Revista ConCiencia EPG–Edición Especial Marzo 2022
ISSN digital: 2517-9896 ISSN impreso: 2517-9896
https://doi.org/10.32654/ConCienciaEPG/Eds.especial-1

Online version of a Theory of Mind test in Mexicans

Versión en-línea de una prueba para evaluar la teoría de la mente en población mexicana

Recepción del artículo: 04.09.21 | Aceptación del artículo: 15.02.22

Bernarda Téllez Alanís[1]
btellez@uaem.mx
https://orcid.org/0000-0003-2940-9283

Rubén Avilés Reyes[2]
ruben.aviles.reyes@uabc.edu.mx
https://orcid.org/0000-0001-5328-3081

Itzel Alonso Carrillo[3]
itzel.alonso@uaem.mx
https://orcid.org/0000-0001-5234-0697

Sandra Verónica Meza Cavazos[4]
sandra.mezacvz@uanl.edu.mx
https://orcid.org/0000-0001-6123-8360

Erwin Rogelio Villuendas González[5]
erwin.villuendas@umich.mx
https://orcid.org/0000-0001-9141-8514

[1]Centro de Investigación Transdisciplinar en Psicología, Universidad Autónoma del Estado de Morelos.
[2]Universidad Autónoma de Baja California
[3]Universidad Autónoma del Estado de Morelos.
[4]Facultad de Psicología. Universidad Autónoma de Nuevo León
[5]Facultad de Psicología. Universidad Michoacana de San Nicolás de Hidalgo

Para referenciar este artículo:

Téllez Alanís, B., Avilés Reyes, R., Alonso Carrillo, I., Meza Cavazos, S. V., & Villuendas González, E. (2022). Online version of a Theory of Mind test in Mexicans. *Revista ConCiencia EPG,* 7 (Edición Especial), 1-22. https://doi.org/CONCIENCIAEPG/ Eds. Especial-1

Autor de correspondencia:
Bernarda Téllez Alanís
btellez@uaem.mx

Abstract

The "Reading the Mind in the Eyes test" (RMET), also known as "The Eyes test", is one of the most widely used tests to study Theory of Mind (ToM) and has been translated to several languages around the world. The aim of this study was to present an online version of the RMET designed for a Mexican population (Barrera-Valdivia et al., 2019) and to explore sex and age-related effects on the performance of the test. 1489 adults (1054 female, 71%) age 18 to 87 (M=30, SD=15) participated in the study. Participants were instructed to choose the word that best described what the person in each picture was feeling among four options. Stimuli consisted of full color images of 45 people's eyes who expressed a variety of mental states. Results showed that 30 stimuli met the selection criteria: ≥50% selected on target and ≤25% on any of the foils. Average accuracy in response was 76% (SD= 11%) with observable differences by sex and age. Response accuracy allowed to classify stimuli into three difficulty levels (easy, intermediate, and difficult). Results showed that the online version of the RMET is an adequate assessment tool for ToM research in Mexican population as well as other Latin Americans communities that share cultural features.

Key words: Theory of Mind test, online test, mental states, eyes test.

Acknowledgement. Thanks to Dr. Roque Méndez from Texas State University for providing us with the test

Revista ConCiencia EPG–Edición Especial Marzo 2022
ISSN digital: 2517-9896 ISSN impreso: 2517-9896
https://doi.org/10.32654/ConCienciaEPG/Eds.especial-1

Resumen

La prueba "Lectura de la mente en los ojos" o la "prueba de las Miradas" es uno de los test más utilizados para estudiar la Teoría de la Mente. De esta prueba existen traducciones en varios idiomas alrededor del mundo. El objetivo de este documento es presentar una versión en línea de la prueba de las Miradas diseñada para población mexicana (Barrera-Valdivia et al., 2019) y explorar los efectos del sexo y la edad en la ejecución de la prueba. Participaron 1489 adultos (1054 mujeres, el 71%) entre 18 y 87 años (edad media ± DE, 30 años ± 15). Los participantes debían elegir una palabra, entre cuatro opciones, la que mejor describiera el estado mental expresado en los ojos de las personas. Los estímulos fueron 45 fotografías a color representando una variedad de estados mentales. Los resultados mostraron que 30 estímulos que cumplieron los requisitos de selección: selección del estímulo blanco ≥.50% y selección ≤25% de cualquiera de los distractores. La ejecución correcta promedio fue del 76% (DE=11%). Se observaron diferencias de acuerdo con el sexo y a la edad. Adicionalmente los estímulos se clasificaron por su nivel de ejecución correcta en tres niveles de dificultad (difícil, medio, fácil). Estos resultados demostraron que esta versión en línea de la prueba de las Miradas es adecuada para realizar investigación sobre teoría de la mente con población mexicana, así como en comunidades latinoamericanas con los que se comparten características culturales.

Palabras clave: Teoría de la mente, prueba en línea, estados mentales, test de las miradas.

Agradecimiento: Al Dr. Roque V. Méndez Canales de la Universidad Estatal de Texas por proporcionarnos la prueba.

Introduction

Premack and Woodruff (1978) suggested that the Theory of Mind (ToM) is the ability to attribute mental states to oneself and others, indicating that it involves a system of inferences based on observable information. These inferences can help us understand and predict behavior not only of those people we interact with, but of ourselves as well. ToM is currently considered a crucial component of social cognition: processes that allow us to behave in a way that takes into account other individuals of our same species (Adolphs, 1999), primarily the complex and flexible

behaviors like the ones observed in primates. According to the social cognition model used for this study, three different components have been recognized: detection of facial emotions, empathy, and ToM (Atenas et al., 2019). These processes are intertwined, and several neural networks have been associated with their functioning (VanOverwalle, 2009; Adams et al., 2009; Fernandez et al., 2018). Social cognition is essential to human bonding: it can determine if a relationship outcome is either healthy and successful, or unsuccessful and unbalanced.

