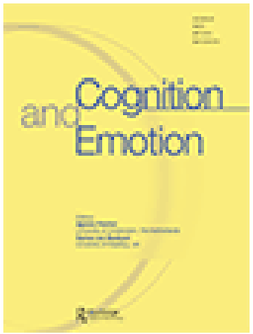

Routledge
Taylor & Francis Group

Cognition and Emotion

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

Through the eyes of a child: preschoolers' identification of emotional expressions from the child affective facial expression (CAFE) set

Vanessa LoBue, Lewis Baker & Cat Thrasher

To cite this article: Vanessa LoBue, Lewis Baker & Cat Thrasher (2017): Through the eyes of a child: preschoolers' identification of emotional expressions from the child affective facial expression (CAFE) set, Cognition and Emotion, DOI: 10.1080/02699931.2017.1365046

To link to this article: http://dx.doi.org/10.1080/02699931.2017.1365046

Published online: 10 Aug 2017.

COGNITION AND EMOTION, 2017
https://doi.org/10.1080/02699931.2017.1365046

Routledge
Taylor & Francis Group

BRIEF ARTICLE

Through the eyes of a child: preschoolers' identification of emotional expressions from the child affective facial expression (CAFE) set

Vanessa LoBue[a], Lewis Baker[a] and Cat Thrasher[b]

[a]Psychology Department, Rutgers University, Newark, NJ, USA; [b]Psychology Department, University of Virginia, Charlottesville, VA, USA

ABSTRACT

Researchers have been interested in the perception of human emotional expressions for decades. Importantly, most empirical work in this domain has relied on controlled stimulus sets of adults posing for various emotional expressions. Recently, the Child Affective Facial Expression (CAFE) set was introduced to the scientific community, featuring a large validated set of photographs of preschool aged children posing for seven different emotional expressions. Although the CAFE set was extensively validated using adult participants, the set was designed for use with children. It is therefore necessary to verify that adult validation applies to child performance. In the current study, we examined 3- to 4-year-olds' identification of a subset of children's faces in the CAFE set, and compared it to adult ratings cited in previous research. Our results demonstrate an exceptionally strong relationship between adult ratings of the CAFE photos and children's ratings, suggesting that the adult validation of the set can be applied to preschool-aged participants. The results are discussed in terms of methodological implications for the use of the CAFE set with children, and theoretical implications for using the set to study the development of emotion perception in early childhood.

ARTICLE HISTORY
Received 16 September 2016
Revised 27 June 2017
Accepted 31 July 2017

KEYWORDS
Faces; emotional facial expressions; children's faces; emotional faces

The perception of human emotions has been an important topic of psychological research for decades. Most studies to date on human emotion perception generally rely on controlled stimulus sets of posed faces. Thus, there are currently dozens of validated facial expression sets available for use in scientific research, creating an easy and controlled way of examining our responses to human facial expressions. However, these sets come with some limitations. First and foremost, most available stimulus sets of emotional facial expressions only capture the expressions of one particular age group—adults. Second, most stimulus sets typically contain only highly iconic, high intensity exemplars.

Based on these limitations, a new stimulus set of emotional facial expressions was recently made available for use in scientific research—The Child Affective Facial Expression (CAFE) set. Unlike all existing stimulus sets, the CAFE set features photographs of 2- to 8-year-old children ($M = 5.3$ years; $R = 2.7$–8.7 years) posing for 6 emotional facial expressions—sadness, happiness, surprise, anger, disgust, and fear—plus a neutral face. The full set features 90 female models and 64 male models (27 African American, 16 Asian, 77 Caucasian/European American, 23 Latino, and 11 South Asian). Although CAFE only includes seven putatively "basic emotions," appealing to a classic discrete emotions perceptive (e.g. Ekman & Friesen, 1978), it also contains natural variation between exemplars to allow researchers from other perspectives to identify faces that are reminiscent of more subtle forms, or faces that are blends of multiple emotional expressions (Coan & Gottman, 2007; Keltner & Buswell, 1996). To allow for such variation, the CAFE set contains 1192 exemplars.

