Cognitive Neuropsychiatry, 2013
http://dx.doi.org/10.1080/13546805.2013.823859

Routledge
Taylor & Francis Group

The Reading the Mind in the Eyes test: validation of a French version and exploration of cultural variations in a multi-ethnic city

Marie Prevost[a]*, Marie-Eve Carrier[b], Gabrielle Chowne[a], Phyllis Zelkowitz[b], Lawrence Joseph[c] and Ian Gold[a]

[a]*Leacock Building, Room 908, Departments of Philosophy and Psychiatry, 855 Sherbrooke West, McGill University, Montreal, Quebec H3A 2T7, Canada;* [b]*Department of Psychiatry and Lady Davis Institute for Medical Research, Jewish General Hospital, 3755 Côte Ste-Catherine Road, Montreal, Quebec, H3T 1E2, Canada;* [c]*Division of Clinical Epidemiology, McGill University Health Centre, Royal Victoria Hospital, 687 Pine Avenue West, V Building, Room V2.10, Montreal, Quebec, Canada, H3A 1A1*

Introduction. The first aim of our study was to validate the French version of the Reading the Mind in the Eyes test, a theory of mind test. The second aim was to test whether cultural differences modulate performance on this test.
Methods. A total of 109 participants completed the original English version and 97 participants completed the French version. Another group of 30 participants completed the French version twice, one week apart.
Results. We report a similar overall distribution of scores in both versions and no differences in the mean scores between them. However, 2 items in the French version did not collect a majority of responses, which differed from the results of the English version. Test-retest showed good stability of the French version. As expected, participants who do not speak French or English at home, and those born in Asia, performed worse than North American participants, and those who speak English or French at home.
Conclusions. We report a French version with acceptable validity and good stability. The cultural differences observed support the idea that Asian culture does not use theory of mind to explain people's behaviours as much as North American people do.

Keywords: theory of mind; empathy; translation; Francophone; culture; language

Introduction

The Reading the Mind in the Eyes Task (RMET; Baron-Cohen, Jolliffe, Mortimore, & Robertson, 1997) was initially created to fill a gap in the evaluation of theory of mind in autistic adults. Theory of mind, the ability to attribute or infer beliefs, intentions and desires to others (Premack & Woodruff, 1978), has mostly been evaluated using tests that are easily passed by most adults, with or without autism, who have a mental age greater than 6. Baron-Cohen et al. (1997) thus created the RMET to explore more subtle theory of mind biases in adults. The task involves looking at pictures of strangers' faces and choosing one of two words that best describes what the person in the picture is feeling or thinking. Because the test had various psychometric problems, it was revised by the same team (Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001). The test now has 36 pictures of the

*Corresponding author. Email: marie.prevost@mail.mcgill.ca

eye regions of males and females, and participants have to choose the best word of four that describes what the person is feeling or thinking. Originally in English, the RMET has been translated into several languages, including Turkish (Kelemen, Keri, Must, Benedek, & Janka, 2004), Japanese (Kunihira, Senju, Dairoku, Wakabayashi, & Hasegawa, 2006), German (Voracek & Dressler, 2006), and Swedish (Hallerbäck, Lugnegård, Hjärthag, & Gillberg, 2009). One of the aims of the present study is to validate the French version of this test by comparing the performance of English speakers and French speakers using the original RMET and the French RMET respectively.

The RMET is largely used in research addressing social cognition deficits. High-functioning autistic patients perform worse than others on the RMET (Baron-Cohen et al., 2001; Murphy, 2006). The observation of such bias was not possible with classical first- or second-order false-belief tasks, which are not difficult enough to differentiate populations with more subtle social deficits. The RMET has also been repeatedly used in schizophrenia patients who show a deficit in mental-state recognition (Hirao et al., 2008; Irani et al., 2006; Kelemen et al., 2005; Kettle, O'Brien-Simpson, & Allen, 2008; Russell et al., 2000). A French version of the test will be useful for researchers and clinicians working with Francophone populations of autistic and schizophrenia patients. Apart from being a useful tool to explore possible social deficits in patients with various mental health disorders and neurological pathologies, the RMET has introduced more ecologically valid stimuli than those usually used for the study of social cognition.

The need for more ecological stimuli and methods of testing has been made clear in recent years as evidence has accumulated that one's thought and behaviour is likely to differ between real social interactions and laboratory tasks (Chatel-Goldman, Schwartz, Jutten, & Congedo, 2013). One question about ecological validity is whether culture influences our understanding of others' mental states. Surprisingly little is known about this, even though it is *prima facie* plausible that culture would modulate the predictions, and the explanation, of others' behaviour (Lillard, 1998, 1999). The importance of this question goes beyond the scope of research. At a time when information and goods are exchanged everyday between people from all around the world, and when people easily move from one continent to another, adjustments have to be made to accommodate all. Showing that we do not understand people in a similar manner according to our cultural background will help to better shape these accommodations.

Previous studies have made cross-cultural comparisons using tests made for their specific culture. Shahaeian, Peterson, Slaughter, and Wellman (2011) explored theory of mind development in children in Iran and children in Australia. Adams et al. (2010) observed that white American students performed better on the RMET with the original pictures than did visiting Japanese students. Other studies have showed that cultural background modulates the way we attribute causes to behaviours and events (Mason & Morris, 2006). Montreal is a culturally diverse city, which makes it convenient for exploring cultural questions. Another aim of the present study, therefore, was to explore the influence of culture on theory-of-mind performance in people living in a cultural environment that is different to the one they were born into. More specifically, we investigated how country and continent of birth predicted RMET performance, and whether native language, when other than French or English, influenced that performance.

Methods

Translation

We first translated the English version of the test into French, and then a professional translator reverse translated it into English. The reverse translation was compared to the original version of the English test. Problematic items were reviewed and revised.

It is essential in any translation that translated words are understood in the same manner as in their original language. Although several adjectives used in the RMET refer to basic emotional states such as happy, sad, scared, and angry, and are thus easily translated and understood cross-culturally, more complex mental states, such as "flustered" or "aghast", require words with narrower meanings, which may not have an exact parallel in the target language. Translating the target word and foil words may also alter the level of difficulty of the terms and alter performance in the target language. A further complication is raised by having French as the target language. Different dialects of French are spoken in different regions, including Quebec. Our goal was to create a single translation that would be understood equally well by natives of both France and Quebec. As a result, a native of France and a native French speaker of Quebec jointly carried out the translation. The translators reached an agreement on the French equivalents of the English terms. If a reverse translation diverged from the original English, the French terms were modified. The French version is given in Appendix 1.