Test development is relevant within the context of research in cognitive psychology, cognitive neurosciences, neuropsychology, neuropsychiatry, among other disciplines, to allow for the assessment of social cognition processes, and specifically, ToM. The tests used most often to assess ToM are the "False Belief task" by Wimmer and Perner (1983) and Baron-Cohen et al. (1985), the "Strange Stories task" by Happé (1994), the "Faux pas" test by Baron-Cohen et al. (1999), and the "Reading the Mind in the Eyes test" (RMET) by Baron-Cohen et al. (1997, 2001). Other tests that have also been used are the "Yoni Task", by Shamay-Tsoory and Aharon-Peretz (2007) and the "Movie for Assessment of Social Cognition" (MASC) from Dziobek et al. (2006). Many studies and tests have been created to better understand the dimensions of ToM, as portrayed by Wellman et al. (2001) and VanOverwalle (2009). However, not many instruments have been developed or adapted to measure these processes in a Mexican population.

The RMET, published by Baron-Cohen et al. in 1997, was developed to test a person's ability to detect emotional and mental states out of photographs of other people's eyes alone. The original test involved showing 25 black and white pictures taken out of magazines with male and female models posing different facial expressions. Participants would choose one of two possible responses with an opposite mental state term. (e.g., serious–playful). Afterwards, Baron-Cohen et al. (2001) improved the test by including more items (they tested 40 and selected 36), with four possible responses (reducing chance performance); included complex mental states, not only basic emotions; and excluded items that could be resolved with the direction of the eyes (ignoring-attentive). The authors also included the same number of male and female faces and added a glossary with the meanings of the mental states used in the test so the

participants could consult it if they had any questions about any of the terms used.

The RMET assesses ToM and has been extensively used worldwide. The main findings include performance differences between sexes. Females have been shown to perform better (Kirkland et al., 2013), aside from than individuals with autism tend to have lower performance (Baron-Cohen et al., 1997, 2001; Peñuelas-Calvo et al., 2019). Similarly, persons with major depressive disorder may display lower performance compared to those not depressed (Nejati, 2018) like those with schizophrenia (Charernboon & Patumanond, 2017), or alcohol dependence (Maurage et al., 2011), among others. It is important to develop this cognitive-emotional ability to recognize the mental states in others because it allows or promotes a social and empathic interaction much-needed in today's world. Regarding the reliability of the RMET, no indicators were calculated in either of the original versions of Baron-Cohen et al. (1997, 2001), nor in the version by Barrera-Valdivia (2019). On the other hand, various investigations have analyzed the psychometric properties of the RMET and have reported internal consistency indicators of reliability under .60. A meta-analysis of the RMET has recently been published by Kittel et al. (2021) who reported an internal consistency of .73 including 21 subsamples (n = 4305) with a range between .45 to .96, in addition 50% presented indicators below .70. The convergent validity of the test has been demonstrated with other ToM tests such as the Faux Pas Test (Torralva et al., 2009; Ferguson & Austin, 2010) and the Cognitive Empathy Test (Olderback et al., 2015).

Spanish language adaptations of the test have been developed for Spain (Huerta-Ramos et al., 2016), Argentina (Roman et al., 2012) and Chile (Maureira et al., 2019). Barrera-Valdivia et al. (2019) adapted a version of this test to a sample of Mexican students using full color, high resolution photographs. The models on the photographs were actors with Mexican physiognomy of different ages. These authors made a comparison between the words and the images of the original English test and proposed a set of images and words in Spanish for the Mexican version. They reported that the suggested version was favorable at 65% when compared to the images of the English version. They stated that it is, therefore, better to use the adapted version in the context of its population than the original in English, as several studies have shown individuals are able to recognize facial

Revista ConCiencia EPG–Edición Especial Marzo 2022
ISSN digital: 2517-9896 ISSN impreso: 2517-9896
https://doi.org/10.32654/ConCienciaEPG/Eds.especial-1

expressions from members of their own culture better than from individuals from other cultures (Elfenbein & Ambady, 2003). The same happened with the execution on the RMET by Adams et al. (2009).

The main purpose of this study was to develop an online adaptation of the Barrera-Valdivia et al. (2019) version of the RMET, which is in itself an adaptation of the stimuli from the RMET by Baron-Cohen et al. (2001). However, this study uses an updated set of full color pictures. The task consisted of 45 trials in which a picture of eyes-only area was shown, and participants were instructed to choose, among the four words written below the image, the one that best represented the gaze's expression. Models for all the pictures were Mexican. Therefore, it shows how many and which stimuli accomplish the two selection criteria previously proposed by Baron-Cohen et al. (2001). These authors stated which response options should be assigned: A) over 50% of the participants select the target word correctly and B) none of the three foils is selected by more than 25%. This strategy prevents having response options where a foil performs above chance, considering chance is 25% in a forced-choice of 4 items. This also guarantees that the target word is twice the value of chance, based on data obtained from the wide sample of participants. As indicated by Chapman and Chapman (1978), multiple-choice tests are expected to have higher reliability and observed-score variance, ranging between chance and 100%, for accuracy levels. Additionally, this study explored if there is a difference in performance according to sex and age, as reported by previous literature (sex: Baron-Cohen et al., 1997; Baron-Cohen et al., 2001; Kirkland et al., 2013; age: Lee et al., 2021).

It is important for both research and clinical evaluation purposes to have modern instruments adapted to any population according to its particular cultural and social characteristics in order to obtain the cross-cultural comparisons according to the International Test Commission (2017). This study aims to contribute with a test that is useful in a Mexican context and, among the Latin American population that, in general, share similar facial features and culture.

Furthermore, this study was conducted during the COVID-19 lockdown in Mexico. The decision to use a web application was made to avoid the spread between researchers and participants, as well as to evaluate the viability of the virtual modality for this test, and thereby contribute to the practice of tele-research

Revista ConCiencia EPG–Edición Especial Marzo 2022
ISSN digital: 2517-9896 ISSN impreso: 2517-9896
https://doi.org/10.32654/ConCienciaEPG/Eds.especial.1

and tele-psychology. Within the COVID-19 pandemic context, it is generally advised to use online testing to reduce social contact and minimize contagion (Bilder et al., 2020). Also, online tests help reduce travel times for patients, increasing the geographical scope of application, and diminishing costs. Data that has been obtained through online based instruments is considered to be efficient, effective and reliable, similar traditionally obtained data (British Psychological Society, 2020; Elosua, 2020).

