Elena Huerta-Ramos[a,b,c,d*],
Marta Ferrer-Quintero[a,b,c,g*],
Juana Gómez-Benito[h],
Fermín González-Higueras[g],
Daniel Cuadras[c],
Ángel Luis del Rey-Mejías[d,e,f],
Judith Usall[a,b,c],
Susana Ochoa[a,b,c,d]

TRANSLATION AND VALIDATION OF BARON COHEN'S FACE TEST IN A GENERAL POPULATION FROM SPAIN

*Both authors contributed equally to the development of this manuscript.

a) Institut de Recerca de Sant Joan de Déu. Edifici Docent. Santa Rosa, 39-57 08950 Esplugues del Llobregat (Barcelona). Spain.
b) Parc Sanitari Sant Joan de Déu, Sant Boi de Llobregat, Spain. Carrer del Doctor Antoni Pujadas, 42, 08830 Sant Boi de Llobregat.
c) Fundació Sant Joan de Déu, Esplugues de Llobregat, Spain. Edifici Docent. Santa Rosa, 39-57 08950 Esplugues del Llobregat (Barcelona). Spain.
d) CIBERSAM (Centro de Investigación Biomédica en Red - Salud Mental). Av. Monforte de Lemos, 3-5. Pabellón 11. Planta 0 28029 Madrid (Spain).
e) Hospital Clínico San Carlos, Madrid, Spain. C/ Profesor Martín Lagos, s/n. 28040 Madrid (Spain).
f) Comunidad Terapéutica de Jaén. Avda. de Madrid, s/n, Jaén, 23008 (Spain)
g) Departamento de Metodología, Facultad de Psicología, Universidad Complutense de Madrid. Campus de Somosaguas, 28223, Pozuelo de Alarcón. Madrid, Spain
h) Departament de Psicologia Social i Psicologia Quantitativa, Universitat de Barcelona. Passeig de la Vall d'Hebron, 171, 08035 Barcelona, Spain.

Elena Huerta-Ramos designed the Project. Marta Ferrer-Quintero, Fermín González-Higueras and Ángel Luis del Rey-Mejías collected data. Marta Ferrer-Quintero and Elena Huerta-Ramos wrote the manuscript. Daniel Cuadras performed the statistical analysis. Susana Ochoa, Judith Usall and Juana Gómez-Benito provided methodological, statistical and theoretical insights to the work.

ABSTRACT

Introduction. Facial emotion recognition is considered the foundation of effective social functioning, but it has been found impaired in several clinical populations. However, there are few validated tests to measure the ability. To the best of our knowledge, there is no validated measure in a Spanish population. We translated and validated Baron Cohen's Face Test in a general Spanish population.

Methods. The test was administered to 211 (63.3% female) healthy volunteers between 19 and 70 years of age. We used tetrachoric matrices to obtain item per item test-retest reliability and internal consistency. We used confirmatory factor analysis to test for unidimensionality. We used Pearson correlations to examine associations between variables.

Results. The mean score was 18 (SD=1.38). Cronbach's alfa was 0.75. Guttman Lambda 3 indexes yielded 17 out of 20 items to have excellent test-retest reliability. Gender or age differences in performance were not found. The test seems to comply with a one-dimensional structure: CFI=0.889; TLI=0.873 and RMSEA=0.047.

Conclusions. Baron Cohen's Face Test could be a valid measure of FER, although it is not sensitive to age or gender. Because it presents a certain ceiling effect, it could not be appropriate to detect excelling performance.

Keywords. Baron Cohen Face Test, facial emotion recognition, validation, psychometric properties, general population

Correspondence authors:
Elena Huerta-Ramos: mehuerta@pssjd.org
Marta Ferrer Quintero: mf.quintero@pssjd.org

TRADUCCIÓN Y VALIDACIÓN DEL TEST DE CARAS DE BARON COHEN EN POBLACIÓN ESPAÑOLA

RESUMEN

Introducción. El reconocimiento facial de emociones es esencial en el funcionamiento social adecuado. Se han encontrado déficits en muchas poblaciones clínicas. Hay pocos tests validados que midan esta habilidad y ninguno en población española. En este trabajo, tradujimos y validamos el Test de Caras de Baron Cohen en población general española.