Participants

Three groups of participants completed the task. One group consisted of 97 Francophone mothers (mean age = 31.3 years old, $SD = 4.6$) and a second group of 139 Anglophone mothers (mean age = 31.6, $SD = 4.4$), recruited at the Jewish General Hospital and the Maison des Naissances, in Montreal, as part of a larger study. The third group consisted of 30 Francophone students recruited in Montreal (23 women, mean age = 23.2 years old, $SD = 6.9$).

Participants from the two first groups were financially compensated for their time, as part of a larger study, but not for the purpose of the validation in particular. Students completed the French version of the RMET without financial compensation.

All participants gave written consent to participate in the study, which was approved by the Research Ethics Board of McGill University.

Procedure

Participants completed a socio-demographic questionnaire, which included questions about their country of birth, the number of years living in Canada, ethnicity, the language spoken at home, age, and the highest academic degree completed.

Testing followed the format of the standard version of the test by Baron-Cohen et al. (2001). Before beginning the test, an oral description was given to participants. Participants were presented with a single image of a person's eyes and their immediate eye region, along with four descriptive words. They were then asked to choose the word that they believed best described the emotional or mental state of the person in the image. This procedure was carried out for the 1 practice picture and

the 36 test pictures. A score of 1 is given for each correctly chosen target word and a score of 0 if a foil word is chosen. The total score thus ranges from 0 to 36. Participants took as much time as they wanted to complete the test.

A glossary of all the mental state terms was included in order to avoid problems of comprehension affecting an individual's score. Participants were encouraged to use it if they did not understand a word. The glossary was translated by the two translators of the test.

For test-retest analysis, the RMET was administered once and re-administered a second time one week later, following the same procedure.

Analysis

Validity was estimated by comparing scores from the original version of the test to scores from the French version. Overall mean scores and item-specific results were analysed. In each case, we calculated the difference in response rates between English and French versions. Internal consistency was tested using Cronbach's alpha. We report Bootstrap 95% Confidence Interval based on 1000 samples for each alpha.

Test-retest reliability was evaluated using Spearman's correlation. The percentages of participants answering correctly each time across each item, and overall, were calculated, along with the distribution of the total scores. Single Score Intraclass Correlation (ICC), oneway model, was used to evaluate consistency from test to retest. In addition, we report the Bland-Altman plot, which gives a visual representation of the agreement between scores at test and retest.

Items 13 and 23 were not validated (see the Results section below). As a result, all analyses concerning cultural differences excluded items 13 and 23 from the total score of the French version of the RMET. Therefore, we report the percentage of correct answers instead of mean scores, so that results are comparable between populations for the reader. Cultural differences on the RMET mean scores were tested by comparing groups: Canadian born versus non-Canadian born; North Americans versus South Americans, European, African, and Asians; French-speaking or English-speaking at home versus speaking another language at home; Caucasians versus non-Caucasians. Correlations were used to estimate the association between RMET performance and years spent in Canada. In addition, to evaluate whether the frequency of use of the RMET words would modulate performance, frequency of use for target words was calculated (Kucera & Francis, 1967; New, Pallier, Ferrand, & Matos, 2001). Mean scores for target words with high frequency of use were compared to mean score for target words with low frequency of use using t-tests in each group.

We report 95% confidence intervals for all parameter estimates.

Results

Validity

The mean scores were 25.6 ($SD = 5.4$) and 24.8 ($SD = 3.8$) for the Anglophone and the Francophone participants, respectively. The mean difference between the two groups was thus -0.8 ($CI = [-2.1; 0.3]$). Internal consistency was evaluated using

Cronbach's alpha. For the English version, Cronbach's alpha was .77 (*CI* = [.69; .81]) and for the French version, it was .53 (*CI* = [.34; .65]).

Distributions are reported in Table 1 and Figure 1. In the original English version of the RMET, target words for 3 items had a frequency inferior to 50% (Items 7, 19 and 35), meaning that less than half the participants chose that word as the target word. Similar results were observed for the same items in the French version.

Table 1. Distribution for each item (percentage).

	English (N = 139)				French (N = 97)			
Item	Word 1	Word 2	Word 3	Word 4	Word 1	Word 2	Word 3	Word 4
1	**76**	8	11	5	**84**	12	4	0
2	7	**68**	1	23	5	**70**	1	24
3	4	4	**85**	6	2	2	**93**	3
4	1	**53**	14	32	4	**57**	13	26
5	14	14	**69**	3	9	11	**71**	8
6	3	**84**	11	2	2	**80**	13	4
7	10	32	**47**	10	16	39	**33**	11
8	**81**	10	6	2	**68**	15	14	2
9	6	6	1	**86**	9	5	4	**81**
10	**58**	27	14	2	**60**	24	12	4
11	6	15	**72**	7	4	12	**57**	27
12	18	8	**70**	4	16	8	**75**	0
13	16	**60**	6	18	33	**34**	3	30
14	9	17	7	**68**	7	6	2	**85**
15	**69**	4	15	12	**84**	2	11	3
16	1	**66**	3	29	0	**79**	3	18
17	**55**	31	12	3	**48**	39	8	4
18	**65**	19	4	13	**86**	11	0	3
19	14	22	14	**50**	15	29	12	**43**
20	6	**83**	10	0	5	**92**	3	0
21	12	**79**	8	1	4	**86**	10	0
22	**85**	2	2	11	**87**	1	5	7
23	7	9	**53**	31	5	13	**37**	44
24	**75**	12	4	9	**84**	10	1	5
25	4	17	7	**72**	1	19	4	**76**
26	12	10	**60**	18	7	3	**68**	22
27	2	**62**	20	16	3	**49**	28	20
28	**68**	1	24	7	**73**	5	14	7
29	19	4	11	**66**	3	25	6	**66**
30	6	**82**	8	4	8	**80**	6	5
31	12	**66**	14	8	6	**69**	14	10
32	**71**	9	9	12	**80**	5	6	8
33	4	19	5	72	9	18	13	**60**
34	4	20	**59**	17	2	29	**63**	6
35	24	**50**	10	16	30	**47**	14	8
36	1	4	**81**	14	2	3	**71**	24

Notes: Target items are in bold. Items whose target word received less than 50% of answers are shaded grey. Items that scored differently in the French version compared to the original version are shaded darker grey.