Method

Participants: For this assessment, 1489 Mexican participants logged in, gave informed consent, and performed the test in its entirety. Only 9 did not give consent after logging in, and therefore, did not perform the test. It involved the participation of 1054 females (71%) and 435 males (29%) ages 18 to 87 (M=mean, SD=standard deviation, M=30, SD=15). Demographic data of the participants are shown in Table 1.

Table1

Participant´s demographics: age, educational attainment, occupation, and location (n= 1489).

Age range (years)	% Age	Education	% Education	Occupa-tion	% Occupation	Location	% Location
18-20	37	Postgraduate	11	Student	57:0	Morelos	21.0
21-30	29	Bachelor's Degree	63	Employed	23.8	Baja California	19.7
31-40	13	High school	21	Homemaker	2.4	Michoacán	17.5
41-50	8	Secondary or less	5	Retired or pensioner	2.0	Nuevo León	15.3
51-60	7			Unemployed	3.3	Other 28 states	21.8
>60	6			No data	11.6	Abroad	1.5
						No data	3.1

Instruments: An on-line version of Barrera-Valdivia et al. (2019) adaptation of RMET by Baron-Cohen et al. (2001). The test consists of 45 trials in which a full color photograph depicting only the eyes region of a face is presented, along with four words (one target, three foils) that describe different mental states. Participants must

choose the word that best represents the emotion depicted through the gaze in the picture. The images were displayed until a response was registered. Response time for each item, however, were not registered. Physical properties (such as size, brightness, ambiance lightness, distance) of the displayed images were neither registered nor controlled, as they depended upon the device used to respond to the test. Models for all the photographs were Mexican adults (19 female and 26 male) of different ages (young adults n=16, middle-aged n=19; older adults = 10). The entire test was administered through Google Forms. For test inquiries, please contact Dr. Roque V. Mendez Canales at rm04@txstate.edu

Procedure: A national call addressed to Mexican adults was made through social media. As the test was to be performed on-line, subjects (mostly students) were required to be able to use a device with internet access. In the beginning section of the test, subjects were informed about the characteristics of the study and asked to give an informed consent if they decided to participate. The test included 5 demographic questions, and then proceeded with the 45 trials. Google Forms does not allow to register test response times, so subjects were asked to provide the time at the beginning and once they finished. Participants were also asked which device they used to perform the test. The test stayed online from February 15 through July 31, 2021.

Data Analysis: The percentage of selection of each target word and their foils was calculated. The stimuli were selected based on the following criteria: ≥50% responses on the target and ≤25% responses on any one of the foils (Baron-Cohen et al., 2001). Mean values from the 45 faces and the 28 stimuli that met the criteria proposed by Baron-Cohen et al. (2001) were obtained. Furthermore, a repeated measures ANOVA was used to assess the execution, using the total stimuli of the beginning, middle and end (15 stimuli each) and to determine if there were effects of fatigue by the end of the test. After analyzing the results, the variability in the percentage of correct answers observed ranged between 54% and 95%. For this reason, the faces were grouped into three levels of difficulty according to the percentage of selection of the target stimulus (Chapman y Chapman, 1978): difficult (51-70%), intermediate (71-80%), and easy (81-100%). A t-test was performed to compare the differences between the sexes. An ANOVA test was used

to analyze the factors of difficulty by sex. Finally, a 6 x 2 ANOVA was conducted to explore age and sex effects on accuracy using a subsample of 92 random participants from each age group except the older adults' group that already contained 92 participants.

Results

The average time used to solve the test was 12.5 minutes (SD=5.6; ranging between 4 and 40 minutes, n=1286). Results showed that 65% of the participants performed the test on a cell phone device, 26% on a computer, 1% on a tablet, and 8% did not specify. Only 30 stimuli met the criteria of selection ≥50% on the target and selection ≤25% on any one of the foils (Table 2), resulting in 11 female and 19 male faces.

Table 2

Percentage of selection.

Number	Target	Target %	Foil 1 %	Foil 2 %	Foil 3 %	Sex (M Male, F Female)
1	Paniqueado	94.2	1.3	1.7	2.9	M
2	Desilusionado	91.1	0.5	4.2	4.2	M
3	Cauteloso	87.3	0.7	10.9	1.0	M
4	Hostil	50.4	37.5	1.9	10.1	M
5	Dudoso	65.0	26.6	6.3	2.1	F
6	Decidido	65.1	22.9	8.4	3.6	M
7	Preocupado	89.3	4.85	0.9	5.0	M
8	Desafiante	95.0	3.3	0.9	0.9	M
9	Inquieto	49.2	13.5	36.4	0.9	F
10	Culpable	74.1	20.3	1.1	4.4	M
11	Fascinado	44.4	32.4	14.2	9.1	M
12	Seguro	62.1	5.0	31.7	1.2	F
13	Juguetón	74.7	4.4	15.2	5.7	M
14	Impactante	85.6	12.1	1.1	1.2	F
15	Deseo	50.6	31.3	16.7	1.4	F
16	Insistente	35.4	5.4	25.5	33.6	M
17	Preocupado	86.0	0.7	3.1	10.1	M
18	Fantaseando	76.0	13.8	4.4	5.8	F
19	Desanimado	69.1	15.2	3.2	12.4	M
20	Preocupado	66.5	3.4	21.2	8.9	F
21	Arrepentido	45.7	5.5	14.6	34.2	M
22	Ofendido	24.4	8.5	33.3	33.7	M
23	Escéptico	71.3	3.6	21.2	4.0	M
24	Deprimido	54.4	13.1	8.4	24.1	M
25	Indeciso	41.1	42.5	3.2	13.5	F
26	Considerando	46.3	44.7	4.9	44.1	M
27	Decidido	74.1	5.7	18.3	1.9	F
28	Indeciso	40.9	12.4	17.5	29.2	F
29	Amistoso	62.3	2.3	12.6	22.8	M

