cogent psychology

COGNITIVE SCIENCE & NEUROSCIENCE | RESEARCH ARTICLE

Comparison between the Short Story Task and the Reading the Mind in the Eyes Test for evaluating Theory of Mind: A replication report

M. Giordano[1]*, G. Licea-Haquet[1], E. Navarrete[1], E. Valles-Capetillo[1], F. Lizcano-Cortés[1], A. Carrillo-Peña[1] and A. Zamora-Ursulo[1]

*Corresponding author: M. Giordano, Departamento de Neurobiología Conductual y Cognitiva, Instituto de Neurobiología, UNAM, Blvd. Juriquilla 3001, Juriquilla, Qro 76230, Mexico
E-mail: giordano@unam.mx

Abstract: Introduction: The ability to attribute emotional states, beliefs, and intentions to others has been termed Theory of Mind (ToM), mentalizing, and mind reading. The purpose of this study was to find an instrument to measure ToM in the Mexican population, that would yield similar results to those obtained in other cultures, and could discriminate between individuals. To achieve this objective, we replicated a study which compared two measures of ToM in a sample of English-speaking, neurologically intact adults.

Methods: A sample of young Mexican adults (n = 118) was evaluated on the Reading the Mind in the Eyes Test (RMET) and a test that uses naturalistic narrative stimuli, the Short Story Task (SST), and on tests of general cognitive ability, executive functions, and empathy.

Results: We found a significant correlation between the ToM tests, and both tests correlated with verbal ability, general cognitive ability, and empathy, similar to what was seen in a previous study. Both tests discriminated between individuals and were challenging enough that we found no perfect scores.

Conclusions: These results show that both the RMET, which taps into emotion recognition and its categorization with language, and the SST, which relies on

ABOUT THE AUTHORS

M. Giordano is a full professor at the Instituto de Neurobiologia, UNAM. Currently, her group is interested in understanding the cognitive and neural basis of pragmatic language comprehension. Pragmatic language is used in everyday interactions to communicate intentions, it includes, but is not restricted to figurative language, as in metaphors, and indirect speech. There is limited knowledge about the cognitive operations that are involved in pragmatic language comprehension, and most of that information comes from studies in clinical populations. She uses psychometric and neuroimaging tools to answer these questions.

M. Giordano

PUBLIC INTEREST STATEMENT

The ability to attribute emotional states, beliefs, and intentions to others has been termed Theory of Mind (ToM), mentalizing, and mind reading. Since human beings are a social species, social competence is evolutionarily adaptive and indeed essential for adequate interactions. ToM may be one of three central processes that are part of social information processing, the other two being social perception and action observation. ToM is likely to involve a variety of more basic sub-processes that have yet to be described. Psychometric tools that are reliable and can discriminate among neurologically intact adults from different cultural backgrounds are necessary and useful for research in ToM. The purpose of this study was to find a reliable instrument to measure ToM in the Mexican population, that would yield similar results to those obtained in other cultures, and could discriminate between individuals.

cogent •• psychology

narrative fiction to test the ability to interpret mental states, show concurrent validity in a sample of neurologically intact young adults from a Latin-American culture; these tests may be useful in the clinical setting and for basic research into ToM.

Subjects: Behavioral Sciences; Cognitive Psychology; Cognition & Emotion; Cognitive Neuropsychology; Behavioral Neuroscience

Keywords: Social cognition; empathy; verbal competence; verbal fluency; Spanish

The ability to attribute emotional states, beliefs, and intentions to others has been termed Theory of Mind (ToM) (Schurz, Radua, Aichhorn, Richlan, & Perner, 2014), mentalizing (Frith & Frith, 2006), and mind reading (Turner & Felisberti, 2017). ToM was initially used by Premack and Woodruff in 1978 (Premack & Woodruff, 1978) to describe that non-human primates appeared to anticipate the actions of a human faced with a problem. The concept can also be traced to Daniel Dennett's philosophical proposal of the intentional stance (Dennett, 1987).

This ability changes across the lifespan, is associated with effective social interactions, and prosocial behavior (Turner & Felisberti, 2017), and is central for utterance interpretation (Cummings, 2017). Deficits in ToM are present in various disorders including autism spectrum disorder, psychiatric disorders such as schizophrenia (Turner & Felisberti, 2017), and after brain lesions (Shamay-Tsoory & Aharon-Peretz, 2007).

ToM is a multidimensional construct (Turner & Felisberti, 2017), and it has been argued that it needs to be reformulated and deconstructed (Schaafsma, Pfaff, Spunt, & Adolphs, 2015). Some of the proposed elements of ToM include perceptual discrimination and categorization of the socially relevant stimuli and interoceptive signals elicited by those stimuli, semantic or conceptual knowledge, and executive and motivational processes. Social content and causal inference are distinctive features of ToM (Schaafsma et al., 2015). Two ToM components have been identified as fundamental predictors of well-being, the affective, or "hot" component involving the comprehension of emotional states; and the cognitive, or "cold" component involving the attribution of thoughts and beliefs (Henry, Cowan, Lee, & Sachdev, 2015).

Due to the multidimensional nature of ToM, many different tests have been designed, a recent review categorized the existing tests for ToM in three classes: emotion recognition tests; cognitive and affective mentalizing tasks that measure attribution of beliefs, intentions, desires, and emotions; and multidimensional measures that combine these features (Turner & Felisberti, 2017). A meta-analysis using a different approach divided the studies into story-based and nonstory-based (Mar, 2011).