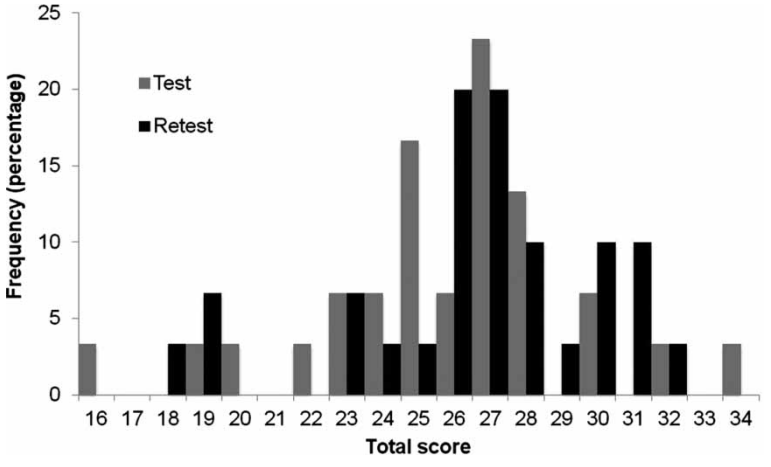

Figure 1. Distribution for the total scores on the original English version (in grey) and on the French version (in black).

In the French version, target words for 4 additional items had a frequency inferior to 50% (items 13, 17, 23 and 27). Results were inconclusive in comparing the original RMET and the French version scores for items 17 (mean difference = 6.2%, *CI* = [−7; 19]) and 27 (mean difference = 12.4%, *CI* = [−0.4; 25.2]), where the target words received the majority of responses. However, scores were different between the two versions of the test for items 13 (mean difference = 25.7%, *CI* = [13.2; 38.2]) and 23 (mean difference = 15.4%, *CI* = [2.7; 28.1]), where the target words did not receive the majority of responses.

Test-retest reliability

Distributions were similar for both tests. There was no difference between the total scores obtained the first time (mean = 25.7, *SD* = 3.7) and those obtained on retest one week later (mean = 26.6, *SD* = 3.6; mean difference = 0.9, *CI* = [−2.7; 1.0]). As expected, there was a strong positive correlation between the test and the retest scores ($r = .7$, *CI* = [.5; .8]). The Intraclass Correlation Coefficient (ICC) was of .70 (*CI* = [.46; .84]). As an additional test of agreement between test and retest, the Bland-Altman plot is reported in Figure 2 (the 95% limits for the range of possible error, as indicated in the figure, was [−6.23; 4.57]).

The distribution of participants' choices for each item is reported in Table 2. Items 7, 13, and 17 did not receive the majority of responses on the same word from test to retest. For item 7, only 36.6% of participants chose the target word both times (*CI* = [19.9; 56.1]). For item 13, nobody chose the target word both times (*CI* = [0; 9.5%]). For item 17, only 20% chose the target word both times (*CI* = [7.7; 38.6]). In addition, less than 50% of the participants chose the target word both times for items 10, 19, 23, 25, 29, 33, and 35 (see Table 3). However, the target words for these items received the majority of responses both times, as shown in Table 2.

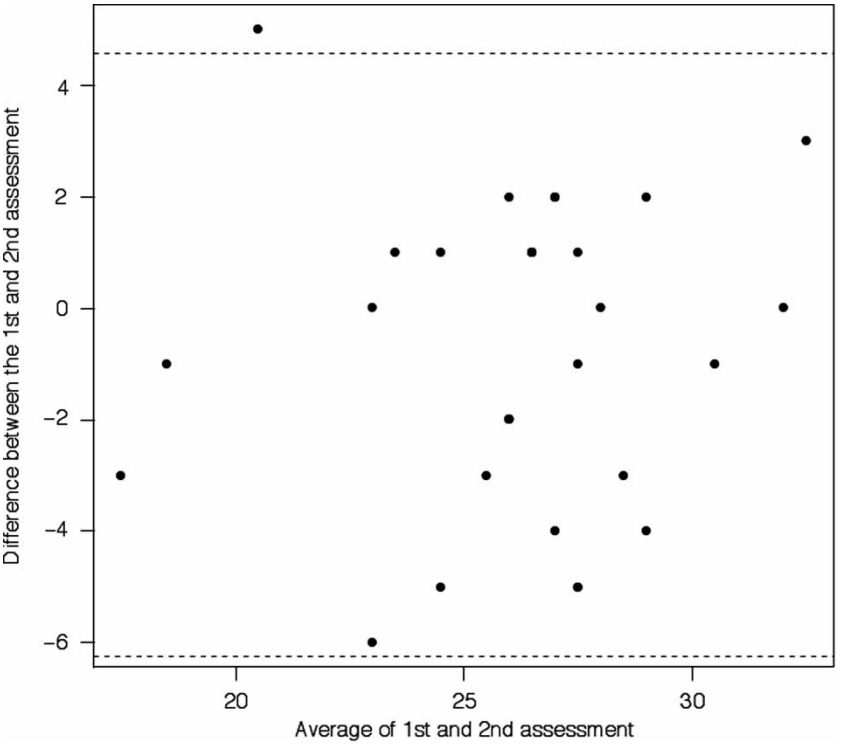

Figure 2. Bland-Altman Plot of the stability of the test at one-week interval.