Métodos. El test fue administrado a 211 personas (63,3 % mujeres) sanas de entre 19 y 70 años de edad. Usamos matrices tetracóricas para obtener la fiabilidad test-retest y la consistencia interna. Se realizó un análisis factorial confirmatorio para comprobar la unidimensionalidad del test. Utilizamos correlaciones de Pearson para examinar asociaciones entre variables.

Resultados. La media en el estudio fue de 18 (DE = 1,38). Se obtuvo un alfa de Cronbach de 0,75. Calculamos los índices Guttman Lambda 3 para cada ítem. 17 de 20 ítems obtuvieron una estabilidad test-retest excelente. No encontramos asociaciones entre el rendimiento y el género, la edad o el nivel académico. El test presentó una estructura unidimensional (CFI = 0,889; TLI = 0,873 y RMSEA = 0,047).

Conclusiones. El Test de Caras de Baron Cohen puede ser útil como instrumento de medida a pesar de no ser sensible al género y a la edad. Puesto que presenta un efecto techo, no resulta un instrumento adecuado para obtener medidas precisas del funcionamiento superior de esta habilidad.

Actas Esp Psiquiatr 2021;49(3):106-13 | ISSN: 1578-2735

INTRODUCTION

Facial emotion recognition (FER) is instrumental in competent social functioning. It has been thoroughly explored both in healthy subjects[1–3] and in different clinical populations[4–7].

Despite an extensive corpus of literature on the topic, most instruments to measure FER have poor or unknown psychometric properties. There are numerous validated datasets of pictures (see[8–14] for different examples). Depending on their particularities, these datasets can offer precise control of variables such as age, gender, ethnicity, and ecological validity, but they are very large, and they have caused methodological differences across studies that have limited comparability.

There is a myriad of tasks that measure FER, but their psychometric properties are generally poor[15] or proper validations are lacking. Some of them are the Ekman-60 Faces Test[16] , the Japanese and Caucasian Brief Affective Recognition Test (JACBARTT)[17], the Facial Emotion Recognition Test (FERT)[18], the Bell Lysaker Emotion Recognition Test (BLERT)[19] the Reading the Mind in the Eyes Test (RMET)[20] or The Videotest of Emotion Recognition[21].

Reviewing all the tasks and tests that measure FER is beyond the scope of this work, but we suggest consulting Passarelli et al, 2018[18] for a complete review.

Some of these tasks have been validated in some countries but not in others and some have been extensively used in research without having appropriate validations. Furthermore, most of them a relatively long or may appear as unnecessarily thorough in certain settings. A short version of The Assessment of Social Inference Test (TASIT)[22] has proven to be reliable as a screening measure[23], but there is little literature on short but sound tasks that are sensitive to use in clinical practice on other domains of social cognition.

In healthy population, there is sound evidence for females performing better than men since childhood and through adulthood[24,25] and for age to be a moderating variable. Despite inconsistency in some results, meta-analytic findings point to decay in the ability with age in all emotions but an unimpaired recognition of disgust, consistent with the natural aging brain[26,27].

The influence of the academic background in performance has been less studied than other demographic variables. The only study specifically examining a possible association found a significant positive correlation between educational level and the FERT and a significant interaction between age and schooling, favouring younger more educated subjects[28].

To the best of our knowledge, there are very few validated FER tests in Spanish population. The RMET has recently been validated in Spanish population[29,30], but its psychometric properties are not excellent. Another well-validated test in the Spanish population is the The Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT)[31], but it must be noted that the MSCEIT assesses all domains of social cognition instead of just FER.