Recently, LoBue and Thrasher (2015) examined adults' accuracy of identifying each face in the CAFE set, and published basic descriptive data on their accuracy of identifying each emotion category across

CONTACT Vanessa LoBue vlobue@psychology.rutgers.edu

the set of child faces. However, although the CAFE set was designed to be used in studies with child participants, there is still no existing data on how accurately *children* identify the faces in the CAFE set. This is important, as basic descriptive statistics on how accurately children are able to identify each category of emotional expression in the CAFE set could be used to create practical recommendations for how the set might be best utilised with child participants.

Given the methodological limitations of studying children, it would be nearly impossible to validate each of the 1192 individual items in the CAFE set with a sample of child participants. However, it is feasible to determine whether the adult validation of CAFE by LoBue and Thrasher (2015) can be applied to children. We can do this by examining children's identification of a subset of the CAFE faces and to then compare the category level (e.g. fear, anger) accuracy scores of children to those of adults. Accordingly, the goals of the current investigation were two-fold. First, we sought to provide descriptive data of children's accuracy of identifying a subset of faces in the CAFE set across each emotion category contained. Second, in order to determine whether the adult validation of the CAFE set can be generalised to child participants, we sought to use adults' accuracy (from LoBue & Thrasher, 2015) to predict accuracy scores collected from our child sample. Together, these data can be used to make practical recommendations about how to use the CAFE set to study emotion perception in children.

Method

In the current research we examined 3- to 4-year-olds' identification of children's faces from the CAFE set. This age group was chosen because it is the youngest age group with significant representation in the set (e.g. there is only a single 2-year-old model). To ensure that participants could accurately and appropriately use the target emotion labels, children were first asked to identify iconic emoticons that each represented one of the 7 basic emotional expressions featured in the CAFE set. Then, children were given a random sample of 50 faces from the CAFE set and were asked to identify the emotion depicted in each.

Participants

LoBue and Thrasher (2015) validated the entire collection of 1192 photos from the CAFE set with 100 adult raters, revealing robust differences in labelling accuracy between emotional categories (Cohen's effect size $f = .814$; Cohen, 1988). To find *a priori* Power (Champely, 2012), we calculated the n-participants necessary to generate differences between 7 group means at the .05 level with an effect size of .814. This analysis revealed that the necessary sample size required to reach Power of .99 based on the adult ratings (who identified all 1192 images) was only seven participants. However, as mentioned above, having a sample of children identify all 1192 images is untenable. Further, given that children are notoriously more variable than adults in their responses, it was reasonable to assume that effect sizes might fall to less than half of adult scores. To be conservative, a power analysis assuming an effect size of one-third of adults (based now on Cohen's effect size $f = .271$) revealed that 55 participants were needed to reach an *a priori* Power of .99.

Thus, a total of 58 children were tested (28M, 30F; $m = 4.0$ years; $r = 3.1$–4.9 years). Children were recruited from preschools in suburban communities within the New York/New Jersey metropolitan area. The Rutgers University IRB approved all procedures, and parents gave informed consent for participation. Based on parent reports of race/ethnicity, approximately 41% of the children were Hispanic, 33% Caucasian, 10% African American, 10% Asian, and 5% Pacific Islander. We ran post-hoc Power analyses to verify our estimates on the collected sample of 58 participants. Regressions revealed a significant difference in accurate labelling by emotional category ($F_{6,\ 336} = 27.4$, $p < 0.001$, $\eta^2 = 0.324$). Using this metric, we calculated an effect size, *f*, of .347 and post-hoc Power > .999.

Materials

The materials were the 1192 photos in the Child Affective Facial Expression Set (CAFE). As mentioned above, CAFE is a validated stimulus set of 2- to 8-year-old children ($M = 5.3$ years; $R = 2.7$–8.7 years) posing for 6 emotional facial expressions—sadness, happiness, surprise, anger, disgust, and fear—plus neutral (see Figure 1). For consistency with other face sets, with the exception of surprise, each expression is depicted with mouths open and with mouths closed, and open-mouthed disgust faces include a tongue protrusion (LoBue & Thrasher, 2015).

