Croatian Adaptation of the Revised Reading the Mind in the Eyes Test (RMET)

Marina Kotrla Topić

Institute of Social Sciences Ivo Pilar, Osijek, Croatia

Marina Perković Kovačević

Department of Psychiatry, University Hospital Centre, Osijek, Croatia

Abstract

The aim of this research was to translate and adapt the revised version of the "Reading the mind in the eyes test" (Baron-Cohen et al., 2001) to the Croatian language, and to provide preliminary data on its reliability, factor structure and convergent validity in a healthy population of Croatian students. After translation and adaptation, the Croatian version of the RMET was administered to 146 undergraduate and graduate students (84 female and 62 male participants). Together with the RMET, we administered the Emotional Empathy Scale (Raboteg-Šarić, 1993). Results show low internal consistency reliability of the Croatian adaptation of the RMET and adequate reliability measured with maximal reliability H coefficient. Confirmatory factor analysis marginally supports the unidimensional model. Convergent validity was marginally confirmed by a significant positive correlation between REMT and empathy. Additionally, we created a short version of the RMET, showing adequate fit indices, but containing only seven items. Internal consistency reliability and composite reliability for this scale were satisfactory. We propose further investigation of psychometric properties of the Croatian adaptation of the RMET with research in general, more representative population. We also propose investigating test-retest reliability, as well as discriminant validity of the test.

Keywords: social cognition, theory of mind, Reading the Mind in the Eyes Test, Croatian adaptation of the RMET

Introduction

In research on social cognition, the theory of mind refers to our understanding of mental states – beliefs, desires, intentions, thoughts, perceptions, etc. (Premack & Woodruff, 1978). It also refers to our understanding of emotions. Theory of mind is

✉ Marina Kotrla Topić, Institute of Social Sciences Ivo Pilar, Šamačka 9/II, 31000 Osijek, Croatia. E-mail: *marina.kotrla.topic@pilar.hr*

defined as a metarepresentational ability – representation of a representation. People use this ability to attribute mental states to themselves and to others. Based on these metarepresentations, we try to explain and predict human behaviour. If someone goes to the kitchen and reaches for chocolate from a cupboard, we assume they want some chocolate (desire) and they believe there is chocolate in the kitchen cupboard (belief), so we explain their behaviour based on desires and beliefs that we attribute to them. This kind of reasoning is something most healthy adults do automatically and unconsciously, and it is considered necessary for functioning in the social world. However, there are people who have difficulties in this domain. Research shows that persons with schizophrenia and autism have impaired ability to correctly attribute mental states to others (Brune, 2005; Beaumont & Newcombe, 2006).

Theory of mind ability is not something we are born with, but rather something we acquire in the early years of childhood. There are numerous studies on which particular abilities develop at which particular time (for review see Baron-Cohen, Tager-Flusberg, & Cohen, 2000; Flavell, 2004). We now know that from birth, babies show particular interest in human faces and voices (Morton & Johnson, 1991), and early on, they interact differently with people and with objects (Legerstee, 1991). Soon, they discover the intentionality or "aboutness" – the meaning people attribute to objects; they label them, like or dislike them and comprehend them in other ways. By their first birthday, babies learn about objects by reading their parents' emotional reactions to them (social referencing) and avoid the ones for which they read a negative emotional reaction (Moses, Baldwin, Rosicky, & Tidball, 2001). The biggest sprout in theory of mind abilities occurs between two and five years of age. At the age of two, children predict someone's behaviour based on their emotions and desires (Wellman & Woolley, 1990), and they gradually become better at recognizing and labelling basic emotions based on facial expressions (Widen & Russell, 2003). Understanding desire and emotion precedes understanding of cognitive mental states. At around the age of four or five, children understand how people acquire information and knowledge, and they begin to understand false beliefs. False beliefs pertain to a situation in which we predict people's behaviour based on their beliefs about reality, and not the reality itself (Wimmer & Perner, 1983). Theory of mind development continues throughout school age, when children begin to understand the second-order false beliefs (Perner & Wimmer, 1985), and even later, when people develop their understanding of deception, irony and sarcasm (Dews et al., 1996).

With most research in the theory of mind focusing on young children and their developing abilities, or lack thereof, there is a number of measures developed to assess those abilities in the early years of life (Šakić, Kotrla Topić, & Ljubešić, 2012). On the other hand, for a long time, it was a challenge to measure theory of mind in adults. Most tests include short stories about an event of some kind that ends in an ambiguous way (Happe, 1994). Participants are then asked a control question about physical events and an experimental question about what the character in the story

meant, intended or understood. Also, there are computerized tests of emotion recognition (Tottenham et al., 2009). However, what has proved to be particularly difficult is to develop a test that is easily administered and that can detect subtle deficits in social understanding in adults with typical intelligence (Baron-Cohen, Wheelwright, Hill, Raste, & Plum, 2001). In 1997, Baron-Cohen, Jolliffe, Mortimore, and Robertson issued the first version of "Reading the Mind in the Eyes Test" (RMET). Described as an adult test of social sensitivity, the test consisted of a series of 25 photographs of the eye region of the face. For each photograph, participants had to choose one of the two possible word descriptors of what the person in the photograph was thinking or feeling. The logic behind these tasks was that a person had to know terms for various mental states and what those terms meant. The following step was to connect those terms to emotional states presented in the photographs of the eye region. Although the test proved to be a success in terms of measuring social sensitivity, showing differences in mind-reading ability between healthy participants and participants with high functioning autism and Asperger syndrome, it has certain psychometric limitations (Baron-Cohen et al., 2001). Therefore, in 2001, Baron-Cohen et al. presented a revised version of the RMET which showed better reliability and validity than the first version (Baron-Cohen et al., 2001). The revised version of the RMET (Baron-Cohen et al., 2001) had 36 items, and the number of possible answers was increased from 2 in the original version to 4. It is easily administered and easy to score. Also, it is freely available for everyone to use. It has so far been translated into many languages including French (Prevost et al., 2014), Italian (Vellante et al., 2013), German (Pflatz et al., 2013), Portuguese (Sanvicente-Vieira et al., 2014), Spanish (Fernandez-Abascal et al., 2013), Turkish (Girli, 2014; Yildirim et al., 2011), Japanese (Kunihira, Senju, Dairoku, Wakabayashi, & Hasegawa, 2006), Swedish (Hallerback, Lungnegard, Hjarthag, & Gillberg, 2009), Romanian (Miu, Pana, & Avram, 2012), Persian (Khorashad et al., 2015), etc. and used in numerous studies with both non-clinical and clinical population (see Vellante et al., 2013 for a review). Despite its frequent use, however, there are not many studies reporting data on the psychometric properties of the test. This is evident even in the case of translations and adaptations of the RMET to various languages (e.g. Sanvicente-Vieira et al., 2014). In their review of the psychometric properties of the RMET, Vellante et al. (2013) also highlight that many of the studies do not report any information on the test reliability. There is no clear reason why this information is so often missing. Furthermore, to the best of our knowledge, there are only two studies reporting factorial analyses of the RMET – one proposing a single factor solution (Vellante et al., 2013) and one reasoning against it (Olderback et al., 2015).