Baron Cohen's Face Test (BCFT)[20] is a FER test originally developed in the construction of the RMET. This test is free for research and clinical use and it can be downloaded from the Autism Research Centre (https://www.autismresearchcentre.com/arc_tests). According to this website, this test has been translated to more than ten languages, but despite thorough searches, we have not been able to find literature on its validation or reference values in any healthy or clinical population.

It is possible that clinicians and researchers could be using this test with no knowledge of its psychometric properties. Furthermore, cultural and language backgrounds have been reported to influence neuropsychological measures[2], which raises awareness on validating tests in different cultural and ethnic populations.

Our aim with this study is to validate Baron Cohen's Face Test in a Spanish population, to investigate the psychometric properties of the instrument and to explore its sensibility to demographic variables.

METHODS

The Face Test was translated and adapted following the ITC guideline[33].

Because the test does not include any complete sentences but isolated words, a team of three people (bilingual English-Spanish, native Spanish and native English speakers) translated them attending to the frequency of use, European-Spanish forms and adapting the options of response to the stimuli's gender. A pilot version of the test was reviewed by peers so flaws could be detected and corrected.

Inclusion criteria included participants between 18 and 70 years of age who had signed the electronic informed consent. Exclusion criteria contemplated mental illness at the moment of the study, severe chronic mental illness, intellectual disability, developmental disorders, brain injury or dementia.

We used a snowball sampling method carried out by three researchers in three hospitals in three different

regions in Spain Parc Sanitari Sant Joan de Déu, Comunidad Terapéutica de Jaén and Hospital Clínico San Carlos. This approach was chosen to ensure we reached the maximum number of community participants from various regions in Spain. The first participants were chosen from the immediate social and working circles of the researchers and were sent an online survey to be completed from any electronic device, containing the tests listed below, a demographic form and questions on the participant's history of mental health. First participants were encouraged to disseminate the survey. Responses were inspected individually. Subjects using psychoactive medication at the time of the assessment were excluded, with the exception of the use of benzodiazepines as muscle relaxers.

The sample was recruited from July 2016 to January 2017 (including re-test). Three months after the first administration, 37 participants were sent a re-test online survey that included Baron-Cohen's Face Test and the Eyes Test. Participants were selected based on a randomized list. 24 participants replied to this survey.

Data were allocated in a server of the hospital that complies with all the safety requirements for the storage of health and research data.

Baron Cohen's Face Test[20]: consists of 20-items showing pictures of an actress displaying an emotion. Participants must choose which emotion the actress is feeling between two different choices of response. Half of the items display basic emotions, whilst the other half displays complex mental states. Examples of stimuli are displayed in figure 1.

The Reading the Mind in the Eyes Test[20]: The Eyes Test consists of 36 pictures of facial affect circumscribed to the eye region. Subjects must choose the emotion the eyes depict amongst four different response options. A glossary is provided and encouraged to use if the subject does not know the meaning of them.

Data Analysis

Data analyzed with IBM SPSS Statistics 22 and R[34]. We used tetrachoric correlation matrices to calculate Cronbach's alpha, test-retest reliability, and factor analysis. This approach was used to ensure a better adjustment to binary items. We used Cronbach's Alfa to test internal consistency. The intraclass correlation coefficient is not the optimal approach for binary items; therefore, we used Guttman Lambda 3 to test item by item time stability and Pearson correlations to test the full test time stability. Convergent validity was examined with a Pearson correlation between the total score in the Face Test and the total score in the Eyes Test. We used t-tests for independent measures to calculate differences between two samples. We used ANOVA tests to find mean differences among groups. We used Pearson correlations to test correlations between variables.

All the tests were run with a 95% confidence interval.

Ethics

The study was designed according to the World Medical Association Declaration of Helsinki[35] and it was approved by the Ethics Committee of Parc Sanitari Sant Joan de Déu as the coordinator center.

Figure 1 Stimuli 3, "surprised" *vs.* "happy" and 4, "angry" *vs.* "scared".