Procedure

The study began with an experimenter asking each child to label the seven basic emotions included in

Figure 1. Samples from the child affective facial expression (CAFE) set. *Source*: LoBue and Thrasher (2015).

the CAFE set using a series of schematic emoticons from Gao and Maurer (2010) (see Figure 2). For each emoticon, the child was asked: "Can you tell me what you think this person is feeling?" If the child answered incorrectly, the experimenter provided the child with the correct label. This was done to ensure that the children were able to use the correct emotion labels for each of the target emotion categories. Their responses were recorded.

Next, the experimenter introduced the child to the CAFE faces: "I brought some pictures of children to show you. For each picture, I want you to tell me how the child in the picture might feel." The child was then asked to spontaneously name the emotion in a series of photographs. If children failed to produce an emotion label spontaneously, the experimenter referred the child to the schematic photographs: "Does the child in the picture feel the same way as one of these?" No feedback was given. EPrime was used to randomly select a set of 50 photos from the total of 1192 in the full set, and this procedure was repeated for all 50 photographs. The procedure lasted approximately 30 min. With 58 participants identifying 50 facial expressions each, we collected a total of 2900 data points.

Results

Preliminary analyses

We ran several preliminary analyses to ensure that our methodology was successful in accurately capturing children's emotion perception. First, we ran a set of analyses on the effect of trial number to examine

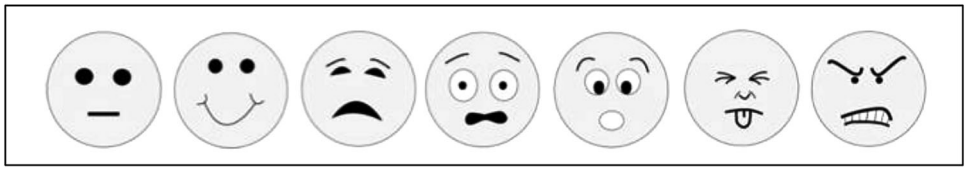

Figure 2. Emoticons from Gao and Maurer (2010).

whether a set of 50 trials exceeded the limits of children's attention. A simple logistic regression model revealed no effect of trial number on children's classification accuracy ($z = -.641$, $p = .521$). An additional model tested the interaction between emotion and trial number, again revealing no significant effect of trial number on classification accuracy ($z = -.776$, $p = .443$), and no interaction with the emotional type of the stimulus (highest $z = 1.433$, $p = .152$).

Second, we examined whether there were any effects of age on categorisation accuracy. A simple linear regression model predicting mean categorisation accuracy from participant age found no significant age effect ($t_{56} = .190$, $p = .850$, $R^2 < .0001$). A second analysis which included emotional category of the stimulus as a within-subjects variable found a significant effect of emotion ($F_{6,336} = 27.47$, $p < .0001$, $\eta^2 = .325$), but no significant effect of age ($F_{1,56} = .025$, $p = .876$, $\eta^2 < .0001$), and no significant interaction ($F_{6,336} = 1.110$, $p = .356$, $\eta^2 = .019$). Variations in age within our sample range (median: 3.925 years; minimum: 3.088; maximum: 4.885 years) did not predict categorisation of the stimuli. Thus, age was not included as a variable in any of the subsequent analyses.

Next, we ran a set of analyses on children who needed to make reference to the emoticons in order to identify the faces in the CAFE set. As mentioned above, if children failed to produce an emotion label spontaneously for each CAFE face, the experimenter referred the child to the schematic photographs and allowed them to make a simple comparison. Interestingly, most of the children tested were able to categorise the CAFE faces verbally without making a single reference to the emoticons ($n = 47$), while others ($n = 11$) needed to refer to the emoticons in every single trial. We ran preliminary analyses examining differences in accuracy scores between children who responded using emoticons and children who never used them. A 2 between (Emoticon Usage) × 7 within (Emotion) mixed design ANOVA predicting mean categorisation accuracy revealed a significant effect of emotion ($F_{6,336} = 27.297$, $p < .0001$, $\eta^2 = .325$), but no effect of emoticon usage ($F_{1,56} = 0.882$, $p = .352$), and no significant interaction ($F_{6,336} = 0.752$, $p = .608$), suggesting no differences between response type on emotional categorisation. Thus, all participants were collapsed for subsequent analyses.

