



# “Reading the Mind in the Eyes Test”: Translated and Korean Versions

Se Jun Koo<sup>1,2</sup>, Ye Jin Kim<sup>1</sup>, Jung Hwa Han<sup>1,3</sup>, Eunchong Seo<sup>1,4</sup>, Hye Yoon Park<sup>1,4</sup>,  
Minji Bang<sup>5</sup>, Jin Young Park<sup>1,6</sup>, Eun Lee<sup>1,4</sup>, and Suk Kyoan An<sup>1,2,4</sup> ✉

<sup>1</sup>Section of Self, Affect and Neuroscience, Institute of Behavioral Science in Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea

<sup>2</sup>Graduate Program in Cognitive Science, Yonsei University, Seoul, Republic of Korea

<sup>3</sup>Department of Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea

<sup>4</sup>Department of Psychiatry, Yonsei University College of Medicine, Severance Hospital, Seoul, Republic of Korea

<sup>5</sup>Department of Psychiatry, CHA Bundang Medical Center, CHA University, Seongnam, Republic of Korea

<sup>6</sup>Department of Psychiatry, Yonsei University College of Medicine, Yongin Severance Hospital, Yongin, Republic of Korea

**Objective** The Reading the Mind in the Eyes Test (RMET) was developed by using Caucasian eyes, which may not be appropriate to be used in Korean. The aims of the present study were 1) to develop a Korean version of the RMET (K-RMET) by using Korean eye stimuli and 2) to examine the psychometric properties of the Korean-translated version of the RMET and the K-RMET.

**Methods** Thirty-six photographs of Korean eyes were selected. A total of 196 (101 females) healthy subjects were asked to take the Korean-translated version of the RMET and K-RMET. To assess internal consistency reliability, Cronbach's alpha coefficients were computed, and test-retest reliability was assessed by the intraclass correlation coefficient (ICC) and Bland-Altman plots. Confirmatory factor analysis (CFA) and item analysis were also conducted.

**Results** Internal consistency, measured by Cronbach's alpha, was 0.542 for the Korean-translated version of the RMET, and 0.540 for the K-RMET. Test-retest reliability (n=25), measured by the ICC, was 0.787 for the Korean-translated version of the RMET, and 0.758 for the K-RMET. In CFA, the assumed single and 3-factor model fit indices were not good in the both types of RMETs. There was difficulty in discrimination in nine items of the Korean-translated version of the RMET and 10 items of the K-RMET.

**Conclusion** The psychometric properties of both the Korean-translated version of the RMET and the K-RMET are acceptable. Both tests are applicable to the clinical population, as well as the general population in Korea. **Psychiatry Investig 2021;18(4):295-303**

**Key Words** Theory of mind, Social cognition, Reliability, Item difficulty, Psychometric properties.

## INTRODUCTION

Humans can guess other people's intentions, beliefs, and emotions through verbal and non-verbal information processing even if they do not have direct experience. These processes are referred to as theory of mind (ToM), defined as the ability to understand and attribute to the mental states of others,<sup>1,2</sup> which is considered a crucial component of social interaction.<sup>3</sup> Possessing high levels of ToM is advantageous for

achieving one's goal by grasping the other person's intentions and adequately responding in social interactions.<sup>4</sup> Conversely, ToM deficit was strongly associated with social dysfunction in the clinical population.<sup>5-8</sup>

The “Reading the Mind in the Eyes” test (RMET) was developed to provide enriched information about ToM deficit. The initial version of the RMET was published in 1997,<sup>9</sup> while Baron-Cohen and colleagues revised it in 2001.<sup>10</sup> The revised version consisted of 36 photographs and presented stimuli around the eyes. Participants were asked to choose which of the four options (i.e., one target word and three foils) described the intentions and emotions of the person in the photographs. Since some of the options had similar emotional valence (i.e., positive, negative, or neutral), participants were required to carefully examine each word and photograph and distinguish subtle differences among words for selecting a target word.

Received: July 28, 2020 Revised: December 4, 2020

Accepted: December 28, 2020

✉ Correspondence: Suk Kyoan An, MD, PhD

Department of Psychiatry, Yonsei University College of Medicine, 50-1 Yonsei-ro, Seodaemun-gu, Seoul 03722, Republic of Korea

Tel: +82-2-2228-1585, Fax: +82-2-313-0891, E-mail: [ansk@yuhs.ac](mailto:ansk@yuhs.ac)

© This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

The eye stimuli is similar to the Facial Action Coding System (FACS)<sup>11</sup> in that it uses facial expression stimuli. However, the RMET differs in that participants only look at the eyes of the person in the picture and infer that person’s relatively complex mental states.<sup>10</sup> More specifically, participants should know the semantics of the option words and map these terms around that person’s eyes. At the unconscious and automatic level, participants should match these eye stimuli with the representation stored in their memory and determine which word is the closest to their eye expressions. For this reason, the RMET is regarded as a test that measures advanced ToM rather than facial emotion recognition. Thus, the RMET has been translated into different languages and extensively used to measure ToM in various clinical populations,<sup>12–19</sup> typical children, and adult populations.<sup>20–23</sup>

The original version of the RMET was initially designed to measure ToM in Caucasian populations; thus, the pictures were initially extracted from Caucasian photographs. If the same eye stimuli are conducted for non-Caucasians, differences in race and ethnicity can affect the test performance in various ethnic groups. For example, in a study comparing RMET scores by country,<sup>24</sup> the average score of Ethiopian medical students was below 22, which was lower than those of Western students who averaged 25–28 points. In comparison, there are studies<sup>25,26</sup> comparing different versions of RMET using pictures of their respective ethnic groups that show a different result. For example, Caucasian American and Japanese students responded more accurately to a set of eye stimuli that matched their own ethnic group,<sup>25</sup> and a partial in-group advantage was also found in Antillean Dutch, Moroccan Dutch, and Dutch samples.<sup>26</sup> Since RMET infers another person’s mental state by using limited information about the eyes and surroundings without other contextual information, presenting Caucasian stimuli to East Asians such as Koreans may lead to more difficulty on the part of the participant.

Given these findings, when conducting RMET research on the Korean population, it is expected that ecological validity can be improved if Korean pictures were used as well. Thus, the aims of the present study were 1) to develop the Korean RMET (K-RMET) by using Korean eye stimuli based on the development process of the original study,<sup>10</sup> and 2) to examine the psychometric properties, such as internal consistency, test–retest reliability, confirmatory factor analysis (CFA), distribution of responses and item analysis of the K-RMET and the original RMET, as translated into Korean, with the Korean population.

## METHODS

### Participants

A total of 196 Korean healthy late adolescents and early adults (101 female, 95 male) were recruited through online job advertisements. Their mean age was 23.02 (SD=2.61), while years of education was 14.41 (SD=1.40). Based on the Mini International Neuropsychiatric Interview (MINI) participants with past or current psychiatric illnesses were excluded. In order to perform the test–retest reliability, some of the participants (n=25) were asked to retake the test. The first test took place during May to July 2018; the retest was conducted from April to October 2019 (test–retest interval: mean=13.96, SD=1.70; range=11–17 months). All participants provided their written informed consent, and the study was approved by the Institutional Review Board of the Severance Hospital (IRB No: 4-2014-0744).

### Measures

For the translation of optional words and a glossary of mental state terms in RMET, a researcher (HJH) with a master’s degree in social psychology, who is fluent in English and Korean, initially translated the RMET to Korean. Thereafter, three experts (psychiatrists ASK and BM and clinical psychologist KSJ) reviewed each word repeatedly until they reached a unanimous agreement regarding the translation.