RESULTS

Initially, 286 people started responding to the online survey. A total of 8 subjects were excluded due to having a mental disorder. Other 67 subjects started responding to the survey but did not finish so their data could not be analyzed. The final sample included 211 participants, 134 females and 77 males between 19 and 70 years of age.

Table 1 Demographic information (n=211)

Age in years (Mean, SD)	40,01 (13,07)
Gender (%)	
Female	63,3
Male	36,7
Formal education (%)	
Primary	5,8
Secondary	16,4
3-year university degree	13
5-year university degree	39,1
Masters	20,3
PhD	5,3

Participants covered a scope of 15 different regions in the Spanish territory.

Normative data and psychometric properties

Our sample had a mean score of 18(SD=1.38), distributed between a minimum score of 14 and a maximum score of 20.

To test internal consistency, we calculated Cronbach's alfa based on the tetrachoric correlation matrix, which yielded a value of 0.75 showing good internal consistency. This value would not be significantly increased by removing any item. We calculated the convergent validity by correlating the total score of the BCFT with the total score of the Eyes Test, which showed a low but significant correlation (r=0.192, p<0.005).

Test-retest stability as measured by the Pearson correlation between the first and the second application of the tests was r=0.372, p=0.088. Subsequently, we used the tetrachoric correlation matrix to obtain the Guttman lambda 3 indexes for each item. This index is equivalent to Cronbach's alfa [36]. Table 2 below exposes the agreement indexes for test-retest reliability.

Table 2 Agreement indexes for test-retest reliability

Guttman Lamda 3 Values	Items
> 0,8	1, 2, 3, 4, 5, 6, 8, 9, 10, 11, 13, 14, 15, 16
0,60 - 0,8	7, 17, 19
0,40 - 0,60	12, 20
0,20 - 0,40	18

DIMENSIONAL STRUCTURE

A one-factor solution was tested, as this is the model suggested by the author.

We performed a one-factor confirmatory factor analysis (CFA) for all the items. We did not include item 1 because it had a negative variance and 3 and 4, that remained constant.

We used the diagonally weighted least squares (DWLS) estimator, with which we obtained: Comparative Fit Index (CFI)=0.889; Tucker-Lewis Index (TLI)=0.873 and Root Mean Square Error of Approximation (RMSEA)=0.047.

Moderating variables: gender, age and academic level

No differences between genders were found (t(209)=0.865; p=0.388). Correlations between the total score in the test and age were not significant (r=-0.075, p=0.280).

We performed a one way ANOVA between the academic level and the total score in the test. We did not find differences between groups (F(5,200)=1.886; p=0.098).

We found a weak but significant correlation between performance in BCFT and academic level (r=0.153; p=0.028).

DISCUSSION

This work presents the validation of BCFT together with its reference values in a healthy Spanish population, obtains its psychometric properties and discusses moderating variables.

As for BCFT, this result is consistent with results found in general population, which find that the ability works generally well and reaches high performance with relatively small standard deviations [2,26]. Our sample reached similar scores than the original test [20], who reported a mean of

9.13(SD=0.96) in basic emotions and a mean of 9.38(SD=0.62) in complex emotions. However, the test presented a ceiling effect in our sample. In this case, this effect could have happened because the test is short and the items only have two choices, facilitating the chances of guessing the right emotion even if the ability is impaired.

Regarding its psychometric properties, internal consistency reached a value of α=0.75, which would not be significantly increased by removing any item.

Other tasks have yielded similar internal consistency: the JACBART has yielded Cronbach's alfa between 0.86 and 0.92[17], the Videotest of Emotion Recognition reported two Cronbach's alfa , one for the accuracy index (0.74) and the other one for the sensitivity index (0.79)[21]. The validation of the MSCEIT in Spanish population obtained an alfa of 0.80[31]. In Spanish population, the RMET yielded an alfa of 0.56[30].