The Revised version of the RMET had already been translated to Croatian as well (Barać & Vulić-Prtorić, 2016), but this translation is not publicly available. Aiming to explore the psychometric characteristics of the translated version, the test was administered to 97 female psychology students. It showed low reliability, as well

as low convergent validity, and the authors propose that the ecological validity of the test could be enhanced by more thoughtful translation of the descriptors, as well as by letting the participants use a glossary during testing. Therefore, we decided to go through a new process of translation of the RMET and the accompanying glossary to the Croatian language. The aim of this research was to translate and adapt the revised version of the RMET to the Croatian language and to provide preliminary data on its reliability, factor structure, and convergent validity (through correlation with a self-reported measure of empathy), in a healthy population of Croatian students of both genders.

Methods

Participants

Participants were 146 undergraduate and graduate students from the J.J. Strossmayer University in Osijek, Croatia. There were 84 psychology students (78 female and 6 male) and 62 students of electrical engineering, computer science and information technology (6 female and 56 male). In total, there were 84 female and 62 male participants, all native speakers of Croatian, with a mean age of 21.45 years ($SD = 2.06$; range from 19 to 33, Median = 22). All the participants were volunteers, and after the purpose of the study was explained to them, they signed the informed consent form. The study was approved by the Ethics Committee of the Institute of Social Sciences Ivo Pilar.

Procedure

The testing took place at the University over a course of several days, and it was part of a larger research project on empathy and reading. Participants were tested in small groups of up to 25 people to ensure they had enough peace and privacy. All the participants received a booklet containing a series of questionnaires, including the Croatian version of the RMET and Emotional Empathy Scale. The purpose of the study was explained to them and they were assured that anonymity and confidentiality of the information provided would be protected. All the participants first filled out a short questionnaire prepared for the purpose of this research, providing us with data about their gender, age, study major and year of study. After that, the Emotional Empathy Scale and the Croatian version of the RMET were administered.

The Croatian version of the revised adult RMET was administered using a 38-page booklet. The first page contained instructions, while the second one contained one test item that was used for demonstration. The following 36 pages contained test items. On each page there was one photograph of the eye region surrounded with four mental state descriptors. Participants were instructed to make a choice between

the four descriptors and circle the one they think best describes what the person in the picture is thinking or feeling. Additionally, they were presented with a glossary containing definitions of 79 words (including synonyms where possible) and examples of those words being used in a sentence. They were encouraged to use the glossary whenever they felt it would help them better understand a certain term. The participants were not timed and were instructed to take their time and decide carefully on the correct descriptor.

Measures

Before translating the *Reading the Mind in the Eyes Test* (RMET) to Croatian, we contacted the Autism Research Centre (ARC, Cambridge, United Kingdom) to obtain permission to translate and adapt the test. We then proceeded with the translation. Two researchers with proficiency in both English language and theory of mind research independently translated all the items from the original version of the test into the Croatian language. In the next step, the translations were compared and all the different translation options were discussed. As a result, the researchers constructed a unique version containing all the descriptors for which they both agreed they represent the best semantic and conceptual translation of the original items.

We did not use the back translation method because some of the adjectives in the English version of the test that pertain to complex mental states were difficult to translate to Croatian using just one word (e.g. "aghast"). Therefore, it would be hard to expect that translating back to English would result in the exact same adjective as the original version. We took special care that the translated target words and foils are as similar in meaning to the original version as possible and that we keep the same level of difficulty at the same time.

The descriptors in the final version mostly consisted of one word, an adjective, with the exception of six which consisted of two words, one of which was an adverb (e.g. one of the foils in item 1 was "with boredom"). The use of such phrases was necessary to keep the translated words as close in meaning as possible to their English counterparts, but at the same time to grasp the nuances of the Croatian language. All the descriptors are in neuter grammatical gender.

To diminish vocabulary limitations, the RMET was accompanied by a glossary containing definitions of 79 words, each of which was exemplified with a sentence containing the word. The characters in those sentences have Croatian names.

Participants were given one point for each correctly chosen descriptor and the total score on the test was the total number of correctly identified descriptors with the maximum score being 36.

To access the tendency of emotional reactions to other people's emotional experiences, we used the *Emotional Empathy Scale* (Raboteg-Šarić, 1993). The scale consists of 19 items which describe how people feel as a reaction to emotional states

of others or to disturbing life situations. The participants' task was to estimate the degree to which each statement can be applied to them, using a 5-point scale with a predefined range (1 – *it doesn't describe me well*, 5 – *it describes me very well*).

Statistical Analysis

Data were analyzed using the Statistical Package for Social Sciences (SPSS) for Windows, version 21 (IBM Corp., 2012). All the tests were two-tailed and conducted at the 5% level of statistical significance. Since the distribution of results differed from normal both on Croatian adaptation of the RMET and of the Emotional Empathy Scale, Spearman rho correlation coefficients were used to calculate the correlations between measures and Mann Whitney U-test was used to evaluate gender differences in those variables.