In the original study, the authors collected photos of actors and actresses from the magazine, and in each photograph, only around the eyes (from just above the eyebrows to the bridge of the nose) were cut to the same size and used as the stimuli of the RMET. To develop the K-RMET, 146 Korean pictures that were judged to be consistent with the eye stimuli of the original study, age (young, middle, elder), gender, pupil orientation, and facial expressions near the eyes were initially selected from the web search engines. Thereafter, two experts (ASK and PJY), who developed the Korean facial expressions of emotion (KOFEE),<sup>27</sup> reviewed the candidate photographs and 41 photos were selected for the pilot test. The pilot test was conducted on another 25 participants (15 female, 10 male). The two measures of accuracy were 75% for the Korean-translated version of the RMET and 71% for the K-RMET. The accuracy of both tests was comparable to those of previous studies (accuracy range=68–78%).<sup>10,28–30</sup>

After the pilot test, 36 items were finally selected for measurement (for details, Supplementary Table 1 in the online-only Data Supplement, for the use of K-RMET, request to the authors). Each subject was asked to respond to the RMETs in an isolated space, and the photos were presented at a resolution of 425×170 pixels on a 17-inch screen using Inquisit 3.0 (Millisecond Software LLC, Seattle, WA, USA). Participants

were asked to look at each set of eyes and choose from four options (1 target, 3 foils) that would best describes what the person in the picture is thinking or feeling. One practice item was provided to help participants become familiar with the assignment, and there was no time limit. The target word was scored as 1 point, and the foil was scored as 0; the total score ranged from 0 to 36 points. As with the original paper, in order to minimize the impact of individual vocabulary on the test, a glossary of mental state terms was provided to each participant during all measurement periods, to be consulted at any time during the test. The retest was conducted in the same way as the first test procedure.

### Statistical analysis

Data were analyzed using the Statistical Package for the Social Sciences (SPSS), version 25 for Windows (IBM Corporation, Armonk, NY, USA). To perform test-retest reliability, Intraclass correlation coefficients (ICC) with 95% confidence intervals were calculated based on the criteria proposed by Koo and Li,<sup>31</sup> which use a mean-rating ( $\kappa=2$ ), absolute-agreement, 2-way random-effects model (ICC estimates: Excellent: 0.9–1; Good: 0.75–0.9; Moderate: 0.5–0.75; Poor: 0–0.5).<sup>31</sup> To obtain additional agreement, the Bland-Altman approach was employed.<sup>32,33</sup> The Bland-Altman plot was calculated and visualized using blandr<sup>34</sup> in RStudio for Windows version 1.2.1335 (RStudio: Integrated Development for R., RStudio, Inc., Boston, MA, USA). In order to examine the factor structure of the measured data, CFA was performed on the single factor model and the three factor model proposed by Harkness et al.,<sup>35</sup> based on established criteria [chi-square (CMIN)/df ratio (CMIN/df)<2,<sup>36</sup> root mean square error of approximation (RMSEA)≤0.06, standardized root mean squared residual (SRMR)≤0.08, comparative fit index (CFI)≥0.95, Tucker-Lewis index (TLI)≥0.95<sup>37</sup>] using AMOS 25 (IBM Corp., Armonk, NY, USA). All tests were two-tailed and conducted at 5% level of statistical significance.

## RESULTS

### Task performances

Figure 1 showed the distribution of the total scores of the Korean-translated version of the RMET and K-RMET. The mean scores were 25.65 (SD=3.41) and 26.72 (SD=3.38) for the Korean-translated version of the RMET and K-RMET, respectively. There was no gender difference (mean male=25.46; SD=3.61; mean female=25.82; SD=3.21) in the Korean-translated version of the RMET scores [ $t(194)=0.74$ ;  $p=0.463$ ], whereas female subjects (mean=27.24; SD=3.09) had significantly higher scores than male subjects (mean=26.17; SD=3.60) in the K-RMET scores [ $t(194)=2.24$ ;  $p=0.027$ ].

## Reliability analysis

### Internal consistency

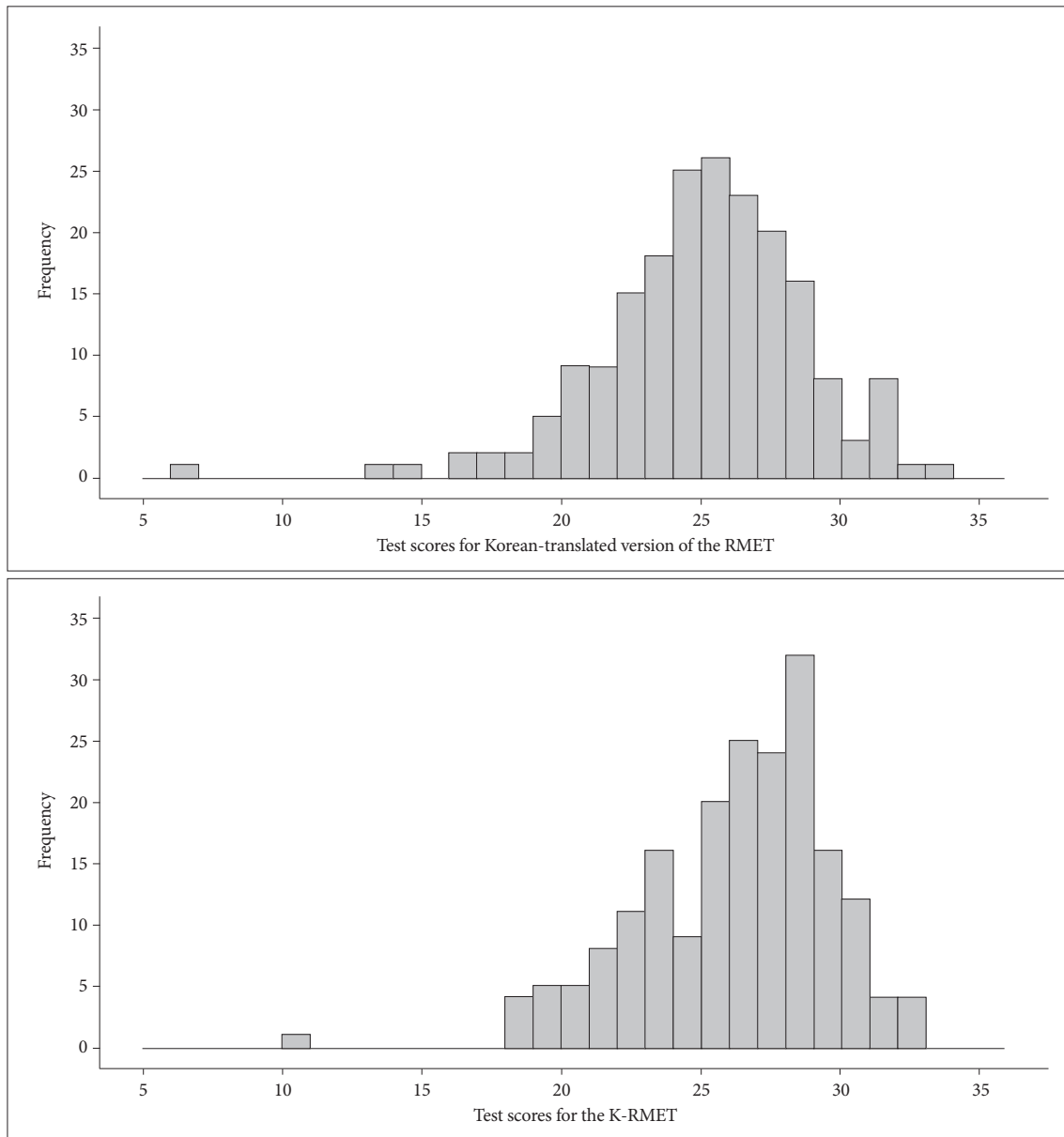
Cronbach's alpha was 0.542 for the Korean-translated version of the RMET and 0.540 for the K-RMET. Additionally, Cronbach's alpha was not found to be improved by excluding any item to improve alpha value in both the Korean-translated version of the RMET and the K-RMET.