BCFT reached a slightly lower value that can be considered adequate.

Literature reporting both internal consistency and test-retest reliability is very scarce. Test-retest reliability using the JACBART is between 0.44 for anger and 0.72 for sadness (computed with t-tests), which suggests a practice effect[17]. To the best of our knowledge, the only test-retest reliability data published for Spanish population is a test-retest study of the Reading the Mind in the Eyes Test, as studied with the Bland-Altman method, which yielded a score of 0.63[29].

Test-retest reliability using the whole test yielded a moderate but significant correlation (r=0.372, p=0.088). To further evaluate test-retest reliability, we calculated the Guttman L3 index for each item to test time stability, as this is a more appropriate approach for binary items. Guttman L3 index is equivalent to Cronbach's alfa and can be interpreted likewise [36]. In our sample, 17 out of 20 items had a value of over 0.6; which allows us to conclude that the test has excellent stability over time. The discrepancy between the test-retest correlation using the whole test and the item by item test-retest may be due to a smaller sample size at retest.

The test was designed following a one-factor model. We only explored a one-factor solution as exploring further factor solutions was beyond the scope of this project. Our results barely reached values to assume adjustment to a one-factor model. Future studies with this test should test for other factor solutions and confirm our findings.

To the best of our knowledge, this is the first published validation of this test. Future studies on this test should report psychometric properties to draw further conclusions.

We did not find gender differences in performance. This is a surprising finding since a female superiority is found consistently across studies with a small but similar size effect[18,25,31]. Lyusin et al (2016) did not find gender differences either in their 7-item task. In their case, the authors attribute it to their more ecological paradigm, which they speculate could be balancing female's advantage[21]. However, it could be possible that a female superiority can only be detected with longer tests.

Interestingly, a cross-cultural study comparing performance in FERT in Brazilian and French population did not find gender differences in the Brazilian sample, but found a female superiority in the French sample[28].

These findings highlight that some aspects of the interaction between FER and demographic variables may not be static across cultures. As a reason for these findings, the authors suggest that perceived gender stereotypes and gender equality measures could influence facial emotion recognition[28].

Similarly, we did not find an influence of age consistent with the literature[26,27]. This is an expected finding in our study since our sample is clustered in the middle age and we lack sufficient subjects of young and old age.

We did not find any mean differences between different academic levels and performance. This is not consistent with some of the literature[16,28]. According to Lindquist (2014), semantic memory plays a crucial role on labelling emotions, regardless of their valence[37]. We speculate although a broader education could provide more opportunities to interact with different emotional contexts and access to more emotional vocabulary, BCFT does not represent a semantic challenge.

Taken together, it seems that although BCFT presents adequate psychometric properties, it has a very low ceiling and it is not sensitive to the demographic factors that, according to the literature, play an important role in the recognition of facial expressions of emotion. These drawbacks may hamper this test's ability to detect subtle decay in FER. However, from our data, it can be derived that because subjects in the general population reach ceiling performance, scores lower than a standard deviation truly reflect deficits in FER. However, this interpretation should be taken with caution until future research examines sensibility and specificity in different pathological populations.

This work should be interpreted in light of some strengths and weaknesses:

As for strengths, this is the first validation published for BCFT, and one of the first validations of a FER test for

Spanish population. This work could offer a framework for clinicians and researchers already using this test, and for other teams developing normative measures of FER for Spanish population.

We were able to reach subjects from all the Spanish territory, which increases the validity of our results.

As for limitations, it must be noted that even if a translation is accurate, cultural differences and familiarity with the words can alter the difficulty of the item or induce different responses to the test. This is of particular importance regarding that the test was designed for British population and adapting it to Spanish may have influenced our subjects' responses. We believe this could be especially true for the ten items assessing complex mental states.

Besides, this test is short, and it displays maximum intensity of the emotion. This is very likely to diminish the test's ability to detect excelling performance or subtle decay in FER.

