PSYCHOMETRIC EVALUATION OF THE 'READING THE MIND IN THE EYES' TEST WITH SAMPLES OF DIFFERENT AGES FROM A POLISH POPULATION

Kamila JANKOWIAK-SIUDA[1], Simon BARON-COHEN[2], Wojciech BIAŁASZEK[1],
Aleksandra DOPIERAŁA[1], Agata KOZŁOWSKA[1], Krystyna RYMARCZYK[1]

[1] Department of Experimental Neuropsychology, SWPS University of Social Sciences and Humanities
Chodakowska 19/31, 03-815 Warsaw, Poland
E-mail: kjankowiak-siuda@swps.edu.pl, wbialaszek@swps.edu.pl, agulczynska@swps.edu.pl, akozlowska2@swps.edu.pl, krymarczyk@swps.edu.pl

[2] Autism Research Centre, Department of Psychiatry, University of Cambridge
Douglas House, 18b Trumpington Road Cambridge CB2 8AH, UK
E-mail: s.baron-cohen@psychol.cam.ac.uk

Abstract: The 'Reading the Mind in the Eyes' Test (RMET) is a test of a Theory of Mind, i.e., the ability to infer the states of minds of other people. The purpose of this study was to evaluate a Polish adaptation of the RMET. The sample consisted of 447 participants, aged 18-85. The internal consistency of the RMET was 0.668; the upper confidence interval was 0.718. The score in the Polish version of the RMET was positively correlated with the English version. Test-retest stability was acceptable, with ICC = 0.886. The correlation of RMET and the cognitive empathy measure confirms the theoretical assumptions. There were significant gender differences in RMET scores: women had higher scores than men. Elderly groups of participants differ statistically from younger groups of participants in the RMET. The Polish version of the RMET showed satisfactory psychometric parameters, comparable to those of the original version.

Key words: RMET, mind reading, Theory of Mind (ToM), cognitive empathy, sex differences

Introduction

The 'Reading the Mind in the Eyes' Test (RMET) is a test of the Theory of Mind (ToM), the ability to infer the states of minds of other people (Baron-Cohen, 2001). ToM includes the recognition of emotional infor-

mation from the face, voice, and body (Tager-Flusberg & Sullivan, 1994). Theory of Mind allows for the acquiring of knowledge about other people and for constructing ideas about what people think and feel. The ToM allows for the constructing of beliefs about other people, making it easier to understand another person's motives and intensions (Baron-Cohen, 2001). The key aspect of the ToM is an ability to take the perspective of other persons, which is what makes this theoretical construct close to cognitive empathy (Philips et al., 2002).

The human face is one of the most important social stimuli with which we deal every day. From the emotional expression on a face, we are able to assess whether a person is friendly or hostile, and we can infer a wide range of mental states. These inferences enable a rapid response in different social situations (Itier & Batty, 2009). The eye area is the most important area of the face in allowing others to recognize facial expressions and their underlying emotions (Baron-Cohen, 1994). We devote more time to the eye area than to other parts of the face (Itier, Villate, & Ryan, 2007; McKelvie, 1976; Fraser, Craig, & Parker, 1990; Itier & Batty, 2009; Althoff & Cohen, 1999; Baron-Cohen, Baldwin, & Crowson, 1997). Other studies also supported the neural basis of this behavior. The superior temporal sulcus (STS) and the superior temporal gyrus (STG) specialize in the perception of the eyes and face. Both are activated by the visual stimuli of the eyes during performance on the RMET (Baron-Cohen, Ring, Wheelwright, Bullmore, Brammer, Simmons, & Williams, 1999; Itier, Alain, Sedore, & McIntosh, 2007; Moor, Op de Macks, Güroğlu, Rombouts, Van der Molen, & Crone, 2012).

Eyes and gaze play an important role in social interactions. Avoiding or failing to maintain eye contact and difficulty in joint attention (Baron-Cohen, 1987) are associated with impaired social communication and impaired ability to read the mental states of others, which occur commonly in people with autism (Baron-Cohen, Jolliffe, Mortimore, & Robertson, 1997).

The original version of the RMET was developed for adults in 1997 and was subsequently modified by adding more response options and improving the psychometric properties of the tool (Baron-Cohen, Jolliffe et al., 1997; Baron-Cohen et al., 2001). The test consists of 36 questions concerning choosing the right emotion or thought corresponding to a particular pair of eyes shown on a picture. A greater proportion of correct answers corresponds to a higher test score.

The RMET assesses the ability to read emotional states of other people from the expression around their eyes (Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001). This tool was created to identify subtle deficits in autism and Asperger's syndrome (Baron-Cohen et al., 2001). Currently, RMET is often applied to study individual differences such as sexual or age differences in the ToM (Baron-Cohen et al., 2001, and Bailey et al., 2008, respectively). The results revealed that women score higher in the original RMET than men (e.g., Baron-Cohen et al., 2001). These findings seem related to the results of the research, which has shown that empathy is more developed in women than in men (Goldenfeld et al., 2005; Baron-Cohen, Richler, Bisarya, Gurunathan, & Wheelwright, 2003; Geary et al., 1998; MaCoby et al., 1999). From earliest infancy, girls spend more time looking at faces, particularly the eyes, whereas boys turn their attention to moving objects (Connellan, Baron-Cohen, Wheelwright, Batki, & Ahluwalia, 2008). Women interpret all nonverbal messages more accurately on the basis of facial expression (e.g., the eyes) and intonation and are better at evaluating emotional states of other people (Baron-Cohen et al., 2003; Hall, 1978).