### Test-retest reliability

Some of the participants ( $n=25$ ) were asked to retake the test. In the Korean-translated version of the RMET, the mean score for the test was 26.24 (SD=3.49) and 26.68 (SD=3.57) for the retest. In the K-RMET, the mean score for the test was 26.12 (SD=3.57) and 27.16 (SD=3.69) for the retest. Using two related sample Wilcoxon Signed Ranks tests, there were no significant differences in both the Korean-translated version of the RMET ( $Z=-0.60$ ,  $p=0.548$ ) and the K-RMET ( $Z=-1.87$ ,  $p=0.062$ ) scores. The ICCs were 0.787 with a 95% CI (0.519, 0.906) for the Korean-translated version of the RMET, and 0.758 with a 95% CI (0.462, 0.892) for the K-RMET. To visualize additional agreement between the test and retest, the Bland-Altman plots were reported in Figure 2. In the Korean-translated version of the RMET, the mean difference (test-retest) was -0.44 with 95% CI ranging from -1.67 to 0.79. The upper limit of agreement was 5.39 with 95% CI ranging from 3.26 to 7.51. The lower limit of agreement was -6.27 with 95% CI ranging from -8.39 to -4.14. In the K-RMET, the mean difference was -1.04 with 95% CI ranging from -2.34 to 0.26. The upper limit of agreement was 5.12 with 95% CI ranging from 2.87 to 7.37. The lower limit of agreement was -7.20 with 95% CI ranging from -9.45 to -4.95.

### Confirmatory factor analysis

In the full-scale single-factor, CMIN/df, RMSEA, and SRMR indicated satisfactory fit for both the Korean-translated version of the RMET [ $\chi^2(594)=738.24$ ;  $p<0.001$ ; CMIN/df=1.24; RMSEA=0.035; 90% CI: 0.026–0.043; SRMR=0.071] and the K-RMET [ $\chi^2(594)=694.63$ ;  $p=0.003$ ; CMIN/df=1.17; RMSEA=0.029; 90% CI=0.018–0.038; SRMR=0.067], but CFI and TLI were not good for both the Korean-translated version of the RMET (CFI=0.435; TLI=0.400) and the K-RMET (CFI=0.532; TLI=0.504). In the emotional valence 3-factor model proposed by Harkness et al.,<sup>35</sup> CMIN/df, RMSEA, and SRMR indicated excellent fit in both the Korean-translated version of the RMET [ $\chi^2(591)=710.72$ ;  $p<0.001$ ; CMIN/df=1.20; RMSEA=0.032; 90% CI: 0.022–0.041; SRMR=0.070] and the K-RMET [ $\chi^2(591)=686.47$ ;  $p=0.004$ ; CMIN/df=1.16; RMSEA=0.029; 90% CI: 0.017–0.038; SRMR=0.067], but CFI and TLI were not good in both the Korean-translated ver-



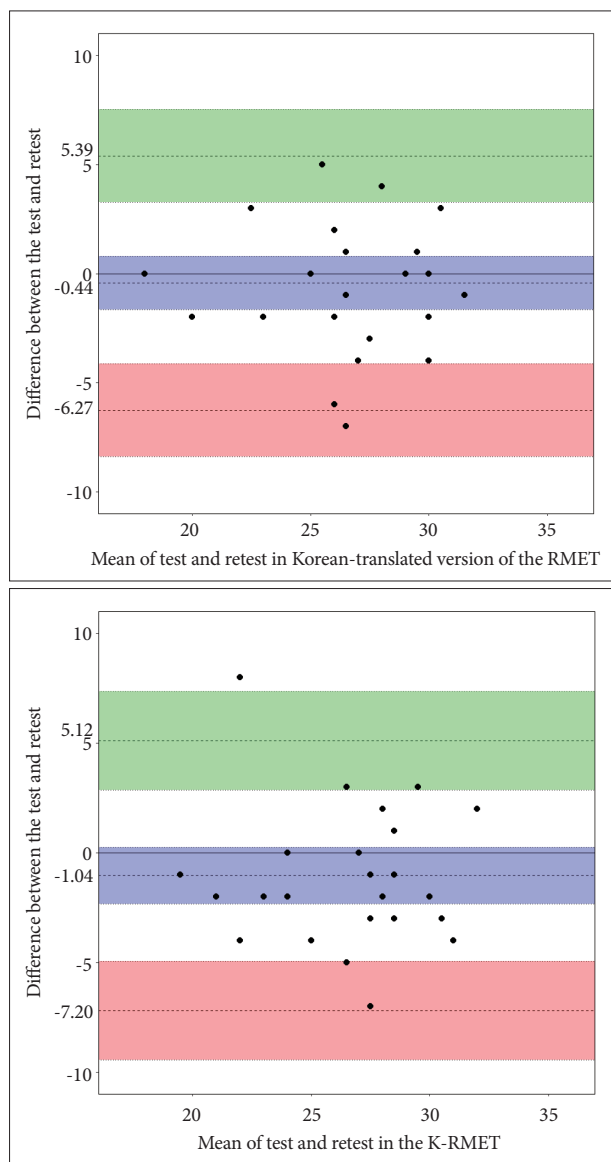
**Figure 1.** The distribution of the total scores for the Korean-translated version of the RMET (top) and the K-RMET (bottom). RMET: The Reading the Mind in the Eyes Test.

sion of the RMET (CFI=0.531; TLI=0.500) and the K-RMET (CFI=0.556; TLI=0.527).

**Item analysis**

Table 1 showed the participants’ choice of answers in each item on the Korean-translated version of the RMET and K-RMET. According to the original developmental process of the RMET, the authors presented two criteria in item selection.<sup>10</sup> The first criteria is that participants should select at least 50% of the target word, and the second criteria is that the rate of selecting the foil word in each item should be less than 25%. In the Korean-translated version of the RMET, on three items (2, 5, and 23), the target word was selected by

fewer than 50% of the participants, while on nine items (2, 3, 5, 12, 14, 16, 23, 25, and 27), one of the foils was chosen by more than 25%. In the K-RMET, on five items (2, 3, 10, 14, and 28), the target word was selected by fewer than 50% of the participants, while on 10 items (2, 3, 10, 11, 14, 20, 21, 22, 27, and 28), one of the foils was chosen by more than 25%. Overall, nine items (2, 3, 5, 12, 14, 16, 23, 25, and 27) of the Korean-translated version of the RMET, and 10 items (2, 3, 10, 11, 14, 20, 21, 22, 27, and 28) of the K-RMET had difficulty in discrimination.



**Figure 2.** The Bland-Altman plots of the Korean-translated version of the RMET (top) and the K-RMET (bottom) assessments. Mean differences with 95% confidence interval (cobalt blue), upper limit of agreement with 95% confidence interval (vine green), and lower limit of agreement with 95% confidence interval (coral) are displayed. RMET: The Reading the Mind in the Eyes Test.

## DISCUSSION

The RMET is one of the well-established tests for measuring ToM, and has been widely used in various countries. To the best of our knowledge, this is the first study to develop eye stimuli of the same gender and mental state as the original paper using Korean photographs, and verify the psychometric properties of the Korean-translated version of the RMET and the K-RMET at the same time. The main findings showed that mean accuracy rates of the Korean-translated version of the RMET and the K-RMET were comparable to those of the pre-

vious studies, and test-retest reliability and item analysis of the both types of the RMETs were acceptable.

