

Construction of face databases for tasks to recognize facial expressions of basic emotions: a systematic review

Daiene de Moraes Fabrício^{1,2} , Bianca Letícia Cavalmoretti Ferreira^{2,3} , Madson Alan Maximiano-Barreto^{1,2} ,
Monalisa Muniz^{1,4} , Marcos Hortes Nisihara Chagas^{1,2,3,5} 

ABSTRACT. Recognizing the other's emotions is an important skill for the social context that can be modulated by variables such as gender, age, and race. A number of studies seek to elaborate specific face databases to assess the recognition of basic emotions in different contexts. **Objectives:** This systematic review sought to gather these studies, describing and comparing the methodologies used in their elaboration. **Methods:** The databases used to select the articles were the following: *PubMed, Web of Science, PsycInfo, and Scopus*. The following word crossing was used: "*Facial expression database OR Stimulus set AND development OR Validation*." **Results:** A total of 36 articles showed that most of the studies used actors to express the emotions that were elicited from specific situations to generate the most spontaneous emotion possible. The databases were mainly composed of colorful and static stimuli. In addition, most of the studies sought to establish and describe patterns to record the stimuli, such as color of the garments used and background. The psychometric properties of the databases are also described. **Conclusions:** The data presented in this review point to the methodological heterogeneity among the studies. Nevertheless, we describe their patterns, contributing to the planning of new research studies that seek to create databases for new contexts.

Keywords: Facial Expression; Validation Study; Emotions; Facial Recognition; Psychometrics.

CONSTRUÇÃO DE BANCOS DE FACES PARA TAREFAS DE RECONHECIMENTO DE EXPRESSÕES FACIAIS DE EMOÇÕES BÁSICAS: UMA REVISÃO SISTEMÁTICA

RESUMO. Reconhecer as emoções do outro é uma habilidade importante para o contexto social, que pode ser modulada por variáveis como sexo, idade e raça. Vários estudos buscam elaborar bancos de faces específicos para avaliar o reconhecimento de emoções básicas em diferentes contextos. **Objetivos:** Esta revisão sistemática buscou reunir esses estudos, descrevendo e comparando as metodologias utilizadas em sua elaboração. **Métodos:** As bases de dados utilizadas para a seleção dos artigos foram: *PubMed, Web of Science, PsycInfo e Scopus*. Foi utilizado o seguinte cruzamento de palavras: "*facial expression database OR stimulus set AND development OR validation*". **Resultados:** O total de 36 artigos mostrou que a maioria dos estudos utilizou atores para expressar as emoções, que foram suscitadas de situações específicas para serem o mais espontâneas possível. Os bancos de faces foram compostos principalmente de estímulos coloridos e estáticos. Além disso, a maioria dos estudos buscou estabelecer e descrever padrões para registrar os estímulos, como a cor das roupas utilizadas e o fundo. As propriedades psicométricas dos bancos de faces também são descritas. **Conclusões:** Os dados apresentados nesta revisão apontam para a heterogeneidade metodológica entre os estudos. Apesar disso, descrevemos seus padrões, contribuindo para o planejamento de novas pesquisas que buscam criar bancos de faces específicos para novos contextos.

Palavras-chave: Expressão Facial; Estudo de Validação; Emoções; Reconhecimento Facial; Psicometria.

This study was conducted by the Study and Research Group on Mental Health, Cognition and Aging – ProVive, Universidade Federal de São Carlos, São Carlos, SP, Brazil.

¹Universidade Federal de São Carlos, Departamento de Psicologia, São Carlos SP, Brazil.

²Universidade Federal de São Carlos, Grupo de Pesquisa em Saúde Mental, Cognição e Envelhecimento, São Carlos SP, Brazil.

³Universidade de São Paulo, Departamento de Neurociências e do Comportamento, Ribeirão Preto SP, Brazil.

⁴Universidade Federal de São Carlos, Grupo de Pesquisa de Inteligência Emocional, São Carlos SP, Brazil.

⁵Instituto Bairral de Psiquiatria, Itapira SP, Brazil.

Correspondence: Daiene de Moraes; Email: daienemoraes1@gmail.com.

Disclosure: The authors report no conflicts of interest.

Funding: This study was financed in part by the Brazilian fostering agencies: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES [Coordination for the Advancement of Higher Education Personnel]), finance code 001). DMF is a recipient of a scholarship from the CAPES (grant: # 88887.338752/2019-00) and MAMB is a recipient of a scholarship from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP [State of São Paulo Research Assistance Foundation]), process: 20/04936-4).

Received on April 27, 2022; Received in its final form on August 01, 2022; Accepted on August 23, 2022.



INTRODUCTION

Emotions play an important role in society life, as they enable interaction among people. According to the evolutionary theories, all emotions derive from a set of basic emotions common to both humans and animals and which are genetically determined^{1,2}. One of the ways for us to recognize the other's emotion is through facial expressions, since the face is one of the most expressive visual stimuli in society life³. The ability to recognize emotions through the face can already be perceived in newborns, a fact that justifies the innate nature of this skill⁴.

From a study using a systematized task, Ekman and Friesen⁵ postulated six basic emotions, which are related to evolutionary adaptations and can be universally recognized, namely, happiness, sadness, fear, disgust, surprise, and anger. In addition, they identified that the cultural aspects did not modulate the way in which these emotions were expressed⁵. Thus, the evidence indicated that all human beings had the same movements of the facial muscles under certain circumstances^{6,7}, turning the ability to express emotions into a behavioral phenotype.

However, a number of studies began to notice that, within this phenotype common to human beings, some variables could modulate the way to recognize these facial expressions, such as cultural context⁸, age⁹, gender¹⁰, and race¹¹. Taking these variables into account, several studies started to construct and validate specific face databases to assess the ability to recognize emotions through facial expressions¹²⁻¹⁶ since, when selecting a set of facial expression stimuli, it is necessary to consider characteristics of the model that are expressing the emotions, as well as who will recognize them.

Therefore, the existing facial expression databases present great diversity with regard to the physical characteristics of those who express the emotions, the way in which emotions are induced during the construction of the image database, and how they are presented in the validation stage¹²⁻¹⁴. Despite the methodological differences across the studies, they follow important standards for the construction and validation of the series of stimuli. Comparing the methodology used by the studies in the creation of these databases, regardless of the characteristics of who expresses the stimuli, can contribute to the planning of new research studies that seek to create face databases for new contexts. Thus, the objective of this systematic review was to gather studies that constructed face databases to assess the recognition of facial expressions of basic emotions, describing and comparing the methodologies used in the stimuli construction phase.

METHODS

Search strategies and eligibility criteria

The search strategy for this systematic review was created and implemented prior to study selection, in accordance with the checklist presented in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)¹⁷. The databases used to select the articles were the following: *PubMed*, *Web of Science*, *PsycInfo*, and *Scopus*. The following word crossing was used: "Facial expression database OR Stimulus set AND development OR Validation." The searches were conducted from June to December 8, 2021.

The lists of references of the selected articles were also researched for additional sources. The inclusion criteria were surveys that constructed face databases to assess the recognition of basic emotions, published in original articles or disclosed on official websites, without language or time restrictions. Letters to the editor, books and book chapters, reviews, comments, notes, errata, theses, dissertations, and bibliographic/systematic reviews were excluded. In addition, it is worth noting that only the construction stage of the databases was included in this review.

Therefore, additional studies conducted after construction, such as normative data, were not contemplated in the analysis.

Study selection

All the articles found in the databases were saved in the *Rayyan* electronic reference manager. After removing duplicate articles and according to the inclusion criteria of this study, all articles were evaluated by two independent researchers (DF and BF) through their titles and abstracts. In this stage, the researchers classified the articles as "yes," "no," or "perhaps." Subsequently, the researchers reached consensus as to whether the articles recorded as "perhaps" should be included in the review.

After the inclusion of these studies, three researchers (DM, BF, and MB) read the articles in full and extracted information such as year of publication and study locus, name of the database built, characteristics of the participants who expressed the emotions (number of participants, place of recruitment, gender, age and race), basic emotions expressed, and final total of stimuli included in the database and their specific characteristics (Table 1)^{12-16,25-63}. Subsequently, the methodological characteristics of the databases were collected, such as the method used to elicit the emotions, patterns in the capture of stimuli, criteria used in the validation stage, sample characteristics in the validation stage, and psychometric qualities assessed (Table 2)^{12-16,25-63}.

Table 1. General characteristics of face databases.

Authors and year of publication	Study location	Name of the built database	Theoretical reference	Characteristics of participants who expressed emotions	Basic emotions expressed	Total of stimuli	Specific characteristics of stimuli
Benda and Scherf. (2020) ²⁵	The United States	Complex Emotion Expression Database (CEED)	Recognition of complex emotions in young people (Empirical)	8 professional actors - Age: 20.9 years; SD=3.1 - Sex: M=4; F=4 - Race: Caucasians (n=4) and Black (n=4)	1) Happiness 2) Sadness 3) Fear 4) Disgust 5) Anger 6) Surprise	243 images	- Black and white - Static
Chung et al. (2019) ²⁶	South Korea	Yonsei Face Database (YFace DB)	Basic emotions (Ekman and Friesen, 1975) ⁵⁶	74 local community and university volunteers - Age: 19-40 years - Sex: M=37; F=37 - Race: Koreans	1) Happiness 2) Sadness 3) Fear 4) Disgust 5) Anger 6) Surprise 7) Neutral	1,480 stimuli	- Colorful - Static and dynamic *Open and closed mouth *Varied intensities
Conley et al. (2018) ¹⁶	The United States	The racially diverse affective expression (RADIATE)	Racial heterogeneity in emotion recognition (Empirical)	109 community adults - Age: 18-30 years - Sex: M=53; F=56 - Race: Asian (n=22), Black/African-Americans (n=38), Caucasians (n=28), Hispanic (n=20) and others (n=1)	1) Happiness 2) Sadness 3) Fear 4) Disgust 5) Anger 6) Surprise 7) Neutral	1,721 images	- Colorful and black and white - Static *Open and closed mouth
Dalrymple et al. (2013) ²⁷	The United States	The Dartmouth Database of Children's Faces	Recognition of emotions in children (Empirical)	80 community children - Age: 9.84 years; SD=2.33 - Sex: M=40; F=40 - Race: Caucasians	1) Happiness 2) Sadness 3) Fear 4) Disgust 5) Anger 6) Surprise 7) Neutral	964 images	- Colorful - Static *Happiness with closed mouth and happiness showing teeth
Donadon et al. (2019) ²⁸	Brazil	Baby Faces	Ekman's Neurocultural Theory (1972) ⁵⁷	20 babies - Age: 9 months; SD=1.5 - Sex: M=10; F=10 - Race: Caucasians (n=66%), Black (n=17%), and Japanese (n=17%)	1) Happiness 2) Sadness 3) Fear 4) Anger 5) Surprise 6) Neutral	57 images	- Colorful - Static
Ebner et al. (2010) ¹³	Germany	Faces--a life-span Database of Facial Expressions	Age differences in emotion recognition (Ruffman et al., 2008) ⁵⁸	179 actors and extras recruited from a modeling agency - 61 young (24.3 years; SD=3.5) - 60 middle-age (49.0 years; SD=3.9) - 58 elderly (73.2 years; SD=2.8) - Sex: M=86; F=85 - Race: Caucasians (n=179)	1) Happiness 2) Sadness 3) Fear 4) Disgust 5) Anger 6) Neutral	2,052 images	- Colorful - Static
Egger et al. (2011) ²⁹	The United States	NIMH Child Emotional Faces Picture Set (NIMH-ChEFS)	Recognition of emotions in children (Empirical)	59 child actors - Age: 13.6 years - Sex: M=20; F=39 - Race: ND	1) Happiness 2) Sadness 3) Fear 4) Anger 5) Neutral	482 images	- Colorful - Static *Two directions of gazing: direct and avoided

Continue...