Cultural differences

We investigated whether continent or country of birth, ethnicity, language spoken at home, and years spent in Canada influenced performance on the RMET. A total of 88 participants were born outside Canada, and 148 were born in Canada. Out of the 236 participants, 47 did not speak French or English at home, and data are missing for 1 participant. Mean education levels were not different between participants born outside Canada (mean = 15.2 years, $SD = 2.6$) and participants born in Canada (mean = 15.4 years, $SD = 2.8$; mean difference = −0.2 years, $CI = [-0.9; 0.5]$), or between participants not speaking either English or French at home (mean = 14.5 years, $SD = 3.0$) and those speaking these languages at home (mean = 15.4 years, $SD = 2.6$; mean difference = 0.9, $CI = [-0.0; 1.9]$). Amongst the non-Canadian born participants, 5 were from North America (USA, $n = 5$), 15 were from South America (Brazil, $n = 1$; Colombia, $n = 2$; El Salvador, $n = 1$; Haiti, $n = 4$; Martinique, $n = 1$; Mexico, $n = 2$; Peru, $n = 1$; St Vincent and the Grenadine, $n = 1$; Venezuela, $n = 2$), 29 were from Europe (Belgium, $n = 2$; Bosnia-Herzegovina, $n = 1$; Bulgaria, $n = 1$; Croatia, $n = 1$; England, $n = 1$; France, $n = 10$; Germany, $n = 1$; Italy, $n = 2$; Poland, $n = 2$; Romania, $n = 6$; Switzerland, $n = 2$), 12 were from Africa (Algeria, $n = 3$; Guinea, $n = 1$; Madagascar, $n = 1$; Morocco, $n = 3$; Congo, $n = 1$; Senegal, $n = 1$; Tunisia, $n = 2$), and 27 were from Asia (Afghanistan, $n = 1$; Bangladesh, $n = 2$; China, $n = 1$; India, $n = 2$; Israel, $n = 4$; Kazakhstan, $n = 1$; Philippines, $n = 5$; Russia, $n = 2$;

Table 2. Distribution for each item (percentage).

	Test				*Retest*			
Item	*Word 1*	*Word 2*	*Word 3*	*Word 4*	*Word 1*	*Word 2*	*Word 3*	*Word 4*
1	**87**	3	10	0	**93**	3	3	0
2	7	**73**	10	10	7	**73**	3	17
3	0	7	**80**	13	0	3	**83**	14
4	3	**73**	7	17	10	**70**	7	3
5	0	17	**83**	0	10	13	**77**	0
6	17	**73**	7	3	0	**87**	13	0
7	10	43	**40**	7	3	33	**53**	10
8	**83**	7	0	10	**90**	10	0	0
9	13	3	0	**83**	10	7	0	**83**
10	**50**	17	13	20	**50**	30	13	7
11	3	0	**90**	7	3	3	**83**	10
12	20	0	**80**	0	10	7	**80**	3
13	27	**23**	0	50	23	**20**	7	50
14	3	4	0	**93**	7	3	0	**87**
15	**80**	3	10	7	**87**	3	0	10
16	7	**77**	0	16	0	**83**	7	10
17	**23**	27	13	37	**30**	20	23	27
18	**97**	3	0	0	**97**	3	0	0
19	13	30	7	**50**	7	27	3	**63**
20	13	**70**	17	0	3	**80**	17	0
21	13	**70**	17	0	7	**83**	10	0
22	**90**	0	7	3	**83**	0	7	10
23	0	20	**43**	37	7	10	**47**	36
24	**83**	3	4	10	**74**	13	3	10
25	13	30	10	**47**	20	13	10	**57**
26	7	3	**73**	17	7	0	**77**	16
27	3	**73**	13	10	3	**77**	13	7
28	**87**	7	3	3	**93**	7	0	0
29	10	33	10	**47**	7	20	17	**56**
30	3	**93**	4	0	0	**93**	3	4
31	3	**80**	10	7	3	**77**	13	7
32	**67**	10	7	16	**83**	0	7	10
33	10	7	3	**80**	7	40	0	**53**
34	3	7	**83**	7	3	0	**87**	10
35	23	**63**	7	7	30	**60**	7	3
36	0	7	**76**	17	0	3	**87**	10

Notes: Target items are in bold. Items for which the target word did not collect the majority of answers both times are shaded grey.

Saudi Arabia, $n = 1$; South Korea, $n = 1$; Sri Lanka, $n = 1$; Syria, $n = 3$; Taiwan, $n = 1$; Ukraine, $n = 1$; Vietnam, $n = 1$).

Participants who were born outside Canada performed worse (mean = 64.2%, $SD = 14.4$) than Canadian-born participants (mean = 68.9%, $SD = 11.9$; difference = 4.7, $CI = [1.3; 8.1]$). This effect was due to participants born in Asia performing worse (58.2%, $SD = 15.9$, $n = 27$) than North American participants (68.8%, $SD = 12.1$, $n = 153$; mean difference = 10.6, $CI = [4.1; 17.2]$), whereas

Table 3. Percentage of participants who chose the target word both at first test and one week later.

Item	*n*	*N*	*Percentage*	*95% Confidence interval, lower limit*	*95% Confidence interval, upper limit*
1	25	30	83.33%	65.28%	94.36%
2	19	30	63.33%	43.86%	80.07%
3	21	30	70.00%	50.60%	85.27%
4	18	30	60.00%	40.60%	77.34%
5	23	30	76.67%	57.72%	90.07%
6	20	30	66.67%	47.19%	82.71%
7	11	30	36.67%	19.93%	56.14%
8	24	30	80.00%	61.43%	92.29%
9	21	30	70.00%	50.60%	85.27%
10	9	30	30.00%	14.73%	49.40%
11	23	30	76.67%	57.72%	90.07%
12	21	30	70.00%	50.60%	85.27%
13	0	30	0.00%	0.00%	9.50%
14	24	29	82.76%	64.23%	94.15%
15	21	30	70.00%	50.60%	85.27%
16	20	30	66.67%	47.19%	82.71%
17	6	30	20.00%	7.71%	38.57%
18	28	30	93.33%	77.93%	99.18%
19	11	30	36.67%	19.93%	56.14%
20	19	30	63.33%	43.86%	80.07%
21	21	30	70.00%	50.60%	85.27%
22	23	30	76.67%	57.72%	90.07%
23	8	30	26.67%	12.28%	45.89%
24	20	30	66.67%	47.19%	82.71%
25	8	30	26.67%	12.28%	45.89%
26	19	30	63.33%	43.86%	80.07%
27	20	30	66.67%	47.19%	82.71%
28	25	30	83.33%	65.28%	94.36%
29	12	30	40.00%	22.66%	59.40%
30	26	30	86.67%	69.28%	96.24%
31	18	30	60.00%	40.60%	77.34%
32	19	30	63.33%	43.86%	80.07%
33	14	30	46.67%	28.34%	65.67%
34	23	30	76.67%	57.72%	90.07%
35	14	30	46.67%	28.34%	65.67%
36	21	30	70.00%	50.60%	85.27%

