

**Evaluating the Discriminant Validity of the Metaphors Test**  
**Daniel N. Erosa, Matthew Roe, and Kimberly A. Barchard**



University of Nevada, Las Vegas

**Reference:** Erosa, D. N., Roe, M., & Barchard, K. A. (2014, April). Evaluating the discriminant validity of the Metaphors Test. Poster presented at the Western Psychological Association annual convention in Portland, OR.

**Contact Information:** Kimberly A. Barchard, Department of Psychology, University of Nevada, Las Vegas, 4505 S. Maryland Parkway, P.O. Box 455030, Las Vegas, NV, 89154-5030, USA, barchard@unlv.nevada.edu

**Abstract**

The Metaphors Test (Barchard, Hensley, Anderson, & Walker, 2013) is a new test of emotion perception in which test takers indicate the extent to which various emotions are conveyed by metaphors. In order for the Metaphors Test to be considered a valid test of emotion perception, it must have discriminant validity. Discriminant validity occurs when a test has small or zero correlations with tests of unrelated constructs. The Big Five-Factor Model (McCrae & John, 1992) is one of the most well-known frameworks for personality. It contains five dimensions: Openness, Conscientiousness, Extroversion, Agreeableness, and Neuroticism. Based upon theory and previous research, the Metaphors Test should have low correlations with the Big Five. In this study, the scores from the Metaphors Test were correlated with scales from the Five-Factor Test (Goldberg, 1992). A total of 181 adults were recruited through mTurk to complete this study online. Conscientiousness had a moderate positive correlation with the Metaphors Test, suggesting that some participants were not taking this study seriously. Future research should take steps to ensure that all of the data that are analyzed are based upon participants who attended to the study materials – for example, by removing data from respondents that did not take sufficient time or by using a pre-screened participant panel. Agreeableness had a moderate-to-high correlation with the Metaphors Test. This might suggest that proportion consensus scoring (of any attribute) is influenced by the tendency to care about what other people think. Future research on emotion perception might benefit from focusing on tests with veridical scoring keys, such as the new Measure of Emotional Connotations (Barchard, Kirsch, Anderson, Grob, & Anderson, 2012).

**Introduction**

The ability to perceive emotion in written materials is becoming more and more important as society develops new forms of communication. Technology and social media are two examples of these changes. Without the normal social cues that generally tell us how someone is feeling, such as body language and tone of voice, we must rely on information generated through written language. Verbal tests of emotion perception measure the ability to successfully decipher the emotional connotations of written language.

The Metaphors Test (Barchard et al., 2013) is a new test of emotion perception. The test includes ten metaphors with three emotions for each. Participants are asked to rate the extent to which the metaphor conveys each of the emotions. The Metaphors Test is unique in that the item stems do not include any explicit emotion words. For example, the Metaphors Test contains the item stem, "His face is like a ray of sunshine", which does not include any explicit emotional words (e.g., happy, sad) but still conveys emotions.

Other measures of this skill exist. The first of these is the Emotional Accuracy Research Scale (EARS; Mayer & Geher, 1996). The items on EARS are descriptions of specific situations experienced by eight individuals. Respondents are asked to estimate the emotional state (feelings) of the eight individuals. The thoughts in regards to the situation were then placed in a scale and were followed by twelve pairs of dichotomous mood items. Respondents choose between pairs that the respondent felt more strongly towards after reading the thought sample. The second test is the Stories task, which is one part of the Multifactor Emotional Intelligence Scale (Mayer et al., 2000). Like the EARS, the stimuli consist of descriptions that were provided by real people. However, on the Stories task, respondents provide ratings on seven emotion scales.

Many of the items on the EARS and the Stories Test contain explicit emotion words (e.g., jealous). Thus, these tests measure both denotative and connotative knowledge. Denotative knowledge is knowledge of dictionary meanings (e.g., “sad” is similar in meaning to “unhappy”). Connotative knowledge is knowledge of meanings that are not in a dictionary (e.g., “home” has more positive connotations than “house”). Thus, the EARS and the Stories Test do not provide pure measures of the ability to perceive the emotional *connotations* of written language. The advantage of the Metaphors Test over these existing measures is that the item stems do not include any explicit emotion words.