30	Aliviado	52.6	5.1	13.2	29.1	F
31	Preocupado	91.9	0.8	2.1	5.2	F
32	Desafiante	88.0	8.1	0.4	3.5	M
33	Decepcionado	58.2	13.6	16.9	11.3	M
34	Interesado	76.3	11.5	6.3	5.9	F
35	Serio	62.0	7.7	10.5	19.8	M
36	Cauteloso	75.3	10.7	11.3	2.8	F
37	Interesado	48.0	3.0	19.1	30.0	F
38	Coqueteando	78.5	10.6	9.5	1.4	F
39	Reflexivo	82.5	6.7	9.0	1.8	F
40	Seguro	62.2	13.1	21.2	3.5	F
41	Serio	78.3	11.6	0.7	9.4	M
42	Conmovido	61.2	13.8	23.0	1.9	M
43	Desconfiado	55.2	10.3	2.5	32.0	M
44	Nervioso	69.3	16.6	6.0	8.1	F
45	Suspicaz	73.1	4.2	16.3	6.4	M

Only 30 stimuli met the criteria of selection ≥50% on the target and selection ≤25% on any one of the foils. Those that did not meet the criteria are marked in gray.

On average, subjects responded accurately in 29 out of the 45 stimuli (64%, SD=9%), ranging from 13 to 41 correct responses. The mean correct execution with the selected 30 stimuli was 23 correct responses (76%, SD=11%), ranging from 7 to 30 correct responses. To rule out the effect of mental fatigue, we compared accuracy within the beginning (M=72%, SD=11.4), middle (M=56%, SD=13) and end (M=71%, SD=13.6) stages of the test, and differences were found ($F_{2,\ 2672}$ =833, p<.001, η^2=.24). Although post hoc comparisons showed differences between midpart and both the beginning and end (but not between beginning and end), highest accuracy was observed both at the beginning and the final part of the test, deeming out fatigue as an explanation. Two thirds of the stimuli that were discarded due to low accuracy (below 50% correct responses) were in the midpart of the test (see table 2).

On the other hand, after observing that the selected items differ widely in percentage of responses obtained, we decided to group them into 3 levels. This was under the assumption that the items with a higher percentage of correct responses are easier compared to those where the percentage of correct responses was lower and are therefore considered more difficult (Chapman and Chapman, 1978). The 30 selected stimuli were grouped in three levels of difficulty according to the accuracy of the responses: difficult (51-70% accuracy), intermediate (71-80% accuracy) and easy (81-100% accuracy), resulting in 10 items in each level (Table 3).

Revista ConCiencia EPG–Edición Especial Marzo 2022
ISSN digital: 2517-9896 ISSN impreso: 2517-9896
https://doi.org/10.32654/ConCienciaEPG/Eds.especial-1

Table 3

Levels of difficulty.

	Original Number	Percentage of selection of the target stimulus	Sex M Male F Female
Easy	8	95.0	M
	1	94.2	M
	31	91.9	F
	2	91.1	M
	7	89.3	M
	32	88.0	M
	3	87.3	M
	17	86.0	M
	14	85.6	F
	39	82.5	F
Intermediate	38	78.5	F
	41	78.3	M
	34	76.3	F
	18	76.0	F
	36	75.3	F
	13	74.7	M
	10	74.1	M
	27	74.1	F
	45	73.1	M
	23	71.3	M
Difficult	44	69.3	F
	19	69.1	M
	20	66.5	F
	6	65.1	M
	29	62.3	M
	40	62.2	F
	35	62.0	M
	42	61.2	M
	33	58.2	M
	24	54.4	M

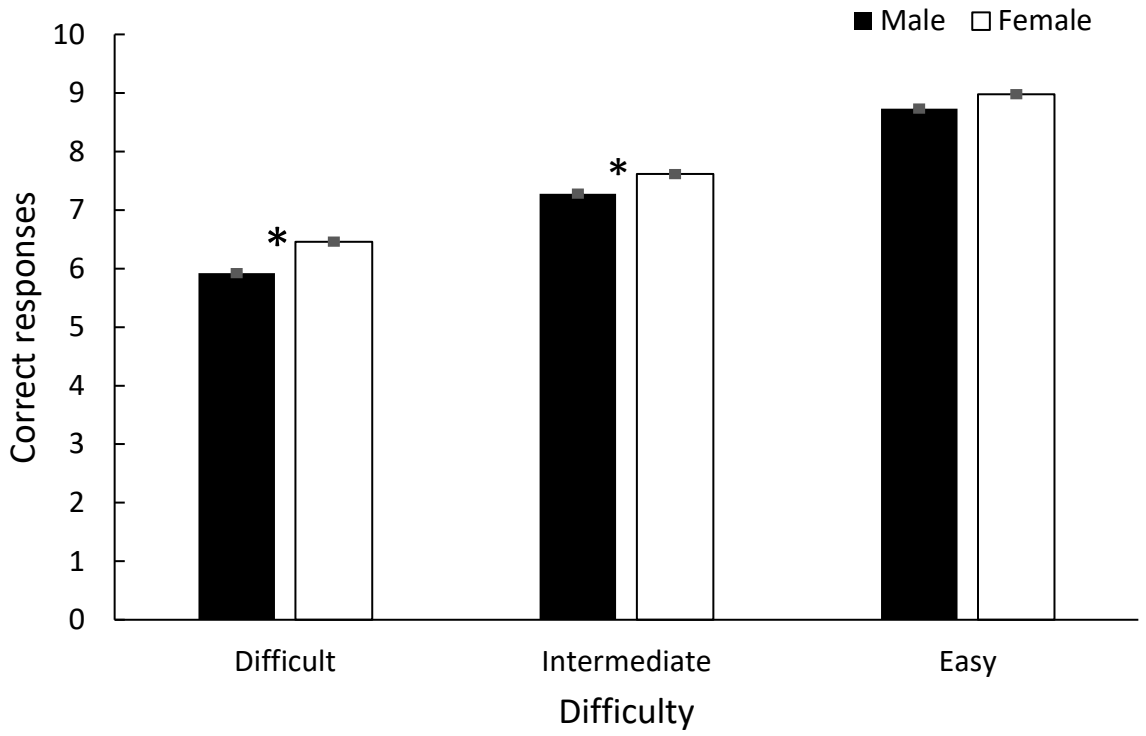
Levels of difficulty according to the percentage of correct selection of the target stimulus: difficult (50-70%), intermediate (71-80%) and easy (81-100%), with 10 items in each level.