Note: Items for which less than 50% of participants chose the target word both times are shaded grey.

performance of participants born in South America (65.7%, $SD = 13.8$, $n = 15$), in Europe (68.7%, $SD = 11.1$, $n = 29$), and in Africa (63.6%, $SD = 13.4$, $n = 12$) did not differ from that of North Americans. Participants who did not speak English or French at home also performed worse (mean = 57.9%, $SD = 16.6$) than participants who spoke English or French at home (mean = 69.4%, $SD = 10.9$; difference = 11.5, $CI = [7.5; 15.4]$). Among the participants born outside Canada, 44 were of Caucasian ethnicity and 44 were of other ethnicities. Caucasian participants did not perform

conclusively better (mean = 66.9%, $SD = 13.2$) than participants of other ethnicities (mean = 61.4%, $SD = 15.1$; difference = 5.5, $CI = [-0.5; 11.5]$). In the group of participants born outside Canada, there was no correlation between the number of years spent in Canada and performance on the RMET ($n = 88$, $r = .1$).

Items whose target word had a low frequency of use ($n = 15$ in English, mean frequency = 3.0, $SD = 2.4$, $n = 18$ in French, mean frequency = 2.4, $SD = 1.4$) were contrasted against those with a high frequency of use ($n = 21$ in English, mean frequency = 40.8, $SD = 41.2$, $n = 18$ in French, mean frequency = 17.7, $SD = 21.6$). There was no difference between scores for high frequency targets (English: 59.6%, $SD = 16.6$; French: 56.7, $SD = 13.6$) and low frequency targets (English: 58.9%, $SD = 25.2$; French: 64.4, $SD = 20.5$) in the participants who did not speak English or French at home (English $N = 37$: mean difference = 0.5, $CI = [-5.9, 6.9]$; French $N = 10$: mean difference = -7.7, $CI = [-23.5, 7.9]$).

Discussion

The present study aimed at validating a French version of the Reading the Mind in the Eyes test originally created by Baron-Cohen et al. (2001) and investigating the effect of culture on participants' performance on this test. The RMET has been widely used as an index of theory of mind, or empathy, both in clinical and healthy populations. The RMET is itself difficult to validate; there are numerous tests of theory of mind or empathy, but they all vary greatly in their design, content, and target population. We did not collect participants' scores on others tests and do not, therefore, report data addressing convergent/divergent validity, discriminant validity, or predictive validity. Because the RMET has been shown to be associated with some theory of mind tasks (de Achával et al., 2010; McGlade et al., 2008), but not others (Boisseau, 2010; Dziobek et al., 2006), and associated with IQ in some studies (Ahmed & Miller, 2011; Golan & Baron-Cohen, 2006; Kenyon et al., 2012), but not others (Baron-Cohen et al., 2001; Chapman et al., 2006; Kelemen et al., 2004), the best strategy for validation was to compare the new French version to the original English version.

Thus, to validate a French version, we compared the scores of a Francophone population using a new French translation to that of an Anglophone population using the original version, tested in the same conditions. We found that distributions are similar in the English and the French versions, and the mean total scores were not different between the Francophone and the Anglophone populations, suggesting that the translation exhibited a satisfactory validity.

However, as measured with Cronbach's alpha, internal consistency was poor in the French version, whereas it was acceptable in the English version. Previous studies that tested the internal consistency of the RMET (Harkness, Jacobson, Duong, & Sabbagh, 2010; Vellante et al., 2012; Voracek & Dressler, 2006) found a Cronbach's alpha as low as in our version; only one study reported a good Cronbach's alpha (Dehning et al., 2012). This suggests that there is considerable variability in RMET scores across populations. As a result, limits of internal consistency are likely due to the original test rather than the translation. In addition, target words for 3 items were chosen less than 50% of the time in both the English and the French versions, again suggesting that the issue is not in the translation. In their original paper, Baron-Cohen et al. (2001) validated all the items, including these 3 conflicting ones. The unexpected

results reported here might be due to the quality of the printed pictures. It may also be due to the relative homogeneity of the participant groups made up of mothers.

In the French version, 4 additional items showed too broad a distribution of answers between the target and the foil words. However, only 2 of these items (items 13 and 23) had clearly worse scores than the English version. Thus our French version has good validity, except for these items, which should be removed or retested using a different translation. For item 13, we tested another French translation of the target word, replacing "prévoyant" with "anticipant", in 38 female participants, without change in the results. The distribution of answers for this item is quite balanced between the target word and 2 foil words (the other foil word received less than 10% of the total answers for this item), suggesting that participants are choosing randomly. A possible explanation is that the meaning of the facial expression in picture 13 might be culturally specific and might, therefore, be a poor match with the French translation of the English word. With respect to item 23, a different pattern emerged: the majority of Anglophones chose the target word "defiant", and one third chose the foil word "curious". The majority of Francophones, however, chose "curieux" (curious), closely followed by the target word "provoquant" (defiant). In this case, it seems that there is some agreement between the groups about the possible target words, but there was a slight mismatch between defiant and the French translation. Interestingly, the words in this picture were not the ones that were controversial during the process of translation and reverse translation. In any case, pictures 13 and 23 ought to be considered carefully and scored separately when using the French version.

Test-retest showed that item 13 remains problematic, as the target word did not receive the majority of answers with this new population of students, and none of those who selected the target words during the first testing phase chose it again one week later. In addition, for a number of items, participants who choose the target word the first time did not choose it the second time, even though overall, the target words received the majority of responses. This overall good stability was confirmed by the Bland-Altman plot and the Intraclass Correlation Coefficient. It thus seems that the stability of the test is not linear: participants obtained on average the same score from one week to the other, but they did not answer all the same questions correctly.