We assessed the internal factor structure of the RMET by confirmatory factor analysis (CFA) using tetrachoric correlation matrices with a mean and variance adjusted weighted least squares (WLSMV) estimation method by Mplus 8.1 (Muthén & Muthén, 1998-2017). Because the data were categorical (e.g. the answers were coded as true or false), the WLSMV estimator was used instead of maximum likelihood. As model fit indices, we used: (a) Sattora-Bentler scaled chi-square ($\chi 2$) (Satorra & Bentler, 2001); (b) the root mean squared error of approximation (RMSEA; Steiger, 2000), where values less than .05 were taken as good fit, and .05-.08 ones as moderate fit; (c) the comparative fit index (CFI) and Tucker-Lewis Index (TLI) where values between .90 and .95 indicated acceptable, and values above .95 indicated good fit (Hu & Bentler, 1999); and (d) standardized root mean square residual (SRMR) smaller than .08 as indicating good fit (Hu & Bentler, 1999).

Results

Reading the Mind in the Eyes Test – Croatian Version

Mean result on the Croatian version of the RMET was 25.75 (*SD* = 3.82). Minimum result was 14 and the maximum result 34. Modal value was 27, with 21 participants achieving this result. Distribution of all the scores is presented in Figure 1. Distribution differs from normal with Skewness = -.517 and Kurtosis = .095. The Shapiro-Wilk test = .969, *df* = 130, *p* = .005 points to the same conclusion regarding distribution.
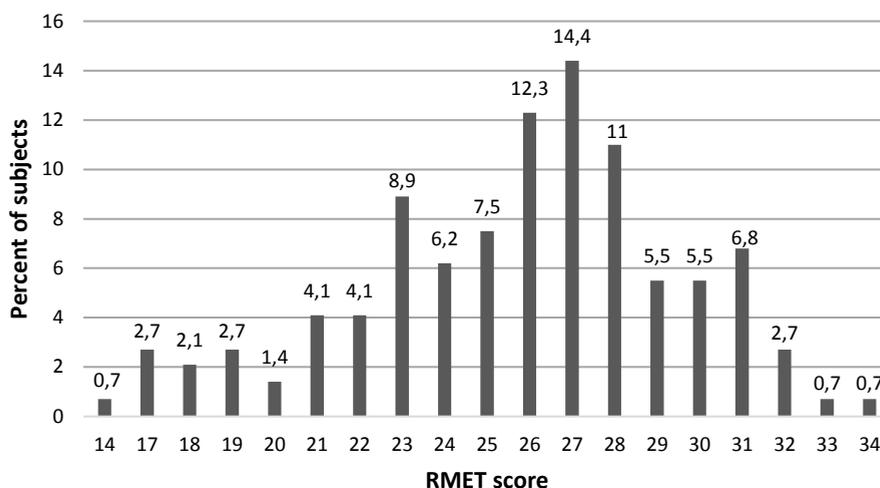
Figure 1. Distribution of total scores on the Croatian version of the RMET.

The percentages of participants who selected one of the four possible descriptors in each item of the Croatian version of the RMET are presented in Table 1. Single item analysis shows that all the items except one were correctly answered by more than 50% of participants. The exception is item 2, which is also the only item that had a higher percentage of the participants choosing a different answer than the correct one (35.6% compared to 33.6%). The item with the highest frequency of correct answers is item 36 with 92.6% correct answers.

Table 1.

Percentages of Participants who Selected One of the Four Possible Descriptors in each Item of the Croatian Version of the RMET (Correct Answers are Indicated in Bold)

Item no.	Answer A	%	Answer B	%	Answer C	%	Answer D	%
1	**Playful**	**51.4**	Comforting	15.1	Arrogant	28.1	Bored	5.5
2	Terrified	35.6	**Upset**	**33.6**	Irritated	7.5	Annoyed	23.3
3	Joking	1.4	Flustered	0.7	**Desire**	**76.0**	Convinced	21.9
4	Joking	0	**Insisting**	**78.6**	Amused	0.7	Relaxed	20.7
5	Irritated	8.9	Sarcastic	17.1	**Worried**	**71.2**	Friendly	2.7
6	Aghast	0.7	**Fantasizing**	**58.9**	Impatient	32.2	Alarmed	8.2
7	Apologetic	6.2	Friendly	24.7	**Uneasy**	**51.4**	Dispirited	17.8
8	**Despondent**	**89.7**	Relieved	5.5	Shy	4.1	Excited	0.7
9	Annoyed	6.8	Hostile	9.6	Horrified	13.0	**Preoccupied**	**70.5**
10	**Cautious**	**62.3**	Insisting	27.4	Bored	7.5	Aghast	2.7
11	Terrified	7.5	Amused	3.4	**Regretful**	**84.2**	Flirtatious	4.8
12	Indifferent	16.4	Embarrassed	1.4	**Skeptical**	**80.8**	Dispirited	1.4