Finally, we examined whether children's accuracy was significantly better if they were first able to accurately identify and label the corresponding emoticons before identifying the faces in the CAFE set. Based on a series of one-way ANOVAs on average accuracy to identify each category of emotion based on whether or not children could accurately identify the corresponding emoticon, children were significantly more accurate at identifying happy faces ($n = 54$; $m = .81$) if they correctly identified the happy emoticon before testing than if they did not ($n = 4$, $m = .53$; $F_{1,57} = 11.0$, $p = 0.002$, $R^2 = 0.164$). Surprisingly, this was not the case for any of the other facial expressions, p's > 0.05. For these target expressions, children's accuracy in identifying the CAFE faces did not differ by whether or not they correctly identified the corresponding emoticon before testing (see Table 1). These findings suggest that for most target

Table 1. Number of children who correctly/incorrectly identified the emoticons for each emotion category, the percent (mean) of CAFE expressions they identified correctly, and the results of a series ANOVAs comparing these means for children who identified the emoticons correctly versus incorrectly for each emotion category.

		N	Mean	*F*	R^2	*p*
Angry	Correct	50	0.64	0.40	0.007	0.531
	Incorrect	8	0.58			
Disgust	Correct	23	0.42	1.70	0.030	0.193
	Incorrect	35	0.54			
Fear	Correct	36	0.30	1.50	0.026	0.223
	Incorrect	22	0.23			
Happy	Correct	54	0.81	11.00	0.164	0.002
	Incorrect	4	0.53			
Neutral	Correct	15	0.43	0.10	0.002	0.752
	Incorrect	43	0.45			
Sad	Correct	49	0.47	0.10	0.002	0.770
	Incorrect	9	0.44			
Surprised	Correct	36	0.65	3.01	0.053	0.083
	Incorrect	22	0.49			

expressions, the ability to identify and use emotion labels for highly iconic emoticons did not necessarily indicate that children would be more accurate at identifying the real faces in the CAFE set.

Classification rates

Accuracy scores were calculated by obtaining the percent of trials in which participants correctly identified each category of facial expressions. The means for each of the seven expressions for both children and adults (from LoBue & Thrasher, 2015) are listed in Table 2. There was substantial variability across the faces, with a mean of 54% (95% CI = [.519,.556]) accuracy across all of the photographs in the set (Cohen's $\kappa = .459$; compare to overall adult accuracy of.66, 95% CI = [.660,.665], and $\kappa = .6036$). Classification rates for all categories are displayed in Table 3. All categories were correctly classified above chance (lowest correct classification rate: fear, $m = .26$, $t_{57} = 4.19$, $p < 001$, $d = .55$). Further, children misclassified angry expressions as disgust significantly more often than can be predicted by chance responding alone ($m = .27$; $t_{57} = 3.66$, $p < 001$, $d = .48$). Likewise, children significantly misclassified fear as surprised ($m = .31$; $t_{57} = 3.65$, $p < .001$, $d = .48$). There were no other significant misclassifications. This pattern of errors is consistent with previous research (e.g. Gao & Maurer, 2010).

A repeated measures ANOVA controlling for subject as a random effect on the average accuracy

Table 2. Mean accuracy scores for identification of each emotion category in the CAFE set by children (current data set) and adults (data from LoBue & Thrasher, 2015).

	Mouth	N	Child Mean	Child SD	Child SEM	Adult Mean	Adult SD	Adult SEM
Angry	Closed	265	0.710	0.453	0.028	0.650	0.250	0.015
	Open	197	0.560	0.497	0.035	0.670	0.203	0.015
	Both	462	0.650	0.478	0.022	0.660	0.232	0.011
Disgust	Closed	228	0.200	0.402	0.027	0.550	0.193	0.013
	Open	238	0.740	0.442	0.029	0.740	0.098	0.006
	Both	466	0.470	0.500	0.023	0.650	0.180	0.008
Fear	Closed	206	0.330	0.470	0.033	0.480	0.199	0.014
	Open	166	0.180	0.386	0.030	0.380	0.145	0.011
	Both	372	0.260	0.440	0.023	0.440	0.184	0.010
Happy	Closed	301	0.890	0.313	0.018	0.930	0.029	0.002
	Open	220	0.630	0.485	0.033	0.750	0.181	0.012
	Both	521	0.780	0.415	0.018	0.850	0.151	0.007
Neutral	Closed	311	0.560	0.497	0.028	0.860	0.089	0.005
	Open	229	0.290	0.454	0.030	0.410	0.229	0.015
	Both	540	0.440	0.497	0.021	0.670	0.278	0.012
Sad	Closed	177	0.580	0.495	0.037	0.750	0.157	0.012
	Open	118	0.360	0.481	0.044	0.450	0.222	0.021
	Both	295	0.490	0.501	0.029	0.630	0.235	0.014
Surprise	Both	244	0.610	0.488	0.031	0.700	0.118	0.008
TOTAL	**Closed**	1488	0.570	0.495	0.013	0.720	0.232	0.006
	Open	1412	0.500	0.500	0.013	0.600	0.230	0.006
	Both	2900	0.540	0.499	0.009	0.660	0.238	0.004