Age difference is another aspect of individual variances obtained in the RMET results. Older adults score lower than younger adults in RMET (Bailey et al., 2008). This aligns with the results of studies that re-

vealed problems with recognizing emotions from the upper part of a face, the near-eyes region, in people over 62 years of age (Philips et al., 2002; Bailey & Henry, 2009).

Taking into account how important the ability is to recognize emotions from faces, both for the ToM and for empathizing, it seems valuable to have a tool for both researchers and practitioners to measure that ability. From the scientific point of view, preparation of the Polish version of the RMET allows for the running of studies on Polish populations, thus making comparative analysis between different populations in the world. Thus, the aim of the present study is to evaluate the psychometric values of the Polish version of the RMET.

Methods

Participants

The study involved three samples. Group A ($N = 24$, 17 women, 7 men) consisted of participants fluent in Polish and English, age 28-47 years, who analyzed the linguistic terms of RMET translation. Group B comprised 325 people (161 women and 164 men, aged 18-45 years). This sample was used to test the scale and its internal consistency coefficient, and in a subset of the sample, its temporal stability ($N = 60$, 28 females and 32 males, aged 21-41 years), as well as its correlation with other questionnaires. The third group (group C) consisted of 98 people (49 women, 49 men, aged 25-85 years) and was used only to test the hypothesis of the existence of a relation between RMET and age. The samples were drawn from students and staff at a university, and staff (but not patients) in a nursing home and a psychiatric hospital.

Measures

The Reading the Mind in the Eyes Test (RMET). The test consisted of showing 36 images of pairs of eyes from adult men and women. Around each picture were four adjectives. Respondents selected the adjective that best described what the person in the picture was feeling or thinking, and (in the control conditions) judged the gender of the person in the photo. For every correct answer, the participant received one point. After summing the points, the result for the RMET was obtained, with a maximum of 36 points available (Baron-Cohen, 2001). Validation studies of the original version of the test showed that it has reliability acceptable for a measure for group comparisons and experimental $\alpha = .63$ for RMET (Harkness, Jacobson, Duong, & Sabbagh, 2010).

Empathy Sensitivity Scale (ESS). The scale is a paraphrase of the Interpersonal Reactivity Index (IRI) by Davis (Davis, 1980). It is a self-report questionnaire including 28 statements with a five-point response scale, comprising the following subscales: Perspective-taking (PT) – the ability and tendency to spontaneously adopt another's point of view; Empathic Concern (EC) – the tendency to empathize with people experiencing failure and loss; Personal Distress (PD) – the propensity to experience feelings in the context of strong negative experiences of others; and Fantasy (F) – the ability to be moved by fictional, imaginary events (feelings and actions of characters from books or movies). The last subscale, Fantasy, was excluded from the Polish adaptation of the Scale because it is the least theoretically grounded subscale, often overlooked in studies (e.g., Davis & Oathout, 1987; Davis, Hall, & Meyer,

2003). Examples of statements in the scale are: "Sometimes I try to understand my friends better by imagining how things look from their point of view." (PT); "Reluctantly, I give emotional support to people in difficult situations." (EC); and "Finding myself in a situation of emotional tension scares me." (PD). The tool has satisfactory reliability and theoretical validity (Cronbach's alpha for the EC and PD is 0.78 and 0.74 for PT) (Kaźmierczak et al., 2007).

Psychological Gender Inventory (PGI). PGI is a tool to assess the psychological sex of the individual. Psychological gender is the spontaneous willingness to use the gender dimension in relation to oneself and others. The PGI consists of 35 items (15 - Femininity scale, 15 - Masculinity scale, 5 - neutral positions). The subject is asked to mark the degree to which s/he agrees with the statements, using a five-point scale. Test items reflect the cultural stereotypes of masculinity and femininity. The tool has satisfactory internal consistency coefficient (Femininity scale - 0.79, Masculinity scale - 0.78) (Kuczyńska, 1992).

Emotional Intelligence Scale-Faces (EIS-F). The scale measures the ability to recognize facial expressions. The test consists of 18 photographs of faces (half male, half female). Individual photographs are assigned sets of six names of emotions. The participant has to decide in each case whether the face shown in the photograph expresses those emotions. The total number of test items is 108 (18 photographs x 6 emotions). The scale has high reliability, Cronbach's alpha ranges from 0.77 to 0.87, depending on gender and age (Matczak, Piekarska, & Studniarek, 2005).

Empathy Quotient (EQ-S) [short version] (Jankowiak-Siuda, Kantor-Martynuska,

Siwy-Hudowska, submitted). The Empathy Quotient-Short (EQ-S) (Wakabayashi et al., 2006) is a shortened version of the scale established for measuring empathy (Baron-Cohen & Wheelwright, 2004). The scale consists of 22 statements that describe how an individual behaves towards other people. Among these statements, there are some that determine the ability to put ourselves into the shoes of another person, to anticipate and understand what they might feel, think, or do (cognitive empathy), and to generate emotional responses to others (emotional empathy), as well as those that combine cognitive and emotional empathy (multidimensional empathy).