However, little research on the validity of the Metaphors Test has been done at this point. Therefore, the purpose of this study is to contribute to our knowledge of the Metaphors Test by examining its discriminant validity. The ability to perceive emotions is one aspect of emotional intelligence. To demonstrate that emotional intelligence is a new and useful construct, it is important that it is distinct from well-known constructs such as the Big Five personality traits (Joseph & Newman, 2010). The Big Five traits are Openness, Conscientiousness, Extroversion, Agreeableness, and Neuroticism (McCrae & John, 1992). The Metaphors Test is expected to have low to moderate correlations with each of these dimensions, because it should be measuring different constructs.

Barchard et al. (2013) examined the relationship between the Metaphors Test and the Big Five personality dimensions using a sample of 353 university students. Only the correlations for openness and agreeableness were statistically significant, and both of these were small. The purpose of the current study is to replicate those results using a non-student sample.

## **Methodology**

### **Participants**

A total of 181 individuals were recruited through Amazon’s Mechanical Turk (mTurk) system. mTurk is a crowdsourcing website that allows requesters (such researchers), to create Human Intelligence Tasks (HITs). HITs can be used to promote and advertise studies posted on mTurk. Workers are individuals that complete the HITs. In a research study, the workers are the participants. In general, mTurk HITS provide low monetary compensation. Participants in this study were paid 10 cents.

Participants ranged in age from 20 to 68 (mean 31.05, standard deviation 10.83). Of these, 44.8% were female. Participants lived in the following countries: 86.2% India, 9.9% United States, .6% Russia, and 3.5% other. Participants had a variety of first languages: 29.3% Tamil, 27.1% English, 22.7% Malayalam, 8.3% Hindi, and 12.7% other. Ethnically, participants identified themselves as follows: 78.8% Asian, 11.7% White, 4.5% Indian, 3.4% American Indian or Alaska Native, and 2.8% other.

## Measures

### *The Metaphors Test*

The Metaphors Test (Barchard et al., 2011) was designed to measure the ability to perceive emotion in written language. The test contains ten metaphors, with three emotions each, for a total of 30 items. Participants are instructed to indicate the extent to which each metaphor conveys the given emotions. An example item is given in Figure 1. The test is scored using proportion consensus scoring.

	Not at all	A little	Moderately	A lot	Extremely
His face is a ray of sunshine.					
uplifted	1	2	3	4	5
embarrassed	1	2	3	4	5
admiration	1	2	3	4	5

Figure 1. *Example Item from the Metaphors Test*

### *International Personality Item Pool*

The International Personality Item Pool (IPIP) was developed as a measure of personality that could be accessed at no cost (Ehrhart, Roesch, Ehrhart, & Kilian, 2008). The IPIP Big Five Personality Test (Goldberg, 1992) is a 50-item questionnaire in which the participants are asked how accurately each item describes them right now. The test contains ten items for each of the Big Five traits. Participants respond to each item using a five-point scale (1 = Very Inaccurate, 5 = Very Accurate).

## Results

The Metaphors Test correlated significantly with four of the five scales of the Five-Factor Test. The only exception was extroversion. The results are shown in Table 1. Most of these correlations are small to moderate. However, the correlation with agreeableness ( $r(179) = .58$ ,  $p < .01$ ) might be considered large.

## Conclusion

The original study on the Metaphors Test (Barchard et al., 2013) showed that it had small correlations with each of the Big Five personality traits. Only two dimensions had significant correlations and both of these correlations were small. They were openness ( $r(351) = .26$ ,  $p < .001$ ) and agreeableness ( $r(351) = .24$ ,  $p < .05$ ).

Two of the five correlations were similar in the current study: the correlation for extraversion was still very small and the correlation for openness was still small but significant. However, the other correlations were larger in this study than in the previous one. Restriction of range in the student sample might account for the difference in the size of the correlations. This study found a moderate correlation between the Metaphors Test and conscientiousness. This correlation might be due to the sample that was used in the current study. Participants for this study were recruited through mTurk. These participants are paid a small amount in return for completing the study. It could be that some participants did not attend carefully to the study materials. Future research should ensure that all participants attend to the study materials.