Although using an online survey allowed us to recruit a sample from different regions in Spain, losing face-to-face contact with the subjects could have hampered our results.

Finally, we failed to recruit a balanced sample, neither in gender nor age or education. We believe our female sample is big enough to detect an advantage in performance, however, it has a small proportion of old and very young people and it over-represents subjects with higher education.

In spite of its limitations, we believe this work gives a valuable resource to researchers and clinicians in Spain. Further studies examining sensitivity-specificity in different populations or with subjects in old age and with less academic background are recommended.

Acknowledgments. We want to acknowledge Cátedra UAM-ASISA (Gestión Sanitaria y Economía de la Salud) for funding this project. We want to acknowledge CIBERSAM (Centro de Investigación Biomédica en Red – Salud Mental) for earmarking funds for this project. We are grateful to the volunteers that took the time to participate in this research.

Declaration of interest statement. None of the authors have interests to disclose.

REFERENCES

1. Christensen KJ, Riley BE, Heffernan KA, Love SB, Sta. Maria MEM. Facial Recognition Test in the Elderly: Norms, Reliability and Premorbid Estimation. Clin Neuropsychol (Neuropsychology, Dev Cogn Sect D) [Internet]. 2002 Feb 1 [cited 2017 Nov 6];16(1):51–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11992226
2. Kessels RPC, Montagne B, Hendriks AW, Perrett DI, de Haan EHF. Assessment of perception of morphed facial expressions using the Emotion Recognition Task: normative data from healthy participants aged 8-75. J Neuropsychol [Internet]. 2014 Mar [cited 2016 Sep 27];8(1):75–93. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23409767
3. Rodger H, Vizioli L, Ouyang X, Caldara R. Mapping the development of facial expression recognition. Dev Sci. 2015;18(6):926–39.
4. Kohler CG, Walker JB, Martin EA, Healey KM, Moberg PJ. Facial emotion perception in schizophrenia: a meta-analytic review. Schizophr Bull [Internet]. 2010 Sep [cited 2016 Oct 6];36(5):1009–19. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19329561
5. Samamé C, Martino DJ, Strejilevich SA. Social cognition in euthymic bipolar disorder: systematic review and meta-analytic approach. Acta Psychiatr Scand [Internet]. 2012 Apr [cited 2016 Oct 13];125(4):266–80. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22211280
6. Pringle A, Harmer CJ, Cooper MJ. Biases in emotional processing are associated with vulnerability to eating disorders over time [Internet]. Eating Behaviors Jan, 2011 p. 56–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21184974
7. Dalili MN, Penton-Voak IS, Harmer CJ, Munafò MR. Meta-analysis of emotion recognition deficits in major depressive disorder. Psychol Med [Internet]. 2015 Apr [cited 2016 Oct 6];45(6):1135–44. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25395075
8. Dalrymple KA, Gomez J, Duchaine B, Anastasi J, Rhodes M, Wright D, et al. The Dartmouth Database of Children's Faces: Acquisition and Validation of a New Face Stimulus Set. Urgesi C, editor. PLoS One [Internet]. 2013 Nov 14 [cited 2016 Aug 18];8(11):e79131. Available from: http://dx.plos.org/10.1371/journal.pone.0079131
9. Ekman P, Friesen W V. Pictures of Facial Affect [Internet]. Consulting Psychologists Press, editor. Palo Alto, CA; 1976 [cited 2017 Apr 18]. Available from: https://books.google.es/books/about/Pictures_of_Facial_Affect.html?id=gbfMSgAACAAJ&tredir_esc=y
10. Langner O, Dotsch R, Bijlstra G, Wigboldus DHJ, Hawk ST, van Knippenberg A. Presentation and validation of the Radboud Faces Database. Cogn Emot [Internet]. 2010 Dec [cited 2016 Nov 10];24(8):1377–88. Available from: http://www.tandfonline.com/doi/abs/10.1080/02699930903485076
11. Marchewka A, Żurawski Ł, Jednoróg K, Grabowska A. The Nencki Affective Picture System (NAPS): Introduction to a novel, standardized, wide-range, high-quality, realistic picture database. Behav Res Methods [Internet]. 2014 Jun [cited 2017 Nov 6];46(2):596–610. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23996831