The Reading the Mind in the Eyes test (RMET; (Baron-Cohen, Jollifle, Mortimore, & Robertson, 1997)), has been widely used. It involves recognition of mental state terms, probes non-automatic processes, and is sensitive to variation in neurologically typical adults. The revised version of the RMET includes 36 items with four response options per item (Baron-Cohen, Wheelwriht, Hill, Raste, & Plumb, 2001). Some studies have found correlations between RMET and measures of empathy or IQ (reviewed in (Vellante et al., 2013)).

Naturalistic narrative stimuli are an alternative that allows ToM targets to be integrated into a context (Turner & Felisberti, 2017), an example is the Short Story Task (SST), a test developed by Dodell-Feder, Lincoln, Coulson, and Hooker (2013). This task requires reading a fictional story written by Ernest Hemingway about two characters in a romantic relationship and includes questions that assess explicit mental state reasoning and comprehension. The test demonstrated

cogent •• psychology

sensitivity to variations among neurologically typical adults and concurrent validity with the RMET and the Interpersonal Reactivity Index (IRI) (Dodell-Feder et al., 2013).

Because of our interest in the cognitive basis of pragmatic language and the central role that ToM plays in utterance interpretation (Cummings, 2017), our purpose was to find a dependable measure of ToM for neurologically intact Spanish-speaking, Mexican adults, that would yield similar results to those obtained in other cultures, and could discriminate between individuals. Although there may be a universal cognitive substrate for ToM, this needs empirical confirmation. Cultural tools and practices shape brain processes (Kitayama & Park, 2010), there is variation in human psychology and behavior across human populations (Henrich, Heine, & Norenzayan, 2010), and radical differences between languages (Evans & Levinson, 2009). For these reasons we needed to compare the performance of our participants in these ToM tests with that of participants from a different culture that speaks a different language. To achieve this objective, we replicated the study by Dodell-Feder et al. (2013), and correlated the RMET (Baron-Cohen et al., 2001) and the SST (Dodell-Feder et al., 2013) between them, and also with empathy and general intelligence. Also, we measured the internal consistency of these tests and their correlation with executive functions.

Empathy and general intelligence were measured to make our study comparable to that of Dodell-Feder et al. (2013). Empathy refers to the ability to recognize and identify what someone else is feeling (cognitive component), and to share that emotional state (emotional component) (Lucas-Molina, Pérez-Albéniz, Ortuño-Sierra, & Fonseca-Pedrero, 2017). To measure empathy, we used the Interpersonal Reactivity Index (IRI), a self-report instrument consisting of four separate subscales that have been widely used to test both components of empathy (Lucas-Molina et al., 2017). The fantasy subscale of the IRI was found to correlate with the ToM measures (Dodell-Feder et al., 2013), we expected to see similar results in our sample. Executive functions were measured to identify if the performance on the ToM tests was related to general domain functions, such as inhibitory control, working memory, verbal fluency, and mental flexibility. Besides, these measures together with general intelligence allowed us to describe our sample more fully.

1. Methods

Participants (n = 118, 73 females) were young (23.03 ± 3.61 years of age) Mexican students. Inclusion criteria were having Spanish as their maternal language, being currently enrolled in a bachelor's or graduate degree program, between 18–35 years of age, and signing the informed consent form. Exclusion criteria included having been diagnosed with a neurological or psychiatric disorder, and scores above the norm in the Symptom Checklist 90 (SCL-90; (Cruz-Fuentes, López-Bello, Blas-García, González-Macías, & Chávez-Balderas, 2005)). The study was approved by the institutional ethics committee and complied with the federal guidelines of the Mexican Health Department (http://www.salud.gob.mx/unidades/cdi/nom/compi/rlgsmis.html), which agree with international regulations.

1.1. Intelligence and executive functions tests

The Wechsler Adult Intelligence Scale (WAIS-IV) was used to obtain the Perceptual Reasoning Index (PRI), and Verbal Comprehension Index (VCI). Together, these indices are used to calculate the General Ability Index (GAI). Executive functions (inhibitory control, working memory, verbal fluency, and mental flexibility) were evaluated using the *Batería Neuropsicológica de Funciones Ejecutivas y Lóbulos Frontales (BANFE)* (Flores Lázaro, Ostrosky Shejet, & Lozano Gutiérrez, 2014).

1.2. Theory of mind and empathy tests

The IRI was used to measure affective and emotional empathy (Davis, 1980), it is a self-report questionnaire that includes four subscales: Fantasy (FS) determines the identification with fictional characters; Perspective Taking (PT) relates to the consideration of others' viewpoints; Empathic Concern (EC) measures feelings of compassion and concern for others in need; and Personal Distress (PD), relates to the reaction of discomfort to the distress of others (Lucas-Molina et al., 2017). A recent study has confirmed the construct validity and equivalence between genders of

the four-factor structure of the IRI in a large sample of Spanish-speaking college students (Lucas-Molina et al., 2017).