Table 3. Proportion of children's responses by emotion category.

		Children's classification						
		Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise
Emotion displayed	Angry	.65*	.02	.05	.10	.02	.04	.12
	Disgust	.27*	.47*	.04	.05	.04	.07	.05
	Fear	.10	.03	.26*	.13	.09	.08	.31*
	Happy	.02	.01	.02	.78*	.03	.02	.12
	Neutral	.05	.02	.06	.12	.44*	.11	.19
	Sad	.16	.07	.11	.03	.08	.49*	.06
	Surprise	.05	.01	.09	.17	.04	.02	.61*

Note: The correct emotion category is displayed on the left, and the proportion of children who chose each response category is displayed across the top. Correct responses are highlighted in gray. Overall accuracy of.538 (95% CI = [.519,.556]), Cohen's $\kappa = .459$. Means noted by * are above chance ($p < .05$).

scores for each of the 7 target emotions revealed significant differences in accuracy by emotional category ($F_{6,\ 342} = 27.4$, $p < 0.001$, $\eta^2 = 0.230$). Comparisons indicated that children identified happy faces significantly more accurately than all other faces (lowest $t_{342} = 3.64$, $p = .001$ $d = .482$). All other comparisons between emotions were statistically significant with Bonferroni correction at $p < .05$ except for the following: There was no significant difference between the identification of angry versus surprised faces ($t_{342} = 0.924$, $p = 0.178$, $d = .122$); there was no significant difference between the identification of disgust versus sad faces ($t_{342} = 0.332$, $p = 0.370$, $d = .044$) or disgust versus neutral faces ($t_{342} = 0.715$, $p = 0.234$, $d = .095$). Fearful faces were identified significantly less accurately than all other faces (lowest $t_{342} = 4.02$, $p < .001$ $d = .534$). This pattern is similar to what LoBue and Thrasher (2015) reported with adults.

Next, we tested whether mouth position played a role in categorisation accuracy, consistent with previous research (Widen & Naab, 2012). "Surprised" emotional faces only had open-mouthed exemplars and were removed from analysis. Furthermore, since participants saw a random selection of stimuli from the CAFE set, not all participants saw both open and closed exemplars for the remaining emotional categories. These participants ($n = 16$) were removed from analysis. A repeated-measures ANOVA predicting accuracy from emotion and mouth expression found significant main effects of emotion ($F_{5,\ 205} = 25.850$, $p < .0001$, $\eta^2 = .387$) and mouth expression ($F_{1,\ 41} = 16.930$, $p < .0001$, $\eta^2 = .292$), and a significant interaction effect ($F_{5,\ 205} = 28.381$, $p < .0001$, $\eta^2 = .409$). As illustrated in Table 2, subjects demonstrated overall greater accuracy for Closed versus Open faces, with the exception of significantly greater accuracy for Open-Disgust than Closed-Disgust faces.

Comparison to adult ratings

In order to examine how children's identification of the CAFE faces might vary based on the difficulty of the target expressions, we compared children's accuracy in the current study to adults' validity scores reported with the CAFE set (LoBue & Thrasher, 2015). These scores were obtained from data on 100 college-aged adults who identified each emotion in the CAFE set (http://databrary.org/volume/30). Although validity scores are not necessarily a direct measure of iconicity, in this case, they can be used as a measure of *difficulty*. LoBue and Thrasher (2015) applied a one-parameter logistic or *Rasch model* to the adult data to calculate a difficulty score (b_i), along with fit statistics (in-fit and out-fit), for each photograph in order to indicate whether the faces varied substantially within emotional category. Results indicated that almost all of the faces in the set were reliable despite varying degrees of difficulty. Importantly, the difficulty scores (b_i) were highly correlated with the accuracy scores, $r = -0.858$, $p < 0.001$, which are easier to interpret than b_i. Thus, here we will compare children's accuracy scores to those of adults' as an index of difficulty.