Item no.	Answer A	%	Answer B	%	Answer C	%	Answer D	%
13	Decisive	2.7	**Anticipating**	**77.4**	Threatening	2.1	Shy	17.8
14	Irritated	9.7	Disappointed	6.2	Depressed	8.3	**Accusing**	**75.9**
15	**Contemplative**	**61.0**	Flustered	13.0	Encouraging	12.3	Amused	13.7
16	Irritated	4.8	**Thoughtful**	**63.7**	Encouraging	1.4	Sympathetic	30.1
17	**Doubtful**	**65.1**	Affectionate	20.5	Playful	7.5	Aghast	6.8
18	**Decisive**	**91.1**	Amused	2.1	Aghast	4.1	Bored	2.7
19	Arrogant	15.1	Grateful	15.8	Sarcastic	6.8	**Tentative**	**62.3**
20	Dominant	17.1	**Friendly**	**74.7**	Guilty	8.2	Horrified	0
21	Embarrassed	4.1	**Fantasizing**	**88.4**	Confused	4.8	Panicked	2.7
22	**Preoccupied**	**77.4**	Grateful	1.4	Insisting	3.4	Imploring	17.8
23	Content	2.1	Apologetic	7.5	**Defiant**	**59.6**	Curious	30.8
24	**Pensive**	**82.9**	Irritated	8.2	Excited	0.7	Hostile	8.2
25	Panicked	3.4	Incredulous	14.4	Despondent	13.7	**Interested**	**68.5**
26	Alarmed	4.8	Shy	2.1	**Hostile**	**76.0**	Anxious	17.1
27	Joking	0	**Cautious**	**73.1**	Arrogant	17.9	Reassuring	9.0
28	**Interested**	**76.0**	Joking	1.4	Affectionate	13.7	Contented	8.9
29	Impatient	6.8	Aghast	5.5	Irritated	16.4	**Reflective**	**71.2**
30	Grateful	0.7	**Flirtatious**	**82.9**	Hostile	12.3	Disappointed	4.1
31	Ashamed	7.5	**Confident**	**65.8**	Joking	0.7	Dispirited	26.0
32	**Serious**	**78.1**	Ashamed	3.4	Bewildered	15.1	Alarmed	3.4
33	Embarrassed	4.1	Guilty	27.4	Fantasizing	8.2	**Concerned**	**60.3**
34	Aghast	3.4	Baffled	18.5	**Distrustful**	**66.4**	Terrified	11.6
35	Puzzled	13.0	**Nervous**	**63.0**	Insisting	11.6	Contemplative	12.3
36	Ashamed	1.4	Nervous	0.7	**Suspicious**	**92.5**	Indecisive	5.5

Reliability Analyses for the Croatian Adaptation of the RMET

Internal consistency was measured with Cronbach's Alpha and it was .54 with all 36 items. Reliability was also measured by maximal reliability H for confirmatory factor analysis (CFA) testing the unidimensional model (see the following paragraph). Maximal reliability H coefficient was .74.

Factorial Analysis of the Croatian Adaptation of the RMET

To test the unidimensional model, we ran a confirmatory factor analysis (CFA) using WLSMV estimator, as explained in the paragraph on Statistical analysis. Goodness of fit indices were as follows: $\chi^2 = 690.98$, $df = 528$, $p > .05$, CFI = .506, TLI = .474, RMSEA = 0.040 (0.028-0.050). Since factor loadings for most items were low and the goodness of fit indices also points to a poor model, we further explored the possibility of creating a shortened version of the test. The model was created by successively removing items with low saturation until reaching acceptable goodness of fit indices. What remained were seven items (items 11, 12, 15, 18, 20, 28 and 34) that showed maximal interrelations and shared most of the common

variance. Factor loadings of these items are presented in Table 2. Goodness of fit indices of the short version were as follows: $\chi^2 = 23.06$, $df = 14$, $p > .05$, CFI = .904, TLI = .856, RMSEA = 0.067, probability RMSEA (<=.05) = .258.

Composite reliability coefficient of the short version of the RMET was .79, and internal consistency measured with Cronbach's Alpha was .61. The correlation between the short version and the original version was .75.

Table 2.

Standardized Factor Loadings of Items in the Short Version of the RMET (items 11. 12. 15. 18. 20. 28 and 34)

Item number	Factor loadings	Standard error	p - value
RMET_11	.673	.134	.000
RMET_12	.586	.132	.000
RMET_15	.539	.131	.000
RMET_18	.589	.167	.000
RMET_20	.557	.124	.000
RMET_28	.585	.126	.000
RMWT_34	.618	.130	.000

Emotional Empathy Scale

The mean result of the Emotional Empathy Scale was $M = 24.94$, $SD = 4.77$. Internal consistency was measured with Cronbach's Alpha and it was .88. The values of Shapiro-Wilk test = .950, $df = 130$, $p = .000$ indicates that distribution of scores differed from normal.

Correlations between RMET and Empathy

To analyze the convergent validity of the RMET we calculated the correlations between the RMET and self-reported measure of empathy. Since the distribution of the RMET and emotional empathy scores differed from normal, we proceeded with nonparametric correlation analysis. Spearman rho correlation coefficient between the RMET and emotional empathy was .19, $p = .032$, pointing to a weak but significant positive relation between these variables. The short version of the RMET did not significantly correlate with the self-reported measure of empathy (Spearman rho = .16, $p = .074$).

Gender Differences

To investigate possible gender differences in these measures, we used the Mann Whitney U-test. The results (Table 3) show a significant difference in both the RMET (and its short version) and emotional empathy in favour of female participants.

Table 3.

Results of the Mann Whitney U-Test for Gender Differences in the RMET, Emotional Empathy and the Short Version of RMET

Variable	Gender	*N*	Mean Rank	Sum of Ranks	M W U-test	*p*
RMET	Female	84	88.89	7467.0	1311.000	.000
	Male	62	52.65	3264.0		
Emotional empathy	Female	82	78.89	6469.0	870.000	.000
	Male	48	42.63	2046.0		
Short RMET	Female	84	85.42	7175.0	1603.0	.000
	Male	62	57.35	3556.0		

Discussion

The purpose of this study was to create a Croatian version of the RMET (Baron-Cohen et al., 2001), which would allow comparisons with the results from other countries. Furthermore, it would provide researchers in Croatia with a theory of mind measure for adults without and potentially with specific disorders, such as schizophrenia, autism spectrum disorders, eating disorders, etc. If proven reliable, this measure would be useful in clinical practice, as well as for scientific purposes.

Our results show that the mean result on the Croatian version of the RMET was 25.75 (*SD* = 3.82), which is somewhat lower compared to the student group of participants in the original study (*M* = 28.00, *SD* = 3.50) (Baron-Cohen et al., 2001). Some previous studies in other languages also report similar mean scores (Vellante et al., 2013).

