

ORIGINAL PAPER

Translation and validation of the Multifaceted Empathy Test into Croatian in a sample of health professionals

Helena Štrucelj^{1,2}, Ljerka Ostojić^{2,3,4}, Mladenka Tkalčić³

¹Department of Public Health, Faculty of Health Studies, University of Rijeka, Rijeka, Croatia

²Centre for Mind and Behaviour, University of Rijeka, Rijeka, Croatia

³Department of Psychology, Faculty of Humanities and Social Sciences, University of Rijeka, Rijeka, Croatia

⁴Division of Cognitive Sciences, Faculty of Humanities and Social Sciences, University of Rijeka, Rijeka, Croatia

Received October 8, 2025; Accepted August, 10, 2026. Copyright: This is an open access article under the CC BY-NC-4.0 license.

Abstract

Aim: The aim was to translate the Multifaceted Empathy Test (MET-core-2) into Croatian using the back translation method and validate it on a sample of health professionals. **Design:** A cross-sectional correlational study. **Methods:** The sample consisted of 179 health professionals (age range 21–58 years). The Empathy Quotient questionnaire was used to test convergent validity, while the Big Five Inventory and the Dirty Dozen questionnaire were used to test divergent validity. **Results:** Our results suggested similar psychometric properties of the Croatian version of the MET-core-2 and two comparative studies: high reliability of the emotional empathy scales (Cronbach's alpha > 0.950) and low reliability of the cognitive empathy scales (Cronbach's alpha from 0.273 to 0.411). The results for convergent and divergent validity of emotional empathy were in line with predicted patterns based on previous studies using empathy questionnaires, whereas some deviations from expected patterns were obtained for cognitive empathy. These discrepancies can be explained by differences in how empathy is assessed using performance-based tests versus self-report questionnaires. **Conclusion:** Overall, our findings suggest that the Croatian version of the MET-core-2 is suitable for research use among health professionals.

Keywords: back translation method, cognitive empathy, emotional empathy, health professionals, Multifaceted Empathy Test.

Introduction

Empathy – the ability to adapt one's behavior in response to the thoughts and feelings of others (Baron-Cohen & Wheelwright, 2004) – has been shown to play an important role in the mental health and job satisfaction of health professionals and is also associated with improved treatment outcomes and higher patient satisfaction (Delgado Bolton et al., 2022; Derksen et al., 2013; Nembhard et al., 2023). Empathy consists of an emotional and cognitive component (Dziobek et al., 2008). Emotional empathy refers either to experiencing arousal when observing another person's emotions (indirect emotional empathy) or to one's own emotions being triggered when observing another's emotions (direct emotional empathy). Cognitive empathy refers to the understanding of another person's thoughts and feelings (Dziobek et al., 2008; Guadagni et al.,

2014). To illustrate, a nurse may understand that a patient is afraid of being separated from their family while in intensive care (cognitive empathy), feel concern for the patient (emotional empathy), and occasionally comfort them with a short conversation (adapting one's behavior to the thoughts and feelings of others). It is difficult for a single measure to capture all of these components of empathy, such that no gold standard instrument for measuring empathy exists (de Lima & Osorio, 2021).

Self-report scales, including the Empathy Quotient (Baron-Cohen & Wheelwright, 2004), are the most commonly used measures to assess empathy (de Lima & Osorio, 2021). Such questionnaires usually require participants to report their experience of empathy and emotional arousal when another person experiences an emotional state (emotional empathy), as well as to report how well they are able to recognize the emotional state of others (cognitive empathy). In contrast, performance-based empathy tests may be more effective in reducing socially desirable responses associated with self-report

Corresponding author: Helena Štrucelj, Department of Public Health, Faculty of Health Studies, University of Rijeka, Viktora Cara Emina 5, 51000 Rijeka, Croatia; email: helena.strucelj@uniri.hr

measures. These tests require participants to complete tasks in which they are presented with images or videos of individuals in various emotional states, with responses evaluated as correct or incorrect (Dziobek et al., 2008). In addition, even when tests require participants to self-report the intensity of empathy toward another person, they may be more effective than questionnaires in reducing socially desirable responses. Unlike questionnaires, which typically ask participants to consider general situations, these tests ask participants to report the intensity of empathy of specific individuals depicted in the task (Dziobek et al., 2008). A further benefit of empathy tests is that they allow for direct manipulation of the context in which another person's emotions are depicted, thereby enhancing the ecological validity of the measure (Foell et al., 2018). A review of most commonly used measurements of empathy by de Lima and Osorio (2021) mentions the Multifaceted Empathy Test (Dziobek et al., 2008) as an example of a test that assesses both emotional and cognitive empathy. Originally developed in German (Dziobek et al., 2008), the test has been translated into English and Brazilian Portuguese and validated in samples of adults from the general population in the United States (Foell et al., 2018) and young adults in Brazil (Fernandes Vieira de Lima et al., 2023). It is primarily used to differentiate neurotypical from neurodivergent groups (e.g., those on the autism spectrum), but also to measure empathy in neurotypical populations (Dziobek et al., 2008).