The participant is asked to agree or disagree with each statement by selecting one of four possible answers ("definitely yes", "probably yes", "probably not", "definitely not") (Wakabayashi et al., 2006). Validation studies showed that this tool has high reliability ($\alpha = 0.88$). Scoring is calculated as follows: "probably yes" = one point; "definitely yes" = two points; and the other answers = zero points. This is so for items 1, 2, 5, 6, 7, 8, 10, 11, 13, 15, 16, 17, 18, 19, 20, and 21. Item numbers 3, 4, 9, 12, 14, and 22 are scored differently. For answers "probably not" one point is given, and for the "definitely not" answer, two points are given. The remaining positive responses score zero points. After summing up the points, the total EQ-S is obtained, with a maximum of 44 points (Baron-Cohen, Richler, Bisary, Gurunathan, & Wheelwright, 2003; Baron-Cohen & Wheelwright, 2004).

Procedure

The adaptation of the RMET was carried out in accordance with principles of transla-

tion, demonstrating accuracy of translation, that is, maintaining accuracy to the original version of the questionnaire, while allowing modifications required by the given language (Zawadzki & Hornowska, 2008). The original was translated into Polish and then back-translated into English by bilingual speakers for verification. The original scale was translated independently by five translators: three psychologists fluent in English and two English philologists. After an analysis of selected adjectives, for which there were significant differences (items 2, 8, 11, and 27), they were returned to the translators for another attempt. Discrepancies were discussed, followed by three psychologists choosing the most appropriate version.

To check the consistency of both language versions, correlation coefficients were calculated, and the significance of differences between the results of both versions of the questionnaire was tested. The next step was to analyze the reliability, consistency, and temporal stability of the tool and then to examine the convergent validity. Correlation of RMET results with the following scales: SWE, PGI, EIS-F, and IE were calculated. Finally, it was checked whether the RMET results were dependent upon the age of the subjects.

Reliability analysis was tested on the basis of an internal consistency, i.e., Cronbach's alpha. Intergroup comparisons performed to assess the significance of differences when comparing groups were conducted using the t-Student test for independent samples. To calculate the relationship between the questions within scales, as well as the relationship between scales, r-Pearson correlations were used. The statistical package IBM SPSS Statistics was used to calculate and analyze the validation of the Polish adaptation of the tool.

Results

Analysis of Equivalence of Two RMET Language Versions

The language equivalence analysis was performed on the sample of 24 bilingual participants (aged 24-45 years). All participants first completed the English version, and then, after four weeks, the Polish version. To verify the results from the two language versions, we calculated the Pearson's correlation coefficient. Scores obtained from the English and Polish versions of RMET were significantly correlated ($r = .731$; $p < .001$; CI [.430; 1]). This indicates that there is high consistency between the two language versions. Moreover, there were no statistically significant differences between means in both measurements ($t(23) = 0.332$; $p = .743$; Cohen's $d = 0.068$). Mean difference between Polish and English versions was equal to 1.167 (95% CI [-.873; 1.206]) points on the scale, with $SD = 2.461$. Although we are aware that it would be methodologically more correct if one half of the sample completed the English version first and then the Polish version, whereas the second half of the sample completed the versions in reverse order, we believe that a four-week-long delay between first testing and second testing is long enough to lessen the impact of repeated testing on the results. In conclusion, the analysis showed that the Polish version is equivalent to the original RMET version.

Test-Retest Reliability and Internal Consistency of RMET

The internal consistency analysis was based on Cronbach's alpha coefficient for

dichotomous data (KR20). This analysis was performed on a sample of $N = 325$ participants. Cronbach's alpha for this scale was .668 (95% CI [0.614; 0.718]), which can be interpreted as satisfactory for our scientific purposes. Cronbach's alpha was not markedly improved by the removal of any item from the score. In the next step of analysis, we tested test-retest reliability. The analysis was performed on the sample of 60 subjects aged 21-41 years (28 females and 32 males). The correlation between two measurements separated by four weeks was $r = .886$, $p < .001$, 95% CI [.765; 1]. The time stability of the RMET is very satisfactory.

Descriptive Statistics of the Results of RMET

The main sample with which internal consistency of the test was assessed consisted of 325 adult participants (161 females, 164 males), aged 18-45 years (mean 27.848, $SD = 7.485$). In this sample, 90 participants completed only the RMET, and the remaining 235 completed the RMET and other questionnaires, which are described below. Descriptive statistics for the RMET scores are presented in Table 1. We also include results separately for female and male participants because of significant gender differences on the RMET (for significance test-

ing, see the section on sex differences provided below).

Theoretical Validity

Theoretical validity of the scale was checked on the subsamples of sample B (sex differences and convergent and divergent validity) and C (RMET score and age). The results obtained show similarities with the original version (Baron-Cohen, 2001).