First, we measured the relation between child scores and adult scores continuously based on a logistic regression predicting child categorisation accuracy for each stimulus image by adult categorisation accuracy. Adult accuracy significantly predicted child accuracy ($z = 20.93$, $p < .0001$, $OR = 61.45$). This demonstrates an exceptionally strong predictive ability of adult ratings on child ratings and suggests that the adult validation of the CAFE set also can be applied to child participants.

We next measured the relation between child and adults scores categorically by dividing images into five bins based on the average accuracy of 100 *adults'* identification of the same photos in LoBue and Thrasher (2015)—photos that adults identified with 0%–20% accuracy, 20%–40% accuracy, 40%–60% accuracy, 60%–80% accuracy, and 80%–100% accuracy. When examining children's accuracy according to these bins, children's scores again follow a similar pattern to those of adults'. In other words, children were least accurate in the bin where adults were the least accurate (0%–20%) and they were the most accurate in the bin where adults were the most accurate (80%–100%). Table 4 lists children's average accuracy by emotion category for each of these bins.

To examine whether children's scores differed significantly across bins, we ran a mixed effects model on the item-level data within the five bins. The results demonstrated a significant difference between bins, $F_{4,\ 2895} = 148.0$, $p < 0.001$. Post-hoc comparisons (*Tukey-d*) indicated that there were significant differences between accuracy scores in each bin ($p's < 0.001$) with the exception of 0%–20% and 20%–40%, $p = 0.305$ (see Figure 3). In other words, there were significant gains in children's accuracy as adults' accuracy increased from 20%–40% to 40%–60%, from 40%–60% to 60%–80%, and from 60%–80% to 80%–100%. However, children's scores did not increase significantly as adults' accuracy increased from 0%–20% to 20%–40%.

Table 4. Children's accuracy scores (from current data set) by emotion category based on adult accuracy (0%–100%, from LoBue & Thrasher, 2015).

Range of adult accuracy		Child mean accuracy	Number of photos	Std. Dev. of Mean	Std. Error of mean
Angry	0%–20%	.23	26	.430	.084
	20%–40%	.43	53	.500	.069
	40%–60%	.53	64	.503	.063
	60%–80%	.68	146	.469	.039
	80%–100%	.80	173	.403	.031
Disgust	0%–20%	.14	7	.378	.143
	20%–40%	.08	50	.274	.039
	40%–60%	.30	98	.459	.046
	60%–80%	.59	233	.493	.032
	80%–100%	.64	78	.483	.055
Fear	0%–20%	.06	31	.250	.045
	20%–40%	.17	143	.381	.032
	40%–60%	.32	127	.469	.042
	60%–80%	.37	62	.487	.062
	80%–100%	.67	9	.500	.167
Happy	0%–20%	.	0	.	.
	20%–40%	.38	13	.506	.140
	40%–60%	.35	37	.484	.080
	60%–80%	.57	60	.500	.065
	80%–100%	.86	411	.346	.017
Neutral	0%–20%	.12	51	.325	.046
	20%–40%	.15	78	.363	.041
	40%–60%	.45	44	.504	.076
	60%–80%	.49	123	.502	.045
	80%–100%	.58	244	.494	.032
Sad	0%–20%	.06	18	.236	.056
	20%–40%	.24	42	.431	.067
	40%–60%	.38	50	.490	.069
	60%–80%	.57	109	.498	.048
	80–100%	.70	76	.462	.053
Surprise	0%–20%	.	0	.	.
	20%–40%	.	0	.	.
	40%–60%	.45	58	.502	.066
	60%–80%	.68	133	.470	.041
	80%–100%	.64	53	.484	.067

Finally, we ran a series of one-sample t-tests to compare children's accuracy scores in each bin to chance (1/7, or 0.14). The results indicated the children performed at a better rate than chance (p's < 0.001) in all but the lowest bin (0%–20%: $t_{132} = -.696$, $p = 0.488$). In other words, in the lowest bin, where adults were only 0%–20% accurate when identifying individual photographs in the CAFE set, children's responses did not differ from chance.