Furthermore, there were significant differences between sexes (t_{1487}= 5.6, p < .001, *d*=0.32) with a small effect, female (M=29.2, standard error=SE, SE=.12) performed better than male (M=27.8, SE=.21) by an average of one correct response. To verify if the detected execution levels presented differences within the execution, a mixed ANOVA was performed with two factors: difficulty (3) and sex (2). Results showed significant differences with a large effect according to

difficulty ($F_{2,2974}$=1228, p< .001, η^2=.30) and sex ($F_{1,487}$=37.9, p< .001, η^2=.01) with a small effect, and the interaction between difficulty and sex ($F_{2,2974}$=3.89, p=.021, η^2=.01) with a small effect. Females performed with a better execution (M=7.8, SE=.04) compared to males (M=7.4, SE=.04). According to the post hoc analysis, there were also differences between the levels of difficulty (easy M=8.9, SE=.04; intermediate M=7.5, SE=.04, and difficult M=6.3, SE=.04). Also, the post hoc analysis confirmed that females executed better in intermediate and difficult levels (Figure 1).

Figure 1

*Mean and standard error on correct responses according to level of difficulty by sex (*p<.05).*

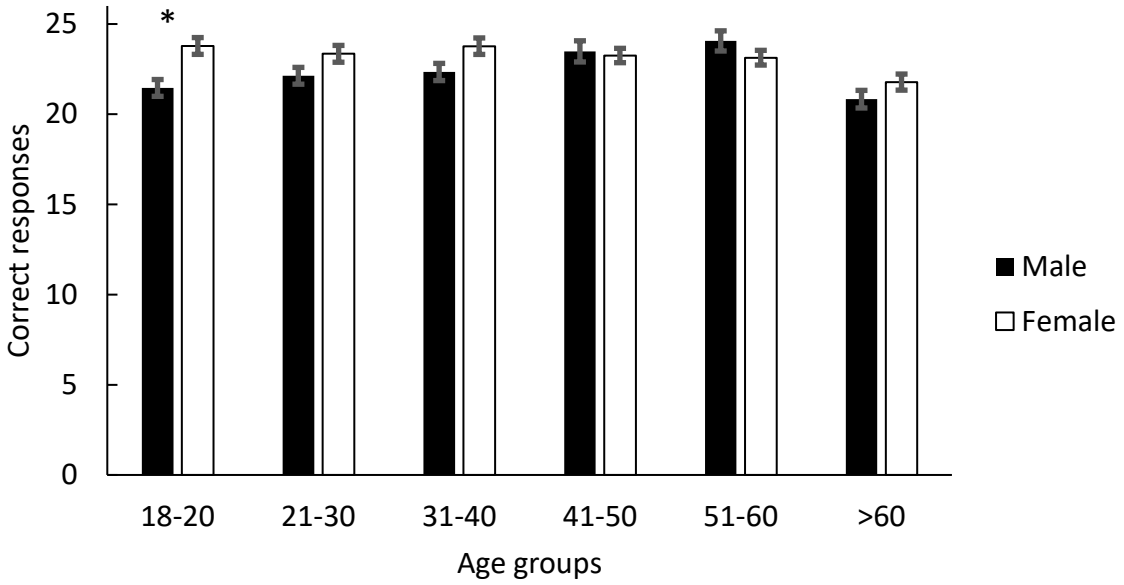


To perform the analysis based on age group by sex, 92 participants were selected at random from each group to create subsamples. According to the ANOVA test, there were significant differences in age ($F_{5, 540}$=5.8, p< .001, η^2=.05), in sex ($F_{1, 540}$=8.3, p= .004, η^2=.014), and in the interaction of sex and age ($F_{5, 540}$=3, p= .011, η^2=.03), all tree resulting with a small effect. The difference in sex can also be seen in this analysis with a smaller subsample (males, M=22.4, SE=.21; females, M=23.2, SE=.18).

The post hoc analysis showed that the group of older adults (M=21.3, SE=.34) was different to all the groups (21-30: M=22.7, SE=.34; 31-40: M=23.1, SE=.34; 41-50: M=23.3, SE=.34; 51-60: M=23.5, SE=.34) except the youngest (18-20, M=22.6, SE=.34). And the interaction of sex by age post hoc indicated that the differences between male and female only transpired in the young adult's group (Figure 2).

Figure 2

*Mean and standard error on correct responses by sex and age group, n=92 per group (*p<.05)*



Discussion

The analysis of the proposed online RMET showed that 30 stimuli have adequate characteristics to be recognized properly, with an average response accuracy of 70%. This makes it a simple task without ceiling effects and with executions above chance level. Likewise, this is a quick version (the average time was 12.5 minutes) in which accuracy did not decline by the end of the test, which is why we can infer the participants would not feel fatigued or bored. In relation to mental fatigue, Van Cutsem et al. (2017) indicated that tasks associated to mental fatigue last over 30 minutes. And because this is a self-applied, automated test, it can be performed at any moment in the context of research or

Revista ConCiencia EPG–Edición Especial Marzo 2022
ISSN digital: 2517-9896 ISSN impreso: 2517-9896
https://doi.org/10.32654/ConCienciaEPG/Eds.especial.1

clinical evaluation, as long as the participant possesses an electronic device with access to the internet.

Furthermore, when performing the analysis of the correct execution, we defined groups of items with three different levels of difficulty (easy, intermediate, and difficult) which can be used for further analysis in particular samples. The original test was not designed to have levels of difficulty; however, after analyzing the results from the correct responses, a wide range of percentages of selection of the target word was perceived, which indicates that some stimuli are easier than others. The difficulty in the task can vary by 3 factors: quality of the photograph where the mental state is expressed; quality of the target word assigned to the face; and the level of semantic similarity or difference between the foil and target.