Sex Differences

There were significant gender differences in RMET scores ($t(301) = 4.486$; $p < .001$; Cohen's $d = 0.532$). This difference suggests medium effect in sex differences. Women had significantly higher scores than males (see Table 1).

RMET Score and Age

The following analyses were performed on sample C, which consisted of 98 participants (49 women; 49 men) aged 25-85 years. There were 40 younger participants (25-34 years) and 58 older participants (70-85 years) in the sample. The older group of participants had lower scores ($M = 19.017$; $SD = 3.882$) on RMET than the younger group ($M = 23.725$; $SD = 3.850$). The difference is statistically

Table 1 *Descriptive statistics of RMET*

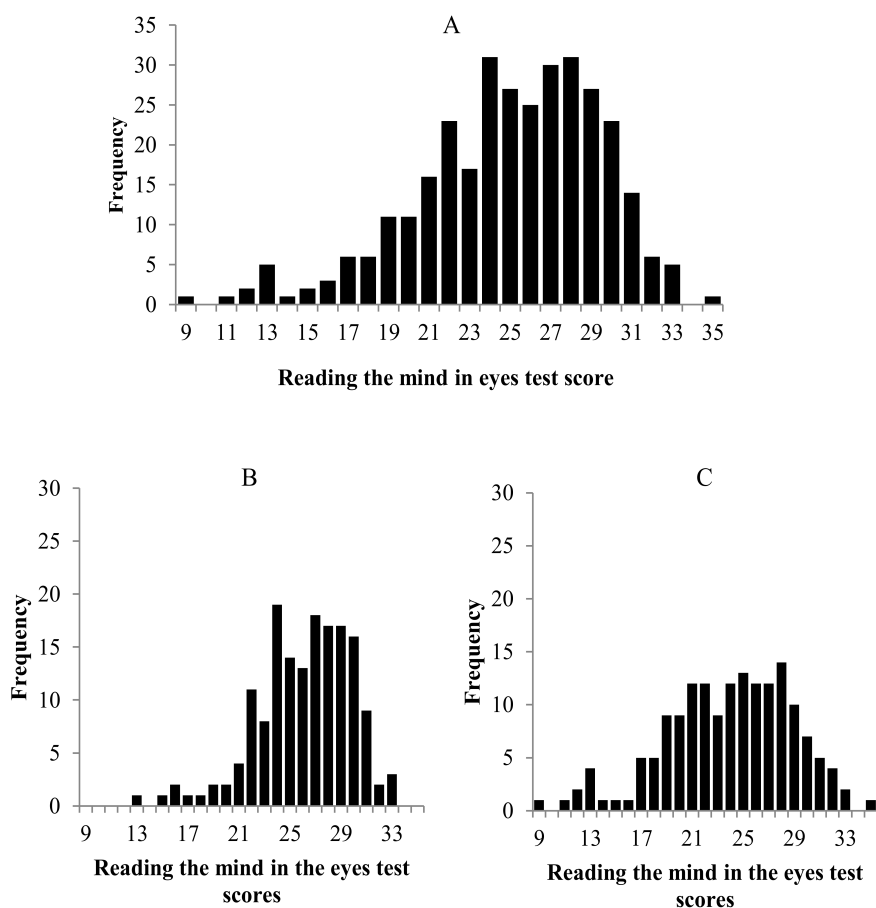| Statistic	Whole sample	Females ($n = 161$)	Males ($n = 164$)
Mean	24.982	26.112	23.872
Standard deviation	4.525	3.691	4.982
Skewness	-0.716	-0.753	-0.468
Kurtosis	0.506	0.831	-0.027
Minimum	9	13	9
Maximum	35	33	35

Figure 1 Distribution of RMET scores, (36 items) (panel A: whole sample (N = 325); panel B: females (N = 161); panel C: males (N = 164)

significant (*t*(96) = 5.921l; *p* < .001; Cohen's *d* = 1. 209), and the effect size can be classified as large.

Convergent and Divergent Validity of the Polish RMET Adaptation

Out of all participants from sample B, 175 subjects (88 women, 87 men) completed ESS scale (Kaźmierczak, Plopa, & Retowski, 2007) and RMET. Table 2 shows Pearson correlation coefficients of RMET and ESS subscales.

The analysis revealed that RMET correlates only with the subscale of Perspective-taking. Out of the 175 subjects who were described in abovementioned analysis, 135 participants (68 women, 67 men) also completed PGI scales. During the next analysis,

Table 2 *Correlations of RMET and ESS*

EES subscales	Correlation with significance level and 95% confidence interval
Empathic Concern	$r = 0.073$; $p = .339$, CI [-.077; .222]
Personal Distress	$r = -0.026$; $p = .728$, CI [-.176; .124]
Perspective-taking	$r = 0.301$; $p < .001$, CI [.158; .444]

Table 3 *RMET and PGI correlations*

PGI scales	Correlation with significance level and 95% confidence interval
Feminity scale	$r = 0.069$; $p = .428$; CI [-.102; .240]
Masculinity scale	$r = -0.044$; $p = .616$; CI [-.215; .128]

it was checked whether RMET correlated with two PGI scales (Table 3). As can be seen in Table 3, RMET correlated neither with the Feminity scale, nor with the Masculinity scale.