Given these limitations, we excluded items 13 and 23 when we turned to the question of cultural variations on the RMET. As expected, participants who were born outside of Canada, or who did not speak French or English at home, performed worse on the RMET than Canadian-born participants, or those speaking English, French, or both languages at home. As there were no differences in level of education between these groups, we concluded that these findings indeed reflected a cultural difference with respect to theory of mind. In particular, participants born in Asia performed worse than participants born in North America, but no difference was observed for those born in South America, Europe, and Africa, suggesting that Asians may approach some theory of mind tasks differently than members of other cultures. Previous research has shown that people from different cultures might not make as much mental state inferences as Westerners and might draw on different sources of information to explain behaviour. People from East Asia are more likely to depend on social context than on internal state to explain behaviour (Choi, Dalal, Kim-Prieto, & Park, 2003; Morris & Peng, 1994). Our data show that participants born in Asia performed worse than the North American participants on the RMET. This may be because they tend to rely more on environmental and contextual cues

and thus have less practice in making mental state attributions from faces (Masuda et al., 2008; Masuda & Nisbett, 2001).

There were no differences between Caucasian and non-Caucasian participants among those born outside Canada. It thus seems that the differences in performance observed in the present study do not reflect a difference in perception of facial indices that would differ across ethnicities, as suggested in a previous study (Rule et al., 2010). The idea that variations in performance on the RMET were due to linguistic competence was not supported by our data. There was no difference between performance on items with a high frequency of use and those with a low frequency of use in the participants who did not speak English or French at home. Finally, we did not observe an association between the number of years spent in Canada and RMET performance. With time, foreign-born participants would be expected to become more familiar with their new language. The absence of learning curve on performance at the RMET thus suggests that language is not a primary contributor to performance and further confirms the stability of the test. Taken together, these results show that cultural differences in theory of mind are thus not due to difference in facial appearance or to language difficulties.

Limitations

Even though the RMET is a more ecological test than its predecessors, it is not without weaknesses (see Johnston, Miles, & McKinlay, 2008). Because the pictures are taken from magazines, the actual mental states represented are unknown. Competence in theory of mind is instead operationalised as agreement with the majority. Since as a matter of definition, most healthy human beings have a satisfactory capacity for theory of mind, this is likely to be adequate for most subjects. The RMET cannot, however, identify individuals who are exceptionally sensitive to other people's mental states. Despite this caveat, as a measure of theory of mind, the RMET best satisfies the desiderata of being quick, easy to use, more ecological than text alone, and more subtle than basic emotional perception or valence discrimination—but not so easy as to lead to a ceiling effect.

Another potential limitation of the present study is that participants for the validation of the test were all women. Some studies have observed differences between male and female participants on this test (Alaerts, Nackaerts, Meyns, Swinnen, & Wenderoth, 2011; Carroll & Chiew, 2006; Dehning et al., 2012; Hallerbäck et al., 2009; Vellante et al., 2012; Voracek & Dressler, 2006; Yildirim et al., 2011), suggesting that women perform better at the RMET than men. Testing people who are inherently very good at this test and who might easily overcome subtle difficulties added by the translation might thus bias the validation. However, this is very unlikely, since most of the literature reports no difference between male and female participants (Ahmed & Miller, 2011; Baron-Cohen et al., 2001; Chapman et al., 2006; Cook & Saucier, 2010; Harkness et al., 2010; Kettle et al., 2008; Kunihira et al., 2006; Serafin & Surian, 2004; Smeets, Dziobek, & Wolf, 2009; Stanford, Messinger, Malaspina, & Corcoran, 2011; Valla et al., 2010).

Conclusions

The present study reported data showing that a satisfying French version of the RMET is now available. Because translation will always slightly distort an original

version, researchers should be careful when using items 13 and 23 of the French RMET. We also showed that immigrants born in Asia performed worse than Canadian-born participants, adding to the literature suggesting that Asian people do not use the theory of mind as much as Western people do. Interestingly, speaking a language other than French or English at home predicted worse performance than speaking French or English at home, but this was independent of vocabulary proficiency, in accordance with a well-established association between language development and theory of mind capacities.

Acknowledgements

We thank Dr Martin Gibert and Antoine C-Dussault for their valuable help with recruitment.

Funding

This work was supported by the CIHR Team in Perinatal Mental Health.

References

Adams, R. B., Jr., Rule, N. O., Franklin, R. G., Jr., Wang, E., Stevenson, M. T., Yoshikawa, S., ... Ambady, N. (2010). Cross-cultural reading the mind in the eyes: An fMRI investigation. *Journal of Cognitive Neuroscience*, *22*, 97–108. doi:10.1162/jocn.2009.21187

Ahmed, F. S., & Miller, S. L. (2011). Executive function mechanisms of theory of mind. *Journal of Autism and Developmental Disorders*, *41*, 667–678. doi:10.1007/s10803-010-1087-7

Alaerts, K., Nackaerts, E., Meyns, P., Swinnen, S. P., & Wenderoth, N. (2011). Action and emotion recognition from point light displays: An investigation of gender differences. *PLOS ONE*, *6*(6): e20989. doi:10.1371/journal.pone.0020989

Baron-Cohen, S., Jolliffe, 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*, 813–822. doi:10.1111/j.1469-7610.1997.tb01599.x

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*, 241–251. doi:10.1111/1469-7610.00715

Boisseau, E. (2010). *La théorie de l'esprit chez les schizophrènes paranoïdes* [The theory of mind in paranoid schizophrenia patients] (Unpublished PhD thesis). Université du Québec à Montréal, Canada.