With this distinction of difficulties, the execution of those items that get close to the ceiling effect can be analyzed to show the difference from those that are higher than chance, and to then detect if these dissimilarities between the samples studied are generalized or happen in one level or the other. Chapman and Chapman (1978) indicate this in their analysis about the comparison between groups of patients with schizophrenia and a control group of participants. With regard to the groups of items based on the level of difficulty, 10 items were assigned to each group, making it possible to make comparisons between them, specifically in sex and age, where significant interactions were found.

Another one of the objectives was to explore the presence of differences according to sex. Previous studies have confirmed the existence of these differences (Baron-Cohen et al., 2001; Kirkland et al., 2013). However, we must acknowledge that the obtained effect in all of the analysis of this study is low, where the difference lies in higher numbers in females by an average of a single response. This small effect had already been reported in the meta-analysis performed by Kirkland et al. (2015) in which the moderating variables were language, country, group of investigators, availability of data, but found no differences between them. The significant interactions found in this study between age and sex are discussed later in this section.

The analysis to compare sex and difficulty showed that there were differences between the way males and females responded to the “intermediate” and “difficult” levels of difficulty but on the level “easy” it showed no difference.

Categorizing by levels of difficulty can be used again in future shorter versions of this test to include easy and difficult items and to avoid polarization to either end.

On the other hand, according to Moor et al. (2012), even though the differences in age groups have been less explored, there have been reports that teenagers get less correct responses than adults; and Lee et al. (2021) indicated that in elderly people, the performance begins to decline as the age increases. In their study, the researchers applied a version with 10 items and evaluated people between 66 and 105. The data obtained in this present study shows that the cognitive-emotional ability assessed remains stable throughout adulthood and then declining after 60 years old. We also found that females performed the task better than males, but only in the youngest age group (18-20), indicating that as age increases, the differences disappear. This finding suggests that differences related to sex may interact with age, which is why there should be more attention brought to these variables. For example, Olderback et al. (2015) did not find differences between the sexes in a sample with age of 33 years old (SD=11).

It is important to mention that there is a controversy if whether the detection of mental states by looking at gazes is in fact part of the ToM; because the original test was designed as a direct assessment of this process to explore if there was an alteration in patients with autism (Baron-Cohen et al., 2001). However, according to authors like Craig et al. (2004), when analyzing this test, they have found that it is more related to the processes of recognition of emotion and empathy.

We consider that, in effect, emotional recognition in faces is an essential component to the resolution of the RMET and it is important that this link is conserved so that the task can be performed correctly. However, it requires a semantic understanding of the words used for the response options and the ability to locate the relationship between them and the gaze presented. Also, there are empathic processes that can be activated to facilitate the recognition of the mental states and then be able to attribute it to the image. This attribution corresponds specifically to the process of the ToM.

Additionally, in the beginning of this document, we pointed out that social cognition components were intertwined and even overlapped. In the future, comprehensive studies should be developed to contribute to the theoretical

differentiation of the components of social cognition and to select more specific tests.

On the other hand, there is an advantage in possessing an instrument that presents the gazes of people with similar features distinctive from the culture where the test is being performed. As shown by Adams et al. (2009), persons are able to recognize mental states better in individuals from their own culture. In said study, the authors compared North American and Japanese participants and presented two versions of the RMET: their own and a foreign test. They found that both groups had a better execution and greater brain activation in the superior temporal sulcus from both hemispheres when presented with the version that matched their culture. This suggests that the version from this study will be very useful for research and clinical evaluation in Mexicans, but also extend its use in some Latin-American countries that share the same cultural and physical characteristics, and language.

The increase use of virtual environments for investigation, also known as teleresearch, has increased due to the pandemic because it has been used with the purpose of reducing the spread of COVID-19. Information technologies have also been developed and shared for this reason. These tools had already been implemented but are now a priority because they grant a fast, effective path to approach the population. In this study, the call was made through social media, which allowed for the quick recruitment of a large sample. Previously, Casler et al. (2013) reported that data obtained from virtual platforms or social media were similar to the ones face-to-face in a lab. Even the RMET has been previously applied through an online survey with comparable results (Olderback et al., 2015).

Finally, it is important to recognize the limitations of this study, including that the 30 stimuli selected did not have a similar male to female proportion. There are mostly males and future proposals should maintain a balance. The objective of the study did not include validation analysis; however, we consider that internal consistency, and convergent and divergent validity calculations should be made for future investigations to have more robust tests for our context. Also, there are the inconveniences related to virtual self-applied tests because we do not know if the participants received any help, if there were any distractions, or lack of concentration all throughout, as well as the time they spent

on each item and other physical variables such as the distance between the subject and the stimuli, brightness of the screen, etc. This can only be resolved if this study was compared to one performed in person, inside well-controlled spaces, free of distractions, noise, or help. Another limitation is that it is a study performed in participants with a high education. It is necessary to broaden these investigations in a sample with a more diverse schooling.

Additionally to this, because it was an adaptation and not a normative study, no health history information that we know can interfere with the perception of emotional states, such as mental illnesses (depression, addictions or schizophrenia), was recorded. These aspects have not been explored either by other studies when analyzing the properties of the RMET (Baron-Cohen et al., 2001; Lee et al., 2020; Pfaltz et al., 2013). However, we suggest that future studies include at least a few questions about the presence of psychiatric conditions, like the one done by Baron-Cohen et al. (1997), or about the lack of neurological problems, especially important in the elder population. For the online versions, it is recommended to include attention check questions as part of a strict requirement to guarantee the involvement of the participants (Olderback et al., 2015).

Conclusion

This version of the RMET with 30 full color photographs may be used in online research during this period of time, when we must follow sanitary contingency and online jobs predominate. Even if it is addressed to a Mexican population, it can be used by communities that share the same physiological and cultural characteristics. The detection of different levels of difficulty between the items proposed will make applying the test possible to evaluate different hypotheses according to the level of complexity in specific samples, allowing to visually address in a deeper way the study of the ToM.