A systematic review of studies on empathy in health professionals published between 1971 and 2021 found that 74.5% (350 of 470 analyses) used questionnaires, 0.2% (1 of 470) used functional magnetic resonance imaging, and 25.3% (119 of 470) used other measurement approaches (Nembhard et al., 2023). The authors noted that these alternative methods consisted of game-based tasks in which participants were asked to play the role of a caregiver. Thus, none of the studies included in this systematic review appear to have included an empathy test. In contrast, empathy tests have been used in combination with empathy questionnaires in two studies examining the relationship between sleepiness and sleep deprivation and empathy in emergency medical technicians (Amalean et al., 2021) and in medical students and nurses (Sack et al., 2019), as well as in one study investigating the relationship between sociodemographic and psychological predictors of empathy in nurses (Onofrei et al., 2023). Thus, to date empathy tests appear not to be regularly used to complement questionnaires in the study of empathy of health

professionals, highlighting the need to validate existing tests for wider use with this specific group of professionals for whom an understanding of different aspects of empathy is crucial (Delgado Bolton et al., 2022; Derksen et al., 2013; Nembhard et al., 2023). Consequently, in this study we translated the Multifaceted Empathy Test (MET-core-2; Foell et al., 2018) from English into Croatian and validated the Croatian version in a sample of health professionals.

Given that there exist no known arguments as to why the psychometric properties of the MET-core-2 should differ between the general population and the population of health professionals, we expected that the obtained psychometric parameters would be similar to those obtained by Foell et al. (2018) and Fernandes Vieira de Lima et al. (2023): i.e., higher reliability for emotional empathy and lower reliability for cognitive empathy. To assess convergent validity, performance on the MET-core-2 was correlated with scores on the Empathy Quotient (Baron-Cohen & Wheelwright, 2004). We expected positive correlations between the two measures of emotional empathy, as well as between the two measures of cognitive empathy. In the assessment of divergent validity, we expected relationships between the scores on the MET-core-2 and the scores for the different personality factors of the Big Five (McCrae & Costa, 1987) and the Dark Triad model (Paulhus & Williams, 2002). In terms of the Big Five model, for both emotional and cognitive empathy, we expected a negative correlation with Neuroticism but a positive correlation with the other four factors (Guilera et al., 2019; Martos Martínez et al., 2021; Song & Shi, 2017; for an opposing view, see Schreckenbach et al., 2018). In terms of the Dark Triad model, we expected a negative correlation between emotional empathy and Psychopathy and Machiavellianism but no correlation with Narcissism (Heym et al., 2019; Wertag & Hanzec, 2016). Cognitive empathy was not expected to correlate with any of the Dark Triad traits (Duradoni et al., 2023). Finally, we expected that both women and men would have similar scores (Dehning et al., 2013; Guadagni et al., 2014; Rodríguez-Nieto et al., 2022). Given the expected upper age rank of our sample to be around 55 years, age was expected to be positively correlated with emotional empathy but negatively correlated with cognitive empathy (Beadle et al., 2015; Sommerlad et al., 2021).

Aim

The aim of the present study was to translate the MET-core-2 (Foell et al., 2018) from English into Croatian and to validate the Croatian version for future use with health professionals.

Methods

Design

This was a cross-sectional correlational study. The study is reported in accordance with the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies (von Elm et al., 2008).

Sample

The convenience sample included nurses, physiotherapists, and midwives, all of whom were recruited from part-time health undergraduate and graduate students through the faculty where they studied, regardless of gender, age, work institution, and workplace. Participants were not asked to indicate their profession and their study program. However, from the recruitment process, data are available for the tested participants at the group level: 16.9% of the sample were undergraduate students of nursing, 5.5% were undergraduate students of midwifery, 42.6% were graduate students of nursing, 5.5% were graduate students of midwifery, and 29.5% were graduate students of physiotherapy and clinical nutrition. The participants were tested in 2023. Participation was not compensated.

The sample size of $n = 153$ was calculated a priori using G*Power 3 software. 1.9.6 (Faul et al., 2009) with $\alpha = 0.05$, $\beta = 0.20$ and $r = 0.20$. It was planned to recruit more participants in anticipation of the possibility of incomplete questionnaires. Participants were recruited through their classes, and all students who wished to participate were included, resulting in a total sample size of 183.

The study was approved by the Faculty of Humanities and Social Sciences of the University of Rijeka [640-01/22-01/63, Number 2170-24-02-22-2].

Data collection

First, participants were informed about the study, its ethical approval, the voluntary nature of participation, ways to maintain anonymity, and how to withdraw from the study, after which they signed a written informed consent form. Then, participants completed the questionnaires using the paper-and-pencil method and finally, they

completed the MET-core-2 using a combined computer and paper-and-pencil method. Data collection was conducted in groups of two to four participants while the researcher was present and available to answer any questions. Testing took around 20–30 minutes per participant.

Measures

The MET-core-2 (Foell et al., 2018) was used to measure cognitive and emotional empathy. The stimuli consist of 40 color photographs of people of different ages and of both sexes in different emotional states and situational contexts. Twenty photographs illustrate positive, and 20 photographs illustrate negative emotions. Each photograph is presented twice – once in the cognitive empathy task and once in the emotional empathy task – resulting in a total of 80 items administered to participants. The cognitive and emotional empathy tasks alternate in blocks of ten trials each, with the order of stimuli being randomized within a block. In the cognitive empathy task, participants are presented with a photograph and asked to answer the question “What emotion does this person feel?” by selecting one of four available answers (one correct answer and three distractors). Participants have at their disposal a dictionary explaining all emotion terms used for the possible answers, including synonyms, similar terms, and example sentences in which they are used. The proportion of correct answers is scored separately for positive (CEpos) and negative emotions (CEneg), and for total cognitive empathy (CE). In the emotional empathy task, participants are presented with the photographs and asked to self-assess “How much do you empathize with this person?” on a scale from ‘1 – not at all’ to ‘9 – very much’. Here, the average score is calculated separately for positive (EEpos) and negative emotions (EEneg), as well as for total emotional empathy (EE). To test the predictions regarding the validation of the MET-core-2, we conducted analyses for each of the subscales and total scales.