12. Vaiman M, Wagner MA, Caicedo E, Pereno GL. Development and validation of an Argentine set of facial expressions of emotion. Cogn Emot [Internet]. 2017 Feb 17 [cited 2017 Nov 6];31(2):249–60. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26479048
13. 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 [Internet]. 2015 Jan 5 [cited 2017 Nov 6];5:1516. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25601846
14. 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 [Internet]. 2009 Aug 15 [cited 2019 May 2];168(3):242–9. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0165178108001480
15. Pinkham AE, Penn DL, Green MF, Buck B, Healey K, Harvey PD. The Social Cognition Psychometric Evaluation Study: Results of the Expert Survey and RAND Panel. Schizophr Bull [Internet]. 2014 Jul [cited 2016 Aug 1];40(4):813–23. Available from: https://academic.oup.com/schizophreniabulletin/article-lookup/doi/10.1093/schbul/sbt081
16. Dodich A, Cerami C, Canessa N, Crespi C, Marcone A, Arpone M, et al. Emotion recognition from facial expressions: a normative study of the Ekman 60-Faces Test in the Italian population. Neurol Sci [Internet]. 2014 Jul 19 [cited 2018 Jan 4];35(7):1015–21. Available from: http://link.springer.com/10.1007/s10072-014-1631-x
17. Matsumoto D, LeRoux J, Wilson-Cohn C, Raroque J, Kooken K, Ekman P, et al. A New Test to Measure Emotion Recognition Ability: Matsumoto and Ekman's Japanese and Caucasian Brief Affect Recognition Test (JACBART). J Nonverbal Behav [Internet]. 2000 [cited 2019 May 20];24(3):179–209. Available from: http://link.springer.com/10.1023/A:1006668120583
18. Passarelli M, Masini M, Bracco F, Petrosino M, Chiorri C. Development and validation of the Facial Expression Recognition Test (FERT). Psychol Assess [Internet]. 2018 Nov [cited 2019 Apr 16];30(11):1479–90. Available from: http://www.ncbi.nlm.nih.gov/pubmed/30024180
19. Bryson G, Bell M, Lysaker P. Affect recognition in schizophrenia: a function of global impairment or a specific cognitive deficit. Psychiatry Res [Internet]. 1997 Jul 4 [cited 2017 Apr 24];71(2):105–13. Available from: http://www.sciencedirect.com.sire.ub.edu/science/article/pii/S0165178197000504?via%3Dihub
20. Baron-Cohen S, Wheelwright S, Jolliffe and T. Is There a &tquot;Language of the Eyes&tquot;? Evidence from Normal Adults, and Adults with Autism or Asperger Syndrome. Vis cogn [Internet]. 1997 Sep [cited 2016 Oct 26];4(3):311–31. Available from: http://www.tandfonline.com/doi/abs/10.1080/713756761
21. Lyusin, Dmitry; Ovsyannikova V. Measuring two aspects of emotion recognition ability: Accuracy vs. sensitivity. Learn Individ Differ [Internet]. 2016 Dec 1 [cited 2019 May 20];52:129–36. Available from: https://www-sciencedirect-com.sire.ub.edu/science/article/pii/S1041608015001090
22. McDonald S, Flanagan S, Rollins J, Kinch J. TASIT: A new clinical tool for assessing social perception after traumatic brain injury. J Head Trauma Rehabil [Internet]. [cited 2018 Nov 20];18(3):219–38. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12802165
23. McDonald S, Honan C, Allen SK, El-Helou R, Kelly M, Kumfor F, et al. Normal adult and adolescent performance on TASIT-S, a short version of The Assessment of Social Inference Test. Clin Neuropsychol [Internet]. 2018 May 19 [cited 2019 Sep 3];32(4):700–19. Available from: https://www.tandfonline.com/doi/full/10.1080/13854046.2017.1400106
24. McClure EB. A meta-analytic review of sex differences in facial expression processing and their development in infants, children, and adolescents. Psychol Bull [Internet]. 2000 May [cited 2016 Sep 27];126(3):424–53. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10825784
25. Thompson AE, Voyer D. Sex differences in the ability to recognise non-verbal displays of emotion: A meta-analysis. Cogn Emot [Internet]. 2014 Oct 3 [cited 2016 Aug 8];28(7):1164–95. Available from: http://www.tandfonline.com/doi/abs/10.1080/02699931.2013.875889
26. Lawrence K, Campbell R, Skuse D. Age, gender, and puberty influence the development of facial emotion recognition. Front Psychol [Internet]. 2015 Jun 16 [cited 2016 Aug 16];6:761. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26136697
27. 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.
28. de Souza LC, Bertoux M, de Faria ÂRV, Corgosinho LTS, Prado AC de A, Barbosa IG, et al. The effects of gender, age, schooling, and cultural background on the identification of facial emotions: a transcultural study. Int Psychogeriatrics [Internet]. 2018 Dec 25 [cited 2019 Mar 13];30(12):1861–70. Available from: https://www.cambridge.org/core/product/identifier/S1041610218000443/type/journal_article
29. Fernández-Abascal EG, Cabello R, Fernández-Berrocal P, Baron-Cohen S. Test-retest reliability of the "Reading the Mind in the Eyes" test: a one-year follow-up study. Mol Autism [Internet]. 2013 Sep 11 [cited 2018 Apr 4];4(1):33. Available from: http://molecularautism.biomedcentral.com/articles/10.1186/2040-2392-4-33
30. Redondo I, Herrero-Fernández D. Validation of the Reading the Mind in the Eyes Test in a healthy Spanish