References

Adams Jr., R.B., Rule, N.O., Franklin Jr. R.G., Wang, E., Stevenson, M.T., Yoshikawa, S., Nomura, M., Sato, W., Kveraga, K., & Ambady, N. (2009). Cross - cultural Reading the Mind in the Eyes: An fMRI investigation. *Journal of Cognitive Neuroscience* *22*(1),97-108. https://doi.org/10.1162/jocn.2009.21187

Adolphs R. (1999). Social Cognition and the Human Brain. *Trends in Cognitive Sciences,* *3*(12), 469-479. https://doi.org/10.1016/S1364-6613(99)01399-6

Atenas, T. L., Ciampi Diaz, E., Venegas Bustos, J., Uribe San Martín, R., & Cárcamo Rodríguez, C. (2019). Cognición Social: Conceptos y Bases Neurales. *Revista chilena de neuro-psiquiatría*, *57*(4), 365-376. https://doi.org/10.4067/S0717-92272019000400365

Baron-Cohen, S. Leslie, A. M., & Frith, U. (1985). Does the autistic child have a 'theory of mind'? *Cognition*; *21*(4), 37-46. https://doi.org/10.1016/0010-0277(85)90022-8

Baron-Cohen, S., O'Riordan, M., Stone, V., Jones, R., & Plaisted, K. (1999). Recognition of faux pas by normally developing children and children with Asperger syndrome or high-functioning autism. *Journal of Autism and Developmental Disorders; 29*(5) 407-18. https://doi.org/10.1023/A:1023035012436

Baron-Cohen, S., Wheelwright, S. & Jolliffe, T. (1997). Is there a "language of the eyes"? Evidence from normal adults and adults with autism or Asperger syndrome, *Visual Cognition 4*(3), 311–331. https://doi.org/10.1080/713756761

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y. & Plumb. I. (2001). The "Reading the Mind in the Eyes" Test Revised Version: A Study with Normal Adults, and Adults with Asperger Syndrome or High-functioning Autism, *The Journal of Child Psychology and Psychiatry and Allied* Disciplines 42(2), 241-251. https://doi.org/10.1111/1469-7610.00715

Barrera-Valdivia, P., Méndez, R. V., Czyewska, M., Mendoza-Meraz, G., & Cordero-Hidalgo, A. (2019). Desarrollo de un instrumento para evaluar la empatía emocional en escolares mexicanos. *Tecnociencia Chihuahua*, *13*(3), 149-156. https://vocero.uach.mx/index.php/tecnociencia/article/view/473

Bilder, R. M., Postal, K. S., Barisa, M., Aase, D. M., Cullum, C. M., Gillaspy, S. R., ... & Woodhouse, J. (2020). Inter Organizational Practice Committee recommendations/guidance for teleneuropsychology in response to the COVID-19 pandemic. *Archives of Clinical Neuropsychology*, *35*(6), 647-659. https://doi.org/10.1093/arclin/acaa046

British Psychological Society (2020). *Psychological assessment undertaken remotely*. https://www.bps.org.uk/sites/www.bps.org.uk/files/Policy/Policy%20-%20Files/Psychological%20assessment%20undertaken%20remotely.pdf

Casler, K., Bickel, L., & Hackett, E. (2013). Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. *Computer in Human Behavior.* 29, 2156–2160. https://doi.org/10.1016/j.chb.2013.05.009

Chapman, L. J., & Chapman, J. P. (1978). The measurement of differential deficit. *Journal of psychiatric research*, *14*(1-4), 303-311. https://www.bps.org.uk/sites/www.bps.org.uk/files/Policy/Policy%20-%20Files/Psychological%20assessment%20undertaken%20remotely.pdf

Charernboon, T. & Patumanond, J. (2017). Social cognition in schizophrenia. *Mental Illness,* 9(1), 16-19. https://doi.org/10.1108/mi.2017.7054

Craig, J. S., Hatton, C., Craig, F. B. & Bentall, R. P. (2004). Persecutory beliefs, attributions and theory of mind: Comparison of patients with paranoid delusions, Asperger's syndrome and healthy controls. *Schizophrenia Research*, *69*(1), 29-33. https://doi.org/10.1016/S0920-9964(03)00154-3

Dziobek, I., Fleck, S., Kalbe, E., Rogers, K., Hassenstab, J., Brand, M., ... & Convit, A. (2006). Introducing MASC: a movie for the assessment of social cognition. *Journal of Autism and Developmental Disorders*, *36*(5), 623-636. https://doi.org/10.1007/s10803-006-0107-0

Elfenbein, H.A. & Ambady, N. (2003). Universals and cultural differences in recognizing emotions. *Current Directions in Psychological* Science, 12(5), 159-164. https://doi.org/10.1111/1467-8721.01252

Elosua, P. (2020) Aplicación remonta de test: riesgos y recomendaciones. *Papeles del psicólogo*. 42 (1) 33 https://doi.org/10.23923/pap.psicol2021.2952

Ferguson, F. J., & Austin, E. J. (2010). Associations of trait and ability emotional intelligence with performance on Theory of Mind tasks in an adult sample. *Personality and individual differences*, *49*(5), 414-418. https://doi.org/10.1016/j.paid.2010.04.009

Fernández, M., Mollinedo-Gajate, I., & Peñagarikano, O. (2018). Neural circuits for social cognition: implications for autism. *Neuroscience*, *370*, 148-162. https://doi.org/10.1016/j.neuroscience.2017.07.013

Happé, F.G. (1994). An advanced test of theory of mind: understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders*, *24*(2), 129-54. https://doi.org/10.1007/BF02172093

Huerta-Ramos, E., Ferrer-Quintero, M., Gómez-Benito, J., González-Higueras, F., Cuadras, D., del Rey-Mejías, Á. L., ... & Ochoa, S. (2021). Traducción y validación del test de caras de Baron-Cohen en población española. *Actas Espanolas de Psiquiatria*, *49*(3), 106-113. https://medes.com/publication/160545