General discussion

In the current research, we sought to examine preschool children's accuracy in identifying the faces in the Child Affective Facial Expression (CAFE) set. Our goals were to first provide descriptive data of children's accuracy in identifying a subset of faces in the CAFE set across each emotion category contained within the set, and to then determine whether the adult validation of the CAFE set can be generalised to child participants. Our results demonstrate that there is an exceptionally strong relationship between adult ratings of the CAFE photos and children's ratings, suggesting that the adult validation of the set can be applied to preschool-aged participants. Together, these data can be used to make practical

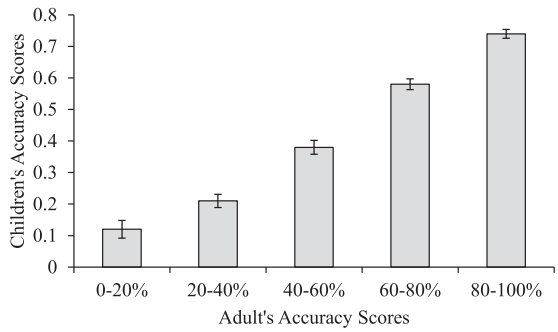

Figure 3. Children's average accuracy scores for individual photographs based on a range of adult accuracy scores on the same photos.

recommendations about how to use the CAFE set to study emotion perception in children.

The first point of consideration for future research using CAFE is that there is substantial variability in children's accuracy scores based both on the emotion category and its difficulty, with trends that mirror data collected from adults (LoBue & Thrasher, 2015). Children were quite accurate (74%) when identifying the easiest and presumably most iconic faces, which adults identified with 80%–100% accuracy. There was a significant decrease in accuracy as adults' accuracy scores decreased from 80%–100% to 60%–80%, from 60%–80% to 40%–60%, and finally from 40%–60% to 20%–40%. There were also significant differences in children's accuracy between emotion categories, identifying the happy faces with the most accuracy, and fear faces with the least. This pattern is similar to what has been reported with adults using the CAFE set (LoBue & Thrasher, 2015). It is possible that the children who posed for the expressions in the CAFE set had the most difficulty posing for fear faces, which is why both children and adults had the most difficulty identifying them. However, high accuracy scores for happy faces and low scores for fear have also been reported in other face sets with adults models (e.g. Tottenham et al., 2009). Thus, an alternative possibility is that fear faces are more difficult to pose and identify for *both* children and adults. Indeed, Gao and Maurer (2010) report a relatively long developmental trajectory for the ability to recognise fear faces, particularly when compared to happy faces, making our results comparable to these reported in previous literature.

A second point that is important to note for future research using CAFE is that preschool-aged children in the current study had substantial difficulty identifying faces in the set that adults recognised with less than 40% accuracy. Although the faces in the lowest (0%–20% correct) adult accuracy range were the only ones that children did not identify more accurately than can be expected by chance, children's accuracy scores the 0%–20% range were not statistically different from the 20%–40% range. Again, children scored well when presented with faces in the 80%–100% adult, so these faces would provide fairly accurate identifications from preschool-aged participants. However, for researchers seeking to draw out more variability in a sample of preschool children, the 60%–80% range might be preferred. It is important to note that here we only tested 3- to 4-year-olds. For older children, we would expect better accuracy scores across the range.

A third point issue worth discussing is that identification of the CAFE faces was not related to the identification of highly iconic emoticons for any of the emotion categories except happiness. Although these emoticons have been used in previous research using a similar procedure (e.g. Gao & Maurer, 2010), it is possible that even highly iconic emoticons are significantly different from actual emotional facial expressions, and thus identification of these emoticons is unrelated to emotional face perception. Future research can examine this relationship more carefully. However, in terms of recommendations for using the CAFE set with children, a pretest with emoticons seems unnecessary, as accuracy in identifying the emoticons was unrelated to accuracy in identifying the CAFE faces.