To test ToM, we used the Argentinian version of the RMET in Spanish (downloaded from the Autism Research Center webpage; http://www.autismresearchcentre.com/arc_tests) (Baron-Cohen et al., 1997, 2001). The test was programmed in PsychoPy (version 1.83.01; (Peirce, 2007)). The test consists of 36 photographs of eyes expressing different mental states, they were displayed in the center of the screen, along with four mental state terms in each corner. Participants had to press the letters q, p, z, or m corresponding to each of the corners of the screen, they did not have access to the glossary of terms to increase the difficulty of the task. The other ToM was the SST (Dodell-Feder et al., 2013). The test consists in reading a short story and answering a spontaneous mental state inference question, five reading comprehension questions, and eight questions about the beliefs, emotions, intentions, and desires of the characters in the story. A Spanish translation was obtained and revised by a professor of Linguistics to ensure that it was written in correct Spanish and that it was comprehensible. The task was applied according to the instructions by Dodell-Feder et al. (2013).
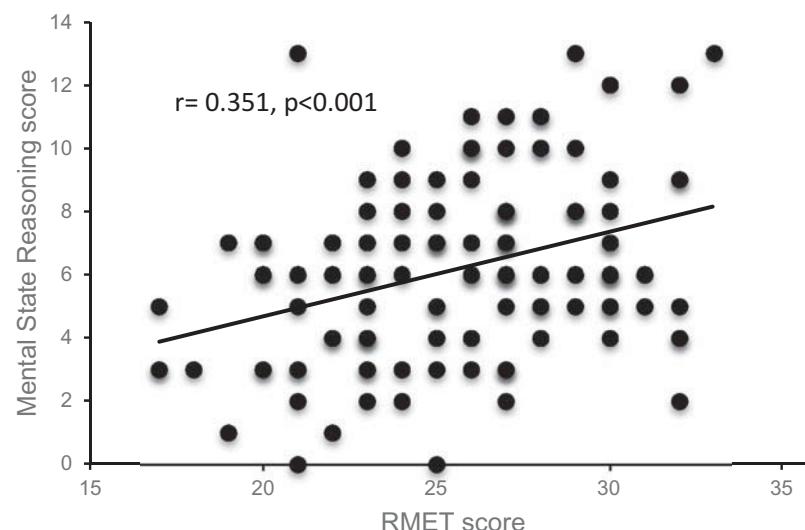
1.3. Statistics
We used SPSS (v. 24.0, IBM Corp.) to calculate descriptive statistics, correlations, and internal consistency measures (Cronbach's alpha), only raw scores were used. To compare between genders, we used t-tests for independent samples. A probability of less than 0.05 was considered significant, and Bonferroni correction was calculated for multiple correlations between tests of empathy and ToM, and between intelligence and executive functions tests.

For the correlation analysis our sample was three times larger than the minimum size (n = 30) for a power of 80% (Sample Size Calculators from UCSF Clinical and Translational Science Institute; calculated using the r = 0.49 value reported by Dodell-Feder et al. (2013)).

2. Results
In agreement with results obtained by previous studies (Dodell-Feder et al., 2013), we found a significant positive correlation ($r(116) = 0.351$, $p < 0.001$) between the RMET and SST (mental state reasoning score; MSR) (Figure 1), better performance in the RMET was associated with a higher score on MSR. The distribution of scores was normal with moderate skewness and kurtosis. No differences in ToM scores between female and male participants were found. The

Figure 1. Scatterplot showing the relation between RMET and MSR scores for the SST (r = 0.351, p < 0.001, n = 116). RMET: Reading the Mind in the Eyes test; SST: Short story task; MSR: mental state reasoning.

only difference between genders was in perceptual reasoning (WAIS-PRI score), which was higher for males (t(115) = 2.175, p < 0.05; 95% CI [.4442, 9.4835]), and in the empathic concern subscale of the IRI (IRI-EC), which was higher for females (t(106) = 2.122, p < 0.05; 95% CI [.1285, 3.7881]). The descriptive statistics of the various tests are presented in Table 1.

In terms of the correlation between RMET with general intelligence and executive functions, performance on this ToM test was significantly correlated with verbal fluency (r(116) = 0.302, p < 0.002), the WAIS-VCI (r(116) = 0.225, p < 0.001), and WAIS-GAI (r(115) = 0.251, p < 0.008). Except for verbal fluency, similar results were found for the MSR/SST, which correlated with WAIS-VCI (r(118) = 0.278, p < 0.003) and WAIS-GAI (r(117) = 0.249, p < 0.008). One participant did not complete the subtests of the WAIS-PRI. To test if verbal comprehension and general intelligence contributed to the association between RMET and MSR scores, we calculated partial correlations. The partial correlations between RMET and MSR/SST were still significant when controlling for verbal fluency (r(113) = 0.335, p < 0.001), WAIS-VCI (r(113) = 0.311, p < 0.002), and WAIS-GAI (r(112) = 0.317, p < 0.002). Slightly lower positive correlations between a measure of intelligence and both ToM tasks were found by Dodell-Feder et al. (2013).

Regarding empathy, and in agreement with previous studies (Dodell-Feder et al., 2013) we found that both ToM tests correlated only with the fantasy subscale of the test of empathy (IRI-FS) (r(108) = 0.212, 0.219, p < 0.030, for RMET and MSR/SST, respectively). The correlations did not reach significance when correcting for the number of comparisons for all tests of social cognition, i.e., empathy and ToM. Importantly, partial correlations between both ToM tests were still significant when controlling for the fantasy subscale of the test of empathy (IRI-FS) (r(105) = 0.297, p < 0.003). The data on the IRI of ten participants was not stored due to hardware malfunction.