Reliability and Factor Structure of the Croatian Adaptation of the RMET

As for internal consistency of the Croatian version of the RMET, Cronbach's Alpha was .54, which points to rather poor reliability. As mentioned earlier, other studies rarely report reliability coefficients, but those that did, show the values of Cronbach's Alpha to be .37 (Khorashad et al., 2015), .53 (Prevost et al., 2013), .58 (Harkness, Jacobs, Duong, & Sabbagh, 2010), .63 in men and .60 in women (Voracek & Dressler, 2006), .60 (Vellante et al., 2013), .70 (Dehning et al., 2012), and .71 (Girli, 2014). These coefficients generally show poor to acceptable reliability. On the other hand, test-retest reliability often shows better results (e.g. .70 in the study of Prevost et al., 2013 and .83 in the study of Vellante et al., 2013), but they are difficult to compare across studies because of different methods that were used, as well as different time intervals between two testings.

Another way of assessing reliability is through maximal reliability H, which was .74 in our research. This type of reliability is similar to maximal weighted internal consistency reliability obtained in Vellante et al.'s (2013) study (.72). Both

ways of calculating the coefficient of internal consistency are based on the use of data of factor loadings instead of raw scores, which tends to increase estimates since it takes into account the relative importance of items in the questionnaire. The estimated reliability of the Croatian adaptation of the RMET reached the level of consensual threshold of .70, which is assumed to represent an adequate level of reliability, but it did not reach far from this value.

There are several possible reasons that could explain rather low reliability indicators, one of which includes small inter-item correlations. Another possibility is that the RMET, in fact, measures more than one factor. The available research on the factorial structure of this test is limited. One of the relevant studies (Vellante et al., 2013) confirmed a unidimensional model, but the other one (Olderbak et al., 2015) suggests that a single factor solution is not a sufficiently representative fit to the data. Our data marginally support the one-dimensional model, as proposed in theory, but just like in Vellante et al. (2013) study, factor loadings are far from optimal, with 19 items not reaching the minimal acceptance threshold of .25 for factor loadings.

Other possible explanations regarding poor reliability, which need to be further investigated, might have to do with the test itself. For example, pictures in the test are black and white photographs, and some of them are extensively shadowed, which might make them harder to evaluate. In fact, Hallerbäck et al. (2009) found that for one item which contained a rather dark photograph of the eye region, the correct answer was chosen by 35.4% of participants in the study, as opposed to 68% of participants in the pilot study in which the photograph was lightened up. Poor reliability could be related to specific translations of the test to various languages as well. As mentioned before, while the terms for some basic emotions are easily translated to most languages, when it comes to more complex emotional states, the task gets increasingly difficult. It might be that, despite the effort that researchers put into finding the right translation, some items become more difficult because of subtle alternations that happen during this process. In the Croatian version of the RMET, there is only one item at which the foil was chosen more often than the correct answer (35.6% of participants chose answer A, and 33.6% the correct answer B). In all the other items the correct answer was chosen by more than 50% of participants, just as in the Fernandez-Abascal et al. (2013) study. This is a good result compared to other attempts of RMET adaptation to new languages in which the number of such items varies from two (Vellante et al., 2013) to seven (Prevost et al., 2014).

Short Version of the RMET

In an attempt to create a better version of the RMET, we created a short version of the test, containing only seven items, which showed adequate fit indices and better reliability indicators than the original version. The target descriptors in these items were as follows: regretful, sceptical, contemplative, decisive, friendly, interested and

distrustful. All the terms were easily translated to Croatian. Four items contained pictures of female faces and three of male faces. Logical analysis of the target descriptor in these items did not bring any meaningful conclusions except that all these terms refer to more cognitive than emotional mental states. But the fact that the short version contains only seven items whose factor loadings were acceptable for this model, compared to 36 in the original test, again raises questions about what the test really measures. Further research should address this question and try to additionally compare the two tests in specific populations.

Convergent Validity

Convergent validity of the Croatian adaptation of the RMET was evaluated by investigating correlations with the Emotional Empathy Scale, since the previous research often showed such correlations. In fact, correlations between the RMET and empathy range from .23 (Voracek & Dressler, 2006) to .56 (Chapman et al., 2006). There are, however, studies that either found no correlations between RMET scores and empathy (Muller et al., 2010) or such correlations were found only for the participants who scored lower on empathy measure than the cutoff score which best differentiates participants with autism from controls (Vellante et al., 2013). In our research, we found a weak but significant positive correlation between the RMET and empathy (.19, $p = .032$), marginally supporting the convergent validity of the Croatian adaptation of the RMET. The short version of the RMET did not significantly correlate with empathy, possibly because the items in the shorter version refer to more cognitive than emotional mental states, as mentioned previously.

Gender Differences

In our research, females scored higher than males on the RMET (as well as on its short version), which is one of the most replicated findings in the previous studies. Vellante et al. (2010) report that female advantage on this test was found in six out of 17 studies and it was later confirmed in their study as well. On the other hand, there are still many studies that did not find any gender differences. Furthermore, some studies report female advantage in the RMET scores, but only for participants with primary education, while for participants with high school education and university degree there were no significant gender differences found (Yildirim et al., 2011).

One specific of our study is that most female participants were also psychology students, while most of the male participants were computer and engineering students. There is a slight possibility that the study program could have an effect on participants' results on RMET and empathy scale, because psychology is in great deal concerned with the study of human emotions and behaviour, while computer studies are not. Based on our study design, we are unable to say is gender a factor that

affected both the choice of the study program and results on RMET and empathy, or did the study process in this program have an additional effect on the results on RMET and empathy. We do however find the latter possibility highly unlikely because some of the students on both studies were in the first year of undergraduate study, so there was in fact not enough time for their study to produce an effect on their theory of mind and empathy, especially if we consider that these constructs are something that develops from early age.

Cultural Specifics

As for cultural specifics, some studies that undertook the translation and adaptation process noted that the test photographs showing the eye region should be adapted to their culture (Sanvicente-Vieira et al., 2014). In fact, Adams et al. (2010) showed evidence of better same- versus different-culture mental state decoding from the eyes). Since all the photographs in the test depict Caucasian male and female actors, and in Croatia most population is Caucasian, we feel the choice of the test photographs did not present a problem in this case.