Table 1. Continuation.

Authors and year of publication	Study location	Name of the built database	Theoretical reference	Characteristics of participants who expressed emotions	Basic emotions expressed	Total of stimuli	Specific characteristics of stimuli
Ekman and Friesen. (1976) ³⁰	The United States	Pictures of Facial Affect (POFA)	Pan-cultural elements in facial expressions of emotions (Ekman et al., 1969) ⁵	10 individuals - Age: ND - Sex: M=4; F=6 - Race: Caucasians and African-American	1) Happiness 2) Sadness 3) Fear 4) Disgust 5) Anger 6) Surprise 7) Neutral	110 images	- Black and white - Static
Fujimura and Umemura. (2018) ³¹	Japan	A facial expression database based on the dimensional and categorical model of emotions	The influence of angles on emotion recognition (Borod et al., 1998) ⁵⁹	8 professional actors - Age: 34.25 years; SD=5.47 - Sex: M=4; F=4 - Race: Japanese	1) Happiness 2) Sadness 3) Fear 4) Disgust 5) Anger 6) Surprise 7) Neutral	920 stimuli	- Colorful - Static and dynamic *Open and closed mouth *Varied angles
Franz et al. (2021) ³²	Germany	Picture-Set of Young Children's Affective Facial Expressions (PSYCAFE)	Recognition of emotions in children (Empirical)	35 children - Age: 4-6 years - Sex: M=14; F=21 - Race: ND	1) Happiness 2) Sadness 3) Fear 4) Disgust 5) Anger 6) Surprise 7) Neutral	104 images	- Colorful - Static *Varied intensities
Garrido et al. (2017) ³³	Portugal	Stills and Videos of facial Expressions (SAVE database)	Recognition of emotions in dynamic stimuli (Empirical)	20 students - Age: 21.75 years; SD=1.97 - Sex: M=12; F=8 - Race: ND	1) Happiness 2) Neutral	120 stimuli	- Colorful - Static and dynamic
Giuliani et al. (2017) ¹⁵	The United States	The DuckEES child and adolescent dynamic facial expressions stimulus set	Recognition of emotions in dynamic stimuli (Empirical)	37 children and teenage actors - Age: 13.24 years; SD=2.09 - Sex: M=15; F=22 - Race: Caucasians (n=89%)	1) Happiness 2) Sadness 3) Fear 4) Disgust 5) Neutral	120 videos	- Colorful - Dynamic
Happy et al. (2015) ³⁴	India	The Indian Spontaneous Expression Database for Emotion Recognition (ISED)	Basic emotions (Ekman and Friesen, 1975) ⁵⁶	50 individuals - Age: 18-22 years - Sex: M=29; F=21 - Race: Indians	1) Happiness 2) Sadness 3) Disgust 4) Surprise	428 videos	- Colorful - Dynamic *Varied intensities
Kaulard et al. (2012) ³⁵	Germany	The MPI Facial Expression Database	Language and emotions (Empirical)	19 native Germans without professional acting experience - Age: 20-30 years - Sex: M=9; F=10 - Race: Caucasians	1) Happiness 2) Sadness 3) Fear 4) Disgust 5) Anger	18800 videos	- Colorful - Dynamic *Varied angles

Continue...

Table 1. Continuation.

Authors and year of publication	Study location	Name of the built database	Theoretical reference	Characteristics of participants who expressed emotions	Basic emotions expressed	Total of stimuli	Specific characteristics of stimuli
Keutmann et al. (2015) ³⁶	The United States	Visual and vocal emotional expressions of adult and child actors	Item Response Theory in face database construction (Empirical)	150 actors (Adults: n=139 and kids: n=11) - Age: 36.1 years; SD=15.6 - Sex: M=73; F=77 - Race: Caucasians (n=98), African-American (n=35), Hawaiian (n=1), mixed (n=1), and others (n=1)	1) Happiness 2) Sadness 3) Fear 4) Anger 5) Neutral	152 stimuli	- Colorful - Static and dynamic *Varied intensities
Kim et al. (2017) ³⁷	South Korea	Korea University Facial Expression Collection – Second Edition (KUFEC-II)	The role of culture in recognizing emotions (Empirical)	57 actors - Age: ND - Sex: M=32; F=36 - Race: Koreans	1) Happiness 2) Sadness 3) Fear 4) Disgust 5) Anger 6) Surprise 7) Neutral	399 images	- Colorful - Static
Langner et al. (2010) ³⁸	Netherlands	Radboud Faces Database	The influence of angles and direction of gaze on emotion recognition (Empirical)	49 young and children Young: 39 Children: 10 - Age: ND - Sex: M=24; F=25 - Race: Caucasians (n=49)	1) Happiness 2) Sadness 3) Fear 4) Disgust 5) Anger 6) Surprise 7) Neutral	5,880 images	- Colorful - Static *Three directions of gaze: front, right, and left *Varied face angles
LoBue and Thrasher. (2015) ¹⁴	The United States	The Child Affective Facial Expression (CAFE)	Recognition of emotions in children's faces of different races (Empirical)	154 children - Age: 5.3 years - Sex: M=64; F=90 - Race: African-Americans (n=27), Caucasians (n=77), Asians (n=16), Latinos (n=23), and South Asia (n=11)	1) Happiness 2) Sadness 3) Fear 4) Disgust 5) Anger 6) Surprise 7) Neutral	1,192 images	- Colorful - Static *Open and closed mouth
Lundqvist et al. (1998) ³⁹	Sweden	Karolinska Directed Emotional Faces (KDEF) Database	-	70 actors - Age: 25 years (20-30 years) - Sex: M=35; F=35 - Race: ND	1) Happiness 2) Sadness 3) Fear 4) Disgust 5) Anger 6) Surprise 7) Neutral	490 images	- Colorful - Static *Varied face angles
Ma et al. (2020) ⁴⁰	China	Han, Hui, and Tibetan Chinese facial expression database	The role of culture in recognizing emotions (Empirical)	630 volunteers - Age: Han (22 years; SD=2.7); Hui (22.8 years; SD=2.4); and Tibet (21.4 years; SD=2.5) - Sex: M=315; F=315 - Race: Chinese from different regions	1) Happiness 2) Sadness 3) Fear 4) Disgust 5) Anger 6) Surprise 7) Neutral	930 images	- Colorful - Static
Ma et al. (2015) ⁴¹	The United States	Chicago Face Database (CFD)	Limitations of existing face databases (Empirical)	158 individuals from the University of Chicago Laboratory and amateur actors - Age: 13.6 years - Sex: M=73; F=85 - Race: Black (n=85) and Caucasians (n=73)	1) Happiness 2) Fear 3) Neutral	158 images	- Colorful - Static *Two directions of gaze: direct and averted

Continue...

Table 1. Continuation.

Authors and year of publication	Study location	Name of the built database	Theoretical reference	Characteristics of participants who expressed emotions	Basic emotions expressed	Total of stimuli	Specific characteristics of stimuli
Maack et al. (2017) ⁴²	Norway	The Tromsø Infant Faces Database (TIF)	Influence of child stimuli on the adult attention system (Brosch et al., 2007; Parsons et al., 2011; Borgi et al., 2014) ⁶⁰⁻⁶²	18 babies - Age: 4-12 months - Sex: M=8; F=10 - Race: Caucasians	1) Happiness 2) Sadness 3) Fear 4) Disgust 5) Anger 6) Surprise 7) Neutral	119 images	- Colorful - Static
Meuwissen et al. (2017) ⁴³	The United States	Developmental Emotional Faces Stimulus Set (DEFSS)	Limitations of existing face databases (Empirical)	116 volunteers 42 children 44 teenagers 30 adults - Age: ND - Sex: M=43; F=73 - Race: White (n=102), non-White (n=15)	1) Happiness 2) Sadness 3) Fear 4) Anger 5) Neutral	404 images	- Colorful - Static
Miner and Park. (2004) ⁴⁴	The United States	A lifespan database of adult facial stimuli	Influence of age on emotion recognition (Empirical)	576 community volunteers - Age: 18-93 years - Sex: M=219; F=357 - Race: Caucasians (n=435), African-American (n=89), and others (n=52)	1) Happiness 2) Neutral	1,142 images	- Colorful - Static
Negrão et al. (2021) ⁴⁵	Brazil	The Child Emotion Facial Expression Set	Recognition of emotions in children (Empirical)	132 children - Age: 4-6 years - Sex: M=42%; F=58% - Race: Caucasian (n=71%), African (n=24%), Asian (5%)	1) Happiness 2) Sadness 3) Fear 4) Disgust 5) Anger 6) Surprise 7) Neutral	971 stimuli	- Colorful - Static and dynamic
Novello et al. (2018) ⁴⁶	Brazil	Youth Emotion Picture Set	Recognition of Facial Emotions in Teens (Empirical)	31 randomly selected volunteers - Age: 17.4 years; SD=2.7 - Sex: M=14; F=17 - Race: Caucasians (n=27), Blacks (n=1), and mixed (n=3)	1) Happiness 2) Sadness 3) Fear 4) Disgust 5) Anger 6) Surprise 7) Neutral	42 images	- Black and white - Static
O'Reilly et al. (2016) ⁴⁷	The United Kingdom	The EU-Emotion Stimulus Set	Limitations of existing face databases (Empirical)	19 actors - Age: 10-70 years - Sex: M=9; F=10 - Race: Caucasians (n=13), Afro-Caribbean/British-Asian (n=2), Blacks (n=2), mixed white/Asian (n=1), Mediterranean/Asian-British (n=1)	1) Happiness 2) Sadness 3) Fear 4) Disgust 5) Anger 6) Surprise 7) Neutral	249 videos	- Colorful - Dynamic
Olszanowski et al. (2015) ⁴⁸	Poland	Warsaw set of emotional facial expression pictures (WSEFEP)	Limitations of existing face databases (Empirical)	30 professional actors - Age: 20-30 years - Sex: M=14; F=16 - Race: Polish	1) Happiness 2) Sadness 3) Fear 4) Anger 5) Surprise 6) Neutral	210 images	- Colorful - Static

Continue...

Table 1. Continuation.