The correlation of RMET, EIS-F and EQ-S was checked using a sample of 60 subjects from sample B (28 women and 32 men), who completed only these three questionnaires. There was a significant correlation between the RMET and the EIS-F ($r = .467$; $p < .001$; 95% CI [.235; .700]). The final analysis showed a non-significant correlation between EQ-S and RMET ($r = .124$; $p = .374$; 95% CI [-.137; .384]). Overall, the analyses showed that the Polish adaptation of the RMET is valid.

Discussion

The ability to recognize emotional expressions (from the face, voice, or body) is the basis of the Theory of Mind, the identification of the mental states of others (Baron-Cohen, 1995; Adolphs, 2009). Psychometric

evaluations of the Reading the Mind in the Eyes' test – Polish version indicates, similarly to the English version, that the Polish adaptation is characterized by satisfactory psychometric properties. In the present study, Cronbach's alpha was .668, whereas maximal weighted internal consistency reliability for the unidimensional model provided a better estimate (.718). This indicates acceptable internal consistency of the RMET in Poland; thus, it is a reliable tool that measures mindreading in adults. Reliability, which was reported in past studies of the RMET, was .58 to .70 (Voracek & Dressler, 2006; Harkness et al., 2010; Dehning, Girma, Gasperi, Meyer, Tesfaye, & Siebeck, 2012; Vellante, Baron-Cohen, Melis, Petretto, Masala, & Preti). Additionally, the test-retest stability of the Polish version of RMET was acceptable, with interclass correlation coefficients equaling 0.886.

There were sex differences in Reading the Mind in the Eyes' test scores: women had higher scores than men, and the effect size was medium. The same results were obtained

by Carroll and Chiew (2006), Voracek and Dressler (2006), Yildirim, Kaşar, Güdük, Ateş, Küçükparlak and Ozalmete (2011), and Vellante et al. (2012), which confirms results of earlier studies (Baron-Cohen & Wheelwright, 2004; Lawrence, Shaw, Baker, Baron-Cohen, & David, 2004; Von Horn, Backman, Davidsson, & Hansen, 2010; Eisenberg & Lennon, 1983). Women tend to be more emotionally expressive (Kring & Gordon, 1998). A substantial body of research has suggested that women tend to perform better than men in identifying and discriminating between different facial emotional expressions (Buck, Miller, & Caul, 1972; Hampson, Van Anders, & Mullin, 2006; McClure, 2000; Thayer & Johnsen, 2000). Women have faster reaction times and a higher rate of correct classification than men (Fischer, Rodriguez, Mosquera, van Vianen, & Manstead, 2004). This study provides new evidence concerning psychological gender and shows that higher femininity and lower masculinity lead to higher capacity for ToM. Participants who scored high on the RMET also scored higher on the Femininity scale of the Psychological Gender Inventory (with statistical tendency) and lower results on the masculinity scale of the inventory.

Some studies report evidence for a correlation between Empathy Quotient and RMET (Ali & Chamorro-Premuzic, 2010; Carroll & Chiew, 2006; Cook &Saucier, 2010; Voracek & Dressler, 2006), although the correlations reported are rather weak, from 0.23 to 0.44. In this study, we did not find a correlation between Empathy Quotient-Short and the Polish version of the RMET. The same result was obtained in the Italian version of RMET (Vellante et al., 2012). This lack of correlation may be related to the fact that the Empathy Quotient is a multi-dimensional measure of empathy. On the basis of the Empathy Quotient results, it is difficult to separate cognitive and affective empathy, with the former being theoretically related to the Theory of Mind (Rogers et al., 2007). Therefore, it is not surprising that the majority of validation studies did not report the analysis of theoretical convergence between the Empathy Quotient and the RMET (Ethiopian version, Dehning et al., 2012; Hungarian version, Kelemen, Keri, Must, Benedek, & Janka 2004; Japanese version, Kunihira, Senju, Dairoku, Wakabayashi, & Hasegawa 2006; Turkish version, Yildirom et al., 2011; Swedish version, Hallerbäck, Lugnegard, Hjarthag, & Gillberg, 2009).

This interpretation also receives some support from the results of our study, which show positive correlations between cognitive empathy measured by the Empathy Sensitivity Scale, and from the RMET results. The Empathy Sensitivity Scale, another test measuring multi-dimensionally understood empathy, includes three scales, separating the dimensions of empathy: Empathic Concern and Personal Distress (affective empathy) and Perspective-taking (cognitive empathy). The only correlation obtained between the RMET results and the Empathy Sensitivity Scale subscales involved the cognitive empathy measure (Perspective-taking subscale). This supports the theoretical assumptions about RMET as a tool specialized in measuring rather a cognitive dimension of empathy, and supports our line of reasoning. Cognitive empathy includes the imagining and understanding of another person's perspectives (their feelings and intentions), enabling the observer to act in a context-specific manner (Jankowiak-Siuda & Zajkowski, 2013). It is worth pointing out that there are neurological reasons the ToM and perspective taking

should be related. During the act of imagining, the most active regions include the medial prefrontal cortex (MPFC) (Decety & Sommerville, 2003; Frith & Frith, 2003; Jeannerod, 2003; Meltzoff & Decety, 2003; Blanke & Arzy, 2005).