sample and women with anorexia nervosa. Cogn Neuropsychiatry [Internet]. 2018 Jul 4 [cited 2019 May 20];23(4):201–17. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29635964

31. Extremera N, Fernández-Berrocal P, Salovey P. Spanish version of the Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT). Version 2.0: reliabilities, age and gender differences. Psicothema [Internet]. 2006 [cited 2018 Sep 25];18 Suppl:42–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17295956
32. Flores I, Casaletto KB, Marquine MJ, Umlauf A, Moore DJ, Mungas D, et al. Performance of Hispanics and Non-Hispanic Whites on the NIH Toolbox Cognition Battery: the roles of ethnicity and language backgrounds. Clin Neuropsychol [Internet]. 2017 May 19 [cited 2019 Sep 3];31(4):783–97. Available from: https://www.tandfonline.com/doi/full/10.1080/13854046.2016.1276216
33. Muñiz J, Elosua P, Hambleton RK, International Test Commission. [International Test Commission Guidelines for test translation and adaptation: second edition]. Psicothema [Internet]. 2013 [cited 2018 Mar 2];25(2):151–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23628527
34. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing [Internet]. Vienna, Austria: ISBN 3-900051-07-0; 2012. Available from: https://www.r-project.org/
35. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA [Internet]. 2013 Nov 27 [cited 2017 Jul 6];310(20):2191–4. Available from: http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.2013.281053
36. Guttman L. A basis for analyzing test-retest reliability. Psychometrika [Internet]. 1945 Dec [cited 2017 Mar 30];10(4):255–82. Available from: http://link.springer.com/10.1007/BF02288892
37. Lindquist KA, Gendron M, Barrett LF, Dickerson BC. Emotion perception, but not affect perception, is impaired with semantic memory loss. Emotion [Internet]. 2014 Apr [cited 2019 May 21];14(2):375–87. Available from: http://doi.apa.org/getdoi.cfm?doi=10.1037/a0035293