Carroll, J. M., & Chiew, K. Y. (2006). Sex and discipline differences in empathising, systemising and autistic symptomatology: Evidence from a student population. *Journal of Autism and Developmental Disorders*, *36*, 949–957. doi:10.1007/s10803-006-0127-9

Chapman, E., Baron-Cohen, S., Auyeung, B., Knickmeyer, R., Taylor, K., & Hackett, G. (2006). Fetal testosterone and empathy: Evidence from the Empathy Quotient (EQ) and the "Reading the Mind in the Eyes" test. *Social Neuroscience*, *1*, 135–148. doi:10.1080/17470910600992239

Chatel-Goldman, J., Schwartz, J.-L., Jutten, C., & Congedo, M. (2013). Non-local mind from the perspective of social cognition. *Frontiers in Human Neuroscience*, *7*:107. doi:10.3389/fnhum.2013.00107

Choi, I., Dalal, R., Kim-Prieto, C., & Park, H. (2003). Culture and judgment of causal relevance. *Journal of Personality and Social Psychology*, *84*, 46–59. doi:10.1037/0022-3514.84.1.46

Cook, C. M., & Saucier, D. M. (2010). Mental rotation, targeting ability and Baron-Cohen's Empathizing–Systemizing Theory of Sex Differences. *Personality and Individual Differences*, *49*, 712–716. doi:10.1016/j.paid.2010.06.010

de Achával, D., Costanzo, E. Y., Villarreal, M., Jáuregui, I. O., Chiodi, A., Castro, M. N., … Guinjoan, S. M. (2010). Emotion processing and theory of mind in schizophrenia patients and their unaffected first-degree relatives. *Neuropsychologia*, *48*, 1209–1215. doi:10.1016/j.neuropsychologia.2009.12.019

Dehning, S., Girma, E., Gasperi, S., Meyer, S., Tesfaye, M., & Siebeck, M. (2012). Comparative cross-sectional study of empathy among first year and final year medical students in Jimma University, Ethiopia: Steady state of the heart and opening of the eyes. *BMC Medical Education*, *12*:34. doi:10.1186/1472-6920-12-34

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*, 623–636. doi:10.1007/s10803-006-0107-0

Golan, O., & Baron-Cohen, S. (2006). Systemizing empathy: Teaching adults with Asperger syndrome or high-functioning autism to recognize complex emotions using interactive multimedia. *Development and Psychopathology*, *18*, 591–617. doi:10.1017/S0954579406060305

Hallerbäck, M. U., Lugnegård, T., Hjärthag, F., & Gillberg, C. (2009). The Reading the Mind in the Eyes Test: Test-retest reliability of a Swedish version. *Cognitive Neuropsychiatry*, *14*, 127–143. doi:10.1080/13546800902901518

Harkness, K. L., Jacobson, J. A., Duong, D., & Sabbagh, M. A. (2010). Mental state decoding in past major depression: Effect of sad versus happy mood induction. *Cognition and Emotion*, *24*, 497–513. doi:10.1080/02699930902750249

Hirao, K., Miyata, J., Fujiwara, H., Yamada, M., Namiki, C., Shimizu, M., … Murai, T. (2008). Theory of mind and frontal lobe pathology in schizophrenia: A voxel-based morphometry study. *Schizophrenia Research*, *105*, 165–174. doi:10.1016/j.schres.2008.07.021

Irani, F., Platek, S. M., Panyavin, I. S., Calkins, M. E., Kohler, C., Siegel, S. J., … Gur, R. C. (2006). Self-face recognition and theory of mind in patients with schizophrenia and first-degree relatives. *Schizophrenia Research*, *88*, 151–160. doi:10.1016/j.schres.2006.07.016

Johnston, L., Miles, L., & McKinlay, A. (2008). A critical review of the eyes test as a measure of social-cognitive impairment. *Australian Journal of Psychology*, *60*, 135–141. doi:10.1080/00049530701449521

Kelemen, O., Erdélyi, R., Pataki, I., Benedek, G., Janka, Z., & Kéri, S. (2005). Theory of mind and motion perception in schizophrenia. *Neuropsychology*, *19*, 494–500. doi:10.1037/0894-4105.19.4.494

Kelemen, O., Keri, S., Must, A., Benedek, G., & Janka, Z. (2004). No evidence for impaired "theory of mind" in unaffected first-degree relatives of schizophrenia patients. *Acta Psychiatrica Scandinavica*, *110*, 146–149. doi:10.1111/j.1600-0047.2004.00357.x

Kenyon, M., Samarawickrema, N., Dejong, H., Van den Eynde, F., Startup, H., Lavender, A., … Schmidt, U. (2012). Theory of mind in bulimia nervosa. *International Journal of Eating Disorders*, *45*, 377–384. doi:10.1002/eat.20967

Kettle, J. W. L., O'Brien-Simpson, L., & Allen, N. B. (2008). Impaired theory of mind in first-episode schizophrenia: comparison with community, university and depressed controls. *Schizophrenia Research*, *99*, 96–102. doi:10.1016/j.schres.2007.11.011

Kucera, H., & Francis, W. N. (1967). *Computational analysis of present day American English*. Providence, RI: Brown University Press.

Kunihira, Y., Senju, A., Dairoku, H., Wakabayashi, A., & Hasegawa, T. (2006). "Autistic" traits in non-autistic Japanese populations: Relationships with personality traits and cognitive ability. *Journal of Autism and Developmental Disorders*, *36*, 553–566. doi:10.1007/s10803-006-0094-1

Lillard, A. (1998). Ethnopsychologies: Cultural variations in theories of mind. *Psychological Bulletin*, *123*, 3–32. doi:10.1037/0033-2909.123.1.3

Lillard, A. (1999). Developing a cultural theory of mind. The CIAO approach. *Current Directions in Psychological Science*, *2*, 57–61. doi:10.1111/1467-8721.00014

Mason, M. F., & Morris, M. W. (2010). Culture, attribution and automaticity: A social cognitive neuroscience view. *Social Cognitive and Affective Neuroscience*, *5*, 292–306. doi:10.1093/scan/nsq034

Masuda, T., Ellsworth, P. C., Mesquita, B., Leu, J., Tanida, S., & Van de Veerdonk, E. (2008). Placing the face in context: Cultural differences in the perception of facial emotion. *Journal of Personality and Social Psychology*, *94*, 365–381. doi:10.1037/0022-3514.94.3.365

Masuda, T., & Nisbett, R. E. (2001). Attending holistically versus analytically: Comparing the context sensitivity of Japanese and Americans. *Journal of Personality and Social Psychology*, *81*, 922–934. doi:10.1037/0022-3514.81.5.922

McGlade, N., Behan, C., Hayden, J., O'Donoghue, T., Peel, R., Haq, F., … Donohoe, G. (2008). Mental state decoding v. mental state reasoning as a mediator between cognitive and social function in psychosis. *The British Journal of Psychiatry*, *193*, 77–78. doi:10.1192/bjp.bp.107.044198