Authors and year of publication	Study location	Name of the built database	Theoretical reference	Characteristics of participants who expressed emotions	Basic emotions expressed	Total of stimuli	Specific characteristics of stimuli
Passarelli et al. (2018) ⁴⁹	Italy	Facial Expression Recognition Test (FERT)	Basic emotions (Ekman e Friesen, 1975) ⁵⁶ and Item Response Theory (Reise and Revicki, 2014) ⁶³	6 professional actors - Age: ND - Sex: M=3; F=3 - Race: ND	1) Happiness 2) Sadness 3) Fear 4) Disgust 5) Anger 6) Surprise 7) Neutral	42 images	- Colorful - Static
Romani-Sponchiado et al. (2015) ⁵⁰	Brazil	Child Emotions Picture Set	Recognition of facial emotions in children (Empirical)	18 children - Age: 6-7 years (6.93 years; SD=0.3); 8-9 years (9.12 years; SD=0.57), and 10-11 years (10.72 years; SD=0.61) - Sex: M=9; F=9 - Race: Caucasians (n=14), African-American (n=3), and Indigenous (n=1)	1) Happiness 2) Sadness 3) Fear 4) Disgust 5) Anger 6) Surprise 7) Neutral	225 images	- Black and white - Static *Varied intensities
Samuelsson et al. (2012) ⁵¹	Sweden	Umeå University Database of Facial Expressions	Limitations of existing face databases (Empirical)	60 community individuals - Age: 17-67 years (30.19 years; SD=10.66) - Sex: M=30; F=30 - Race: Swedes, Central Europe, Arabs, and Asians	1) Happiness 2) Sadness 3) Fear 4) Disgust 5) Anger 6) Surprise 7) Neutral	424 images	- Colorful - Static
Sharma and Bhushan. (2019) ⁵²	India	Indian Affective Picture	Basic emotions (Ekman and Friesen, 1975) ⁵⁶ and limitations of existing face databases (Empirical)	4 professional actors - Age: 25.25 years; SD=3.77 - Sex: M=2; F=2 - Race: Indians	1) Happiness 2) Sadness 3) Fear 4) Disgust 5) Anger 6) Surprise 7) Neutral	140 images	- Colorful - Static *Varied face angles
Tottenham et al. (2009) ¹²	The United States	The NimStim set of facial expressions	Basic emotions (Ekman and Friesen, 1975) ⁵⁶ and limitations of existing face databases (Empirical)	43 professional actors - Age: 21-30 years - Sex: M=25; F=18 - Race: Africans, Europeans, and Latin Americans	1) Happiness 2) Sadness 3) Fear 4) Disgust 5) Anger 6) Surprise 7) Neutral	672 images	- Colorful - Static *Open and closed mouth
Tracy et al. (2009) ⁵³	Canada	Universidade da Califórnia, Davis, Set of Emotion Expressions (UCDS)	Basic emotions (Ekman and Friesen, 1975) ⁵⁶ and limitations of existing face databases (Empirical)	28 community individuals - Age: 27.0 years - Sex: M=14; F=14 - Race: White and African	1) Happiness 2) Sadness 3) Fear 4) Disgust 5) Anger 6) Surprise	73 images	- Colorful - Static

Continue...

Table 1. Continuation.

Authors and year of publication	Study location	Name of the built database	Theoretical reference	Characteristics of participants who expressed emotions	Basic emotions expressed	Total of stimuli	Specific characteristics of stimuli
Vaiman et al. (2017) ⁵⁴	Argentina	Expresiones de Emociones Faciales (FACS)	The role of culture in recognizing emotions (Empirical)	14 Argentines from the community - Age: 25.53 years; SD=8.72 - Sex: M=8; F=6 - Race: ND	1) Happiness 2) Sadness 3) Fear 4) Disgust 5) Anger 6) Surprise 7) Neutral	60 images	- Colorful - Static
Yang et al. (2020) ⁵⁵	China	Tsinghua facial expression database	The role of culture in recognizing emotions (Empirical)	63 young and 47 elderly Chinese natives with an interest in acting Young - Age: 23.82 years; SD=4.18 - Sex: M=32; F=31 - Race: Chinese Elderly - Age: 64.40 years; SD=3.51 - Sex: M=21; F=26 - Race: Chinese	1) Happiness 2) Sadness 3) Fear 4) Disgust 5) Anger 6) Surprise 7) Neutral	880 images	- Colorful - Static

ND: not declared; M: male; F: female; SD: standard deviation. *Additional features of the face database.

Risk of bias

The studies selected in this review are for the construction of face databases. In this sense, the traditional risk of bias tools used in randomized and nonrandomized studies is not applicable. The task elaborated by the studies must offer valid and interpretable data for the assessment of facial recognition of basic emotions of individuals in certain contexts. Therefore, the quality of the studies included can be observed based on the analyses performed for the reliability and validity of the databases elaborated^{18,19}.

Data analysis

We analyzed the psychometric properties assessed by the studies in the stage for the validation of the stimuli (Table 2)^{64,65}. This information is important to assess the quality of the database that was elaborated. Qualitatively, we followed the standards for educational and psychological testing of the American Educational Research Association²⁰ and the stages specified in Resolution 09-2018 of the Brazilian Federal Council of Psychology²¹, which regulates the dimensions necessary for the assessment of psychological tests. Consequently, information based on the analysis of the database items and the measures for validity evidence were obtained (Table 2).

In addition, we sought to identify in Table 2 when the psychometric measure assessed by the studies presented satisfactory indexes. For accuracy, as a reference standard we used the consensus among most of the studies on the construction of face databases that include stimuli with recognition rates $\geq 70\%$. In some cases, the studies established other rates for recognition, which were indicated as symbols in the table.

Since accuracy is a fundamental indicator for stimuli selection and has been widely used as a quality parameter for construction studies, this variable is included in the table as an indicator of both precision and content-based validity evidence, since it is a precision measure that was used to validate the database content. For agreement among the evaluators, the studies generally use Cohen's or Fleiss' kappa indexes. Therefore, we used value $\geq 60\%$ as a reference^{22,23}. For internal consistency, we used Cronbach's alpha value > 0.70 as a reference²⁴.

RESULTS

Selection and presentation of the studies

Figure 1 presents the search and selection process for the 36 articles included in this systematic review^{12-17,25-63}.

Table 2. Methodological characteristics used in the studies to create the databases.

Authors and year of publication	Name of the database elaborated	Method used to elicit the emotions	Patterns in stimulus capture	Criteria used in the validation stage for inclusion of stimuli in the final database	Sample characteristics in the stage for the validation of the stimuli	Psychometric properties assessed
Benda and Scherf. (2020) ²⁵	Complex Emotion Expression Database (CEED)	1) Presentation of an equivalent photograph expressing the emotion 2) Emotions elicited from specific situations	- Background: White - Clothes: ND - Distractors removed: ND	Accuracy ≥50%	796 volunteers recruited through MTurk - Age: 34. years; SD=11.6 - Gender: M=403; F=388 - Race: ND	- Analysis of the items: Accuracy* - Validity evidence: Content-based: Accuracy and error in each item
Chung et al. (2019) ²⁶	Yonsei Face Database (YFace DB)	1) Presentation of an equivalent photograph expressing the emotion 2) Instruction on muscle movement of the emotions based on the FACS 3) Emotions elicited from specific situations	- Background: White - Clothes: Black T-shirt - Distractors removed: Beards, glasses, makeup, and bangs	Accuracy, intensity, and naturalness	212 students from the Seoul University - Age: 18-28 years - Gender: M=97; F=115 - Race: ND	- Analysis of the items: Accuracy [†] - Precision: Accuracy - Validity evidence: Content-based: Accuracy Based on the relationship with other variables: ANOVA for difference in precision between genders of the stimuli and evaluators, t-test for difference in mean accuracy between genders and emotions, and post-hoc Bonferroni analysis for items with significant differences [‡]
Conley et al. (2018) ¹⁶	The racially diverse affective expression (RADIATE)	Presentation of an equivalent photograph expressing the emotion	- Background: White - Clothes: White sheet - Distractors removed: Glasses, headband, hats	Accuracy and Cohen's kappa	662 participants recruited through MTurk - Age: 18-35 years (27.6 years; SD=3.8) - Gender: M=402; F=260 - Race: Asian (n=48), Black/African-American (n=70), Caucasian (n=470), Hispanic (n=63), and others (n=11)	- Precision: Reliability (test-retest) [†] - Validity evidence: Content-based: Accuracy [†] ; Cohen's kappa [†] and variability in precision by race of the model
Dalrymple et al. (2013) ²⁷	The Dartmouth Database of Children's Faces	Emotions elicited from specific situations	- Background: Black - Clothes: Black dresses and black hats - Distractors removed: Glasses and jewelry	Images recognized with ≥70% accuracy	163 students and members of the Dartmouth College academic community - Age: 19.6 years; SD=4.15 - Gender: M=67; F=96 - Race: ND	- Precision: Accuracy [†] and Cohen's kappa among the evaluators [†] - Validity evidence: Content-based: Accuracy and Cohen's kappa among the evaluators Based on the relationship with other variables: ANOVA for difference in precision between gender of the stimuli and evaluators [‡]

Continue...

Table 2. Continuation.

Authors and year of publication	Name of the database elaborated	Method used to elicit the emotions	Patterns in stimulus capture	Criteria used in the validation stage for inclusion of stimuli in the final database	Sample characteristics in the stage for the validation of the stimuli	Psychometric properties assessed
Donadon et al. (2019) ²⁸	Baby Faces	The parents were instructed and trained to provoke the intended emotions	ND	Rasch model to minimize floor and ceiling effects with values from 0.50 to 1.50 Rate of correct answers according to Kringelbach et al. 2008 ⁶⁴	Validation 119 volunteers from the community - Age: 36 years; SD=12.8 - Gender: M=36.1%; F=63.9% - Race: Caucasian (n=69.7%), Black (n=26.1%), and Japanese (n=4.2%) Retest 31 volunteers from the community - Age: 38.06 years; SD=11.57 - Gender: M=35.5%; F=64.5% - Race: Caucasian (n=74%), Black (n=19.5%), and Japanese (n=6.5%)	- Analysis of the items: Adjustment and difficulty of the items by the Rasch model - Precision: Reliability (test-retest) [†] - Validity evidence: Content-based: Accuracy [§] Based on the relationship with other variables: ANCOVA to assess the differences between groups considering the sociodemographic variables (gender, race, schooling level of the adults, and gender and race of the faces in the stimulus) [†]
Ebner et al. (2010) ¹³	Faces--a life-span Database of Facial Expressions	1) Emotion induction through photographs and videos 2) Emotions elicited from specific situations	- Background: Gray - Clothes: Gray T-shirt - Distractors removed: Jewelry, glasses, and makeup	Agreement among evaluators for (1) purity of the facial expression and (2) high intensity facial expression	154 students - Age: 20-81 years - Gender: M=78; F=76 - Race: Caucasian	- Precision: Accuracy [†] and consensus among the evaluators [†] - Validity evidence: Content-based: Accuracy and consensus among the evaluators Based on the relationship with other variables: ANOVA for face age × evaluator's age × emotion expressed [‡]
Egger et al. (2011) ²⁹	NIMH Child Emotional Faces Picture Set (NIMH-ChEFS)		- Background: Gray - Clothes: ND - Distractors removed: ND	The cutoff point for the image to be included was that ≥15 evaluators identified the intended emotion	20 professors and employees of the Duke University Medical Center - Age: 38.3 years - Gender: M=7; F=13 - Race: ND	- Analysis of the items: Accuracy [†] - Difficulty of the items: Intensity and representativeness scores - Precision: Agreement among the evaluators ^{//} - Validity evidence: Content-based: Accuracy and agreement among the evaluators
Ekman and Friesen. (1976) ³⁰	Pictures of Facial Affect (POFA)	Instruction on muscle movement of the emotions based on FACS	ND	ND	ND	ND

Continue...

Table 2. Continuation.