We used MS Power Point 2016 (Microsoft Corporation, Redmond, WA) for stimulus presentation and a paper-and-pencil answer sheet. Randomized combinations of stimuli were created using E-Prime 2.0 (Psychology Software Tools, Inc., Sharpsburg, PA), and the randomized order of assignment of combinations was created using MS Excel 2016 (Microsoft Corporation, Redmond, WA).

Convergent validity assessment

We used the Croatian translation of the short version (Wertag & Hanzec, 2016) of the Empathy Quotient questionnaire (EQ; Baron-Cohen & Wheelwright,

2004) to measure the general empathy quotient and its three dimensions, namely Cognitive Empathy (11 items), Emotional Reactivity (11 items), and Social Skills (6 items). Each of the 28 items about empathic behaviors are self-rated on a scale from '0 – strongly disagree' to '3 – strongly agree'. In the Croatian sample, a three-factor structure has been confirmed, with Cronbach's alpha values ranging from 0.81 to 0.90 for Cognitive Empathy, Emotional Reactivity, and the total score, and 0.40 for Social Skills (Wertag & Hanzec, 2016). In the present study, we only analyzed the Cognitive Empathy and Emotional Reactivity subscales.

Divergent validity assessment

We used the Croatian version (Kardum et al., 2006) of the Big Five Inventory (BFI; Benet-Martinez & John, 1998) to measure the five-factor personality traits Openness to Experience (10 items), Conscientiousness (9 items), Extraversion (8 items), Agreeableness (9 items) and Neuroticism (8 items). Each of the 44 items concerning personality characteristics are self-rated on a scale from '1 – strongly disagree' to '5 – strongly agree'. The five-factor structure has been confirmed in the Croatian sample, with Cronbach's alpha values ranging from 0.72 to 0.83 for the individual scales (Kardum et al., 2006).

We used the Croatian version (Kardum et al., 2023) of the Dirty Dozen questionnaire (DD; Jonason & Webster, 2010) to measure the Dark Triad traits Psychopathy (4 items), Narcissism (4 items), and Machiavellianism (4 items). Each of the 12 statements about typical reactions and behaviors are self-rated on a scale from '1 – strongly disagree' to '9 – strongly agree'. The three-factor structure has been confirmed in the Croatian sample, with Cronbach's alpha values ranging from 0.63 to 0.88 for the subscales and the overall Dark Triad score (Kardum et al., 2023). In this study, we analyzed each of the three subscales separately.

Translation and adaptation of the MET-core-2

The MET-core-2, including participant instructions, emotion labels, and the dictionary, was translated from English into Croatian using the back-translation method. Authors HŠ and MT independently conducted the forward translation into Croatian and compared the translations. There were no major differences between the two translations. Author LJO conducted the back translation into English. Authors HŠ and MT then compared this back translation with the original and made the final adjustments based on consensus.

Data analysis

Statistical analyses were performed using MS Excel 2016 (Microsoft Corporation, Redmond, WA), Statistica 14.0.0.15 (TIBCO Software Inc.), JASP-0.17.3.0.msi., and Psychometrica (Lenhard & Lenhard, 2022). Descriptive data on the sample and the scales included frequencies and percentages for categorical variables and mean (M) or median for numerical variables, depending on the data distribution normality, which was tested using the Shapiro-Wilk's test. Internal consistency of the scales was assessed using Cronbach's α and the average item-total correlation. For binary variables (CEpos, CEneg, and CE), reliability was calculated using the Kuder-Richardson formula (Zaiontz, 2024). Correlation between measures was examined using Spearman's rank correlations. Item difficulty was additionally calculated for cognitive empathy of the MET-core-2 as the proportion of correct responses for each item. Based on the results of Levene's test and Shapiro-Wilk's test, sex differences in cognitive empathy in the MET-core-2 were tested using the Mann-Whitney U-test and in emotional empathy using the Welch t-test. The significance level was set at $\alpha = 0.05$.

Results

Sample characteristics

After excluding four out of 183 participants (due to incomplete questionnaires) the final sample size was $n = 179$ (15.6% men, median age = 27.5 years, $Q = 5.38$, range 21–52 years; and 84.4% women, median age = 31, $Q = 9$, range 21–58 years).

MET-core-2 psychometric parameters

Cognitive empathy. The mean item difficulty was $p = 0.67$ ($SD = 0.23$), ranging from 0.11 to 0.97. Easiest items were Victorious (0.97) and Fearful (0.93), whereas the most difficult items were Amused (0.11) and Crestfallen (0.12).

The average scores for CEpos, CEneg, and CE were in the medium range. Reliability and the average item-total correlation was low for all three scales. There was a low positive correlation between the two subscales and a high positive correlation between each subscale and CE (Table 1, upper part).

Emotional empathy. The average scores for EEpos, EEneg, and EE were in the high medium range. For all three scales, reliability and the average item-total correlation were high. There was a moderate positive correlation between the two

Table 1 Descriptive data and Spearman correlation for the MET-core-2

	M (SD)	Cronbach's alpha	Average item-total correlation		CEneg ^b	CE ^c
CEpos ^a	0.673 (0.110)	0.339	0.269	CEpos ^a	0.207**	0.764***
CEneg ^b	0.692 (0.110)	0.273	0.251	CEneg ^b	-	0.763
CE ^c	0.660 (0.085)	0.411	0.201	CE ^c	-	-
	Median (Q25–Q75)				EEneg^b	EE^c
EEpos ^d	6.100 (4.925–7.200)	0.956	0.721	EEpos ^a	0.489***	0.847***
EEneg ^d	6.500 (5.125–7.400)	0.966	0.763	EEneg ^b	-	0.850***
EE ^f	6.225 (5.087–6.963)	0.965	0.633	EE ^c	-	-

** $p < 0.01$; *** $p < 0.001$; ^acognitive empathy for positive emotions; ^bcognitive empathy for negative emotions; ^ccognitive empathy; ^demotional empathy for positive emotions; ^eemotional empathy for negative emotions; ^femotional empathy

subscales and a high positive correlation between each subscale and EE (Table 1, lower part).