According to the gender of subjects, the overall mean scores and mean scores of the Korean-translated version of the RMET (overall mean=25.65; SD=3.41; mean male=25.46; SD=3.61; mean female=25.82; SD=3.21) and the K-RMET (overall mean=26.72; SD=3.38; mean male=26.17; SD=3.60; mean female=27.24; SD=3.09) were similar to those reported in the original paper, and were within the range of the mean scores of other RMET validation studies (for details, Supplementary Table 2 in the online-only Data Supplement). With regard to accuracy and gender differences, a statistically significant female advantage was observed in most studies of the RMET, including original papers from Baron-Cohen and colleagues.<sup>10,38-40</sup> In the K-RMET, it was also found that females had significantly higher average scores than males. Some studies demonstrated that females were known to recognize faces faster and more accurately than males,<sup>41</sup> and better differentiate subtle emotions.<sup>42</sup> Other studies have attempted to explain that females are better than males in distinguishing positive and negative emotions, due to their “attachment promotion” and “fitness threat” derived from an evolutionary perspective.<sup>43</sup> On the other hand, there was no significant difference in performance between males and females in the Korean-translated version of the RMET, supporting the findings of other studies that report the absence of a female advantage, albeit small in number.<sup>21,44</sup> Taken together, the K-RMET may have the advantage in that it properly reflects that are actually occurring in the research fields, such as gender differences in performance. It is considered that further studies that include various ages and cultures will be needed to reach a firm conclusion on the female advantage.

The internal consistency, measured using Cronbach’s alpha, was found to be relatively not good for the Korean-translated version of the RMET (0.542) and K-RMET (0.540). In addition, internal consistency did not improve by excluding any item to improve alpha value in both the Korean-translated version of the RMET and K-RMET. Similarly, the Cronbach’s alpha values of most previous RMET studies were not good,<sup>24,30,45-47</sup> with the exception of a few studies that showed an internal consistency of greater than 0.7.<sup>21,48</sup> It may be explained that the unacceptable internal consistency reliabilities of various versions of the RMET, including the Korean-translated version of the RMET and K-RMET, may be derived from the characteristics of the samples in each study, or the characteristics of the test itself, such as quality of pictures.

Test-retest reliability, as assessed by the ICC and Bland-Altman plot, were shown to be good with the Korean-translated version of the RMET and K-RMET. The ICCs of both Korean-translated version of the RMET were greater than 0.75,

corresponding to "good" intraclass correlation coefficients according to the criteria proposed by Koo and Li.<sup>31</sup> Furthermore, the Bland-Altman approach revealed that all mean differences were within the upper/lower limit of agreement ex-

cept for one case among the 25 retest participants. These results suggested that learning effect is minimal in both types of the RMETs and the measurement results are stable over time.

A CFA was performed using commonly used model fit in-

**Table 1.** Distribution of responses in percentages (N=196)

Item	Korean-translated version of the RMET				K-RMET					
	1	2	3	4	1	2	3	4		
1	<b>68.9</b>	8.7	9.2	13.3	<b>91.8</b>	3.6	3.6	1.0		
2	*	16.8	<b>20.9</b>	5.6	56.6	*	1.5	<b>5.6</b>	62.8	30.1
3	*	1.0	1.0	<b>64.3</b>	33.7	*	7.1	0.5	<b>45.9</b>	46.4
4		2.6	<b>82.7</b>	10.2	4.6		0.5	<b>83.7</b>	7.7	8.2
5	*	14.3	49.5	<b>36.2</b>	0.0		12.2	10.7	<b>77.0</b>	0.0
6		0.5	<b>90.3</b>	3.1	6.1		0.5	<b>89.3</b>	5.1	5.1
7		4.6	16.3	<b>64.8</b>	14.3		7.1	1.0	<b>90.8</b>	1.0
8		<b>90.8</b>	5.1	0.5	3.6		<b>91.8</b>	3.6	0.5	4.1
9		5.1	10.7	1.0	<b>83.2</b>		1.0	6.6	2.0	<b>90.3</b>
10		<b>50.5</b>	24.0	20.9	4.6	*	<b>38.8</b>	2.0	20.4	38.8
11		8.2	17.3	<b>60.2</b>	14.3	*	26.0	1.5	<b>70.4</b>	2.0
12	*	29.6	4.6	<b>58.7</b>	7.1		14.8	5.1	<b>67.3</b>	12.8
13		3.1	<b>78.6</b>	1.5	16.8		12.8	<b>78.6</b>	1.5	7.1
14	*	37.2	10.7	2.0	<b>50.0</b>	*	39.8	10.7	0.5	<b>49.0</b>
15		<b>95.4</b>	0.0	2.0	2.6		<b>91.3</b>	6.6	0.5	1.5
16	*	2.0	<b>62.8</b>	4.1	31.1		16.3	<b>66.8</b>	3.6	13.3
17		<b>92.9</b>	4.1	1.5	1.5		<b>88.3</b>	6.6	2.6	2.6
18		<b>97.4</b>	2.0	0.0	0.5		73.5	21.4	0.5	4.6
19		17.3	6.1	16.3	<b>60.2</b>		12.2	2.0	16.8	<b>68.9</b>
20		24.5	<b>68.4</b>	6.6	0.5	*	27.0	<b>58.7</b>	3.1	11.2
21		8.7	<b>80.6</b>	9.2	1.5	*	1.0	<b>61.7</b>	31.1	6.1
22		<b>91.8</b>	1.0	2.6	4.6	*	<b>62.2</b>	1.5	25.0	11.2
23	*	10.2	2.0	<b>34.2</b>	53.6		1.5	0.0	<b>96.9</b>	1.5
24		<b>77.6</b>	2.6	2.0	17.9		<b>89.8</b>	3.6	1.5	5.1
25	*	1.0	36.2	1.0	<b>61.7</b>		0.5	6.6	0.5	<b>92.3</b>
26		6.6	0.5	<b>80.6</b>	12.2		4.1	1.0	<b>94.4</b>	0.5
27	*	0.5	<b>61.2</b>	26.0	12.2	*	1.0	<b>50.0</b>	48.0	1.0
28		<b>84.7</b>	5.6	1.5	8.2	*	<b>36.2</b>	10.2	19.4	34.2
29		2.6	1.5	12.8	<b>83.2</b>		1.5	0.5	4.6	<b>93.4</b>
30		7.1	<b>74.0</b>	16.3	2.6		0.5	71.4	23.0	5.1
31		7.7	<b>72.4</b>	9.2	10.7		0.5	<b>88.3</b>	8.2	3.1
32		<b>66.8</b>	3.1	10.7	19.4		<b>79.6</b>	4.1	11.7	4.6
33		7.1	9.2	11.7	<b>71.9</b>		2.0	11.2	7.1	<b>79.6</b>
34		5.1	13.3	<b>73.5</b>	8.2		1.5	2.6	<b>93.9</b>	2.0
35		9.7	<b>68.4</b>	12.8	9.2		19.4	75.5	3.1	2.0
36		0	3.1	<b>96.4</b>	0.5		4.1	1.0	<b>88.8</b>	6.1

The correct answers are shown in bold. \*Item that does not meet the criteria suggested by Baron-Cohen and colleagues: 1) participants should select at least 50% of the target word or 2) the rate of selecting the foil word in each item should be less than 25%. RMET: The Reading the Mind in the Eyes Test

dices. In the single-factor model, fit indices such as CMIN/df, RMSEA, and SRMR were good for both the Korean-translated version of the RMET and the K-RMET, but did not meet the appropriate criteria in CFI and TLI. In the emotional valence 3-factor, overall model fit was slightly improved, but some fit indices were still poor. By looking at the results of previous studies, similar to the results of the present study in the single-factor and 3-factor model, the overall fit indices did not meet the above mentioned established criteria,<sup>22,44</sup> and even a study in the Korean subjects that reported relatively fair model fit, was not seemed to be good in CFI.<sup>49</sup> For ease of understanding, emotions are generally categorized such as positive, negative, and neutral, but in reality, emotions are made up of more complex and subtle combinations of reactions. In line with Vellante et al.,<sup>24</sup> proposed that the RMET may also have factor structures of more diverse dimensions rather than a few categories.