Alternative Classifications

The result of the RMET is calculated as the number of correct answers. Yet, some researchers suggest a different classification system (Fertuck et al., 2009; Harkness, Sabbagh, Jacobson, Chowdrey, & Chen, 2005; Yildirim et al., 2011). Yildirim et al. (2011) propose a system in which each answer gets a certain point ranging from one to four. The correct answer is awarded the highest score, and all the other answers receive a score from one to three, depending on how often they are chosen. The authors suggest that such classification could be beneficial for investigating more subtle differences between participants (Yildirim et al., 2011). Harkness et al. (2005) asked student participants to rate each eye set as negative (e.g. "Upset'"), neutral (e.g. "Reflective"), or positive (e.g. "Friendly'"), by using a 7-point scale (1 = *very negative*, 4 = *neutral*, 7 = *very positive*), in order to additionally explore difficulties in social functioning in dysphoric college students. Using the same algorithm for identifying mental state valence sub-scores, Fertuck et al. (2009) showed that patients with borderline personality disorder perform significantly better than the healthy controls in both total score and the neutral emotional valence items score.

Limitations of the Study

All the participants in the study are students, which means they are at least prospected to achieve a high level of education. In other words, the sample is not representative of the Croatian population. This means that in the general population,

participants might experience difficulties in understanding some of the descriptors. The participants in this study were encouraged to use the glossary for all the terms they were not familiar with, but we have no data indicating how many of them did in fact use the glossary and in what instances.

Conclusion

Our data point to low reliability of the Croatian adaptation of the RMET based on the internal consistency coefficient, and to adequate reliability based on maximal reliability H coefficient. Furthermore, the result of the CFA marginally supports a unidimensional model. Convergent validity was confirmed by a significant positive correlation between the RMET and empathy. A short version of the RMET was created showing adequate fit indices, but containing only seven items. Before making any final judgments on the reliability and validity of the Croatian version of the RMET, we propose further development of this instrument in order to improve its internal factor structure. We also suggest a research with a more representative population; and investigating test-retest reliability, as well as discriminant validity of the test. We also propose taking account of education and gender variables. Furthermore, Croatian adaptation of the RMET and its short version are yet to be administered to populations that show deficits in theory of mind abilities, such as patients suffering from autism or schizophrenia, who might serve as the best subjects in assessing the validity of the RMET.

References

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

Barać, B., & Vulić-Prtorić, A. (2016). What does the Eyes Test really examine? Some methodological difficulties in testing theory of mind using the Eyes Test. *Clinical Psychology*, *9*(2), 1-21. doi:10.21465/2016-KP-2-0002

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., Tager-Flusberg, H., & Cohen, D. J. (Eds.). (2000). *Understanding other minds: Perspectives from developmental cognitive neuroscience* (2nd ed.). New York: Oxford University Press.

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 and Allied Disciplines*, *42*, 241-251. doi:10.1111/1469-7610.00715

Beaumont, R., & Newcombe, P. (2006). Theory of mind and central coherence in adults with High-Functioning Autism or Asperger Syndrome. *Autism, 10*, 365-382. doi:10.1177/1362361306064416

Brune, M. (2005). "Theory of mind" in schizophrenia: A review of the literature. *Schizophrenia Bulleting, 31*, 21-42. doi:10.1093/schbul/sbi002

Chapman, E., Baron-Cohen, S., Auyeung, B., Knickmeyer, R., Taylor, K., & Hackett, G. (2006). Fetal testosterone and empathy: Evidence from the empathy quotient (EQ) and the "reading the mind in the eyes" test. *Social Neuroscience, 1*, 135-148. doi:0.1080/17470910600992239

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

Dews, S., Winner, E., Kaplan, J., Rosenblatt, E., Hunt, M., Lim, K., & Smarsh, B. (1996). Children's understanding of the meaning and functions of verbal irony. *Child Development*, *67*(6), 3071-3085. doi:10.1111/j.1467-8624.1996.tb01903.x

Fernández-Abascal, E. G., Cabello, R., Fernández-Berrocal, P., & Baron-Cohen, S. (2013). Test-retest reliability of the "Reading the Mind in the Eyes" test: A one-year follow-up study. *Autism, 4*, 1-6. doi:10.1186/2040-2392-4-33

Fertuck, E. A., Jekal, A., Song, I., Wyman, B., Morris, M. C., Wilson, S. T., ... Stanley, B. (2009). Enhanced "Reading the Mind in the Eyes" in borderline personality disorder compared to healthy controls. *Psychological Medicine, 39,* 1979-1988. doi:10.1017/S003329170900600X

Flavell, J. H. (2004). Theory-of-mind development: Retrospect and prospect. *Palmer Quarterly, ProQuest Education Journals*, *50*(3), 274. doi:10.1353/mpq.2004.0018

Girli, A. (2014). Psychometric properties of the Turkish child and adult form of "Reading the Mind in the Eyes Test". *Psychology, 5*, 1321-1337. doi:10.4236/psych.2014.511143

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

Harkness, K., Sabbagh, M., Jacobson, J., Chowdrey, N., & Chen, T. (2005). Enhanced accuracy of mental state decoding in dysphoric college students. *Cognition and Emotion, 19*(7), 999-1025. doi:10.1080/02699930541000110

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

Happe, F. G. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders, 24*, 129-154. doi:10.1007/BF02172093

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55. doi:10.1080/10705519909540118

Khorashad, B. S., Baron-Cohen, S., Roshan, G. M., Kazemian, M., Khazai, L., Aghili, Z., ... Afkhamizadeh, M. (2015). The "Reading the Mind in the Eyes" test: Investigation of psychometric properties and test-retest reliability of the Persian version. *Journal of Autism and Developmental Disorders, 45*, 2651-2666. doi:10.1007/s10803-015-2427-4