Convergent and divergent validity

In the assessment of convergent validity for the cognitive empathy scales, only CEneg showed a low positive correlation with Cognitive Empathy in the EQ, while CEpos and CE did not correlate significantly with Cognitive Empathy. In addition, CEneg also showed a low positive correlation with Emotional Reactivity in the EQ.

In the assessment of convergent validity for the emotional empathy scales (EEpos, EEneg, EE), all three showed low to moderate positive correlations with Emotional Reactivity in the EQ. In addition, all three scales also showed low positive correlations with Cognitive Empathy in the EQ (Table 2).

In the assessment of divergent validity regarding the BFI, the three cognitive empathy scales showed

neither the predicted negative correlation with Neuroticism nor a positive correlation with the other four factors (Table 2). Of the predicted positive correlations between the three emotional empathy scales and four BFI factors, only a subset were significant: EEpos and EE were significantly correlated with Openness to Experience and Agreeableness, whereas EEneg was significantly correlated only with Agreeableness. Of the predicted negative correlations between the emotional empathy scales and Neuroticism, these were significant only for EEpos and EE (see Table 2). In the assessment of divergent validity regarding the DD, as predicted, none of the three cognitive empathy scales correlated significantly with the factors of the DD. As predicted, all three emotional empathy scales correlated negatively with Psychopathy and Machiavellianism and did not correlate significantly with Narcissism (Table 2).

Table 2 The Spearman correlation between the validation instruments and the MET-core-2

		CEpos ^d	CEneg ^e	CE ^f	EEpos ^g	EEneg ^h	EE ⁱ
EQ^a	Cognitive Empathy	0.044	0.203**	0.142	0.183*	0.152*	0.160*
	Emotional Reactivity	< 0.001	0.202**	0.134	0.244***	0.375***	0.316***
BFI^b	Openness to Experience	< 0.001	0.122	0.074	0.193*	0.104	0.147*
	Conscientiousness	-0.077	-0.013	-0.047	0.117	0.092	0.084
	Extraversion	0.062	0.036	0.061	0.121	0.020	0.049
	Agreeableness	-0.129	-0.021	-0.095	0.303***	0.265***	0.294***
	Neuroticism	0.034	0.020	0.043	-0.286***	-0.034	-0.150*
DD^c	Psychopathy	0.097	-0.004	0.046	-0.233**	-0.418***	-0.376***
	Machiavellianism	0.123	0.089	0.101	-0.172*	-0.172*	-0.196**
	Narcissism	0.119	0.020	0.077	-0.066	-0.111	-0.107

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; ^aEmpathy Quotient questionnaire; ^bBig Five Inventory; ^cDirty Dozen Questionnaire; ^dcognitive empathy for positive emotions; ^ecognitive empathy for negative emotions; ^fcognitive empathy; ^gemotional empathy for positive emotions; ^hemotional empathy for negative emotions; ⁱemotional empathy

Relationships with age and sex

There were no statistically significant differences between women and men in any of the cognitive empathy scales of the MET-core-2 (CEpos, $U = 2482.0$, $p = 0.140$, $d = 0.220$; CEneg, $U = 2119.5$, $p = 0.984$, $d = 0.003$; CE, $U = 2336.5$,

$p = 0.376$, $d = 0.132$) and in the EEpos subscale of emotional empathy ($t(177) = -1.163$, $p = 0.252$, CI 95% [-1.131, 0.306], $d = -0.241$). Women scored significantly higher than men in EEneg ($t(177) = -3.356$, $p = 0.002$, CI 95% [-1.838, -0.457], $d = -0.662$) and EE ($t(177) = -2.782$, $p = 0.008$, CI 95% [-1.378, -0.218], $d = -0.552$). Age did not

correlate significantly with CEpos ($\rho = -0.088$, $p = 0.243$), CEneg ($\rho = 0.082$, $p = 0.273$), and CE ($\rho = 0.005$, $p = 0.951$) but it did correlate slightly to moderately with EEpos ($\rho = 0.150$, $p = 0.045$), EEneg ($\rho = 0.431$, $p < 0.001$), and EE ($\rho = 0.336$, $p < 0.001$). When sex was partially removed, the correlation between EEpos and age was no longer statistically significant ($\rho = 0.140$, $p = 0.061$), while the other correlations with age remained similar to when sex was not removed (CEpos, $\rho = -0.072$, $p = 0.339$; CEneg, $\rho = 0.084$, $p = 0.267$; CE, $\rho = 0.015$, $p = 0.842$; EEneg, $\rho = 0.410$, $p < 0.001$; EE, $\rho = 0.315$, $p < 0.001$).

Discussion

The first step in this validation study was to obtain the psychometric properties of the Croatian version of the MET-core-2 with health professionals. As expected, item difficulty and reliability of the MET-core-2 obtained in our sample were similar to those of two comparable studies (Fernandes Vieira de Lima et al., 2023; Foell et al., 2018). The average item difficulty in our study was similar to that of Foell et al. (2018). In addition, both our study and that by Fernandes Vieira de Lima et al. (2023) found “amused” and “crestfallen” to be the most difficult items and “victorious” to be among the two easiest items. We found high reliability of the emotional empathy scales and low reliability of the cognitive empathy scales, which aligns with the results of Foell et al. (2018) and Fernandes Vieira de Lima et al. (2023). The low reliability of the cognitive empathy scales may be related to the tasks requiring a certain level of knowledge of emotion vocabulary. Although participants had access to a dictionary of emotions, some emotion terms may have been unfamiliar to them. Foell et al. (2018) reported a positive correlation between performance on the Shipley Test of Vocabulary test and cognitive empathy. In future studies, researchers may wish to systematically examine participants’ use of the emotion dictionary to clarify the extent to which they accurately understand the meaning of the emotion terms provided.