For item analysis, among 36 items, nine items of the Korean-translated version of the RMET and 10 items of the K-RMET had difficulty in discrimination. In previous studies,<sup>21,24,29,30,44,47,50</sup> questionable items in discrimination (e.g., target response rate) were less than 50% (criteria A), or the foils were selected by more than 25% (criteria B),<sup>10</sup> ranging from 3 to 15 out of 36 items (for details, Supplementary Table 3 in the online-only Data Supplement). Meanwhile, some studies reported a version in which several items were deleted because they did not meet their standards in the development process of the RMET.<sup>22,28</sup> In addition, some items that recorded the target word as foils are diversely distributed. Moreover, in most studies, items 7, 17, and 23 did not meet criteria A or B; whereas the Korean-translated version of the RMET met the criteria for all two items except item 23, and the K-RMET met the criteria for all three items. By analyzing the foil patterns based on emotional valence, as proposed by Harkness et al.,<sup>35</sup> the negative valences “doubtful” (item 17) and “defiant” (item 23) were frequently responded with the foils “affectionate” and “curious,” respectively. Furthermore there were differences between the target emotional valence (negative) and the foil (positive or neutral). Similarly, the neutral valence “uneasy” (item 7) was responded with the foil “friendly” in most studies conducted in other countries, and there also was a difference between the target emotional valence (neutral) and the foil (positive). Conversely, for items 2 and 14, unlike previous studies, our study did not meet criteria A or B. However, participants responded to “upset” (item 3) and “accusing” (item 14) with “annoyed” and “arrogant” in the Korean-translated version of the RMET and the K-RMET, respectively, and “irritated” in both. Moreover, there was no difference in the type of the emotional valence between the target and the foil (both are negative).

There have been few studies in which East Asian eye stimuli have been used.<sup>25,51</sup> Moreover, no item analysis on the subject matter has ever been conducted. Thus, there is a limit to direct comparison with our findings. Although it is difficult to conclude that the aforementioned foil patterns are characteristic response patterns of East Asians, some items of the Korean-translated version of the RMET and the K-RMET showed relatively consistent response patterns, which were distinguished from studies conducted in other cultures, follow-up studies to compare emotional response patterns according to culture and race are considered necessary. In addition, since the response patterns are different for each study, it can be considered preferable to use a full version of the test rather than removing specific items, especially for cross-cultural comparison.

The present study had certain limitations. First, for the eye stimuli of the K-RMET, referring to the process of developing emotional stimuli in the original author and other studies, web searching was performed to obtain vivid and natural Korean facial emotional stimuli, which are not artificially made in a laboratory environment. Although only photographs around the eyes of the entire face were used for the development of the test, some of the stimuli were extracted from photos of Korean celebrities. Participants who have identified of the person in the eye stimulus may interpret the intention of the facial expression more positively or negatively based on their familiarity with the person. Second, our study generalized the present findings to all adult ages, since late adolescents to early adults were recruited as participants in the present study, and middle-aged or older adults did not participate. In a group that has a hierarchical culture according to age, such as in Korea, the difference between the age of the person in the eye photo and the age of the participant can affect the participant's perception of the expression of the eyes in the task. In future studies, empirical works reflecting the above-mentioned relative age differences can be investigated for more accurate measurement. Lastly, in the current study, only CFA for the single-factor assumed in the original paper and the 3-factor widely used in the previous studies were performed. By looking at the results of previous studies, there are studies reported on abbreviated items to maximize model fit.<sup>22,44</sup> However, only different combinations of items are presented for each study, and no consensus result has been reported yet. In order to obtain more robust evidence on the validity of the test, follow-up studies related to the number of factors and the total number of questions will be needed.

In summary, since the RMET is easy to use and can be evaluated in a short time, it has been developed and used in various countries. We developed the RMET, as translated into

Korean and the K-RMET by using Korean eye stimuli, resulting to acceptable psychometric properties, such as reliability and item analysis. Future studies should provide additional evidence for convergent and divergent validity through other ToM tests, neurocognitive ability, and personality traits known to be related to social cognition. In addition, it is necessary to expand this study to other clinical populations, such as those with autism spectrum disorder, schizophrenia, and ultra-high risk for psychosis, as well as the general population of various ages and education levels in Korea.

### Supplementary Materials

The online-only Data Supplement is available with this article at <https://doi.org/10.30773/pi.2020.0289>.

### Acknowledgments

This work was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning, Republic of Korea (grant number 2017R1A2B3008214).

### Conflicts of Interest

The authors have no potential conflicts of interest to disclose.

### Author Contributions

Conceptualization: Suk Kyoan An. Data curation: Ye Jin Kim, Eunhong Seo, Hye Yoon Park. Formal analysis: Se Jun Koo, Eunhong Seo, Hye Yoon Park. Investigation: all authors. Methodology: Se Jun Koo. Project administration: Suk Kyoan An, Eun Lee. Resources: Ye Jin Kim, Jung Hwa Han. Software: Se Jun Koo. Supervision: Suk Kyoan An. Validation: Jung Hwa Han, Ye Jin Kim, Minji Bang, Jin Young Park. Visualization: Se Jun Koo. Writing—original draft: Se Jun Koo. Writing—review & editing: all authors.

### ORCID iDs

Se Jun Koo	<a href="https://orcid.org/0000-0002-9051-8684">https://orcid.org/0000-0002-9051-8684</a>
Ye Jin Kim	<a href="https://orcid.org/0000-0003-0597-8190">https://orcid.org/0000-0003-0597-8190</a>
Jung Hwa Han	<a href="https://orcid.org/0000-0002-6776-2427">https://orcid.org/0000-0002-6776-2427</a>
Eunhong Seo	<a href="https://orcid.org/0000-0003-2644-266X">https://orcid.org/0000-0003-2644-266X</a>
Hye Yoon Park	<a href="https://orcid.org/0000-0001-9579-8112">https://orcid.org/0000-0001-9579-8112</a>
Minji Bang	<a href="https://orcid.org/0000-0002-1669-4014">https://orcid.org/0000-0002-1669-4014</a>
Jin Young Park	<a href="https://orcid.org/0000-0002-5351-9549">https://orcid.org/0000-0002-5351-9549</a>
Eun Lee	<a href="https://orcid.org/0000-0002-7462-0144">https://orcid.org/0000-0002-7462-0144</a>
Suk Kyoan An	<a href="https://orcid.org/0000-0003-4576-6184">https://orcid.org/0000-0003-4576-6184</a>