Kunihira, Y., Senju, A., Dairoku, H., Wakabayashi, A., & Hasegawa, T. (2006). 'Autistic' traits in nonautistic Japanese populations: Relationships with personality traits and cognitive ability. *Journal of Autism and Developmental Disorders, 36*, 553-566. doi:10.1007/s10803-006-0094-1

Legerstee, M. (1991). The role of person and object in eliciting early imitation. *Journal of Experimental Child Psychology, 51*(3), 423-433. doi:10.1016/0022-0965(91)90086-8

Miu, A. C., Pana, S. E., & Avram, J. (2012). Emotional face processing in neurotypicals with autistic traits: Implications for the broad autism phenotype. *Psychiatry Research, 198*, 489-494. doi:10.1016/j.psychres.2012.01.024

Morton, J., & Johnson, M. H. (1991). CONSPEC and CONLERN: A two-process theory of infant face recognition. *Psychological Review, 98*(2), 164-181. doi:10.1037/0033-295x.98.2.164

Moses, L. J., Baldwin, D. A., Rosicky, J. G., & Tidball, G. (2001) Evidence for referential understanding in the emotions domain at twelve and eighteen months. *Child Development*, *72*(3), 718-735. doi:10.1111/1467-8624.00311

Muller, F., Simion, A., Reviriego, E., Galera, C., Mazaux, J. M., Barat, M., & Joseph, P. A. (2010). Exploring theory of mind after severe traumatic brain injury. *Cortex, 46*, 1088-1099. doi:10.1016/j.cortex.2009.08.014

Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus user's guide*. Eighth Edition. Los Angeles, CA: Muthén & Muthén. Retrieved from https://www.statmodel.com/download/usersguide/MplusUserGuideVer_8.pdf

Olderbak, S., Wilhelm, O., Olaru, G., Geiger, M., Brenneman, M. W., & Roberts, R. D. (2015). A psychometric analysis of the reading the mind in the eyes test: Toward a brief form for research and applied settings. *Frontiers in Psychology, 6*, 1-14. doi:10.1037/t01093-000

Perner, J., & Wimmer, H. (1985). John thinks that Mary thinks that - Attribution of 2nd-order beliefs by 5-year-old to 10-year-old children. *Journal of Experimental Child Psychology, 39*(3), 437-471. doi:10.1016/0022-0965(85)90051-7

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

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioural and Brain Sciences*, *1*(4), 515-526. doi:10.1017/S0140525X00076512

Prevost, M., Carrier, M., Chowne, G., Zelkowitz, P., Joseph, L., & Gold, I. (2013). The Reading the Mind in the Eyes test: Validation of a French version and exploration of cultural variations in a multi-ethnic city. *Cognitive Neuropsychiatry, 19*(3), 189-204. doi:10.1080/13546805.2013.823859

Prevost, M., Carrier, M. E., Chowne, G., Zelkowitz, P., Joseph, L., & Gold, I. (2014). The Reading the Mind in the Eyes test: Validation of a French version and exploration of cultural variations in multi-ethnicity. *Cognitive Neuropsychiatry, 19*, 189-204. doi:10.1080/13546805.2013.823859

Raboteg-Šarić, Z. (1993). *Empatija, moralno rasuđivanje i različiti oblici prosocijalnog ponašanja.* Zagreb: Odsjek za psihologiju Filozofskog fakulteta u Zagrebu.

Sanvicente-Vieira, B., Kluwe-Schiavon, B., Wearick-Silva, L. E., Piccoli, G. L., Scherer, L., Tonelli, H. A., & Grassi-Oliveira, R. (2014). Revised Reading the Mind in the Eyes Test (RMET) - Brazilian version. *The Revista Brasilieira de Psiquiatria, 36*(1), 60-67. doi:10.1590/1516-4446-2013-1162

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika, 66*(4), 507-514. doi:10.1007/ BF02296192

Steiger, J. H. (2000). Point estimation, hypothesis testing, and interval estimation using the RMSEA: Some comments and a reply to Hayduk and Glaser. *Structural Equation Modeling, 7*(2), 149-162. doi:10.1207/S15328007SEM0702_1

Šakić, M., Kotrla Topić, M., & Ljubešić, M. (2012). Pristupi procjeni teorije uma u dojenačkoj i predškolskoj dobi. *Psihologijske teme, 21*(2), 359-381.

Tottenham, N., Tanaka, J., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., ... Nelson, C. A. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research, 168*, 242-249. doi:10.1016/j.psychres.2008.05.006

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

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

Wellman, H. M., & Woolley, J. D. (1990). From simple desires to ordinary beliefs: The early development of everyday psychology. *Cognition*, *35*(3), 245-275. doi:10.1016/0010-0277(90)90024-E

Widen, S. C., & Russell, J. A. (2003). A closer look at preschoolers' freely produced labels for facial expressions. *Developmental Psychology, 39*(1), 114-128. doi:10.1037/0012-1649.39.1.114

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*(1), 103-128. doi:10.1016/0010-0277(83)90004-5

Yildirim, E. A., Kasar, M., Guduk, M., Ates, E., Kucukparlak, I., & Ozalmete, E. O. (2011). Investigation of the reliability of the "Reading the mind in the eyes test" in a Turkish population. *Turkish Journal of Psychiatry, 22*, 177-186. doi:10.5080/u6500