Also worth mentioning is the correlation obtained between the RMET and Emotional intelligence scale-faces. To the best of our knowledge, only one paper thus far has presented a positive correlation between the RMET and ad hoc emotion recognition task (Alaerts, Nackaerts, Meyns, Swinnen, & Wenderoth, 2011). Emotional intelligence scale-faces is developed to measure the ability to recognize mimic expressions, a component of cognitive skill involved in emotional intelligence. The correlation between the RMET and the scale provides another argument for the conclusion that the Polish version of the RMET is valid and reliable.

As far as the age differences are concerned, the elderly group of participants has lower results on the RMET then a younger group of participants. The current results suggest that late adulthood is associated with a reduced capacity for Theory of Mind (Ligneau-Herve & Muller, 2005; Maylor, Moulson, Muncer, & Taylor, 2002; Bailey & Henry, 2008). To some degree, reduced ToM capacity is related to the less involved social activity of older people, which may be considered a specific type of the ToM and empathizing training (German & Hehman, 2006; Maylor et al., 2002; McKinnon & Moscovitch, 2007; Sullivan & Ruffman, 2004). Moreover, research on the relationship between cognitive and emotional empathy and social functioning in late adulthood (Bailey, Henry, & Hippel, 2008; Bailey & Henry, 2008) has revealed lower results in cognitive empathy tasks in older adults when compared to younger adults. This research is also consistent with the suggestion from neurobiological studies that the prefrontal cortex is important to perspective taking (Jankowiak-Siuda, Rymarczyk, & Grabowska, 2011) and may be vulnerable to age-related decline (Bailey & Henry, 2008).

Although present research suggests that the Polish version of the RMET has good psychometric characteristics, some limitations should to be considered. For example, the internal consistency is acceptable (Cronbach's alpha of .67) and comparable with other language versions of the test, however it allows the RMET to be used for research purposes rather than for individual diagnosis of the capacity for ToM. Additionally, while the retest reliability was high, the period of time between the first and second assessments was relatively short (four weeks).

In conclusion, this study shows that the Polish version of the RMET is a valid and reliable measure of Theory of Mind, which can be used to study a wide range of individual differences in adults (age and sex among other factors), especially for research purposes. The Polish RMET has comparable characteristics to other language versions of the test, so it can be used, for example, for between cultures comparisons. This test could also be used by researchers interested in differences in the Theory of Mind between specific groups, including such clinical samples as people with ASD, various types of psychosis and personality disorders.

References

Adolphs, R. (2009). The social brain: Neural basis of social knowledge. *Annual Reviews of Psychol-*

ogy, *60*, 693-716. doi:10.1146/annurev.psych.60. 110707.163514

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*, e20989.

Ali, F., & Chamorro-Premuzic, T. (2010). Investigating Theory of Mind deficits in nonclinical psychopathy and Machiavellianism. *Personality and Individual Differences*, *49*, 169-174. doi: 10.1162/089892902760807258.

Althoff, R. R., & Cohen, N. J. (1999). Eye-movement-based memory effect: A reprocessing effect in face perception. *Journal of Experimental Psychology Learning Memory and Cognition*, *25*(4), 997-1010. doi:10.1037//0278-7393.25. 4.997

Bailey, P. E., & Henry, J. D. (2008). Growing less empathic with age: Disinhibition of the self-perspective. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *63*(4), 219-226.

Bailey, P. E., Henry, J. D., & von Hippel, W. (2008). Empathy and social functioning in late adulthood. *Aging and Mental Health*, *12*(4), 499-503.

Baron-Cohen, S. (1987). Autism and symbolic play. *British Journal of Developmental Psychology*, *5*(2), 139-148.

Baron-Cohen, S. (1994). How to build a baby that can read minds: Cognitive mechanisms in mind-reading. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition*, *13*, 513-552.

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*: Massachusetts, US: MIT Press/Bradford Books.

Baron-Cohen, S. (2001). Theory of mind in normal development and autism. *Prisme*, *34*, 174-183.

Baron-Cohen, S. (2003). *The essential difference: Men, women and the truth about autism*. New York: Basic Books.

Baron-Cohen, S., & Wheelwright, S. (2004). The Empathy Quotient: An investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders*, *34*, 163-175.

Baron-Cohen, S., Baldwin, D. A., & Crowson, M. (1997). Do children with autism use the speaker's direction of gaze strategy to crack the code of language? *Child Development*, *68*(1), 48-57. doi:10.1111/j.1467-8624.1997.tb01924.x

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., Richler, J., Bisarya, D., Gurunathan, N., & Wheelwright, S. (2003). The Systemising Quotient (SQ): An investigation of adults with Asperger syndrome or high functioning autism and normal sex differences. *Philosophical Transactions of the Royal Society, Series B, Special issue on "Autism: Mind and Brain"*, *358*, 361-374. doi: 10.1098/rstb.2002. 1206

Baron-Cohen, S., Ring, H., Wheelwright, S., Bullmore, E. T., Brammer, M. J., Simmons A., & Williams, S. C. R. (1999). Social intelligence in the normal and autistic brain: An fMRI study. *European Journal of Neuroscience*, *11*(6), 1891-1898. doi: 10.1046/j.1460-9568.1999.00621.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. *Journal of Child Psychology and Psychiatry*, *42*(2), 241-251. doi; 10.1111/1469-7610.00715

Blanke, O., & Arzy, S. (2005). The out-of-body experience: disturbed self-processing at the temporo-parietal junction. *Neuroscientist*, *11*(1),16-24. doi: 10.1177/1073858404270885

Buck, R., Miller, R. E., & Caul, W. F. (1972). Sex, personality, and physiological variables in the communication of affect via facial expression. *Journal of Personality and Social Psychology*, *30*, 587-596.