### REFERENCES

- Premack D, Woodruff G. Does the chimpanzee have a theory of mind? *Behav Brain Sci* 1978;1:515-526.
- Baron-Cohen S, Leslie AM, Frith U. Does the autistic child have a “theory of mind”. *Cognition* 1985;21:37-46.
- Green MF, Penn DL, Bentall R, Carpenter WT, Gaebel W, Gur RC, et al. Social cognition in schizophrenia: an NIMH workshop on definitions, assessment, and research opportunities. *Schizophr Bull* 2008;34:1211-1220.
- Berezkei T. Machiavellian intelligence hypothesis revisited: what evolved cognitive and social skills may underlie human manipulation. *Evolut Behav Sci* 2018;12:32.
- Fett AK, Viechtbauer W, Penn DL, van Os J, Krabbendam L. The relationship between neurocognition and social cognition with functional outcomes in schizophrenia: a meta-analysis. *Neurosci Biobehav Rev* 2011;35:573-588.
- Bora E, Bartholomeusz C, Pantelis C. Meta-analysis of Theory of Mind (ToM) impairment in bipolar disorder. *Psychol Med* 2016;46:253-264.
- Roncone R, Falloon IR, Mazza M, De Risio A, Pollice R, Necozone S, et al. Is theory of mind in schizophrenia more strongly associated with clinical and social functioning than with neurocognitive deficits? *Psychopathology* 2002;35:280-288.
- Kim N, Choi US, Ha S, Lee SB, Song SH, Song DH, et al. Aberrant neural activation underlying idiom comprehension in Korean children with high functioning autism spectrum disorder. *Yonsei Med J* 2018; 59:897-903.
- Baron-Cohen S, Jolliffe T, Mortimore C, Robertson M. Another advanced test of theory of mind: Evidence from very high functioning adults with autism or Asperger syndrome. *J Child Psychol Psychiatry* 1997;38:813-822.
- Baron-Cohen S, Wheelwright S, Hill J, Raste Y, Plumb I. The “Reading the Mind in the Eyes” test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *J Child Psychol Psychiatry* 2001;42:241-251.
- Friesen E, Ekman P. Facial action coding system: a technique for the measurement of facial movement. Palo Alto: Consulting Psychologists Press; 1978;3:5.
- Bora E, Yucel M, Pantelis C. Theory of mind impairment in schizophrenia: meta-analysis. *Schizophr Res* 2009;109:1-9.
- Kettle JW, O'Brien-Simpson L, Allen NB. Impaired theory of mind in first-episode schizophrenia: comparison with community, university and depressed controls. *Schizophr Res* 2008;99:96-102.
- Onuoha RC, Quintana DS, Lyvers M, Guastella AJ. A meta-analysis of theory of mind in alcohol use disorders. *Alcohol Alcoholism* 2016;51: 410-415.
- Peñuelas-Calvo I, Sareen A, Sevilla-Llewellyn-Jones J, Fernández-Berocal P. The “Reading the mind in the eyes” test in autism-spectrum disorders comparison with healthy controls: a systematic review and meta-analysis. *J Autism Devel Disord* 2019;49:1048-1061.
- Richman MJ, Unoka Z. Mental state decoding impairment in major depression and borderline personality disorder: meta-analysis. *Br J Psychiatry* 2015;207:483-489.
- Stewart E, Catroppa C, Lah S. Theory of mind in patients with epilepsy: a systematic review and meta-analysis. *Neuropsychol Rev* 2016;26: 3-24.
- Charernboon T. Negative and neutral valences of affective theory of mind are more impaired than positive valence in clinically stable schizophrenia patients. *Psychiatry Investig* 2020;17:460-464.
- Hur DH, Park JH, Kwon SM, Kim YT, Kwon DH, Cho SN, et al. An investigation into theory of mind of schizophrenia using hinting task and eyes task. *J Korean Soc Biol Ther Psychiatry* 2006;12:215-223.
- Chapman E, Baron-Cohen S, Auyeung B, Knickmeyer R, Taylor K, Hackett G. Fetal testosterone and empathy: evidence from the empathy quotient (EQ) and the “reading the mind in the eyes” test. *Soc Neurosci* 2006;1:135-148.
- Girli A. Psychometric properties of the Turkish child and adult form of “Reading the Mind in the Eyes Test”. *Psychology* 2014;5:1321-1337.
- Olderbak S, Wilhelm O, Olaru G, Geiger M, Brenneman MW, Roberts RD. A psychometric analysis of the reading the mind in the eyes test: toward a brief form for research and applied settings. *Front Psychol* 2015; 6:1503.
- Vogindroukas I, Chelas EN, Petridis NE. Reading the Mind in the Eyes Test (children’s version): a comparison study between children with typical development, children with high-functioning autism and typically developed adults. *Folia Phoniatria et Logopaedica* 2014;66:18-24.
- Vellante M, Baron-Cohen S, Melis M, Marrone M, Petretto DR, Masala C, et al. The “Reading the Mind in the Eyes” test: systematic review of psychometric properties and a validation study in Italy. *Cogn Neuro-psychiatry* 2013;18:326-354.
- Adams Jr RB, Rule NO, Franklin Jr RG, Wang E, Stevenson MT, Yoshikawa S, et al. Cross-cultural reading the mind in the eyes: an fMRI



- investigation. *J Cogn Neurosci* 2010;22:97-108.
26. van der Meulen A, de Ruyter D, Blokland A, Krabbendam L. Cross-cultural mental state reading ability in Antillean Dutch, Moroccan Dutch, and Dutch young adults. *J Cross Cult Psychol* 2019;50:419-440.
  27. Lee SB, Koo SJ, Song YY, Lee MK, Jeong YJ, Kwon C, et al. Theory of mind as a mediator of reasoning and facial emotion recognition: findings from 200 healthy people. *Psychiatry Investig* 2014;11:105-111.
  28. Yildirim EA, Kaşar M, Gündük M, Ateş E, Küçükparlak İ, Özalmete EO. Investigation of the reliability of the “Reading the Mind in the Eyes Test” in a Turkish population. *Turk J Psychiatry* 2011;22:177-186.
  29. Pestana J, Menéres MSSPC, Gouveia MJPM, Oliveira RF. The Reading the Mind in the Eyes Test: a Portuguese version of the adults’ test. *Análise Psicológica* 2018;36:369-381.
  30. Prevost M, Carrier ME, Chowne G, Zerkowicz P, Joseph L, Gold I. The Reading the Mind in the Eyes test: validation of a French version and exploration of cultural variations in a multi-ethnic city. *Cogn Neuro-psychiatry* 2014;19:189-204.
  31. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 2016;15:155-163.
  32. Bland JM, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;327:307-310.
  33. Hamilton C, Stamey J. Using bland– altman to assess agreement between two medical devices- Don’t forget the confidence intervals! *J Clin Monitor Comput* 2007;21:331-333.
  34. Datta D. blandr: a Bland-Altman method comparison package for R. Zenodo. 2017. Available at: <https://cran.r-project.org/web/packages/blandr/README.html>. Accessed June 20, 2020.
  35. Harkness K, Sabbagh M, Jacobson J, Chowdrey N, Chen T. Enhanced accuracy of mental state decoding in dysphoric college students. *Cogn Emot* 2005;19:999-1025.
  36. Schermelleh-Engel K, Moosbrugger H, Müller H. Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods Psychol Res Online* 2003;8:23-74.
  37. Hu Lt, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct Equ Modeling* 1999;6:1-55.
  38. Jankowiak-Siuda K, Baron-Cohen S, Bialaszek W, Dopierala A, Kozłowska A, Rymarczyk K. Psychometric evaluation of the ‘Reading the mind in the eyes’ test with samples of different ages from a Polish population. *Studia Psychologica* 2016;58:18.
  39. Hallerback MU, Lugnegård T, Hjärthag F, Gillberg C. The Reading the Mind in the Eyes Test: test–retest reliability of a Swedish version. *Cogn Neuropsychiatry* 2009;14:127-143.
  40. Kirkland RA, Peterson E, Baker CA, Miller S, Pulos S. Meta-analysis reveals adult female superiority in “Reading the Mind in the Eyes Test”. *North Am J Psychol* 2013;15:121-146.
  41. Hall JK, Hutton SB, Morgan MJ. Sex differences in scanning faces: Does attention to the eyes explain female superiority in facial expression recognition? *Cogn Emot* 2010;24:629-637.
  42. Hoffmann H, Kessler H, Eppel T, Rukavina S, Traue HC. Expression intensity, gender and facial emotion recognition: Women recognize only subtle facial emotions better than men. *Acta Psychol* 2010;135:278-283.
  43. Hampson E, van Anders SM, Mullin LI. A female advantage in the recognition of emotional facial expressions: Test of an evolutionary hypothesis. *Evolut Human Behav* 2006;27:401-416.
  44. Đorđević J, Živanović M, Pavlović A, Mihajlović G, Stašević-Karličić I, Pavlović D. Psychometric evaluation and validation of the Serbian version of ‘Reading the Mind in the Eyes’ test. *Psihologija* 2017;50:483-502.
  45. Harkness KL, Jacobson JA, Duong D, Sabbagh MA. Mental state decoding in past major depression: effect of sad versus happy mood induction. *Cogn Emot* 2010;24:497-513.
  46. Voracek M, Dressler SG. 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. *Pers Individ Diff* 2006;41:1481-1491.
  47. Khorashad BS, Baron-Cohen S, Roshan GM, Kazemian M, Khazai L, Aghili Z, et al. The “Reading the Mind in the Eyes” test: investigation of psychometric properties and test–retest reliability of the persian version. *J Autism Dev Disord* 2015;45:2651-2666.
  48. Dehning S, Girma E, Gasperi S, Meyer S, Tesfaye M, Siebeck M. 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 Med Educ* 2012;12:34.
  49. Lee HR, Nam G, Hur JW. Development and validation of the Korean version of the Reading the Mind in the Eyes Test. *PLoS One* 2020;15:e0238309.
  50. Fernández-Abascal EG, Cabello R, Fernández-Berrocal P, Baron-Cohen S. Test-retest reliability of the ‘Reading the Mind in the Eyes’ test: a one-year follow-up study. *Mol Autism* 2013;4:33.
  51. Xi C, Zhu Y, Zhu C, Song D, Wang Y, Wang K. Deficit of theory of mind after temporal lobe cerebral infarction. *Behav Brain Funct* 2013;9:15.