Morris, M. W., & Peng, K. (1994). Culture and cause: American and Chinese attributions for social and physical events. *Journal of Personality and Social Psychology*, *67*, 949–971. doi:10.1037/0022-3514.67.6.949

Murphy, D. (2006). Theory of mind in Asperger's syndrome, schizophrenia and personality disordered forensic patients. *Cognitive Neuropsychiatry*, *11*, 99–111. doi:10.1080/13546800444000182

New, B., Pallier, C., Ferrand, L., & Matos, R. (2001) Une base de données lexicales du français contemporain sur internet: LEXIQUE™ [A lexical database of contemporary French on the internet: LEXIQUE™]. *L'Année Psychologique*, *101*, 447–462. doi:10.3406/psy.2001.1341

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*, 515–526. doi:10.1017/S0140525X00076512

Rule, N. O., Ambady, N., Adams, R. B., Jr., Ozono, H., Nakashima, S., Yoshikawa, S., & Watabe, M. (2010). Polling the face: Prediction and consensus across cultures. *Journal of Personality and Social Psychology*, *98*, 1–15. doi:10.1037/a0017673

Russell, T. A., Rubia, K., Bullmore, E. T., Soni, W., Suckling, J., Brammer, M. J., … Sharma, T. (2000). Exploring the social brain in schizophrenia: Left prefrontal underactivation during mental state attribution. *American Journal of Psychiatry*, *157*, 2040–2042. doi:10.1176/appi.ajp.157.12.2040

Serafin, M., & Surian, F. (2004). Il test degli occhi: Uno strumento per valutare la "teoria della mente" [The eye test: A tool for evaluating the theory of mind]. *Giornale Italiano di Psicologia*, *4*, 839–860. doi:10.1421/18849

Shahaeian, A., Peterson, C. C., Slaughter, V., & Wellman, H. M. (2011). Culture and the sequence of steps in theory of mind development. *Developmental Psychology*, *47*, 1239–1247. doi:10.1037/a0023899

Smeets, T., Dziobek, I., & Wolf, O. T. (2009). Social cognition under stress: Differential effects of stress-induced cortisol elevations in healthy young men and women. *Hormones and Behavior*, *55*, 507–513. doi:10.1016/j.yhbeh.2009.01.011

Stanford, A. D., Messinger, J., Malaspina, D., & Corcoran, C. M. (2011). Theory of mind in patients at clinical high risk for psychosis. *Schizophrenia Research*, *131*, 11–17. doi:10.1016/j.schres.2011.06.005

Valla, J. M., Ganzel, B. L., Yoder, K. J., Chen, G. M., Lyman, L. T., Sidari, A. P., … Belmonte, M. K. (2010). More than maths and mindreading: Sex differences in empathizing/systemizing covariance. *Autism Research*, *3*, 174–184. doi:10.1002/aur.143

Vellante, M., Baron-Cohen, S., Melis, M., Marrone, M., Petretto, D. R., Masala, C., & Preti, A. (2012). The "Reading the Mind in the Eyes" test: Systematic review of psychometric properties and a validation study in Italy. *Cognitive Neuropsychiatry*, advance online publication. doi:10.1080/13546805.2012.721728

Voracek, M., & Dressler, S. G. (2006). Lack of correlation between digit ratio (2D:4D) and Baron-Cohen's "Reading the Mind in the Eyes" Test, empathy, systemising, and autism-spectrum quotients in a general population sample. *Personality and Individual Differences*, *41*, 1481–1491. doi:10.1016/j.paid.2006.06.009

Yildirim, E. A., Kasar, M., Guduk, M., Ates, E., Kucukparlak, I., & Ozalmete, E. O. (2011). Investigation of the reliability of the "reading the mind in the eyes test" in a Turkish population. *Turk Psikiyatri Dergisi*, *22*, 177–186. Retrieved from http://www.turkpsikiyatri.com/default.aspx?modul=ingilizceOzet&gFPrkMakale=810

Appendix 1: French version of the Reading the Mind in the Eyes test

Exemple	jaloux	**paniqué**	arrogant	haineux
1	**joueur**	réconfortant	irrité	s'ennuyant
2	terrifié	**bouleversé**	arrogant	agacé
3	blagueur	angoissé	**désir**	convaincu
4	blagueur	**insistant**	amusé	détendu
5	irrité	sarcastique	**inquiet**	amical
6	effondré	**rêveur**	impatient	alarmé
7	s'excusant	amical	**mal à l'aise**	démoralisé
8	**découragé**	soulagé	timide	excité
9	agacé	hostile	horrifié	**préoccupé**
10	**prudent**	insistant	s'ennuyant	effondré
11	terrifié	amusé	**plein de regrets**	charmeur
12	indifférent	embarrassé	**sceptique**	démoralisé
13	déterminé	**prévoyant**	menaçant	timide
14	irrité	déçu	déprimé	**accusateur**
15	**contemplatif**	angoissé	encourageant	amusé
16	irrité	**songeur**	encourageant	compatissant
17	**dubitatif**	affectueux	joueur	effondré
18	**déterminé**	amusé	effondré	s'ennuyant
19	arrogant	reconnaissant	sarcastique	**hésitant**
20	dominant	**amical**	coupable	horrifié
21	embarrassé	**rêveur**	confus	paniqué
22	**préoccupé**	reconnaissant	insistant	suppliant
23	content	s'excusant	**provoquant**	curieux
24	**pensif**	irrité	excité	hostile
25	paniqué	incrédule	découragé	**intéressé**
26	alarmé	timide	**hostile**	anxieux
27	blagueur	**prudent**	arrogant	rassurant
28	**intéressé**	blagueur	affectueux	content
29	impatient	effondré	irrité	**réfléchi**
30	reconnaissant	**charmeur**	hostile	déçu
31	honteux	**confiant**	blagueur	démoralisé
32	**sérieux**	honteux	bouche-bée	alarmé
33	embarrassé	coupable	rêveur	**soucieux**
34	effondré	dérouté	**méfiant**	terrifié
35	perplexe	**nerveux**	insistant	contemplatif
36	Honteux	nerveux	**suspicieux**	indécis