The second step in the validation of the MET-core-2 was to test its convergent validity. Here, our results were only partially in line with the predictions regarding positive correlations between the MET-core-2 scales and the corresponding EQ scales. Both the subscale scores and the overall score of emotional empathy were positively correlated with the Emotional Reactivity scale in the EQ. However, for cognitive empathy, only

the MET-core-2 subscale for negative emotions was positively correlated with the Cognitive Empathy scale of the Empathy Quotient (EQ). This discrepancy could be explained by the fact that emotional empathy is assessed more similarly across the two instruments than cognitive empathy. In both measures, emotional empathy relies on self-report: in the MET-Core-2, participants rate their empathic response toward the person depicted in the photograph (Dziobek et al., 2008), and in the EQ, they rate their general tendency to experience emotional reactions to others’ mental states (Lawrence et al., 2004). In contrast, to measure cognitive empathy, the difference between the MET-core-2 as a test of empathy compared to the EQ as an empathy questionnaire is larger: in the MET-core-2, participants have to identify the correct name of the emotion of the person shown in a photograph (Dziobek et al., 2008), while in the EQ, they self-assess the level of appreciation of others’ affective state (Lawrence et al., 2004). It should be emphasized that all significant correlations in our study between the MET-core-2 and the EQ were low, as were those in previous validation studies of the MET-core-2 where convergent validity was assessed by correlating it with the Interpersonal Reactivity Index questionnaire (Fernandes Vieira de Lima et al., 2023; Foell et al., 2018).

The third step in the validation of the MET-core-2 was to test its divergent validity by correlating it with the BFI factors and the Dark Triad traits. In line with our predictions, the subscales and the total score of emotional empathy positively correlated with Openness to Experience and Agreeableness, and negatively with Neuroticism of the BFI. However, the correlations between emotional empathy with Extraversion and Conscientiousness were not significant. Thus, our results are also only partially in line with those from studies in which empathy was measured with questionnaires (Guilera et al., 2019; Martos Martínez et al., 2021; Song & Shi, 2017). There are two aspects to consider when interpreting these differences. The first is that most questionnaires ask participants to self-assess their emotional empathy in general, while the MET-core-2, whilst also relying on self-assessment of emotional empathy, does this for specific situations and specific people shown in a photograph. The second is that questionnaires ask participants to self-assess their usual reaction, while tests ask them to self-assess their current reaction. Thus, the self-assessment of emotional empathy in the MET-core-2 may be affected by

the identification with the person in the photograph, which is not applicable in questionnaires. Contrary to predictions, the cognitive empathy scales in the MET-core-2 were not correlated with the BFI factors. These results thus differ from those obtained in most previous studies, in particular those by Guilera et al. (2019), Martos Martínez et al. (2021), and Song and Shi (2017), in all of which cognitive empathy was measured with different questionnaires, i.e., the Interpersonal Reactivity Index (Davis, 1980) or the Basic Empathy Scale (Jolliffe & Farrington, 2006). However, our results are similar to those reported in a study by Schreckenbach et al. (2018) in which emotion recognition was measured by the Diagnostic Analysis of Nonverbal Accuracy-2, Adult Facial Expressions (DANVA-2-AF) (Nowicki & Carton, 1993). It can thus be assumed that here again, the divergence in results could be explained by the difference when cognitive empathy is measured by empathy questionnaires as opposed to empathy tests. For example, high neuroticism is related to low self-esteem and seeing the negative in ambiguous situations (Amestoy et al., 2023; Hufer-Thamm & Riemann, 2021). Accordingly, people with high neuroticism may underestimate their general ability to recognize and understand the emotional state of others, which is how empathy is measured in questionnaires. Thus, this may explain the negative correlation between neuroticism and cognitive empathy as measured by questionnaires. On the other hand, when it comes to actually recognizing the emotional state of specific others in specific situations, as is the case in the tasks in the MET-core-2, neuroticism may not have such an influence.

In terms of assessing divergent validity of the MET-core-2 by correlating it with the Dark Triad traits, our results followed the predicted patterns: cognitive empathy did not correlate with any of the Dark Triad traits, while emotional empathy was negatively correlated with Psychopathy and Machiavellianism and did not correlate with Narcissism. Consistent with the hypotheses, Psychopathy is the Dark Triad trait most strongly correlated with emotional empathy. This result is in line with Wertag and Hanzec's (2016) finding that the strongest correlation was that between Emotional Reactivity of the EQ and Psychopathy.

Overall, our results regarding the convergent and divergent validity of the MET-core-2 are in line with findings suggesting that the factor structure of empathy is still ambiguous (de Lima & Osorio, 2021).

The final step of our validation study was to examine the relationship of emotional and cognitive empathy in the MET-core-2 with age and sex. As expected, we found no sex differences for the cognitive empathy scales. However, we found sex differences in the emotional empathy total score, which appears to be due to women performing better than men on the negative emotion subscale of emotional empathy. One possible explanation for this result is the bias towards women in our sample size (84.4% women). However, it is also possible that there are some true sex differences in emotional empathy, although more research with non-biased samples is needed since previous studies show inconsistent results in this regard. For example, Song and Shi (2017) reported female sex being a positive predictor of emotional but not cognitive empathy, while Fernandes Vieira de Lima et al. (2023) reported higher scores for women for both emotional and cognitive empathy in a face-to-face but not in an online application of the MET-core-2.