**Supplementary Table 1.** Mental state terms of the Korean-translated version of the RMET (English)

Item	Option 1	Option 2	Option 3	Option 4
1	장난끼 가득한 ( <b>playful</b> )	위로하는 (comforting)	신경질이 난 (irritated)	지루해 하는 (bored)
2	겁에 질린 (terrified)	속상한 ( <b>upset</b> )	오만한 (arrogant)	짜증이 난 (annoyed)
3	농담을 하는 (joking)	안절부절못하는 (flustered)	<b>욕망 (desire)</b>	확신에 찬 (convinced)
4	농담을 하는 (joking)	주장하는 ( <b>insisting</b> )	재미있어진 (amused)	느긋한 (relaxed)
5	신경질이 난 (irritated)	비꼬는 (sarcastic)	걱정하는 ( <b>worried</b> )	친절한 (friendly)
6	기겁을 한 (aghast)	<b>환상에 잠긴 (fantasizing)</b>	안달하는 (impatient)	위협을 느낀 (alarmed)
7	죄송스러워 하는 (apologetic)	친절한 (friendly)	<b>마음이 불편한 (uneasy)</b>	의기 소침한 (dispirited)
8	<b>실의에 빠진 (despondent)</b>	안도한 (relieved)	수줍어 하는 (shy)	흥분한 (excited)
9	짜증이 난 (annoyed)	적대적인 (hostile)	질겁한 (horrified)	<b>골몰하는 (preoccupied)</b>
10	조심하는 ( <b>cautious</b> )	주장하는 (insisting)	지루해 하는 (bored)	기겁을 한 (aghast)
11	겁에 질린 (terrified)	재미있어진 (amused)	<b>후회하는 (regretful)</b>	추파를 던지는 (flirtatious)
12	무관심한 (indifferent)	창피해 하는 (embarrassed)	<b>회의적인 (skeptical)</b>	의기 소침한 (dispirited)
13	단호한 (decisive)	기대하는 ( <b>anticipating</b> )	위협하는 (threatening)	수줍어 하는 (shy)
14	신경질이 난 (irritated)	실망한 (disappointed)	우울한 (depressed)	<b>책망하는 (accusing)</b>
15	<b>생각에 잠긴 (contemplative)</b>	안절부절못하는 (flustered)	격려하는 (encouraging)	재미있어진 (amused)
16	신경질이 난 (irritated)	<b>생각에 빠진 (thoughtful)</b>	격려하는 (encouraging)	동정 어린 (sympathetic)
17	<b>의심스러워 하는 (doubtful)</b>	애정이 넘치는 (affectionate)	장난끼 가득한 (playful)	기겁을 한 (aghast)
18	단호한 ( <b>decisive</b> )	재미있어진 (amused)	기겁을 한 (aghast)	지루해 하는 (bored)
19	오만한 (arrogant)	감사해 하는 (grateful)	비꼬는 (sarcastic)	<b>머뭇거리는 (tentative)</b>
20	지배적인 (dominant)	친절한 ( <b>friendly</b> )	죄책감을 느끼는 (guilty)	질겁한 (horrified)
21	창피해 하는 (embarrassed)	<b>환상에 잠긴 (fantasizing)</b>	혼란스러워하는 (confused)	공황상태에 빠진 (panicked)
22	<b>골몰하는 (preoccupied)</b>	감사해 하는 (grateful)	주장하는 (insisting)	애원하는 (imploring)
23	흡족한 (contented)	죄송스러워 하는 (apologetic)	<b>저항적인 (defiant)</b>	호기심에 찬 (curious)
24	수심에 잠긴 ( <b>pensive</b> )	신경질이 난 (irritated)	흥분한 (excited)	적대적인 (hostile)
25	공황상태에 빠진 (panicked)	믿으려 하지 않는 (incredulous)	실의에 빠진 (despondent)	<b>관심 있어 하는 (interested)</b>
26	위협을 느낀 (alarmed)	수줍어 하는 (shy)	<b>적대적인 (hostile)</b>	불안해하는 (anxious)
27	농담을 하는 (joking)	조심하는 ( <b>cautious</b> )	오만한 (arrogant)	안심시키는 (reassuring)
28	<b>관심 있어 하는 (interested)</b>	농담을 하는 (joking)	애정이 넘치는 (affectionate)	흡족한 (contented)
29	안달하는 (impatient)	기겁을 한 (aghast)	신경질이 난 (irritated)	<b>사색적인 (reflective)</b>
30	감사해 하는 (grateful)	추파를 던지는 ( <b>flirtatious</b> )	적대적인 (hostile)	실망한 (disappointed)
31	수치스러워 하는 (ashamed)	자신 있는 ( <b>confident</b> )	농담을 하는 (joking)	의기 소침한 (dispirited)
32	<b>진지한 (serious)</b>	수치스러워 하는 (ashamed)	얼떨떨한 (bewildered)	위협을 느낀 (alarmed)
33	창피해 하는 (embarrassed)	죄책감을 느끼는 (guilty)	환상에 잠긴 (fantasizing)	<b>걱정스러워 하는 (concerned)</b>
34	기겁을 한 (aghast)	당혹스러워 하는 (baffled)	<b>미심쩍어 하는 (distrustful)</b>	겁에 질린 (terrified)
35	어리둥절한 (puzzled)	<b>긴장한 (nervous)</b>	주장하는 (insisting)	생각에 잠긴 (contemplative)
36	수치스러워 하는 (ashamed)	긴장한 (nervous)	<b>수상쩍어 하는 (suspicious)</b>	우유부단한 (indecisive)