International Test Commission (2017). *The ITC Guidelines for Translating and Adapting Tests* [Second edition]. InTestCom.org.

https://www.intestcom.org/files/guideline_test_adaptation_2ed.pdf

Kittel, A. F. D., Olderbak, S., & Wilhelm, O. (2021). Sty in the Mind's Eye: A Meta-Analytic Investigation of the Nomological Network and Internal Consistency of the "Reading the Mind in the Eyes" Test. *Assessment*, 1073191121996469. https://doi.org/10.1177/1073191121996469

Kirkland, R. A., Peterson, E., Baker, C. A., Miller, S., & Pulos, S. (2013). Meta-analysis Reveals Adult Female Superiority in" Reading the Mind in the Eyes Test". *North American Journal of Psychology*, *15*(1),121-146. https://psycnet.apa.org/record/2013-09240-009

Lee, S., Jacobsen, E. P., Jia, Y., Snitz, B. E., Chang, C. C. H., & Ganguli, M. (2021). Reading the mind in the eyes: a population-based study of social cognition in older adults. *The American Journal of Geriatric Psychiatry*, *29*(7), 634-642. https://doi.org/10.1016/j.jagp.2020.11.009

Lee, H. R., Nam, G., & Hur, J. W. (2020). Development and validation of the Korean version of the Reading the Mind in the Eyes Test. PLoS ONE, 15(8 August 2020). https://doi.org/10.1371/journal.pone.0238309

Maurage, P., Grynberg, D., Noël, X., Joassin, F., Hanak, C., Verbanck, P., ... & Philippot, P. (2011). The "Reading the Mind in the Eyes" test as a new way to explore complex emotions decoding in alcohol dependence. *Psychiatry research*, *190*(2-3), 375-378. https://doi.org/10.1016/j.psychres.2011.06.015

Maureira Cid, F., Bravo Rojas, P., Ramírez Nahuelñir, M. A., & Fuentealba Colil, D. (2019). Baremos del coeficiente intelectual y la detección de la expresión emocional a través de la mirada en estudiantes de educación física de Santiago de Chile. *Revista Electrónica de Psicología Iztacala*, *22*(1), 46-57. http://www.revistas.unam.mx/index.php/repi/article/view/68969

Moor, B. G., de Macks, Z. A. O., Güroğlu, B., Rombouts, S. A., Van der Molen, M. W., & Crone, E. A. (2012). Neurodevelopmental changes of reading the mind in the eyes. *Social*

cognitive and affective neuroscience, *7*(1), 44-52. https://doi.org/10.1093/scan/nsr020

Nejati, V. (2018). Negative interpretation of social cue in depression: Evidence from reading mind from eyes test. *Neurology, Psychiatry and Brain Research*, *27*, 12-16. https://doi.org/10.1016/j.npbr.2017.11.001

O'Keeffe, K., Hodder, S., & Lloyd, A. (2020). A comparison of methods used for inducing mental fatigue in performance research: Individualised, dual-task and short duration cognitive tests are most effective. *Ergonomics*, *63*(1), 1-12. https://doi.org/10.1080/00140139.2019.1687940

Olderbak, S., Wilhelm, O., Olaru, G., Geiger, M., Brenneman, M. W., & Roberts, R. D. (2015). A psychometric analysis of the reading the mind in the eyes test: toward a brief form for research and applied settings. Frontiers in Psychology, 0(OCT), 1503. https://doi.org/10.3389/FPSYG.2015.01503

Pfaltz, M. C., McAleese, S., Saladin, A., Meyer, A. H., Stoecklin, M., Opwis, K., & Martin-Soelch, C. (2013). The Reading the Mind in the Eyes Test: Test-retest reliability and preliminary psychometric properties of the German version. *International Journal of Advances in Psychology*, *2*(1), 1-9. https://doi.org/10.5167/uzh-87335

Peñuelas-Calvo, I., Sareen, A., Sevilla-Llewellyn-Jones, J., & Fernández-Berrocal, P. (2019). The "Reading the Mind in the Eyes" test in autism-spectrum disorders comparison with healthy controls: a systematic review and meta-analysis. *Journal of Autism and Developmental Disorders*, *49*(3), 1048-1061. https://doi.org/10.1007/s10803-018-3814-4

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 1*(4), 515-526. http://dx.doi.org/10.5579/rnl.2012.0108

Roman, N., Rojas, G. & Román, N. (2012). Baremos del test de la mirada en español en adultos normales de Buenos Aires, *Neuropsicología Latinoamericana, 4*(3), 1-5.

http://dx.doi.org/10.5579/rnl.2012.0108

Shamay-Tsoory, S. G., & Aharon-Peretz, J. (2007). Dissociable prefrontal networks for cognitive and affective theory of mind: a lesion study. *Neuropsychologia*, *45*(13), 3054-3067. https://doi.org/10.1016/j.neuropsychologia.2007.05.021

Torralva, T., Roca, M., Gleichgerrcht, E., Bekinschtein, T., & Manes, F. (2009). A neuropsychological battery to detect specific executive and social cognitive impairments in early frontotemporal dementia. *Brain*, *132*(5), 1299-1309. https://doi.org/10.1093/brain/awp041

Van Cutsem, J., K. De Pauw, L. Buyse, S. Marcora, R. Meeusen, and B. Roelands. (2017). "Effects of Mental Fatigue on Endurance Performance in the Heat." Medicine and Science in Sports and Exercise 49(8): 1677–1687. doi: 10.1249/MSS.0000000000001263.

Van Overwalle, F. (2009). Social cognition and the brain: a meta-analysis. *Human brain mapping*, *30*(3), 829-858. https://doi.org/10.1002/hbm.20547

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: the truth about false belief. *Child Development; 72*(3), 655-84. https://doi.org/10.1111/1467-8624.00304

Wimmer, H. & Perner, J. (1983). Beliefs about beliefs: representation and the constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*(1), 103-28. https://doi.org/10.1016/0010-0277(83)90004-5