In terms of internal consistency measures (Cronbach's alpha), RMET had an alpha of 0.530; the mean inter-item correlation was 0.032. For the RMET, we found that most items were answered above chance (25%); item 19 had less than 50% correct responses, followed by items 1, 7, 9

Table 1. Descriptive statistics for psychometric measures				
Cognitive Measure	**Mean**	**SD**	**95% CI**	**N**
MENTAL FLEXIBILITY	54.63	13.48	52.59–57.82	118
INHIBITORY CONTROL	173.47	9.60	172.01–175.74	118
VERBAL FLUENCY	22.27	6.40	20.87–23.29	118
WAIS-VCI	102.97	12.78	100.63–105.30	118
WAIS-PRI	105.26	12.20	103.02–107.49	117
WAIS-GAI	104.38	11.78	102.22–106.53	117
SST-SI	0.19	0.39	0.11–0.26	118
SST-COMP	6.88	2.27	6.54–7.37	118
SST-MSR	6.19	2.87	5.60–6.70	118
SST-TOTAL	13.26	4.22	12.49–14.03	118
RMET	25.63	3.78	24.93–26.32	116
IRIPT	17.52	4.14	16.79–18.37	108
IRIFS	16.30	5.41	15.32–17.39	108
IRIEC	19.83	4.59	18.98–20.74	108
IRIPD	12.22	5.10	11.24–13.21	108

WAIS: Wechsler Adult Intelligence Scale (WAIS-IV); PRI: Perceptual Reasoning Index; VCI: Verbal Comprehension Index; GAI: General Ability Index; SST: Short Story Task; SI: spontaneous inference question (max. score 1); COMP: comprehension questions (max. score 10); MSR: mental state reasoning questions (max. score 16); RMET: Read the Mind in the Eyes Test (max. score 36); IRI: Interpersonal Reactivity Index; PT: perspective taking; FS: fantasy; EC: empathy concern; PD: personal distress.

cogent • psychology

slightly above 50%, mean performance was 71.20% accurate (Table S1). The 95% confidence interval for the mean score was 24.93–26.32 (Table 1) with a perfect score being 36; the lowest score was 17 and the highest 33. These results are like those obtained in other non-clinical populations (Baron-Cohen et al., 2001; Dodell-Feder et al., 2013), and show that there is enough variability in the responses to discriminate between individuals (Figure 2; Table S1). To obtain more information about the items included in the RMET, a correlation between mean response time and mean correct responses per item was calculated. The analysis showed a significant negative correlation between mean response time and mean correct answers per item for the RMET (r (36) = −0.617, $p < 0.001$; Figure 3), indicating that the items in the test are heterogeneous in terms of difficulty. Those that were accurately answered were also answered faster and were probably clear-cut for the participants, in contrast, those that took more time had a lower mean score, and were perhaps more challenging for the participants. As indicated before, performance in this test correlated with verbal fluency, so it is possible that the more challenging items required a finer distinction to be made among the mental state terms.

For the SST internal consistency (Cronbach's alpha) was 0.514 for the comprehension questions and 0.621 for the mental state reasoning questions with an inter-item correlation of 0.166 for this section of the test. Only 18% of the participants made a spontaneous mental state inference, and none received a perfect score in this test. Only 10% of the participants obtained a perfect score of 10 for the comprehension questions. For the mental state reasoning questions, there were no perfect scores; the maximum score was 13 out of 16. The first comprehension question had the lowest mean score, while the first and second items for the mental state reasoning questions had the lowest mean scores, and just as for the RMET, we found that there was enough variability to distinguish between individuals (Figure 4). In contrast with the study by Dodell-Feder et al. (2013), we did not find perfect scores in either of the ToM tests.

3. Discussion

The results of this study replicate those by Dodell-Feder et al. (2013) in a sample selected from the Mexican population with different language and culture, in terms of the convergent validity between a widely-used instrument, the RMET (Baron-Cohen et al., 2001), and an instrument that uses fiction to test ToM, the SST (Dodell-Feder et al., 2013). These two measures shared about 12% of the variance between them and correlated with verbal abilities (WAIS-VCI). RMET scores also

Figure 2. Mean score±standard deviation per item for the Reading the Mind in the Eyes Test. The horizontal line shows the level of response due to chance (25%). Maximum score per item is 1, n = 116 participants.

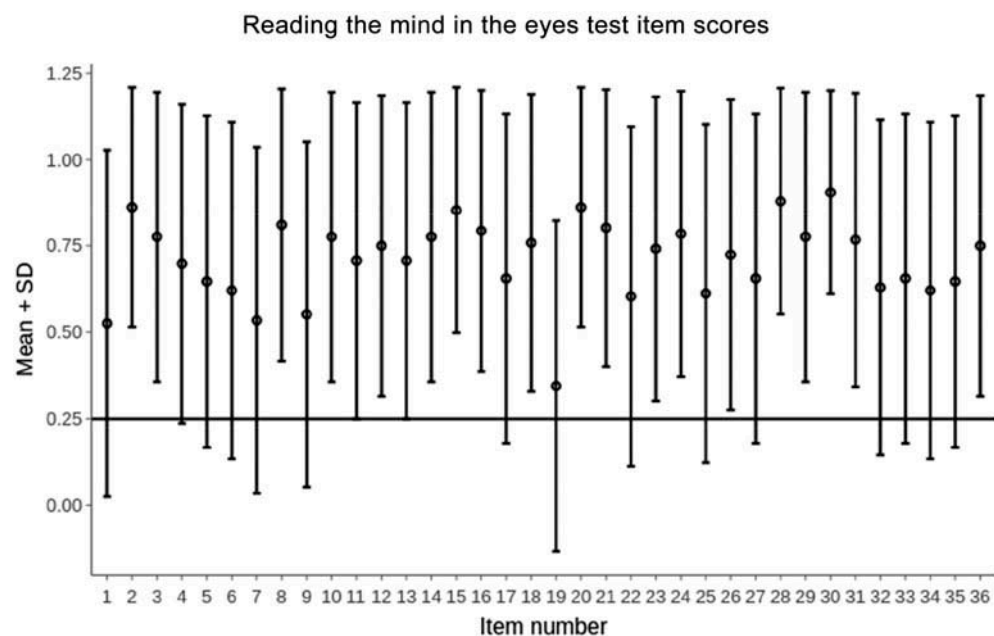Reading the mind in the eyes test item scores