Hrvatska adaptacija Revidirane verzije testa čitanja misli iz očiju

Sažetak

Cilj je ovog istraživanja prijevod i adaptacija revidirane verzije Testa čitanja misli iz očiju (*Reading the mind in the eyes test*, Baron-Cohen i sur., 2001) na hrvatski jezik kako bi se ispitala njegova pouzdanost, faktorska struktura te konvergentna valjanost u populaciji hrvatskih studenata urednog razvoja. Nakon prijevoda i adaptacije hrvatska je verzija Testa čitanja misli iz očiju primijenjena na uzorku od 146 studenata preddiplomskih i diplomskih studija (84 djevojke i 62 mladića). Osim Testa čitanja misli iz očiju primijenjena je i Skala emocionalne empatije (Raboteg-Šarić, 1993). Rezultati pokazuju nisku pouzdanost tipa unutarnje konzistencije te prihvatljivu pouzdanost mjerenu *H*-koeficijentom maksimalne pouzdanosti. Konfirmatorna faktorska analiza granično potvrđuje jednodimenzionalni model. Konvergentna valjanost granično je potvrđena kroz statistički značajnu pozitivnu povezanost Testa čitanja misli iz očiju i empatije. Naposljetku, kreirana je i kratka verzija Testa čitanja misli iz očiju koja se sastoji od svega sedam čestica zadovoljavajućih saturacija. Pouzdanost tipa unutarnje konzistencije te kompozitna pouzdanost ovoga kratkog testa su zadovoljavajuće. U budućim je istraživanjima potrebno dodatno ispitati psihometrijske značajke hrvatske verzije Testa čitanja misli u očima u općoj reprezentativnoj populaciji. Nadalje, potrebno je ispitati test-retest pouzdanost te diskriminativnu valjanost ovog testa.

Ključne riječi: socijalna kognicija, teorija uma, Test čitanja misli iz očiju, hrvatska adaptacija Testa čitanja misli iz očiju

Appendix 1.

Percentages of Participants who Selected One of the Four Possible Descriptors in each Item of the Croatian Version of the RMET (Correct Answers are Indicated in Bold)

Item no.	Answer A	%	Answer B	%	Answer C	%	Answer D	%
1	**Razigrano**	**51.4**	Utješno	15.1	Razdraženo	28.1	S dosadom	5.5
2	Prestravljeno	35.6	**Uzrujano**	**33.6**	Arogantno	7.5	Zlovoljno	23.3
3	Šaljivo	1.4	Uskomešano	0.7	**Požudno**	**76.0**	Uvjereno	21.9
4	Šaljivo	0	**Inzistirajući**	**78.6**	Zabavljen	0.7	Opušteno	20.7
5	Razdraženo	8.9	Sarkastično	17.1	**Zabrinuto**	**71.2**	Prijateljski	2.7
6	Užasnuto	0.7	**Sanjareći**	**58.9**	Nestrpljiv	32.2	Uzbunjeno	8.2
7	Ispričavajući	6.2	Prijateljski	24.7	**Nelagodno**	**51.4**	Potišteno	17.8
8	**Utučeno**	**89.7**	S olakšanjem	5.5	Sramežljivo	4.1	Uzbuđeno	0.7
9	Zlovoljno	6.8	Neprijateljski	9.6	Zaprepašteno	13.0	**Opterećeno**	**70.5**
10	**Oprezno**	**62.3**	Inzistirajući	27.4	S dosadom	7.5	Užasnuto	2.7
11	Prestravljeno	7.5	Zabavljeno	3.4	**Žaleći**	**84.2**	Flertujući	4.8
12	Ravnodušno	16.4	S neugodom	1.4	**Skeptično**	**80.8**	Potišteno	1.4
13	Odlučno	2.7	**S očekivanjem**	**77.4**	Prijeteći	2.1	Sramežljivo	17.8
14	Razdraženo	9.7	Razočarano	6.2	Depresivno	8.3	**Optužujući**	**75.9**
15	**Kontemplativno**	**61.0**	Uskomešano	13.0	Ohrabrujući	12.3	Zabavljeno	13.7
16	Razdraženo	4.8	**Zaokupljen**	**63.7**	Ohrabrujući	1.4	Suosjećajno	30.1
17	**Sumnjičavo**	**65.1**	Privrženo	20.5	Razigrano	7.5	Užasnuto	6.8
18	**Odlučno**	**91.1**	Zabavljeno	2.1	Užasnuto	4.1	S dosadom	2.7
19	Arogantno	15.1	Zahvalno	15.8	Sarkastično	6.8	**Nesigurno**	**62.3**
20	Dominantno	17.1	**Prijateljski**	**74.7**	Krivo	8.2	Zaprepašteno	0
21	S neugodom	4.1	**Sanjareći**	**88.4**	Zbunjeno	4.8	Uspaničeno	2.7
22	**Opterećeno**	**77.4**	Zahvalno	1.4	Inzistirajući	3.4	Preklinjući	17.8
23	Zadovoljno	2.1	Ispričavajući	7.5	**Prkosno**	**59.6**	Znatiželjno	30.8
24	**Sjetno**	**82.9**	Razdraženo	8.2	Uzbuđeno	0.7	Neprijateljski	8.2
25	Uspaničeno	3.4	U nevjerici	14.4	Utučeno	13.7	**Zainteresirano**	**68.5**
26	Uzbunjeno	4.8	Sramežljivo	2.1	**Neprijateljski**	**76.0**	Anksiozno	17.1
27	Šaljivo	0	**Oprezno**	**73.1**	Arogantno	17.9	Pun povjerenja	9.0
28	**Zainteresirano**	**76.0**	Šaljivo	1.4	Privrženo	13.7	Zadovoljno	8.9
29	Nestrpljivo	6.8	Užasnuto	5.5	Razdraženo	16.4	**Zamišljeno**	**71.2**
30	Zahvalno	0.7	**Flertujući**	**82.9**	Neprijateljski	12.3	Razočarano	4.1
31	Posramljeno	7.5	**Samopouzdano**	**65.8**	Šaljivo	0.7	Potišteno	26.0
32	**Ozbiljno**	**78.1**	Posramljeno	3.4	Izbezumljeno	15.1	Uzbunjeno	3.4
33	S neugodom	4.1	Krivo	27.4	Sanjareći	8.2	**Zabrinuto**	**60.3**
34	Užasnuto	3.4	Pogubljeno	18.5	**Nepovjerljivo**	**66.4**	Prestravljeno	11.6
35	Smeteno	13.0	**Nervozno**	**63.0**	Inzistirajući	11.6	Kontemplativno	12.3
36	Posramljeno	1.4	Nervozno	0.7	**Sumnjičavo**	**92.5**	Neodlučno	5.5

* The Table is the same as Table 1 in Results but the descriptors are written in Croatian