Authors and year of publication	Name of the database elaborated	Method used to elicit the emotions	Patterns in stimulus capture	Criteria used in the validation stage for inclusion of stimuli in the final database	Sample characteristics in the stage for the validation of the stimuli	Psychometric properties assessed
Fujimura and Umemura (2018) ³¹	A facial expression database based on the dimensional and categorical model of emotions	1) Emotions elicited from specific situations 2) Instruction on muscle movement of the emotions based on FACS	- Background: White - Clothes: White T-shirt - Distractors removed: Glasses and strong makeup	Agreement among the evaluators Mean of 69% agreement among the evaluators (SD=21%)	39 university students - Age: 21.33 years; SD=2.39 - Gender: M=19; F=20 - Race: Japanese natives	- Precision: Accuracy [†] - Validity evidence: Content-based: Accuracy and confusion matrix of agreement rates for images of dynamic and static expressions of each model
Franz et al. (2021) ³²	Picture-Set of Young Children's Affective Facial Expressions (PSYCAFE)	1) Guidance of emotions in theater workshops 2) Directed Facial Action Task used to guide the movement of anatomical landmarks	- Background: White - Clothes: ND (just face) - Distractors removed: ND (just face)	Step 1 Confirmatory hierarchical cluster analysis by Ward Step 2 Intensity, authenticity, and likeability. Accuracy (77-100%) and AFFDEX Software	Step 1 197 volunteers from the community - Age: 32.9 years; SD=16.1 - Gender: M=33%; F=67% - Race: ND Step 2 44 volunteers from the community - Age: 25.7 years; SD=5.9 - Gender: M=48%; F=52% - Race: ND	- Precision: Accuracy [†] - Validity evidence: Based on the relationship with other variables: Stimulus age × expressed emotion × accuracy
Garrido et al. (2017) ³³	Stills and Videos of facial Expressions (SAVE database)	Emotions elicited from specific situations	- Background: Gray - Clothes: White T-shirt - Distractors removed: Jewelry, glasses, and makeup	Stimuli with an assessment of 2.5 SD above or below the mean	120 university students - Age: 20.62 years; SD=3.39 - Gender: M=22.5%; F=77.5% - Race: Caucasian	- Precision: Accuracy [†] - Validity evidence: Content-based: Accuracy and interest dimensions (valence, excitement, clarity, intensity, appeal, similarity, and familiarity) Based on the relationship with other variables: Accuracy × gender of the model and the participant [¶]
Giuliani et al. (2017) ¹⁵	The DuckEES child and adolescent dynamic facial expressions stimulus set	Emotions elicited from specific situations	- Background: White - Clothes: ND - Distractors removed: ND	Images recognized with ≥70% accuracy	36 volunteers from the Oregon University - Age: 19.5 years; SD=1.95 - Gender: M=14; F=22 - Race: ND	- Precision: Accuracy [†] - Validity evidence: Content-based: Accuracy and Fleiss' kappa [†]

Continue...

Table 2. Continuation.

Authors and year of publication	Name of the database elaborated	Method used to elicit the emotions	Patterns in stimulus capture	Criteria used in the validation stage for inclusion of stimuli in the final database	Sample characteristics in the stage for the validation of the stimuli	Psychometric properties assessed
Happy et al. (2015) ³⁴	The Indian Spontaneous Expression Database for Emotion Recognition (ISED)	Emotion induction through videos	- Background: ND - Clothes: ND - Distractors removed: ND	Agreement among the evaluators (Fleiss' Kappa)	Four trained evaluators - Age: ND - Gender: M=2; F=2 - Race: ND	- Precision: Accuracy [†] - Validity evidence: Content-based: Accuracy and Fleiss' kappa [†]
Kaulard et al. (2012) ³⁵	The MPI Facial Expression Database	Emotions elicited from specific situations	- Background: Black - Clothes: Black cape and hats - Distractors removed: Makeup and beards	Consistency among the evaluators (Fleiss' Kappa)	20 German natives - Age: 19-33 years - Gender: M=10; F=10 - Race: ND	- Precision: Accuracy [†] - Validity evidence: Content-based: Accuracy and Fleiss' kappa [†]
Keutmann et al. (2015) ³⁶	Visual and vocal emotional expressions of adult and child actors	Emotions elicited from specific situations	- Background: Green - Clothes: ND - Distractors removed: ND	Accuracy	510 students, 226 from Drexel University and 284 from the University of Central Florida - Age: ND - Gender: ND - Race: ND	- Analysis of the items: Difficulty analysis and item discrimination by means of the classical test theory - Precision: Accuracy [†] - Validity evidence: Content-based: Accuracy
Kim et al. (2017) ³⁷	Korea University Facial Expression Collection – Second Edition (KUFEC-II)	Instruction on muscle movement of the emotions based on FACS	- Background: Gray - Clothes: Pattern - Distractors removed: Makeup, accessories, and dyed hair	Internal consistency Accuracy	75 evaluators - Age: 19-69 years (26.17 years, SD=5.69) - Gender: M=39; F=36 - Race: ND	- Precision: Accuracy [†] - Validity evidence: Content-based: Accuracy; agreement among the evaluators [†] and scores for purity, valence, and intensity Based on the relationship with other variables: ANOVA to test the effects of gender on recognition [‡] and correlations between the participant's emotional state and task performance [¶]
Langner et al. (2010) ³⁸	Radboud Faces Database	Instruction on muscle movement of the emotions based on FACS	- Background: White - Clothes: Black T-shirt - Distractors removed: Glasses, earrings and makeup	Accuracy	276 students from Radboud University - Age: 21.2 years; SD=4.0 - Gender: M=38; F=238 - Race: ND	- Precision: Accuracy [†] - Validity evidence: Content-based: Accuracy and dimensions of interest (type of expression, intensity, clarity, genuineness, and valence) Based on the relationship with other variables: ANOVA comparing each of the precision variables with age, gender, expression, and gaze direction [‡]

Continue...

Table 2. Continuation.

Authors and year of publication	Name of the database elaborated	Method used to elicit the emotions	Patterns in stimulus capture	Criteria used in the validation stage for inclusion of stimuli in the final database	Sample characteristics in the stage for the validation of the stimuli	Psychometric properties assessed
LoBue and Thrasher. (2015) ¹⁴	The Child Affective Facial Expression (CAFE)	Instruction on muscle movement of the emotions based on FACS was carried out during improvised games	- Background: White - Clothes: White sheet - Distractors removed: ND	Images recognized with ≥60% accuracy	100 undergraduate students from Rutgers University - Age: ND - Gender: M=50; F=50 - Race: African-American (n=17%), Asian (n=27%), White (n=30%), Latin (n=17%), and others (n=9%)	- Analysis of the items: Difficulty of the items: Rasch model - Precision: Test-retest reliability [†] and accuracy [#] - Validity evidence: Content-based: Accuracy
Lundqvist et al. (1998) ³⁹	Karolinska Directed Emotional Faces (KDEF) Database	The participants were free to express the emotion as they wished	Background: Neutral Clothes: Gray T-shirt Distractors removed: Beard, mustache, earrings, glasses, and makeup	ND	ND	ND
Ma et al. (2020) ⁴⁰	Han, Hui, and Tibetan Chinese facial expression database	1) Emotion induction through photographs and videos 2) Instruction on muscle movement of the emotions based on FACS	- Background: Black - Clothes: ND - Distractors removed: Jewelry	Images recognized with ≥60% accuracy	240 volunteers (80 from each study region) - Age: 23 years; SD=1.7 - Gender: M=120; F=120 - Race: Chinese	- Precision: Accuracy ^{**} and method of halves - Validity evidence: Content-based: Accuracy Based on internal consistency: Cronbach's alpha [†]
Ma et al. (2015) ⁴¹	Chicago Face Database (CFD)	1) Emotions expressed from verbal instructions 2) Presentation of an equivalent photograph expressing the emotion	- Background: White - Clothes: Gray T-shirt - Distractors removed: ND	Two independent judges assessed how believable the expression was on a Likert scale from 1 to 9 (1=not at all believable; 9=very believable)	1,087 evaluators (convenience sample) - Age: 26.7 years; SD=10.5 - Gender: M=308; F=552 - Race: White (n=516), Asian (n=117), Black (n=74), bi- or multi-race (n=72), Latin (n=57), others (n=18), and did not report (n=233)	- Precision: Accuracy - Validity evidence: Based on the internal structure: exploratory factor analysis (Varimax rotation) Content-based: Accuracy; agreement among the evaluators [†] and effects of race and gender of the stimuli (criteria for item construction)

Continue...

Table 2. Continuation.

Authors and year of publication	Name of the database elaborated	Method used to elicit the emotions	Patterns in stimulus capture	Criteria used in the validation stage for inclusion of stimuli in the final database	Sample characteristics in the stage for the validation of the stimuli	Psychometric properties assessed
Maack et al. (2017) ⁴²	The Tromsø Infant Faces Database (TIF)	The parents were instructed to elicit the intended emotions with games and specific stimuli	- Background: White - Clothes: White overalls and hat - Distractors removed: ND	The photographs with best agreement among the evaluators were selected Mean classification of clarity and intensity below 2.5 Validation: (a) expression portrayed, (b) clarity of expression, (c) intensity of the expression, and (d) valence of the expression	720 participants - Age: 18-70 years (32.8 years; SD=10.4) - Gender: M=21%; F=79% - Race: ND	- Precision: Accuracy ^{††} - Validity evidence: Content-based: dimensions of interest (type of expression, clarity, intensity, and valence) Based on the relationship with other variables: ANOVA to compare performance × child-rearing stage × gender × mood [¶]
Meuwissen et al. (2017) ⁴³	Developmental Emotional Faces Stimulus Set (DEFSS)	1) Emotions elicited from specific situations 2) Presentation of an equivalent photograph expressing the emotion	- Background: Gray - Clothes: ND - Distractors removed: Jewelry	The images recognized by less of 55% of the evaluators were excluded	228 university students between undergraduate and graduate levels and children preappointed by the family via the Internet - Age: 8-30 years - Gender: M=150; F=254 Race: White (n=81%), non-White (n=17%)	- Precision: Accuracy ^{††} - Validity evidence: Content-based: correct answers by age group, intensity, and emotion
Minear and Park. (2004) ⁴⁴	A life span database of adult facial stimuli	Emotions expressed from verbal instructions	- Background: Gray - Clothes: ND - Distractors removed: ND	ND	ND	ND
Negrão et al. (2021) ⁴⁵	The Child Emotion Facial Expression Set	1) Presentation of an equivalent photograph expressing the emotion 2) Emotions elicited from specific situations	- Background: White - Clothes: White - Distractors removed: ND	Step 1: 100% agreement between two evaluators Step 2: 100% agreement between other two evaluators (two of each step)	Four judges - Age: ND - Gender: ND - Race: ND	- Precision: Accuracy [†] and Cohen's kappa [†] - Validity evidence: Based on the relationship with other variables: accuracy × gender × age [¶] ; emotion × race [‡]

Continue...

Table 2. Continuation.