In terms of the relationship with age, our results were partially in line with the predicted pattern: we found the expected low to moderate positive correlations between age and emotional empathy but did not find the expected significant negative correlation between age and cognitive empathy. Our predictions were based on studies reporting an inverted U-shaped pattern of empathy across the lifespan, with peak levels – subject to some gender differences – occurring around the age of 50 (Sommerlad et al., 2021). However, in the absence of longitudinal studies, it should be noted that the age differences observed in the present study may also reflect cohort effects (Grühn et al., 2008).

Overall, the mean scores of health professionals in our study seem to be similar to those found by Fernandes Vieira de Lima et al. (2023) for adults from the general population and are numerically higher than those reported by Foell et al. (2018) for young adults (18–28 years) from the general population. However, given the complexity of the concept of empathy, and the fact that a direct comparison between health professionals and the general population was not the focus of our study, we would caution against any firm claims regarding empathy levels of health professionals. Some studies report below-average empathic communication in health professionals (Dehning et al., 2013; Moudatsou et al., 2020) and more research is needed to explore different factors that may influence empathy in health professionals.

Finally, our findings illustrated that a combined computer and pen-and-pencil application of the MET-core-2 can be used for research purposes. In addition, having the researcher present was informative since the researcher could observe the metadata of the testing process, such as the time taken, ambiguities in instructions, and the participants' questions.

In combination with previous findings on the psychometric properties of the MET-core-2, our results thus indicate that the MET-core-2 can be used for testing empathy in health professionals and can be widely applied. As such, the MET-core-2 can serve as an important addition to the empathy measures already used to study empathy in this specific group of professionals, where the currently most used methods rely on self-reports in questionnaires (Nembhard et al., 2023). For example, empathy outcomes following medical humanities programs aimed at improving empathy in medical students and health professionals were assessed using a single questionnaire in 12 of 15 studies (Zhang et al., 2023). This suggests that future research would benefit from combining questionnaires with performance-based empathy tests. Integrating self-report measures, empathy tests, and, where applicable, reports from others will be important to capture the multidimensional nature of empathy in health professionals. This, in turn, may improve the design of interventions, education, and support aimed at enhancing mental health and job satisfaction among health professionals (Delgado Bolton et al., 2022; Nembhard et al., 2023; Yilmaz Sezer et al., 2023), as well as promoting patient-centered and team-based healthcare (Ratka, 2018). Our results suggest that the MET-Core-2 may be a suitable instrument for assessing empathy in such contexts.

Limitation of study

The participants in our study were health professionals and as such the results should not be generalized outside this specific population. The participants' professional orientation towards people may have contributed to the observed relationships between their empathy and personality (Guilera et al., 2019). Since participants provided answers using a paper-and-pencil method, there were a number of missing values. Although these were carefully handled before the statistical analyses in this study, fully computerized versions of the test can more easily require participants to answer every question and thus prevent missing data. Since participants provided responses using paper-and-pencil administration, reaction times were

not recorded; therefore, comparisons with studies using fully computerized tasks are not possible for this measure. A final aspect of our study that requires consideration is the unequal percentage of male and female participants. Such a female-biased sample is representative of the population of health professionals but may affect empathy scores (Eurostat, 2021).

Conclusion

The psychometric properties of the Croatian translation of the MET-core-2 can generally be considered good, with the exception of the low reliability of the cognitive empathy scales. These findings are similar to those of previous studies validating the test's translation into English, and Brazilian Portuguese. The convergent and divergent validity of the Croatian translation through examining the correlations with the EQ, the BFI, and the DD are also mostly in line with previous studies and can generally be considered sufficient to good. The Croatian translation of the MET-core-2 can be used in future research on health professionals' empathy to inform interventions and education aimed at enhancing the well-being of health professionals and patient-centered care. However, due to the complexity of the constructs of emotional and cognitive empathy, we recommend combining two or more different measures of empathy in future studies.

Ethical aspects and conflict of interest

Participants provided written informed consent. The study was approved by the Ethics Committee for Scientific Research of the Faculty of Humanities and Social Sciences of the University of Rijeka [640-01/22-01/63, Number 2170-24-02-22-2]. The authors declare no conflicts of interest.

Acknowledgements

The authors would like to thank all participants, as well as Petra Anić, Dražen Domijan, Asmir Gračanin, Igor Kardum, Nada Krapić, Sanda Pletikosić Tončić, and Nataša Šimić for their valuable input and suggestions.

Funding

This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

Author contributions

Conception and design (HŠ, MT), data analysis and interpretation (HŠ, LJO, MT), translation of the MET-core-2 procedure (HŠ, LJO, MT), manuscript draft (HŠ), critical revision of the manuscript (HŠ, LJO, MT), final approval of the manuscript (HŠ, LJO, MT).