cogent ·· psychology

Figure 3. Correlation between mean correct per item (percentage) and mean reaction time (sec) to answer each item for the Reading the Mind in the Eyes Test. A significant negative correlation was found (r = −0.617, N = 36, p < 0.001).
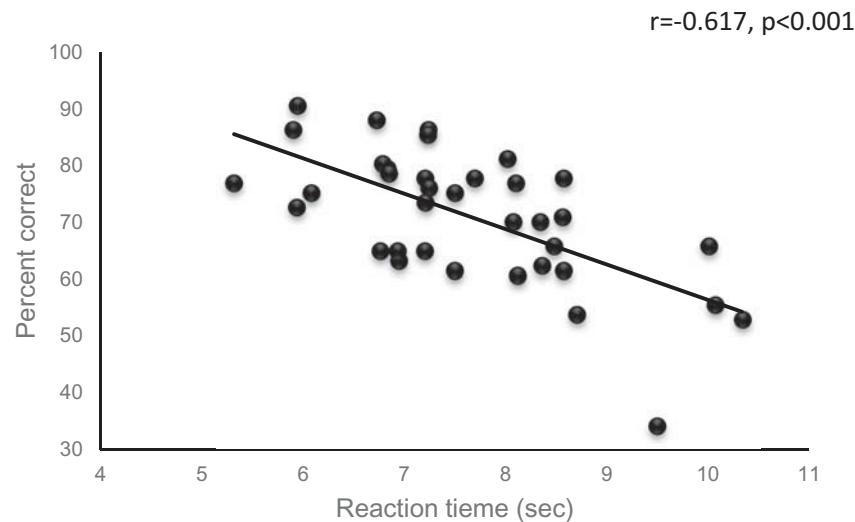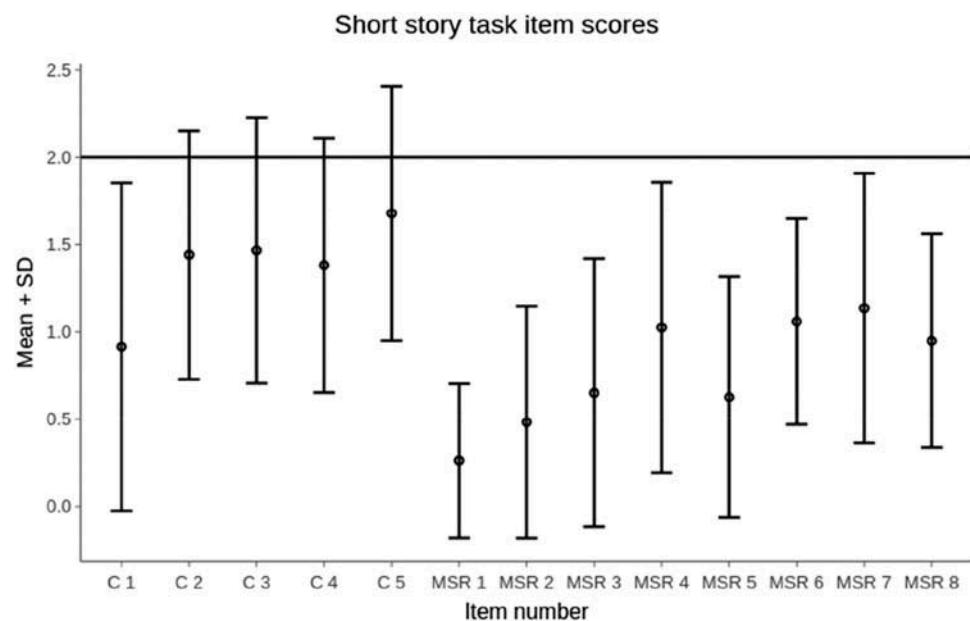

Figure 4. Mean score±standard deviation per item for the Short Story Task. Items C1-C5 are reading-comprehension questions, and items MSR 6-MSR 13 are mental state reasoning questions. Maximum score per item is 2, indicated by the horizontal line, n = 118 participants.

correlated with a measure of verbal fluency indicating the reliance of this test on lexical skills. Scores on both instruments also correlated with intellectual capacity as measured by the General Ability Index from the WAIS-IV, and with the fantasy subscale for a measure of empathy (IRI). The correlation between both measures of ToM was still significant when controlling for those cognitive variables, similarly to what Dodell-Feder et al. (2013) found.

The contribution of this study is the demonstration that both the RMET, that taps into the recognition of emotions and its categorization with language, and the SST, which relies on narrative fiction to test the ability to interpret mental states, are both useful tests for neurologically intact young adults from a Spanish speaking, Mexican culture. Our results, for the most part, reproduce those by the proponents of the SST, thus widening the choices available to researchers working with this kind of population that are interested in studying ToM, its underlying mental operations, and its relation to cognitive functions. We did not find ceiling effects, that is, there were

no perfect scores in either test, and we found that both tests could discriminate between individuals. Although our results argue for the use of both of these tests in the Mexican population, it is essential to make some considerations.