Authors and year of publication	Name of the database elaborated	Method used to elicit the emotions	Patterns in stimulus capture	Criteria used in the validation stage for inclusion of stimuli in the final database	Sample characteristics in the stage for the validation of the stimuli	Psychometric properties assessed
Novello et al. (2018) ⁴⁶	Youth Emotion Picture Set	1) Emotions elicited from specific situations 2) Presentation of an equivalent photograph expressing the emotion 3) Presentation of videos and a game to specifically elicit the emotion of anger	- Background: ND - Clothes: Black cape - Distractors removed: Jewelry	Images recognized with ≥75% accuracy	Adults: 101 volunteers recruited through the snowball method - Age: 18-77 years - Gender: M=31.7%; F=68.3% - Race: ND Adolescents: 54 volunteers from state schools - Age: 12-17 years - Gender: M=40.7%; F=59.3% - Race: ND	- Precision: Accuracy [†] and Cohen's kappa [†] - Validity evidence: Based on the relationship with other variables: comparison of performance by age [¶]
O'Reilly et al. (2016) ⁴⁷	The EU-Emotion Stimulus Set	Emotions elicited from specific situations	- Background: White - Clothes: ND - Distractors removed: ND	Accuracy	1,231 volunteers - Age: 44 years; SD=16.7 - Gender: M=428; F=803 - Race: ND	- Precision: Accuracy ₁ and Cohen's kappa [†] - Validity evidence: Content-based: performance comparison by expression type, valence, and excitation
Olszanowski et al. (2015) ⁴⁸	Warsaw Set of Emotional Facial Expression Pictures (WSEFEP)	Instruction on muscle movement of the emotions based on FACS	- Background: White - Clothes: Black T-shirt - Distractors removed: Beards, mustaches, earrings, and glasses	Agreement in recognition	1,362 participants - Age: 26.6 years; SD=11.6 - Gender: M=261; F=1,101 - Race: ND	- Precision: agreement among the evaluators - Validity evidence: Content-based: purity analysis and intensity coefficient
Passarelli et al. (2018) ⁴⁹	Facial Expression Recognition Test (FERT)	Presentation of an equivalent photograph expressing the emotion	- Background: Black - Clothes: Black T-shirt - Distractors removed: ND	Unidimensional model	794 volunteers from the community - Age: 36.13 years; SD=13.79 - Gender: M=36.2%; F=63.8% - Race: ND	- Validity evidence: Based on the internal structure: factor analysis through the two-parameter Bayesian model - Based on the relationship with other variables; performance comparison between gender and age [‡] - Analysis of the items: Discrimination and difficulty through the Item Response Theory (IRT)

Continue...

Table 2. Continuation.

Authors and year of publication	Name of the database elaborated	Method used to elicit the emotions	Patterns in stimulus capture	Criteria used in the validation stage for inclusion of stimuli in the final database	Sample characteristics in the stage for the validation of the stimuli	Psychometric properties assessed
Romani-Sponchiado et al. (2015) ⁵⁰	Child Emotions Picture Set	Emotion induction through videos	- Background: ND - Clothes: ND - Distractors removed: ND	Images recognized with $\geq 60\%$ accuracy	30 psychologists with experience in child development - Age: ND - Gender: ND - Race: ND	- Precision: Accuracy** and Fleiss' Kappa [†] - Analysis of the items: Accuracy - Validity evidence: Content-based: Fleiss' kappa; chi-square to compare the proportion of posed and spontaneous photographs
Samuelsson et al. (2012) ⁵¹	Umeå University Database of Facial Expressions	Instruction on muscle movement of the emotions based on FACS	- Background: ND - Clothes: ND - Distractors removed: Makeup	Accuracy	526 participants - Age: 18-73 years (37.7 years; SD=13.0) - Gender: M=157; F=369 - Race: ND	- Precision: Accuracy [†] - Validity evidence: - Based on the relationship with other variables; performance comparison by gender and age [†]
Sharma and Bhushan. (2019) ⁵²	Indian Affective Picture	1) Presentation of an equivalent photograph expressing the emotion 2) Emotions elicited from specific situations	- Background: ND - Clothes: ND - Distractors removed: Beards, glasses, and makeup	Accuracy Intensity (9-point scale)	350 undergraduate students - Age: 20.58 years; SD=1.13 - Gender: M=320; F=30 - Race: ND	- Analysis of the items: Accuracy [†] - Validity evidence: Based on the relationship with other variables: t-test to compare men's and women's performance [†]
Tottenham et al. (2009) ¹²	The NimStim set of facial expressions	Emotions expressed from verbal instructions	- Background: ND - Clothes: ND - Distractors removed: Makeup	Validity (accuracy and Cohen's kappa) and reliability	Group 1 47 university students - Age: 19.4 years (SD=1.2) - Gender: M=39; F=47 - Race: European-American (81%), African-American (6%), Asian-American (9%), and Hispanic-American (4%) Group 2 34 volunteers from the community - Age: 25.8 years (SD=4.1) - Gender: M=22; F=12 - Race: European-American (59%), African-American (18%), Asian-American (6%), Hispanic-American (6%), and other races (12%)	- Precision: Accuracy [†] and test-retest [†] - Validity evidence: Content-based: Accuracy and test-retest

Continue...

Table 2. Continuation.

Authors and year of publication	Name of the database elaborated	Method used to elicit the emotions	Patterns in stimulus capture	Criteria used in the validation stage for inclusion of stimuli in the final database	Sample characteristics in the stage for the validation of the stimuli	Psychometric properties assessed
Tracy et al. (2009) ⁵³	University of California, Davis, Set of Emotion Expressions (UCDS)	Instruction on muscle movement of the emotions based on FACS	- Background: Gray - Clothes: White T-shirt - Distractors removed: Jewelry	Accuracy (the most recognized emotion of each expression was included in the final database)	Study 1 175 undergraduate students - Age: ND - Gender: M=35%; F=65% - Race: ND Study 2 234 undergraduate students - Age: ND - Gender: M=21%; F=79% - Race: ND	- Analysis of the items: Accuracy ^{§§} - Validity evidence: Content-based: Accuracy and performance based on race and gender of stimulus
Vaiman et al. (2017) ⁵⁴	FACS	Emotions elicited from specific situations	- Background: Blue - Clothes: White T-shirt - Distractors removed: Hair back (hair up)	Images recognized with ≥70% accuracy	466 students from the Psychology School of the National University of Córdoba. - Age: 20.29 years; SD=4.33 - Gender: M=23%; F=79% - Race: ND	- Precision: Accuracy [†] - Analysis of the items: Discrimination - Validity evidence: Based on the convergent relationship: Descriptive comparison of database performance vs. POFA database performance [†]
Yang et al. (2020) ⁵⁵	Tsinghua facial expression database	1) Emotions elicited from specific situations 2) Instruction on muscle movement of the emotions based on FACS	- Background: White - Clothes: ND - Distractors removed: Tattoos, piercings, jewelry, glasses, and makeup.	Images recognized with ≥70% accuracy	34 young individuals and 31 older adults, Chinese Young individuals - Age: 19-35 years (23.50 years; SD=4.41) - Gender: M=19; F=15 - Race: Chinese Older adults - Age: 58-72 years (65.06 years; SD=3.50) - Gender: M=13; F=18 - Race: Chinese	- Precision: Accuracy [†] and kappa agreement among the evaluators [†] - Validity evidence: Content-based: Accuracy and kappa agreement among the evaluators

ND: not declared; M: male; F: female; MTurk: Amazon Mechanical Turk; FACS: Facial Action Coding System (Ekman and Friesen, 1978)⁶⁵; ANCOVA: analysis of covariance; ANOVA: repeated-measure analysis of variance.

*Only images with ≥50% accuracy were included in the final database; [†]Satisfactory indexes; [‡]There was a significant difference in precision between the analyzed variables; [§]The mean rate of correct identification of the emotions was 62.5%; ^{||}Only images recognized by ≥15 evaluators were included in the final database; [¶]There was no significant difference in precision between the analyzed variables; ^{**}The mean rate of correct identification of the emotions was 66%; ^{***}Only images with ≥60% accuracy were included in the final; ^{††}Accuracy is presented for each emotion and varied from 44 to 100%; ^{†††}Only images recognized by at least 55% of the evaluators were included in the final database. The mean recognition of the final database was 63%; ^{§§}The mean recognition rate of the final database varied from 47 to 94%.

Table 1 presents the general characteristics of the face databases included and Table 2 presents the methodological characteristics used to create each of them.

General characteristics of the face databases included

The articles included were published between 1976 and 2020, the majority dating from 2015 and 2017. Of the 36 articles included, 30.56% were carried out in the

United States. In relation to the theoretical framework used for the construction of the databases, 75% of the studies were empirically based. In other words, the limitations of the databases already built were the basis for this construction.

Most of the articles (61.1%) elaborated databases made up by six basic emotions (i.e., happiness, sadness, fear, anger, disgust, and surprise), as well as neutral faces. Some databases did not neutral faces, or surprise and disgust. Two databases only included happiness and neutral faces, one database only included happiness, fear, and neutral; and another included only happiness, sadness, anger, and surprise.

In relation to the participants, 41.7% of the studies selected resorted to actors (either amateur or professional) to express the emotions. The mean age of the actors varied from 13.24 to 73.2 years, with four studies including different age groups in their databases. Only five of the studies with actors included different races in their samples, and seven studies included any of the specific race, namely, Caucasian, Japanese, Korean, Polish, Indian, or Chinese. Three studies did not report the actors' race.

In relation to the other studies, that is, those that present the basic emotions expressed by community-dwelling individuals, inserted in various contexts, presented ages varying from 4 months to 93 years, and five of these studies included volunteers of different ages. Of these, 10 studies included participants of

different races and the remaining studies included only one race, namely, Korean, Caucasian, Indian, and Chinese. Three studies did not report the participants' race. With regard to the presentation of the stimuli, 86.1% of the studies included colored faces in their databases, four studies used black and white faces, and one study included both colored and black and white faces in its database.

Most of the databases included (75%) present static stimuli, four studies are of dynamic stimuli, and five databases have both static and dynamic stimuli. Five studies presented open and closed mouth expressions, and other studies included additional features such as varying intensities and varying angles. The final total stimuli included in the databases varied from 42 to 18,800.

Methodological characteristics used in the studies

Method used to elicit the emotions

The method used to elicit the emotions varied across the studies. In general, more than one method was used in this stage. Predominantly, 44.4% of the studies used specific situations as one of the ways to elicit the intended emotions, such as "Imagine that you have just won the lottery; imagine that you have just lost a loved one." The studies also used instructions based on the muscle movement of the emotions considering protocols such as the Investigator's Guide for the Facial Action Coding

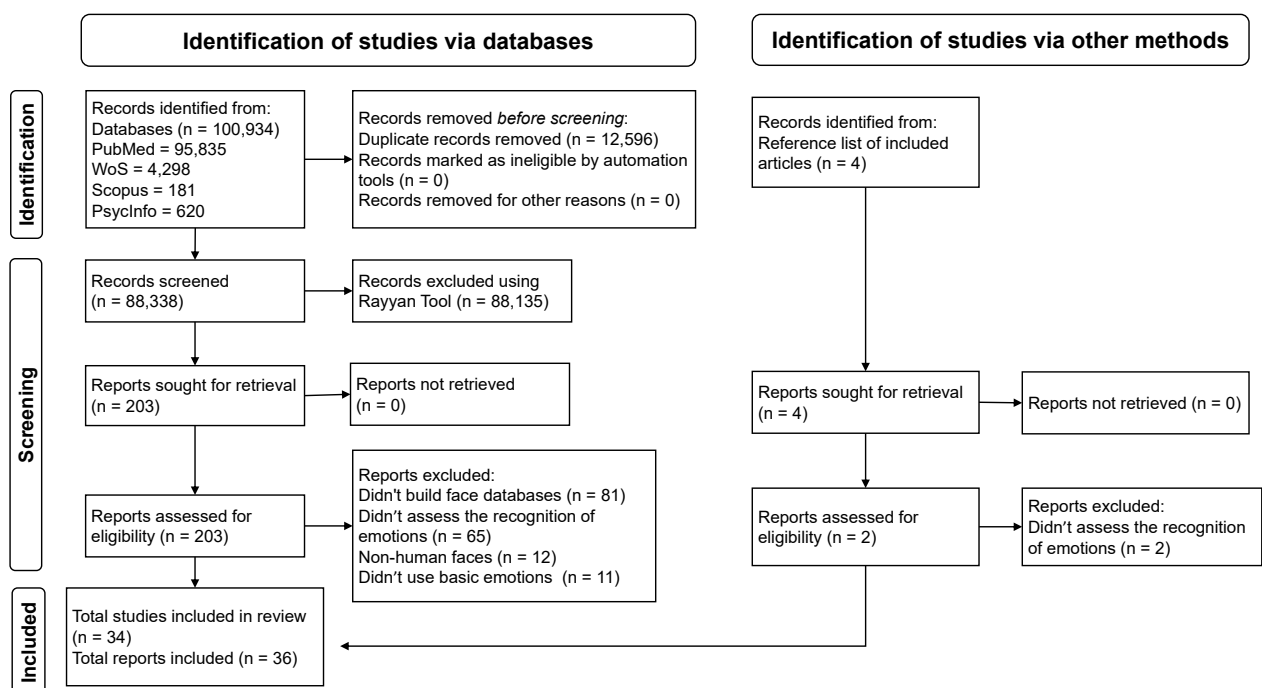


Figure 1. The article selection process according to the PRISMA initiative recommendations¹⁷.