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.

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

Connellan, J., Baron-Cohen, S., Wheelwright, S., Batki, A., & Ahluwalia, J. (2008). Sex differences in human neonatal social perception. *Infant Behavior and Development*, *23*, 113-118.

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/S0163-6383(00)00032-1

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

Davis, M. H., & Oathout, H. A. (1987). Maintenance of satisfaction in romantic relationships: Empathy and relational competence. *Journal of Personality and Social Psychology, 53*(2), 397-410.

Davis, M. H., Hall, J. A., & Meyer, M. (2003). The first year: Influences on the satisfaction, involvement, and persistence of new community volunteers. *Personality and Social Psychology Bulletin, 29*(2), 248-260.

Decety, J., & Somerville, J. A. (2003). Shared representations between self and other: A social cognitive neuroscience view. *Trends in Cognitive Sciences, 7*(12), 527-533. doi:10.1016/j.tics.2003.10.004

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.

Eisenberg, N., & Lennon, R. (1983). Sex differences in empathy and related capacities. *Psychological Bulletin, 94*, 100-131. doi: 10.1037/0033-2909.94.1.100

Fischer, A. H., Rodriguez Mosquera, P. M., van Vianen, A. E. M., & Manstead, A. S. R. (2004). Gender and culture differences in emotion. *Emotion, 4*(1), 87-94. doi: 10.1037/1528-3542.4.1.87

Fraser, I. H., Craig, G. L., & Parker, D. M. (1990). Reaction time measures of feature saliency in schematic faces. *Perception, 19*, 661-673. doi: 10.1068/p190661.

Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society B: Biological Sciences, 358*(1431), 459-473. doi: 10.1098/rstb.2002.1218

Geary, D. C. (1998). *Male, Female: The Evolution of Human Sex Differences*. Washington, DC. 21:. American Psychological Association.

German, T. P., & Hehman, J. A. (2006). Representational and executive selection resources in 'theory of mind': Evidence from compromised belief-desire reasoning in old age. *Cognition, 101*(1), 129-152.

Goldenfeld, N., Baron-Cohen, S., & Wheelwright, S. (2005). Empathizing and systemizing in males, females and autism. *International Journal of Clinical Neuropsychology, 2*, 338-345.

Hall, J. A. (1978). Gender effects in decoding nonverbal cues. *Psychological Bulletin, 85*, 845-857.

Hallerbäck, M. U., Lugnegard, T., Hjarthag, F., & Gillberg, C. (2009). The Reading the Mind in the Eyes Test: Test retest reliability of a Swedish version. *Cognitive Neuropsychiatry, 14*, 127-143.

Hampson, E., van Anders, S. M., & Mullin, L. (2006). A female advantage in the recognition of emotional facial expressions: Test of an evolutionary hypothesis. *Evolution and Human Behavior, 27*(6), 401-416. doi:10.1016/j.evolhumbehav.2006.05.002

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

Itier, R. J., & Batty, M. (2009). Neural bases of eye and gaze processing: The core of social cognition. *Neuroscience and Biobehavioral Reviews, 33*(6), 843-863. doi:10.1016/j.neubiorev.2009.02.004

Itier, R. J., Alain, C., Sedore, K., & McIntosh, A. R. (2007). Early face processing specificity: It's in the eyes! *Journal of Cognitive Neuroscience, 19*(11), 1815-1826. doi:10.1162/jocn.2007.19.11.1815

Itier, R. J., Villate, C., & Ryan, J. D. (2007). Eyes always attract attention but gaze-orienting is task dependent: Evidence from eye movement monitoring. *Neuropsychologia, 45*, 1019-1028.

Jankowiak-Siuda, K., & Zajkowski, W. (2013). A neural model of mechanisms of empathy deficits in narcissism. *Medical Science Monitor, 19*, 934-941. doi: 10.12659/MSM.889593

Jankowiak-Siuda, K., Rymarczyk, K., & Grabowska, A. (2011). How we empathize with others: A neurobiological perspective. *Medical Science Monitor, 17*(1), 18-28. doi:10.12659/MSM.881324

Jeannerod, M. (2003). The mechanism of self-recognition in humans. *Behavioural Brain Research, 142*, 1-15.

Kaźmierczak, M., Plopa, M., & Retowski, S. (2007). Skala wrażliwości empatycznej [Empa-

thy sensibility scale]. *Przegląd Psychologiczny*, *50*(1), 9-24.

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.