It must be noted that the proportion of variance that these two tests shared in our sample was 12% (r = 0.35), while for the sample reported by Dodell-Feder et al. (2013) it was 24% (r = .49). Among the possible reasons for this, are differences between the samples, including cultural differences, differences in the language of the tests, as well as a statistical effect (Button et al., 2013). In terms of the differences between samples, 50% of their participants made at least one spontaneous mental state inference, while few of our participants answered the spontaneous inference question correctly. The average score for mental state reasoning and comprehension were two points above those of the sample in the present study. Cultural differences may also play a role. The fact that the narration is a story written by an American writer, albeit well-known worldwide, it may reflect subtleties of the American culture, not easily understood by a person born and raised in Latin American culture. Nevertheless, the results of both studies coincide in the correlations observed between the SST/MSR and the fantasy subscale of the IRI-FS, and general cognitive ability, which speaks to the fact that the SST/MSR is measuring the same cognitive construct in both samples.

A few studies in children have evaluated the role of bilingualism in the development of ToM and have found that bilingual children outperform monolinguals in typical ToM tasks when controlling for factors such as language proficiency, age, socioeconomic status, and intelligence. The explanations given by the authors vary, some stress the enhancement of executive functions in bilingual children (Goetz, 2003, Kovács, 2009), working memory (Nguyen & Astington, 2014), metalinguistic understanding (Goetz, 2003) and a positive influence on social cognition (Diaz & Farrar, 2018; Goetz, 2003). In this study we did not formally evaluate language proficiency in other natural languages besides Spanish, which was the maternal language of our participants, and all tests were given in Spanish. It could be of interest to evaluate if the advantage observed in bilingual children, extends to adults when they learn a second language later in life.

The SST showed better internal consistency than the RMET, we found an alpha of 0.53 for the RMET, and 0.62 for the MSR/SST. In the previous study, an alpha of 0.54 for MSR/SST was reported (Dodell-Feder et al., 2013). In the case of the RMET, the values of alpha reported vary (see (Vellante et al., 2013)), in a reduced version of the test in a Spanish sample it was 0.56 (Redondo & Herrero-Fernández, 2018). The RMET may be measuring more than one factor, influencing its internal consistency. Some have found it to be unidimensional (Preti, Vellante, & Petretto, 2017), others (Olderbak et al., 2015; Redondo & Herrero-Fernández, 2018), have not found a single factor solution, and instead obtained a shorter version of the test which yielded better internal consistency and homogeneity. We calculated the internal consistency of the 19-item version obtained in the Spanish study (Redondo & Herrero-Fernández, 2018) and found only a marginal increase in internal consistency from 0.53 to 0.58 (data not shown). The main limitation of the present study is the absence of a factor analysis approach that could have contributed to resolving the controversy related to the RMET, and to support that the SST is evaluating only one factor. Although the number of participants was enough to have adequate power for correlation analysis, it was not enough to calculate a factor analysis (Costello & Osborne, 2005).

The heterogeneity between items in the RMET could be related to the words used to describe the mental states in the test. We used the Argentinian version of the test because the terms used were more common in Mexico than those in the Spanish version. Nevertheless, when the mental state terms were searched in a corpus of contemporary Mexican Spanish (Silva-Pereyra, Rodríguez-Camacho, Prieto-Corona, & Aubert-Vázquez, 2013), only 75% of words were found, and their frequencies varied widely. Indeed, it has been suggested that a complete back translation for the RMET is necessary (Redondo and Herrero-Fernández (2018)), and we propose that regional variations in the use of Spanish should be taken into

cogent •• psychology

consideration. In this study, we found a significant negative correlation between mean score and response time for the RMET, evidencing that, as could be expected, participants take longer to answer the more challenging items. Also, performance in this test correlated with verbal fluency, so it is possible that the more challenging items required a finer distinction to be made among the mental state terms. Similar to our findings, item 19 was answered correctly by less than 50% of participants in both Argentinian (Allegri, 2006) and Spanish (Fernandez-Abascal et al., 2013) samples, while in other languages the items with least correct responses are different (for a review, see (Khorashad et al., 2015)).

Our results suggest that the RMET that evaluates emotion recognition, and a fictional story that requires the attribution of mental states to the characters of the story, share a portion of variance indicating that they may be measuring a shared construct, theory of mind. However, theory of mind is a multidimensional construct (Turner & Felisberti, 2017), that could be broken down into basic processes (Schaafsma et al., 2015). In this regard, neuroimaging studies could contribute to characterize the various processes involved. Individual neuroimaging studies and meta-analysis of those studies have provided a statistical map of the possible neural substrate underlying ToM (e.g., Mar, 2011; Schurz et al., 2014). Of particular interest to the present study are the results of the meta-analysis by Schurz et al. (2014). These authors found that tasks such as the RMET, recruit left inferior frontal areas, temporal regions including the temporo parietal junction (TPJ), and medial prefrontal cortex (mPFC). These areas have been implicated in reasoning about others' beliefs and intentions, considered cognitive ToM, as well as in making inferences about their valence, considered motivational ToM (Koster-Hale, Richardson, Velez, Asaba, Young, and Saxe, 2017). Importantly Koster-Hale et al., (2017) found that the brain areas recruited in relation to knowledge about others are different from those involved in other components of linguistic and conceptual processing.

Another meta-analysis (Mar, 2011) is particularly relevant for the present study because the SST uses narrative fiction to test ToM, and Mar's research showed that this particular type of task taps into brain areas involved in ToM. Briefly, Mar (2011) carried out a meta-analysis using the activation likelihood estimation (ALE) approach for neuroimaging studies of story comprehension and studies that used story-based and nonstory-based ToM tests. The conjunction analysis of the three types of tasks showed overlap in areas associated with ToM including mPFC, bilateral posterior superior temporal sulcus (pSTS), TPJ, bilateral middle temporal gyrus, and left inferior frontal gyrus (pars opercularis). Based on those results, Mar (2011) proposed that ToM processes are employed during narrative comprehension, as readers infer the mental states of characters similarly to how they understand them in real-life conspecifics. However, these brain areas include but are not limited to ToM and narrative comprehension; an alternative explanation is that they subserve more general cognitive processes. Processes such as the projection of the self, scene construction, associative processing, or the integration of motivational systems with language/categorization systems appear to share this network (Mar, 2011).