References

- Amalean, A., Pavuluri, H., Schmidt, W. M., & Fowler, L. A. (2021). The effects of sleepiness, time of day, and 12-h shifts on affective and cognitive empathy in emergency medical technicians. *Sleep and Biological Rhythms*, 19, 105–107. <https://doi.org/10.1007/s41105-020-00280-y>
- Amestoy, M. E., D'Amico, D., & Fiocco, A. J. (2023). Neuroticism and stress in older adults: The buffering role of self-esteem. *International Journal of Environmental Research and Public Health*, 20(12), 6102. <https://doi.org/10.3390/ijerph20126102>
- 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(2), 163–175. <https://doi.org/10.1023/b:jadd.0000022607.19833.00>
- Beadle, J. N., Sheehan, A. H., Dahlben, B., & Gutchess, A. H. (2015). Aging, empathy, and prosociality. *The Journals of Gerontology*, 70(2), 213–222. <https://doi.org/10.1093/geronb/gbt091>
- Benet-Martínez, V., & John, O. P. (1998). Los Cinco Grandes across cultures and ethnic groups: multitrait multimethod analyses of the Big Five in Spanish and English. *Journal of Personality and Social Psychology*, 75(3), 729–750. <https://doi.org/10.1037/0022-3514.75.3.729>
- Davis, M. H. (1980). *A multidimensional approach to individual differences in empathy*. JSAS Catalog of Selected Documents in Psychology, 10, 85.
- Dehning, S., Gasperi, S., Krause, D., Meyer, S., Reiß, E., Burger, M., Jacobs, F., Buchheim, A., Müller, N. & Siebeck, M. (2013). Emotional and cognitive empathy in first-year medical students. *International Scholarly Research Notices*, 2013(1), 801530. <https://doi.org/10.1155/2013/801530>
- Delgado Bolton, R. C., San-Martín, M., & Vivanco, L. (2022). Role of empathy and lifelong learning abilities in physicians and nurses who work in direct contact with patients in adverse working conditions. *International Journal of Environmental Research and Public Health*, 19(5), 3012. <https://doi.org/10.3390/ijerph19053012>
- de Lima, F. F., & Osório, F. L. (2021). Empathy: assessment instruments and psychometric quality – a systematic literature review with a meta-analysis of the past ten years. *Frontiers in Psychology*, 12, 781346. <https://doi.org/10.3389/fpsyg.2021.781346>
- Derksen, F., Bensing, J., & Lagro-Janssen, A. (2013). Effectiveness of empathy in general practice: a systematic review. *The British Journal of General Practice*, 63(606), e76–e84. <https://doi.org/10.3399/bjgp13X660814>
- Duradoni, M., Gursesli, M. C., Fiorenza, M., Donati, A., & Guazzini, A. (2023). Cognitive empathy and the Dark Triad: a literature review. *European Journal of Investigation in Health, Psychology and Education*, 13(11), 2642–2680. <https://doi.org/10.3390/ejihpe13110184>
- Dziobek, I., Rogers, K., Fleck, S., Banhemann, M., Heekeren, H. R., Wolf, O. T. & Convit, A. (2008). Dissociation of cognitive and emotional empathy in adults with Asperger syndrome using the Multifaceted Empathy Test (MET). *Journal of Autism and Developmental Disorders*, 38(3), 464–473. <https://doi.org/10.1007/s10803-007-0486-x>
- Eurostat. (2021, March 8). Majority of health jobs held by women. <https://ec.europa.eu/eurostat/web/products-eurostat-news/-/EDN-20210308-1>
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41, 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>
- Fernandes Vieira de Lima, F., Rossi, G., Dos Santos, R. G., & de Lima Osório, F. (2023). Multifaceted Empathy Test (MET): validity evidence for the Brazilian population concerning the computer-based (face-to-face) and online versions. *PLoS ONE*, 18(7), e0284524. <https://doi.org/10.1371/journal.pone.0284524>
- Foell, J., Brislin, S. J., Drislane, L. E., Dziobek, I., & Patrick, C. J. (2018). Creation and validation of an English-language version of the Multifaceted Empathy Test (MET). *Journal of Psychopathology and Behavioral Assessment*, 40(3), 431–439. <https://doi.org/10.1007/s10862-018-9664-8>
- Grühn, D., Rebucal, K., Diehl, M., Lumley, M., & Labouvie-Vief, G. (2008). Empathy across the adult lifespan: longitudinal and experience-sampling findings. *Emotion*, 8(6), 753–765. <https://doi.org/10.1037/a0014123>
- Guadagni, V., Burles, F., Ferrara, M. & Iaria, G. (2014). The effects of sleep deprivation on emotional empathy. *Journal of Sleep Research*, 23(6), 657–663. <https://doi.org/10.1111/jsr.12192>
- Guilera, T., Batalla, I., Forné, C., & Soler-González, J. (2019). Empathy and big five personality model in medical students and its relationship to gender and specialty preference: a cross-sectional study. *BMC Medical Education*, 19(1), 57. <https://doi.org/10.1186/s12909-019-1485-2>
- Heym, N., Firth, J., Kibowski, F., Sumich, A., Egan, V., & Bloxson, C. A. J. (2019). Empathy at the heart of darkness: empathy deficits that bind the Dark Triad and those that mediate indirect relational aggression. *Frontiers in Psychiatry*, 10, 95. <https://doi.org/10.3389/fpsyg.2019.00095>
- Hufer-Thamm, A., & Riemann, R. (2021). On the link of self-esteem, life satisfaction, and Neuroticism. *Journal of Personality*, 89(5), 998–1011. <https://doi.org/10.1111/jopy.12632>
- Jolliffe, D., & Farrington, D. P. (2006). Development and validation of the Basic Empathy Scale. *Journal of Adolescence*, 29(4), 589–611. <https://doi.org/10.1016/j.adolescence.2005.08.010>
- Jonason, P. K., & Webster, G. D. (2010). The dirty dozen: a concise measure of the dark triad. *Psychological Assessment*, 22(2), 420–432. <https://doi.org/10.1037/a0019265>
- Kardum, I., Hudek-Knezevic, J., Marijanović, K., & Shackelford, T. K. (2023). Predicting mate poaching experiences from personality traits using a dyadic analysis. *Journal of Sex Research*, 60(3), 384–398. <https://doi.org/10.1080/00224499.2022.2092586>