System (FACS), others used a photograph as a model, and others elicited the emotions from photographs and/or videos.

Two studies that built faces with infants and children used an instructional protocol, performed by the parents, to elicit the intended emotions. In one study, the individuals could express the emotion any way they wanted. Three studies elicited emotions in the participants through verbal instructions, such as “Make a happy face” and one study used workshops to teach children how to express basic emotions as well as a Directed Facial Action Task used to guide movement of anatomical landmarks.

Recording the stimuli

Most of the studies sought to establish and describe patterns to record the stimuli. For example, the images were photographed against a white background, black, or gray, and the individuals wore black or white garments. In addition, 55.6% of the studies established distractors that should be removed from the volunteers so that the images could be recorded, such as jewelry, accessories, and strong makeup.

Validation stage

The number of participants who validated the faces constructed by the studies varied from 4 to 1,362, and most of the participants who validated the stimuli were inserted in a university context. The way to validate the final stimuli in the database varied across the studies. The majority included recognition accuracy as one of the criteria, with images included reaching recognition percentages from >50 to $\geq 75\%$. The studies also used other criteria to include the stimuli in the final database, such as agreement among the evaluators.

Psychometric properties of the final database

Only one study did not include accuracy as a precision measure. In most of the cases, it was also used to validate the task content and even for item analysis. One study also used the method of halves as a precision measure. In 66.7% of the studies, the stimuli were recognized with $\geq 70\%$ accuracy.

Test-retest reliability was a variable used to assess task precision in four studies, all presenting satisfactory indexes for this dimension. Regarding the measures of validity evidence, 10 studies used Cohen’s kappa or Fleiss’ kappa to validate the task content according to the agreement among the evaluators. All of them presented satisfactory indexes in this dimension. Only one study used Cronbach’s alpha to assess internal consistency, also reporting a satisfactory value.

Six studies analyzed the items’ difficulty. Three studies used Item Response Theory (IRT); one study analyzed difficulty according to the intensity and representativeness scores; one study used the Classical Test Theory (CTT); and one study used discrimination.

Two studies presented validity evidence based on the internal structure. One of them used exploratory factor analysis and the other resorted to factor analysis through the two-parameter Bayesian model. In addition, the other study presented validity evidence based on the convergent relationship, presenting a descriptive comparison of the database built with the POFA bank, with satisfactory indexes.

Fourteen (38.9%) studies presented validity evidence based on the relationship with other variables.

DISCUSSION

The ability to recognize emotional facial expressions can be modulated by variables such as gender, age, and race. In this sense, a number of studies sought to elaborate valid facial expression databases to assess recognition of emotions in specific populations and contexts. However, the methodological heterogeneity among construction studies can make it difficult to create patterns for the construction of these stimuli, regardless of the context and characteristics of who express them. This systematic review sought to gather the studies that built face databases to assess recognition of basic emotions, describing and comparing the methodologies used in its development.

General characteristics of the face databases included

The way to present the stimuli of an emotion recognition test has already been target of discussions among researchers in the area, since a pioneering study showed that the recognition of static and dynamic facial emotional stimuli involves different neural areas⁶⁶. In this review, most of the studies consist of static stimulus databases. The difference in the recognition of static or dynamic stimuli is still an unanswered discussion, given that some studies report a higher rate of recognition of dynamic stimuli^{67,68} while others point to a minimal or no difference in the recognition of these stimuli^{69,70}.

Khosdelazad et al.⁷¹ investigated the differences in the performance of 3 emotion recognition tests in 84 healthy participants. The results point to a clear difference in the performance of tests with static or dynamic stimuli, with the stimuli that change from a neutral face to the intended emotion (dynamic) being the most difficult to be recognized, given the low performance in the test⁷¹. However, it is noteworthy that variables such

as age and schooling also modulated performance in the tests, highlighting the importance of normative data regardless of the type of stimulus chosen⁷¹.

Several stimuli databases for facial expressions of emotions were developed in order to be used in specific populations and cultures⁷². Cultural issues must be taken into account when understanding these emotional expressions, as they can exert an influence on their recognition⁷³. A study that considered ethnicity as an influencing factor in the performance of emotion recognition tasks and compared this ability to identify emotions between Australian and Chinese individuals verified that people perform worse when classifying emotions that are expressed on faces of another ethnicity⁷⁴. In this sense, the cultural characteristics of the stimulus presented can also modulate performance in the test.

In addition to the difference in the pattern of response when recognizing emotions from another culture, studies showed that there is still a difference in the pattern of intensity recognized, regardless of the race or gender of the stimulus presented^{75,76}. This fact happens probably because we manage our emotions according to the our learnings throughout our lives, clearly shaped by the cultural context in which we are inserted^{76,77}. Thus, we learned in certain situations to hide or amplify our emotions, consequently affecting how we recognize emotions and highlighting the clear influences of culture on our social and cognitive abilities^{76,78}.

Furthermore, when we think about the modulating character of the cultural context in the recognition of emotions, it is important to highlight the impact that socioeconomic status can also have on this ability. In particular, some countries and regions with greater socioeconomic disparities may reflect different patterns of cognitive abilities⁷⁹. For example, a large international study investigated, in 12 countries and 587 participants, the influence of nationality on core social cognition skills⁸⁰.

After controlling the analyses for other modulating variables such as age, sex, and education, the results showed that a variation of 20.76% (95%CI 8.26–35.69) in the test score that evaluated emotion recognition can be attributed to the nationality of the individuals evaluated⁸⁰. These results make us reflect on the cultural disparities that exist in underdeveloped countries and how these aspects can influence the social and cognitive variables, as well as the recognition of emotions discussed here.

In addition, aspects related to the participant's profile can also interfere in task performance. Five

studies in this review presented open and closed mouth expressions and other studies included additional features such as varying intensities, gaze directions, and varying angles. These variables can also modulate task performance. Emotions expressed with the mouth open seem to increase the intensity of the emotion perceived by the subject^{81,82}. Consequently, incorporating this face variation to the database can be important to assess the emotion experienced by the individual who recognizes the stimuli. In addition, open-mouthed facial expressions seem to draw more the attention of the respondent than closed-mouthed expressions⁸¹.

Hoffmann et al.⁸³ found a correlation between the intensity and accuracy of recognition of an emotion, where higher intensities were associated with greater accuracy in the perception of the face. However, Wingenbach et al.⁸⁴ did not find effects of the intensity level on expression recognition. Despite the controversial results regarding emotion intensity, it can still be an important variable to be taken into account in the construction of databases in order to compare recognition between different degrees of intensities.

The perception of the emotion expressed can also be modulated by the gaze direction of the person expressing it⁸⁵ so that when gaze is directed at the participant, this recognition is greater than when compared to the look avoided⁸⁶. In addition, photographing the expressions from different angles can increase the ecological validity of the database built³⁸.

Methodological characteristics used in the studies

Method used to elicit the emotions

An important methodological choice in the studies that elaborate face databases is the way in which the stimuli will be elicited and who is going to express them. Our results show that most of the studies included in this systematic review resort to actors (either amateur or professional) to express the emotions. Such methodological choice can be justified by the fact that people who have experience in acting are able to express more realistic emotions than individuals without any experience⁸⁷. Thus, resorting to actors to act out emotions can be advantageous with regard to bringing the emotions expressed to a more real context.

The literature indicates that there are three different ways to induce emotions, namely:

- Posed emotions;
- Induced emotions; and
- Spontaneous emotions^{88,89}.

Posed emotions are those expressed by actors or under specific guidance, tending to be less representative of an emotion expressed in a real context⁸⁹. Induced emotions have a more genuine character than posed emotions, as varied eliciting stimuli are presented to the participant in order to generate the most spontaneous emotion possible⁸⁹. However, it is noteworthy that this way of inducing emotion can also have limitations as to its veracity, since induction is carried out in a context controlled by the researcher⁸⁹. Spontaneous emotions are considered closer to a real-life context. However, due to their observable character, their recording could only be possible when the individuals are not aware that they are being recorded. Thus, any research procedure can bias this spontaneity⁸⁹.

To increase induction effectiveness, the studies use a combination of techniques and procedures to facilitate achievement of the intended emotions. Among the 36 studies analyzed in this review, 44.4% used specific hypothetical situations as one of the ways to elicit the intended emotions, such as “Imagine that you have just won the lottery; imagine that you have just lost a loved one.” Thus, despite induction being generated in a controlled context, using hypothetical everyday situations aims at remedying the limitation of expressions that are not very representative of real life.

Recording the stimuli

All construction studies try to capture stimuli following some kind of pattern. Some explore this pattern more in detail and others are more objective. Despite this, the data included in this review indicate that it is important to standardize the clothes worn by the participants and the background they are positioned against during the capture of stimuli.

In addition, most construction studies have established distractors that should be removed prior to image capture, such as jewelry, accessories, and strong make-up. Our hypothesis is that these distractors could direct the attention of those who respond to the task and exert an impact on recognition performance, since attention can be a modulating variable in emotional tasks⁹⁰.

Validation stage

The way to validate the stimuli in the databases elaborated varies greatly across the studies. Based on the methods used in the construction, the validation criteria are defined. Accuracy is the most used precision indicator in the development and validation of face databases that assess recognition of emotions^{12,13}, which is why it was presented in most of the studies included. Recognition rate $\geq 70\%$ is the most frequently used. However, the

choice of which criterion to adopt at this stage is varied, and it is common to adopt other rates and criteria to validate the database, such as intensity, clarity, and agreement between evaluators.

Psychometric properties of the final database

We seek to follow the standards established by Resolution 09-2018 of the Federal Council of Psychology, which regulates the necessary dimensions for the assessment of psychological tests to verify the psychometric qualities of the databases. Although the studies present construction of tasks and not instruments, recognition of emotions is an important skill that allows for interaction in society and can be used to assess social cognition to predict the diagnosis of mental disorders⁹¹.