Kring, A. M., & Gordon, A. H. (1998). Gender differences in emotion: Expression, experience, and physiology. *Journal of Personality and Social Psychology*, *74*, 686-703. doi:10.1037/0022-3514.74.3.686

Kuczyńska, A. (1992). *Inwentarz do oceny Płci Psychologicznej* [Gender Inventory]. Warszawa: Pracownia Testów Psychologicznych.

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.

Lawrence, E. J., Shaw, P., Baker, D., Baron-Cohen, S., & David, A. S. (2004). Measuring empathy: Reliability and validity of the empathy quotient. *Psychological Medicine*, *34*, 911-919.

Ligneau-Herve, C., & Mullet, E. (2005). Perspective-taking judgments among young adults, middle-aged, and elderly people. *Journal of Experimental Psychology: Applied*, *11*(1), 53-60. doi:10.1037/1076-898X.11.1.53

Maccoby, E. E. (1999). *The two sexes: Growing up apart, coming together*. Cambridge, Massachusetts: Harvard University Press.

Matczak, A., Piekarska, J., & Studniarek, E. (2005). *Skala Inteligencji Emocjonalnej – Twarze (SIE-T): podręcznik* [Emotional Intelligence Scale – Faces: manual]. Warszawa: Pracownia Testów Psychologicznych.

Maylor, E. A., Moulson, J. M., Muncer, A., & Taylor, L. A. (2002). Does performance on theory of mind tasks decline in old age? *British Journal of Psychology*, *93*, 465-485. doi:10.1348/000712602761381358

McClure, E. B. (2000). A meta-analytic review of sex differences in facial expression processing and their development in infants, children, and adolescents. *Psychological Bulletin*, *126*(3), 424-453. doi:10.1037/0033-2909.126.3.424

McKelvie, S. J. (1976). The role of eyes and mouth in the memory of a face. *American Journal of Psychology*, *89*, 311-323.

McKinnon, M. C., & Moscovitch, M. (2007). Domain-general contributions to social reasoning: Theory of mind and deontic reasoning re-explored. *Cognition*, *102*(2), 179-218.

Meltzoff, A. N., & Decety, J. (2003). What imitation tells us about social cognition: A rapprochement between developmental psychology and cognitive neuroscience. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *358*(1431), 491-500. doi:10.1098/rstb.2002.1261

Moor, G. B., Op de Macks, Z. A., Güroğlu, B., Rombouts, S. A. R. B., Van der Molen, M. W., & Crone, E. A. (2012). Neurodevelopmental changes of reading the mind in the eyes. *Social Cognitive and Affective Neuroscience*, *7*, 44-52. doi:10.1093/scan/nsr020

Phillips, L. H., MacLean, R. D. J., & Allen, R. (2002). Age and the understanding of emotions: Neuropsychological and sociocognitive perspectives. *Journal of Gerontology: Psychological Sciences, 57B*, 526-530.

Rogers, K., Dziobek, I., Hassenstab, J., Wolf, O. T., & Convit, A. (2007). Who cares? Revisiting empathy in Asperger syndrome. *Journal of Autism and Developmental Disorders, 7*, 709-715. doi:10.1007/s10803-006-0197-8.PMID 16906462.

Sullivan, S., & Ruffman, T. (2004). Emotion recognition deficits in the elderly. *International Journal of Neuroscience*, *114*, 403-432.

Tager-Flusberg, H., & Sullivan, K. (1994). Predicting and explaining behavior: A comparison of autistic, mentally retarded and normal children. *Journal of Child Psychology and Psychiatry*, *35*, 1059-1075. doi:10.1111/j.1469-7610.1994.tb01809.x

Thayer, J., & Johnsen, B. H. (2000). Sex differences in judgement of facial affect: A multivariate analysis of recognition errors. *Scandinavian Journal of Psychology*, *41*, 243-246. doi:10.1111/1467-9450.00193

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*, *18*(4), 326-354. doi:10.1080/13546805.2012.721728

von Horn, A., Backman, L., Davidsson, T., & Hansen, S. (2010). Empathizing, systemizing and finger length ratio in a Swedish sample. *Scandinavian Journal of Psychology*, *51*, 31-37. doi:10.1111/j.1467-9450.2009.00725.x

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, systemizing, and autism-spectrum quotients in a general population sample. *Personality and Individual Differences*, *41*, 1481-1491. doi:10.1016/j.paid.2006.06.009

Wakabayashi, A., Baron-Cohen, S., Wheelwright, S., Goldenfeld, N., Delaney, J., Fine, D., Weil, L. (2006). Development of short forms of the Empathy Quotient (EQ-Short) and the Systemizing Quotient (SQ-Short). *Personality and Individual Differences*, *41*, 929-940. doi:10.1016/j.paid.2006.03.017

Yildirim, E. A., Kaşar, M., Güdük, M., Ateş, E., Küçükparlak, 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/Turkish Journal of Psychiatry*, *22*, 177-186.

Zawadzki, B., & Hornowska, E. (2008). Psychometria. Konstrukcja i adaptacja testów psychologicznych [Psychometry. Constructing and adapting psychological tests]. In J. Strelau, & D. Doliński (Eds.), *Psychologia* (pp. 847-891). Gdańsk: GWP.