The correct answers are shown in bold

**Supplementary Table 2.** Descriptive statistics of the RMET in past and present studies<sup>1-9</sup>

Country	Sample	Overall mean (SD)	Male mean (SD)	Female mean (SD)
Korea, Korean-translated version of the RMET	196 (101 female)	25.7 (3.4)	25.5 (3.6)	25.8 (3.2)
Korea, K-RMET	196 (101 female)	26.7 (3.4)	26.2 (3.6)	27.2 (3.1)
Canada (Cook & Saucier, 2010)	88 (45 female)	Not reported	26.1 (3.8)	27.3 (4.1)
Germany (Pfaltz et al., 2013)	155 (86 female)	24.5 (3.5)	24.8 (3.2)	24.2 (3.7)
Italy (Vellante et al., 2013)	200 (108 female)	24.8 (4.2)	24.1 (4.7)	25.5 (3.5)
Korea (Lee, Nam & Hur, 2020)	200 (111 female)	26.6 (3.0)	26.5 (3.0)	26.6 (3.1)
Persia (Khorashad et al., 2015)	545 (282 female)	22.8 (3.4)	22.4 (3.6)	23.1 (3.2)
Poland (Jankowiak-Siuda et al., 2016)	325 (161 female)	25.0 (4.5)	23.9 (5.0)	26.1 (3.7)
Portugal (Pestana et al., 2018)	130 (71 female)	23.6 (5.9)	27.8 (2.9)	26.7 (3.8)
Turkey (Girli, 2014)	268 (160 female)	21.6	21.0	22.1
UK (Baron-Cohen et al., 2001)	103 (50 female)	28.0 (3.5)	27.3 (3.7)	28.6

**Supplementary Table 3.** A comparison of first/second frequent foils in different version of the RMET test<sup>3,5,7,8,10-12</sup>

Item	Target	Korean-translated version of the RMET (N=196)	K-RMET (N=196)	English (Prevost et al., 2014; N=139)	French (Prevost et al., 2014; N=97)	Italian (Vellante et al., 2013; N=200)	Persian (Khorashad et al., 2015; N=545)	Portuguese (Pestana et al., 2018; N=130)	Servian (Đorđević et al., 2017; N=260)	Spanish (Fernandez-Abascal et al., 2013; N=358)	Turkish (Girli, 2014; N=268)
1	Playful							Irritated*	Irritated*		Comforting*
2	Upset	Annoyed*	Arrogant*								
3	Desire	Convinced†	Convinced*			Convinced†	Convinced†		Convinced†		Joking†
4	Insisting			Relaxed†	Relaxed†	Relaxed†					
5	Worried	Sarcastic*					Irritated†				
6	Fantasizing										
7	Uneasy			Friendly*	Friendly*	Friendly*	Dispirited	Friendly*	Friendly†		Friendly*
8	Despondent										
9	Preoccupied										
10	Cautious		Aghast*	Insisting†		Insisting†	Insisting*				
11	Regretful		Terrified†		Flirtatious†						
12	Skeptical	Indifferent†									Indifferent†
13	Anticipating				Decisive*		Hard To Check†				
14	Accusing	Irritated†	Irritated*								
15	Contemplative										
16	Thoughtful	Sympathetic†		Sympathetic†							
17	Doubtful			Affectionate†	Affectionate*	Affectionate†	Affectionate†	Affectionate*		Affectionate†	Affectionate†
18	Decisive										
19	Tentative				Grateful†	Grateful†		Grateful†		Grateful*	Grateful*
20	Friendly		Dominant†								
21	Fantasizing		Confused†								
22	Preoccupied		Insisting†								
23	Defiant	Curious*		Curious†	Curious*		Curious†	Curious†	Curious†		Curious*
24	Pensive					Excited†	Irritated†				
25	Interested	Incredulous†					Incredulous*				Incredulous*
26	Hostile										Anxious†
27	Cautious	Arrogant†	Arrogant†		Arrogant*		Arrogant*				
28	Interested		Contented*				Contented*				Joking*
29	Reflective				Aghast†		Aghast*				
30	Flirtatious										
31	Confident										
32	Serious										
33	Concerned						Guilty†			Guilty†	
34	Distrustful				Baffled†						
35	Nervous				Puzzled*	Puzzled*	Puzzled†		Puzzled*		Puzzled*
36	Suspicious						Indecisive†	Indecisive†			

\*The target word was selected by fewer than 50% of the participants, †One of the option words was chosen by more than 25%

## REFERENCES

1. Cook CM, Saucier DM. Mental rotation, targeting ability and Baron-Cohen's empathizing–systemizing theory of sex differences. *Pers Individ Diff* 2010;49:712-716.
2. Pfaltz MC, McAleese S, Saladin A, Meyer AH, Stoecklin M, Opwis K, et al. The Reading the Mind in the Eyes Test: test-retest reliability and preliminary psychometric properties of the German version. *Int J Adv Psychol* 2013;2:1-9.
3. Vellante M, Baron-Cohen S, Melis M, Marrone M, Petretto DR, Masala C, et al. The “Reading the Mind in the Eyes” test: systematic review of psychometric properties and a validation study in Italy. *Cogn Neuropsychiatry* 2013;18:326-354.
4. Lee HR, Nam G, Hur JW. Development and validation of the Korean version of the Reading the Mind in the Eyes Test. *PLoS one* 2020;15:e0238309.
5. Khorashad BS, Baron-Cohen S, Roshan GM, Kazemian M, Khazai L, Aghili Z, et al. The “Reading the Mind in the Eyes” test: investigation of psychometric properties and test–retest reliability of the persian version. *J Autism Dev Disord* 2015;45:2651-2666.
6. Jankowiak-Siuda K, Baron-Cohen S, Bialaszek W, Dopierala A, Kozłowska A, Rymarczyk K. Psychometric evaluation of the ‘Reading the mind in the eyes’ test with samples of different ages from a Polish population. *Studia Psychologica* 2016;58:18.
7. Pestana J, Menéres MSSPC, Gouveia MJPM, Oliveira RF. The Reading the Mind in the Eyes Test: A Portuguese version of the adults’ test. *Análise Psicológica* 2018;36:369-381.
8. Girli A. Psychometric properties of the Turkish child and adult form of “Reading the Mind in the Eyes Test”. *Psychology* 2014;5:1321-1337.
9. Baron-Cohen S, Wheelwright S, Hill J, Raste Y, Plumb I. The “Reading the Mind in the Eyes” test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *J Child Psychol Psychiatry* 2001;42:241-251.
10. Prevost M, Carrier ME, Chowne G, Zekowitz P, Joseph L, Gold I. The Reading the Mind in the Eyes test: validation of a French version and exploration of cultural variations in a multi-ethnic city. *Cogn Neuropsychiatry* 2014;19:189-204.
11. Đorđević J, Živanović M, Pavlović A, Mihajlović G, Stašević-Karličić I, Pavlović D. Psychometric evaluation and validation of the Serbian version of ‘Reading the Mind in the Eyes’ test. *Psihologija* 2017;50:483-502.
12. Fernández-Abascal EG, Cabello R, Fernández-Berrocal P, Baron-Cohen S. Test-retest reliability of the ‘Reading the Mind in the Eyes’ test: a one-year follow-up study. *Mol Autism* 2013;4:33.