The analyses presented by the studies in this stage are also heterogeneous. However, some dimensions presented in the studies become strictly necessary to verify the quality of the database elaborated. With regard to the technical requirements, it is important to evaluate dimensions related to precision and validity evidence of the constructed task^{20,21}. It is worth noting that normative data are also important to assess the quality of the task. However, this variable and other important analyses were not included in this review as they are found in articles published separately.

This review showed that the studies that elaborate face databases for the recognition of emotions present heterogeneous methods. However, similarities between the studies allow us to trace important patterns for the development of these stimuli, such as using more than one method to elicit the most spontaneous emotion possible, standardizing the characteristics of the volunteers for capturing the stimuli, validating the database based on preestablished criteria, and presenting data referring to precision and validity evidence. With regard to future directions related to the research methods, greater standardization of the methods for eliciting and validating emotions would make the choice of the type of task to be used in each context more reliable.

Authors' contributions. DMF: conceptualization, formal analysis, investigation, methodology, project administration, writing – original draft, writing – review & editing; BLCE, MAMB: analysis, methodology, writing – original draft, writing – review & editing; MM: resources, writing – review & editing, supervision, writing – review & editing; MHNC: conceptualization, investigation, methodology, project management, supervision, writing – review & editing.

REFERENCES

1. Darwin C. The expression of the emotions in man and animals. Chicago: University of Chicago Press; 2015.
2. Plutchik R. The nature of emotions: human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American Scientist*. 2001;89(4):344-50.
3. Palermo R, Rhodes G. Are you always on my mind? A review of how face perception and attention interact. *Neuropsychologia*. 2007;45(1):75-92. <https://doi.org/10.1016/j.neuropsychologia.2006.04.025>
4. Pascalis O, Slater A. The development of face processing in early childhood. New York: Nova Science Publishers; 2003.
5. Ekman P, Sorenson ER, Friesen WV. Pan-cultural elements in facial displays of emotion. *Science*. 1969;164(3875):86-8. <https://doi.org/10.1126/science.164.3875.86>
6. Ekman P. Facial expression and emotion. *Am Psychol*. 1993;48(4):384-92. <https://doi.org/10.1037/0003-066X.48.4.384>
7. Schmidt KL, Cohn JF. Human facial expressions as adaptations: evolutionary questions in facial expression research. *Am J Phys Anthropol*. 2001;Suppl 33:3-24. <https://doi.org/10.1002/ajpa.20001>
8. Barrett LF, Mesquita B, Gendron M. Context in emotion perception. *Current Directions in Psychological Science*. 2011;20(5):286-90. <https://doi.org/10.1177/0963721411422522>
9. Ebner NC. Age of face matters: age-group differences in ratings of young and old faces. *Behav Res Methods*. 2008;40(1):130-6. <https://doi.org/10.3758/bm.40.1.130>
10. Chaplin TM, Aldao A. Gender differences in emotion expression in children: a meta-analytic review. *Psychol Bull*. 2013;139(4):735-65. <https://doi.org/10.1037/a0030737>
11. Zebrowitz LA, Kikuchi M, Fellous JM. Facial resemblance to emotions: group differences, impression effects, and race stereotypes. *J Pers Soc Psychol*. 2010;98(2):175-89. <https://doi.org/10.1037/a0017990>
12. Tottenham N, Tanaka JW, Leon AC, McCarry T, Nurse M, Hare TA, et al. The NimStim set of facial expressions: judgments from untrained research participants. *Psychiatry Res*. 2009;168(3):242-9. <https://doi.org/10.1016/j.psychres.2008.05.006>
13. Ebner NC, Riediger M, Lindenberger U. FACES--a database of facial expressions in young, middle-aged, and older women and men: development and validation. *Behav Res Methods*. 2010;42(1):351-62. <https://doi.org/10.3758/BRM.42.1.351>
14. LoBue V, Thrasher C. The Child Affective Facial Expression (CAFE) set: validity and reliability from untrained adults. *Front Psychol*. 2015;5:1532. <https://doi.org/10.3389/fpsyg.2014.01532>
15. Giuliani NR, Flournoy JC, Ivie EJ, Von Hippel A, Pfeifer JH. Presentation and validation of the DuckEES child and adolescent dynamic facial expressions stimulus set. *Int J Methods Psychiatr Res*. 2017;26(1):e1553. <https://doi.org/10.1002/mpr.1553>
16. Conley MI, Dellarco DV, Rubien-Thomas E, Cohen AO, Cervera A, Tottenham N, et al. The racially diverse affective expression (RADIATE) face stimulus set. *Psychiatry Res*. 2018;270:1059-67. <https://doi.org/10.1016/j.psychres.2018.04.066>
17. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71. <https://doi.org/10.1136/bmj.n71>
18. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. *Am J Med*. 2006;119(2):166.e7-16. <https://doi.org/10.1016/j.amjmed.2005.10.036>
19. Pittman J, Bakas T. Measurement and instrument design. *J Wound Ostomy Continence Nurs*. 2010;37(6):603-7. <https://doi.org/10.1097/WON.0b013e31819f0a60>
20. American Educational Research Association, American Psychological Association, National Council on Measurement in Education. Standards for educational and psychological testing. Washington: American Educational Research Association; 2014.
21. Brasil. Conselho Federal de Psicologia. Resolução nº 9, de 25 de abril de 2018. Estabelece diretrizes para a realização de Avaliação Psicológica no exercício profissional da psicóloga e do psicólogo, regulamenta o Sistema de Avaliação de Testes Psicológicos - SATEPSI e revoga as Resoluções nº 002/2003, nº 006/2004 e nº 005/2012 e Notas Técnicas nº 01/2017 e 02/2017 [cited on Dec 01, 2022]. Available from: <https://satepsi.cfp.org.br/docs/ResolucaoCFP009-18.pdf>
22. Cohen J. A coefficient of agreement for nominal scales. *Education and Psychological Measurement*. 1960;20(1):37-46. <https://doi.org/10.1177/001316446002000104>
23. Fleiss JL. Measuring nominal scale agreement among many raters. *Psychological Bulletin*. 1971;76(5):378-82. <https://doi.org/10.1037/h0031619>
24. Cortina JM. What is coefficient alpha? An examination of theory and applications. *J Appl Psychol*. 1993;78(1):98-104. <https://doi.org/10.1037/0021-9010.78.1.98>
25. Benda MS, Scherf KS. The complex emotion expression database: a validated stimulus set of trained actors. *PLoS One*. 2020;15(2):e0228248. <https://doi.org/10.1371/journal.pone.0228248>
26. Chung KM, Kim S, Jung WH, Kim Y. Development and validation of the Yonsei face database (YFace DB). *Front Psychol*. 2019;10:2626. <https://doi.org/10.3389/fpsyg.2019.02626>
27. Dalrymple KA, Gomez J, Duchaine B. The dartmouth database of children's faces: acquisition and validation of a new face stimulus set. *PLoS One*. 2013;8(11):e79131. <https://doi.org/10.1371/journal.pone.0079131>
28. Donadon MF, Martin-Santos R, Osório FL. Baby faces: development and psychometric study of a stimuli set based on babies' emotions. *J Neurosci Methods*. 2019;311:178-85. <https://doi.org/10.1016/j.jneumeth.2018.10.021>
29. Egger HL, Pine DS, Nelson E, Leibenluft E, Ernst M, Towbin KE, et al. The NIMH Child Emotional Faces Picture Set (NIMH-ChEFS): a new set of children's facial emotion stimuli. *Int J Methods Psychiatr Res*. 2011;20(3):145-56. <https://doi.org/10.1002/mpr.343>
30. Ekman P, Friesen WV. Pictures of facial affect. Palo Alto: Consulting Psychologists Press; 1976.
31. Fujimura T, Umemura H. Development and validation of a facial expression database based on the dimensional and categorical model of emotions. *Cogn Emot*. 2018;32(8):1663-70. <https://doi.org/10.1080/02699931.2017.1419936>
32. Franz M, Müller T, Hahn S, Lundqvist D, Rampoldt D, Westermann JF, et al. Creation and validation of the Picture-Set of Young Children's Affective Facial Expressions (PSYCAFE). *PLoS One*. 2021;16(12):e0260871. <https://doi.org/10.1371/journal.pone.0260871>
33. Garrido MV, Lopes D, Prada M, Rodrigues D, Jerónimo R, Mourão RP. The many faces of a face: comparing stills and videos of facial expressions in eight dimensions (SAVE database). *Behav Res Methods*. 2017;49(4):1343-60. <https://doi.org/10.3758/s13428-016-0790-5>
34. Happy SL, Patnaik P, Routray A, Guha R. The Indian spontaneous expression database for emotion recognition. *IEEE Transactions on Affective Computing*. 2015;8(1):131-42. <https://doi.org/10.1109/TAF-FC.2015.2498174>
35. Kaulard K, Cunningham DW, Bühlhoff HH, Wallraven C. The MPI facial expression database--a validated database of emotional and conversational facial expressions. *PLoS One*. 2012;7(3):e32321. <https://doi.org/10.1371/journal.pone.0032321>
36. Keutmann MK, Moore SL, Savitt A, Gur RC. Generating an item pool for translational social cognition research: methodology and initial validation. *Behav Res Methods*. 2015;47(1):228-34. <https://doi.org/10.3758/s13428-014-0464-0>
37. Kim SM, Kwon YJ, Jung SY, Kim MJ, Cho YS, Kim HT, et al. Development of the Korean facial emotion stimuli: Korea university facial expression collection 2nd edition. *Front Psychol*. 2017;8:769. <https://doi.org/10.3389/fpsyg.2017.00769>
38. Langner O, Dotsch R, Bijlstra G, Wigboldus DHJ, Hawk ST, van Knippenberg A. Presentation and validation of the Radboud Faces Database. *Cognition and Emotion*. 2010;24(8):1377-88. <https://doi.org/10.1080/02699930903485076>
39. Lundqvist D, Flykt A, Öhman A. The Karolinska directed emotional faces--KDEF. (CD ROM). Stockholm: Karolinska Institute, Department of Clinical Neuroscience, Psychology Section; 1998.
40. Ma J, Yang B, Luo R, Ding X. Development of a facial-expression database of Chinese Han, Hui and Tibetan people. *Int J Psychol*. 2020;55(3):456-64. <https://doi.org/10.1002/ijop.12602>
41. Ma DS, Correll J, Wittenbrink B. The Chicago face database: a free stimulus set of faces and norming data. *Behav Res Methods*. 2015;47(4):1122-35. <https://doi.org/10.3758/s13428-014-0532-5>
42. Maack JK, Bohne A, Nordahl D, Livsclatter L, Lindahl ÅAW, Øvervoll M, et al. The Tromsø Infant Faces Database (TIF): development, validation and application to assess parenting experience on clarity and intensity ratings. *Front Psychol*. 2017;8:409. <https://doi.org/10.3389/fpsyg.2017.00409>
43. Meuwissen AS, Anderson JE, Zelazo PD. The creation and validation of the developmental emotional faces stimulus set. *Behav Res Methods*. 2017;49(3):960-6. <https://doi.org/10.3758/s13428-016-0756-7>
44. Minear M, Park DC. A lifespan database of adult facial stimuli. *Behav Res Methods Instrum Comput*. 2004;36(4):630-3. <https://doi.org/10.3758/bf03206543>
45. Negrão JG, Osorio AAC, Siciliano RF, Lederman VRG, Kozasa EH, D'Antino MEF, et al. The child emotion facial expression set: a database for emotion recognition in children. *Front Psychol*. 2021;12:666245. <https://doi.org/10.3389/fpsyg.2021.666245>