- Kardum, I., Gračanin, A., & Hudek-Knežević, J. (2006). Odnos crta ličnosti i stilova privrženosti s različitim aspektima seksualnosti kod žena i muškaraca [Relations of personality traits and attachment styles with different aspects of sexuality in men and women]. *Psihologijske teme*, 15(1), 101–128.
- 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(5), 911–920. <https://doi.org/10.1017/s0033291703001624>
- Lenhard, W., & Lenhard, A. (2022). *Computation of effect sizes*. https://www.psychometrica.de/effect_size.html. Psychometrica. DOI: 10.13140/RG.2.2.17823.92329
- Martos Martínez, Á., Pérez-Fuentes, M. D. C., Molero Jurado, M. D. M., Simón Márquez, M. D. M., Barragán Martín, A. B., & Gázquez Linares, J. J. (2021). Empathy, affect and personality as predictors of engagement in nursing professionals. *International Journal of Environmental Research and Public Health*, 18(8), 4110. <https://doi.org/10.3390/ijerph18084110>
- McCrae, R. R., & Costa, P. T., Jr. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, 52(1), 81–90. <https://doi.org/10.1037//0022-3514.52.1.81>
- Moudatsou, M., Stavropoulou, A., Philalithis, A., & Koukoulis, S. (2020). The role of empathy in health and social care professionals. *Healthcare*, 8(1), 26. <https://doi.org/10.3390/healthcare8010026>
- Nembhard, I. M., David, G., Ezzeddine, I., Betts, D., & Radin, J. (2023). A systematic review of research on empathy in health care. *Health Services Research*, 58(2), 250–263. <https://doi.org/10.1111/1475-6773.14016>
- Nowicki, S., Jr., & Carton, J. (1993). The measurement of emotional intensity from facial expressions. *The Journal of Social Psychology*, 133(5), 749–750. <https://doi.org/10.1080/00224545.1993.9713934>
- Onofrei, L., Serban, C. L., Chirita-Emandi, A., Jeleriu, R. M., & Puiu, M. (2023). The impact of theory of mind, stress and professional experience on empathy in Romanian community nurses – a cross-sectional study. *BMC Nursing*, 22(1), 400. <https://doi.org/10.1186/s12912-023-01569-2>
- Paulhus, D. L., & Williams, K. M. (2002). The Dark Triad of personality: narcissism, machiavellianism, and psychopathy. *Journal of Research in Personality*, 36(6), 556–563. [https://doi.org/10.1016/S0092-6566\(02\)00505-6](https://doi.org/10.1016/S0092-6566(02)00505-6)
- Ratka A. (2018). Empathy and the development of affective skills. *American Journal of Pharmaceutical Education*, 82(10), 7192. <https://doi.org/10.5688/ajpe7192>
- Rodríguez-Nieto, G., Mercadillo, R. E., Pasaye, E. H., & Barrios, F. A. (2022). Affective and cognitive brain-networks are differently integrated in women and men while experiencing compassion. *Frontiers in Psychology*, 13, 992935. <https://doi.org/10.3389/fpsyg.2022.992935>
- Sack, B., Broer, K., & Anders, S. (2019). Sleep deprivation selectively enhances interpersonal emotion recognition from dynamic facial expressions at long viewing times: an observational study. *Neuroscience Letters*, 694, 225–230. <https://doi.org/10.1016/j.neulet.2018.10.035>
- Schreckenbach, T., Ochsendorf, F., Sterz, J., Rüsseler, M., Bechstein, W. O., Bender, B., & Bechtoldt, M. N. (2018). Emotion recognition and extraversion of medical students interact to predict their empathic communication perceived by simulated patients. *BMC Medical Education*, 18(1), 237. <https://doi.org/10.1186/s12909-018-1342-8>
- Sommerlad, A., Huntley, J., Livingston, G., Rankin, K. P., & Fancourt, D. (2021). Empathy and its associations with age and sociodemographic characteristics in a large UK population sample. *PLoS ONE*, 16(9), e0257557. <https://doi.org/10.1371/journal.pone.0257557>
- Song, Y., & Shi, M. (2017). Associations between empathy and big five personality traits among Chinese undergraduate medical students. *PLoS ONE*, 12(2), e0171665. <https://doi.org/10.1371/journal.pone.0171665>
- von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C., Vandenbroucke, J. P., & STROBE Initiative. (2008). Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Journal of Clinical Epidemiology*, 61(4), 344–349. <https://doi.org/10.1016/j.jclinepi.2007.11.008>
- Wertag, A., & Hanzec, I. (2016). Factor structure and psychometric properties of the Croatian short version of The Empathy Quotient. *Suvremena psihologija*, 19(1), 101–110. <https://doi.org/10.21465/2016-SP-191-09>
- Yilmaz Sezer, N., Aker, M. N., Öz, B., & Eren, D. (2023). Midwifery students' traumatic childbirth experiences, traumatic childbirth perceptions, and levels of fear of childbirth. *Journal of Midwifery & Women's Health*, 68(2), 248–254. <https://doi.org/10.1111/jmwh.13464>
- Zaiontz, C. (2024). *Cronbach's alpha basic concepts*. Real Statistics Using Excel. <https://real-statistics.com/reliability/internal-consistency-reliability/kuder-richardson-formula-20/>
- Zhang, X., Pang, H. F., & Duan, Z. (2023). Educational efficacy of medical humanities in empathy of medical students and healthcare professionals: a systematic review and meta-analysis. *BMC Medical Education*, 23(1), 925. <https://doi.org/10.1186/s12909-023-04932-8>