Finally, Yang, Rosenblau, Keifer, & Pelphrey (2015) have proposed that ToM is one of three central processes that are part of social information processing, the other two being social perception and action observation. They highlight the role of the pSTS, which is at the intersection of these processes and supports social information processing on different levels, from temporal integration of sensory cues of other's behaviors to representation of a basic form of intentionality. The proposed neural model includes connections between areas of social perception (amygdala, orbital frontal cortex, fusiform gyrus), areas of action observation (inferior parietal lobule, inferior frontal gyrus), and areas of ToM (TPJ, mPFC, anterior temporal lobule, posterior cingulate cortex/precuneus).

Together the theoretical proposals and neural models reviewed above, underscore the complexity and multidimensional nature of ToM, and its relevance as one of the psychological processes that encode socially and emotionally relevant inputs to provide a response that is adequate to the social context. Given the social nature of humans, expertise in social perception is evolutionarily

adaptive, and deficits in this ability seriously affect social competence and hinder valuable interactions (Yang et al., 2015). Psychometric tools that are reliable and do not show ceiling effects in educated, neurologically intact adults from different cultural backgrounds are necessary and useful not only for the clinical setting but also for research in ToM to propose a universal hypothesis regarding the basic operations that it entails, as well as its relation to other cognitive functions.

Conflict of interest
We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

Author details
M. Giordano[1]
E-mail: giordano@unam.mx
ORCID ID: http://orcid.org/0000-0002-4335-7953
G. Licea-Haquet[1]
E-mail: gilian.liha@gmail.com
E. Navarrete[1]
E-mail: edna22@comunidad.unam.mx
E. Valles-Capetillo[1]
E-mail: elizabeth.valless@comunidad.unam.mx
F. Lizcano-Cortés[1]
E-mail: fernando_lizcanoc@hotmail.com
A. Carrillo-Peña[1]
E-mail: alancarrillop@gmail.com
A. Zamora-Ursulo[1]
E-mail: angel.zamora.ursulo@gmail.com
[1] Departamento de Neurobiología Conductual y Cognitiva, Instituto de Neurobiología, UNAM, Blvd. Juriquilla 3001, Juriquilla, Qro 76230, Mexico.

Supplementary material
Supplemental data for this article can be accessed here.

References
Allegri, R. (2006). Reading the mind in the eyes: Control data from normal cohort in Spanish adult population form Buenos Aires Argentina, September 2006. Retrieved from: http://autismresearchcentre.com/arc_tests.

Baron-Cohen, S., Jollife, T., Mortimore, C., & Robertson, M. (1997). Another advanced test of theory of mind: Evidence from very high functioning adults with autism or asperger syndrome. *Journal of Child Psychology and Psychiatry, 38*(7), 813–822.

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. *Journal of Child Psychology and Psychiatry, 42*(2), 241–251.

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience, 14* (365–376), 1–12. doi:10.1038/nrn3475

Costello, A.B, & Osborne, J.W. (2005). Best practices in exploratory factor analysis: four recommendations for getting the most from yous analysis. practical assessment. *Research & Evaluation, 10*(7), 1-9.

Cruz-Fuentes, C. S., López-Bello, L., Blas-García, C., González-Macías, L., & Chávez-Balderas, R. A. (2005). Datos sobre la validez y confiabilidad de la Symptom Check List 90 (SCL 90) en una muestra de sujetos mexicanos. *Salud Mental, 28*(1), 72–81.

Cummings, L. (2017). Cognitive aspects of pragmatic disorders. In L. Cummings (Ed.), *Research in clinical pragmatics* (pp. 587–616). Switzerland: Cham.

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology, 10*, 85–102.

Dennett, D. C. (1987). *The intentional stance* (pp. 1–200). Cambridge, MA: MIT Press.

Diaz, V., & Farrar, M. J. (2018). Do bilingual and monolingual preschoolers acquire false belief understanding similarly? The role of executive functioning and language. *First Language, 38*(4), 382–398. doi:10.1177/0142723717752741

Dodell-Feder, D., Lincoln, S. H., Coulson, J. P., & Hooker, C. I. (2013). Using fiction to assess mental state understanding: A new task for assessing theory of mind in adults. *PLoS One, 8*(11), e81279. doi:10.1371/journal.pone.0081279.s001

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences, 32* (5), 429–492. doi:10.1017/S0140525X0999094X

Fernández-Abascal, E.G, Cabello, R, Fernández-Berrocal, P, & Baron-Cohen, S. (2013). Test-retest reliability of the 'reading the mind in the eyes' test: a one-year follow-up study. *Molecular Autism, 4*, 33. doi:10.1186/2040-2392-4-33

Flores Lázaro, J. C., Ostrosky Shejet, F., & Lozano Gutiérrez, A. (2014). BANFE: Batería Neuropsicológica de Funciones Ejecutivas y Lóbulos Frontales. [Measurement instrument]. Distrito Federal, México: El Manual Moderno.

Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron, 50*(4), 531–534. doi:10.1016/j.neuron.2006.05.001

Goetz, P. J. (2003). The effects of bilingualism on theory of mind development. *Bilingualism: Language and Cognition, 6*(1), 1–15. doi:10.1017/S1366728903001007

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33*(2/3), 1–75. doi:10.1017/S0140525X0999152X

Henry, J. D., Cowan, D. G., Lee, T., & Sachdev, P. S. (2015). Recent trends in testing social cognition. *Current Opinion in Psychiatry, 28*(2), 133–140. doi:10.1097/YCO.0000000000000139

Khorashad, B.S, Baron-Cohen, S, Roshan, G.M, Kazemian, M, Khazai, L, Aghili, Z, Talaei, A, & Afkhamizadeh, M. (2015). The "reading the mind in the eyes" test: investigation of psychometric properties and test–retest reliability of the persian version. *Investigation Of Psychometric Properties and Test-retest Reliabiity Of The Persian Version. Journal Of Autism and Developmental Disorders, 45*, 2651-2666. doi: 10.1007/s10803-015-2427-4

Kitayama, S., & Park, J. (2010). Cultural neuroscience of the self: Understanding the social grounding of the brain. *Social Cognitive and Affective Neuroscience, 5*, 111–129. doi:10.1093/scan/nsq052

Koster-Hale, J., Richardson, H., Velez, N., Asaba, M., & Young, L. & Saxe, R. (2017). Mentalizing regions represent distributed, continuous, and abstract dimensions of others' beliefs. *NeuroImage, 161*, 9–18. doi:http://dx.doi.org/10.1016/j.neuroimage.2017.08.026

Kovács, A. M. (2009). Early bilingualism enhances mechanisms of false-belief reasoning. *Developmental Science, 12*(1), 48.54. doi:10.1111/j.1467-7687.2008.00742.x

Lucas-Molina, B., Pérez-Albéniz, A., Ortuño-Sierra, J., & Fonseca-Pedrero, E. (2017). Dimensional structure and measurement invariance of the Interpersonal Reactivity Index (IRI) across gender. *Psicothema, 29*(4), 590–595. doi:10.7334/psicothema2017.19

Mar, R. A. (2011). The neural bases of social cognition and story comprehension. *Annual Review of Psychology, 62* (1), 103–134. doi:10.1146/annurev-psych-120709-145406

Nguyen, T., & Astington, J. (2014). Reassessing the bilingual advantage in theory of mind and its cognitive underpinnings. *Bilingualism: Language and Cognition, 17* (2), 396–409. doi:10.1017/S1366728913000394

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, 6* (653), 97–14. doi:10.3389/fpsyg.2015.01503

Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. *Journal of Neuroscience Methods, 162*(1), 8–13. doi:10.1016/j.jneumeth.2006.11.017

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *The Behavioral and Brain Sciences, 4*, 515–526.

Preti, A., Vellante, M., & Petretto, D. R. (2017). The psychometric properties of the "Reading the Mind in the Eyes" test: An item response theory (IRT) analysis. *Cognitive Neuropsychiatry*, 1–21. doi:10.1080/13546805.2017.1300091

Redondo, I., & Herrero-Fernández, D. (2018). Validation of the Reading the Mind in the Eyes test in a healthy Spanish sample and women with anorexia nervosa. *Cognitive Neuropsychiatry, 23*(4), 201–217. doi:10.1080/13546805.2018.1461618

Schaafsma, S. M., Pfaff, D. W., Spunt, R. P., & Adolphs, R. (2015). Deconstructing and reconstructing theory of mind. *Trends in Cognitive Sciences, 19*(2), 65–72. doi:10.1016/j.tics.2014.11.007

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: A meta-analysis of functional brain imaging studies. *Neuroscience and Biobehavioral Reviews, 42*, 9–34. doi:10.1016/j.neubiorev.2014.01.009

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. doi:10.1016/j.neuropsychologia.2007.05.021

Silva-Pereyra, J., Rodríguez-Camacho, M. A., Prieto-Corona, D. M. B., & Aubert-Vázquez, E. (2013). *LEXMEX: Diccionario de frecuencias del español de México* (1st ed., pp. 1–70). (U. N. A. de México, Ed.). Tlalnepantla: Edo. de México.

Turner, R., & Felisberti, F. M. (2017). Measuring mind-reading: A review of behavioral approaches to testing cognitive and affective mental state attribution in neurologically typical adults. *Frontiers in Psychology, 8*(e81279), 117–7. doi:10.3389/fpsyg.2017.00047

Vellante, M., Baron-Cohen, S., Melis, M., Marrone, M., Petretto, D. R., Masala, C., & Preti, A. (2013). The "Reading the Mind in the Eyes" test: Systematic review of psychometric properties and a validation study in Italy. *Cognitive Neuropsychiatry, 18*(4), 326–354. doi:10.1080/13546805.2012.721728

Yang, D. Y.-J., Rosenblau, G., Keifer, C., Pelphrey, K.A.. (2015). An integrative neural model of social perception, action observation, and theory of mind. *Neuroscience & Biobehavioral Reviews, 51*, 263–275. doi: http://dx.doi.org/10.1016/j.neubiorev.2015.01.020

cogent •• psychology

Cogent Psychology (ISSN: 2331-1908) is published by Cogent OA, part of Taylor & Francis Group.

Publishing with Cogent OA ensures:

- Immediate, universal access to your article on publication
- High visibility and discoverability via the Cogent OA website as well as Taylor & Francis Online
- Download and citation statistics for your article
- Rapid online publication
- Input from, and dialog with, expert editors and editorial boards
- Retention of full copyright of your article
- Guaranteed legacy preservation of your article
- Discounts and waivers for authors in developing regions

Submit your manuscript to a Cogent OA journal at www.CogentOA.com