46. Novello B, Renner A, Maurer G, Musse S, Arteche A. Development of the youth emotion picture set. *Perception*. 2018;47(10-11):1029-42. <https://doi.org/10.1177/0301006618797226>
47. O'Reilly H, Pigat D, Fridenson S, Berggren S, Tal S, Golan O, et al. The EU-emotion stimulus set: a validation study. *Behav Res Methods*. 2016;48(2):567-76. <https://doi.org/10.3758/s13428-015-0601-4>
48. Olszanowski M, Pochwatko G, Kuklinski K, Scibor-Rylski M, Lewinski P, Ohme RK. Warsaw set of emotional facial expression pictures: a validation study of facial display photographs. *Front Psychol*. 2015;5:1516. <https://doi.org/10.3389/fpsyg.2014.01516>
49. Passarelli M, Masini M, Bracco F, Petrosino M, Chiorri C. Development and validation of the Facial Expression Recognition Test (FERT). *Psychol Assess*. 2018;30(11):1479-90. <https://doi.org/10.1037/pas0000595>
50. Romani-Sponchiado A, Sanvicente-Vieira B, Mottin C, Hertzog-Fonini D, Arteche A. Child Emotions Picture Set (CEPS): development of a database of children's emotional expressions. *Psychology & Neuroscience*. 2015;8(4):467-78. <https://doi.org/10.1037/h0101430>
51. Samuelsson H, Jarnvik K, Henningson H, Andersson J, Carlbring P. The Umeå university database of facial expressions: a validation study. *J Med Internet Res*. 2012;14(5):e136. <https://doi.org/10.2196/jmir.2196>
52. Sharma U, Bhushan B. Development and validation of Indian Affective Picture Database. *Int J Psychol*. 2019;54(4):462-7. <https://doi.org/10.1002/ijop.12471>
53. Tracy JL, Robins RW, Schrier RA. Development of a FACS-verified set of basic and self-conscious emotion expressions. *Emotion*. 2009;9(4):554-9. <https://doi.org/10.1037/a0015766>
54. Vaiman M, Wagner MA, Caicedo E, Pereno GL. Development and validation of an Argentine set of facial expressions of emotion. *Cogn Emot*. 2017;31(2):249-60. <https://doi.org/10.1080/02699931.2015.1098590>
55. Yang T, Yang Z, Xu G, Gao D, Zhang Z, Wang H, et al. Tsinghua facial expression database – a database of facial expressions in Chinese young and older women and men: development and validation. *PLoS One*. 2020;15(4):e0231304. <https://doi.org/10.1371/journal.pone.0231304>
56. Ekman P, Friesen WV. *Unmasking the face: a guide to recognizing emotions from facial clues*. Nova Jersey: Prentice-Hall; 1975.
57. Ekman P. Universals and cultural differences in facial expressions of emotion. In J. Cole (Ed.), *Nebraska Symposium on Motivation*. Lincoln: University of Nebraska Press; 1972. p. 207-82.
58. Ruffman T, Henry JD, Livingstone V, Phillips LH. A meta-analytic review of emotion recognition and aging: implications for neuropsychological models of aging. *Neurosci Biobehav Rev*. 2008;32(4):863-81. <https://doi.org/10.1016/j.neubiorev.2008.01.001>
59. Borod JC, Koff E, Yecker S, Santschi C, Schmidt JM. Facial asymmetry during emotional expression: gender, valence, and measurement technique. *Neuropsychologia*. 1998;36(11):1209-15. [https://doi.org/10.1016/s0028-3932\(97\)00166-8](https://doi.org/10.1016/s0028-3932(97)00166-8)
60. Brosch T, Sander D, Scherer KR. That baby caught my eye... attention capture by infant faces. *Emotion*. 2007;7(3):685-9. <https://doi.org/10.1037/1528-3542.7.3.685>
61. Parsons CE, Young KS, Kumari N, Stein A, Kringelbach ML. The motivational salience of infant faces is similar for men and women. *PLoS One*. 2011;6(5):e20632. <https://doi.org/10.1371/journal.pone.0020632>
62. Borgi M, Cogliati-Dezza I, Brelsford V, Meints K, Cirulli F. Baby schema in human and animal faces induces cuteness perception and gaze allocation in children. *Front Psychol*. 2014;5:411. <https://doi.org/10.3389/fpsyg.2014.00411>
63. Reise SP, Revicki DA. *Handbook of item response theory modeling*. New York: Taylor & Francis; 2014.
64. Kringelbach ML, Lehtonen A, Squire S, Harvey AG, Craske MG, Holliday IE, et al. A specific and rapid neural signature for parental instinct. *PLoS One*. 2008;3(2):e1664. <https://doi.org/10.1371/journal.pone.0001664>
65. Ekman P, Friesen WV. Facial action coding system. *Environmental Psychology & Nonverbal Behavior*; 1978. <https://doi.org/10.1037/t27734-000>
66. Humphreys GW, Donnelly N, Riddoch MJ. Expression is computed separately from facial identity, and it is computed separately for moving and static faces: neuropsychological evidence. *Neuropsychologia*. 1993;31(2):173-81. [https://doi.org/10.1016/0028-3932\(93\)90045-2](https://doi.org/10.1016/0028-3932(93)90045-2)
67. Cunningham DW, Wallraven C. Dynamic information for the recognition of conversational expressions. *J Vis*. 2009;9(13):7.1-17. <https://doi.org/10.1167/9.13.7>
68. Knappmeyer B, Thornton IM, Bühlhoff HH. The use of facial motion and facial form during the processing of identity. *Vision Res*. 2003;43(18):1921-36. [https://doi.org/10.1016/s0042-6989\(03\)00236-0](https://doi.org/10.1016/s0042-6989(03)00236-0)
69. Gold JM, Barker JD, Barr S, Bittner JL, Bromfield WD, Chu N, et al. The efficiency of dynamic and static facial expression recognition. *J Vis*. 2013;13(5):23. <https://doi.org/10.1167/13.5.23>
70. Fiorentini C, Viviani P. Is there a dynamic advantage for facial expressions? *J Vis*. 2011;11(3):17. <https://doi.org/10.1167/11.3.17>
71. Khoshdelazad S, Jorna LS, McDonald S, Rakers SE, Huitema RB, Buunk AM, et al. Comparing static and dynamic emotion recognition tests: performance of healthy participants. *PLoS One*. 2020;15(10):e0241297. <https://doi.org/10.1371/journal.pone.0241297>
72. Ferreira BLC, Fabricio DM, Chagas MHN. Are facial emotion recognition tasks adequate for assessing social cognition in older people? A review of the literature. *Arch Gerontol Geriatr*. 2021;104277. <https://doi.org/10.1016/j.archger.2020.104277>
73. Matsumoto D, Hwang HS, Yamada H. Cultural differences in the relative contributions of face and context to judgments of emotions. *Journal of Cross-Cultural Psychology*. 2012;43(2):198-218. <https://doi.org/10.1177/0022022110387426>
74. Craig BM, Zhang J, Lipp OV. Facial race and sex cues have a comparable influence on emotion recognition in Chinese and Australian participants. *Atten Percept Psychophys*. 2017;79(7):2212-23. <https://doi.org/10.3758/s13414-017-1364-z>
75. Matsumoto D. Ethnic differences in affect intensity, emotion judgments, display rule attitudes, and self-reported emotional expression in an American sample. *Motiv Emot*. 1993;17:107-23. <https://doi.org/10.1007/BF00995188>
76. Engelmam JB, Pogosyan M. Emotion perception across cultures: the role of cognitive mechanisms. *Front Psychol*. 2013;4:118. <https://doi.org/10.3389/fpsyg.2013.00118>
77. Ekman P, Friesen WV. Constants across cultures in the face and emotion. *J Pers Soc Psychol*. 1971;17(2):124-9. <https://doi.org/10.1037/h0030377>
78. Park DC, Huang CM. Culture wires the brain: a cognitive neuroscience perspective. *Perspect Psychol Sci*. 2010;5(4):391-400. <https://doi.org/10.1177/1745691610374591>
79. Daugherty JC, Puente AE, Fasfous AF, Hidalgo-Ruzzante N, Pérez-García M. Diagnostic mistakes of culturally diverse individuals when using North American neuropsychological tests. *Appl Neuropsychol Adult*. 2017;24(1):16-22. <https://doi.org/10.1080/23279095.2015.1036992>
80. Quesque F, Coutrot A, Cox S, de Souza LC, Baez S, Cardona JF, et al. Culture shapes our understanding of others' thoughts and emotions: an investigation across 12 countries. *PsyArXiv*; 2020. <https://doi.org/10.31234/osf.io/tg2ay>
81. Langeslag SJE, Gootjes L, van Strien JW. The effect of mouth opening in emotional faces on subjective experience and the early posterior negativity amplitude. *Brain Cogn*. 2018;127:51-9. <https://doi.org/10.1016/j.bandc.2018.10.003>
82. Horstmann G, Lipp OV, Becker SI. Of toothy grins and angry snarls--open mouth displays contribute to efficiency gains in search for emotional faces. *J Vis*. 2012;12(5):7. <https://doi.org/10.1167/12.5.7>
83. Hoffmann H, Kessler H, Eppel T, Rukavina S, Traue HC. Expression intensity, gender and facial emotion recognition: women recognize only subtle facial emotions better than men. *Acta Psychol (Amst)*. 2010;135(3):278-83. <https://doi.org/10.1016/j.actpsy.2010.07.012>
84. Wingenbach TSH, Ashwin C, Brosnan M. Sex differences in facial emotion recognition across varying expression intensity levels from videos. *PLoS One*. 2018;13(1):e0190634. <https://doi.org/10.1371/journal.pone.0190634>
85. Adams Jr RB, Kleck RE. Effects of direct and averted gaze on the perception of facially communicated emotion. *Emotion*. 2005;5(1):3-11. <https://doi.org/10.1037/1528-3542.5.1.3>
86. Strick M, Holland RW, van Knippenberg A. Seductive eyes: attractiveness and direct gaze increase desire for associated objects. *Cognition*. 2008;106(3):1487-96. <https://doi.org/10.1016/j.cognition.2007.05.008>
87. Scherer KR, Bänziger T. On the use of actor portrayals in research on the emotional expression. In: Scherer KR, Bänziger T, Roesch E, eds. *Blueprint for affectively computing. A sourcebook*. Oxford: Oxford University Press; 2010. p. 166-76.
88. Wu CH, Lin JC, Wei WL. Survey on audiovisual emotion recognition: databases, features, and data fusion strategies. *APSIPA Transactions on Signal and Information Processing*. 2014;3(1):e12. <https://doi.org/10.1017/ATSIP.2014.11>
89. Haamer RE, Rusadze E, Lüsi I, Ahmed T, Escalera S, Anbarjafari G. Review on emotion recognition databases. In: Anbarjafari G, Escalera S, eds. *Human-robot interaction. Theory and application*. London: IntechOpen; 2017. p. 39-63. <https://doi.org/10.5772/intechopen.72748>
90. Srivastava P, Srinivasan N. Emotional information modulates the temporal dynamics of visual attention. *Perception*. 2008;37:1-29.
91. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders*. 5th ed. Washington: American Psychiatric Press; 2013.