Finally, given the variability contained within CAFE, the set might be particularly useful for future work examining the developmental trajectory of children's emotion perception with more variable expressions. We already know that by the time they reach middle childhood, children are just as good as adults at identifying the most basic human emotions (Thomas, De Bellis, Graham, & LaBar, 2007). Although this work is suggestive of rapid maturity in children's ability to recognise emotional facial expressions, the methodologies relied on faces that were generally high in emotional intensity, and represent only the most stereotypical of expressions. A handful of recent studies using facial expressions of more varied emotional intensities has shown that while children between the ages of 7 and 10 are highly accurate at identifying high intensity emotional expressions, there is a much longer developmental trajectory for accuracy in identifying lower intensity faces, and that this trajectory differs for different categories of emotion (e.g. happy versus disgusted) (Gao & Maurer, 2010; Thomas et al., 2007).

The overall accuracy across emotion categories for 3- to 4-year-old children in the current study was relatively low (54%), supporting a longer developmental trajectory for faces that contain more variability, and supporting more dimensional models of emotion perception that predict patterns of emotion differentiation that continue into later childhood (e.g. Widen & Russell, 2010). Interestingly, the overall accuracy rate found here is similar to other studies using the same age range with highly iconic adult faces (e.g. Widen & Russell, 2013). This is surprising given the variability contained within the CAFE set. One possible way to explain this finding is that children might be more accurate at identifying faces of individuals that are their own age (Hills & Lewis, 2011). Indeed, the children tested

here were the same age as many of the models in the CAFE set. However, for communicative purposes, preschool-aged children likely have more experience with the facial expressions of adults than of other children, making this possibility seem unlikely. Importantly, now that we have data verifying the usefulness of the CAFE set with child participants, researchers can explore issues like these in future research.

In conclusion, the current research suggests that preschool-aged children can identify the faces in the CAFE set accurately, and that the adult validation of the set by LoBue and Thrasher (2015) can be applied to child participants. Given the interest in emotion perception among researchers across psychological disciplines, we expect that the CAFE set will be immensely useful in future research focusing on emotion perception in children.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by National Science Foundation [grant number 1247590].

References

Champely, S. (2012). pwr: Basic functions for power analysis. R package version 1.1.1. Retrieved from http://CRAN.R-project.org/package = pwr

Coan, J. A., & Gottman, J. M. (2007). The specific affect (SPAFF) coding system. In J. A. Coan, & J. J. B. Allen (Eds.), *Handbook of emotion elicitation and assessment* (pp. 106–123). New York, NY: Oxford University Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Ekman, P., & Friesen, W. V. (1978). *Pictures of facial affect*. Palo Alto, CA: Consulting Psychologists' Press.

Gao, X., & Maurer, D. (2010). A happy story: Developmental changes in children's sensitivity to facial expressions of varying intensities. *Journal of Experimental Child Psychology*, *107*, 67–86.

Hills, P. J., & Lewis, M. (2011). The own-age face recognition bias in children and adults. *The Quarterly Journal of Experimental Psychology*, *64*, 17–23.

Keltner, D., & Buswell, B. N. (1996). Evidence for the distinctness of embarrassment, shame, and guilt: A study of recalled antecedents and facial expressions of emotion. *Cognition and Emotion*, *10*, 155–172.

LoBue, V., & Thrasher, C. (2015). The child affective facial expression (CAFE) set: Validity and reliability from untrained adults. *Emotion Science*, *5*, 1–8.

Thomas, L. A., De Bellis, M. D., Graham, R., & LaBar, K. S. (2007). Development of emotional facial recognition in late childhood and adolescence. *Developmental Science*, *10*, 547–558.

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., … Nelson, C. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research*, *168*, 242–249.

Widen, S. C., & Naab, P. (2012). Can an anger face also be scared? Malleability of facial expressions. *Emotion (Washington, D.C.)*, *12*, 919–925.

Widen, S. C., & Russell, J. A. (2010). Differentiation in preschooler's categories of emotion. *Emotion (Washington, D.C.)*, *10*, 651–661.

Widen, S. C., & Russell, J. A. (2013). Children's recognition of disgust in others. *Psychological Bulletin*, *139*, 271–299.