This study found a much higher correlation with agreeableness ( $r(179) = .58$ ) than the

Table 1

### *Correlations between the Big Five and the Metaphors Test*

Personality Dimension	Correlation
Neuroticism	.30**
Agreeableness	.58**
Extraversion	-.01
Conscientiousness	.46**
Openness	.33**

\*\*  $p < .001$ .

original study. Perhaps the correlation is due to the way the Metaphors Test is scored. People who obtain high proportion consensus scores are ones who are sensitive to the people around them and understand how other people think. It therefore makes sense that the Metaphors Test would have a moderate-to-high correlation with agreeableness.

A new test has been designed to attempt to solve the problems found on the Metaphors Test: the Measure of Emotional Connotations (MEC; Barchard, Kirsch, Anderson, Grob, & Anderson, 2012). On this test, the stimuli were carefully designed so that the correct answers were known for each item, regardless of the responses of the norm group. A study of the relationship of the MEC to personality (Hensley, Craun, Grob, & Barchard, 2012) found significant but small correlations with openness ( $r(104) = .28, p < .01$ ) and agreeableness ( $r(104) = .28, p < .01$ ).

### References

- Barchard, K. A., Anderson, E. D., Hensley, S., & Walker, H. E. (2011, May). *The Metaphors Test: Using verbal stimuli to measure emotion perception*. Poster presented at the Association for Psychological Science Annual Convention, Washington, DC.
- Barchard, K. A., Hensley, S., & Anderson, E. D. (2013). When proportion consensus scoring works. *Personality and Individual Differences, 55*, 14-18. doi:10.1016/j.paid.2013.01.017
- Barchard, K. A., Hensley, S., Anderson, E. D., & Walker, H. E. (2013). Measuring the ability to perceive the emotional connotations of written language. *Journal of Personality Assessment, 95*, 332-342.
- Barchard, K. A., Kirsch, P. M., Anderson, E., Grob, K., & Anderson, S. (2012). *Measure of Emotional Connotations*. Unpublished psychological test. Available from Kimberly A. Barchard, University of Nevada, Las Vegas, barchard@unlv.nevada.edu
- Ehrhart, K., Roesch, S. C., Ehrhart, M. G., & Kilian, B. (2008). A test of the factor structure equivalence of the 50-Item IPIP five-factor model measure across gender and ethnic groups. *Journal of Personality Assessment, 90*, 507-516. doi:10.1080/00223890802248869
- Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment, 4*, 26-42.
- Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality, 40*, 84-96. <http://dx.doi.org/10.1016/j.jrp.2005.08.007>.
- Hensley, S., Craun, E. A., Grob, K. E., & Barchard, K. A. (2012, May). "I'm Drowning Here!" *Personality traits and connotations in metaphors*. Poster presented at the Association for Psychological Science annual convention, Chicago, IL.
- Joseph, D. L., & Newman, D. A. (2010). Discriminant validity of self-reported emotional intelligence: A multitrait-multisource study. *Educational and Psychological Measurement, 70*, 672-694.
- Mayer, J. D., & Geher, G. (1996). Emotional intelligence and the identification of emotion. *Intelligence, 22*, 89-113. doi:10.1016/S0160-2896(96)90011-2
- Mayer, J. D., Caruso, D. R., & Salovey, P. (2000). Emotional intelligence meets traditional standards for an intelligence. *Intelligence, 27*, 267-298. doi:10.1016/S0160-2896(99)00016-1
- McCrae, R. R., & Costa, P. r. (1997). Personality trait structure as a human universal. *American Psychologist, 52*, 509-516. doi:10.1037/0003-066X.52.5.509
- McCrae, R., & John, O. (1992). An introduction to the five-factor model and its applications. *Journal of Personality, 60*, 175- 215. doi: 10.1111/j.1467-6494.1992.tb00970.x
- Phillips, M. L., Drevets, W. C., Rauch, S. L., & Lane, R. (2003). Neurobiology of emotion perception i: the neural basis of normal emotion perception. *Biological Psychiatry, 54*, 504 - 514. doi: 10.1016/S0006-3223(03)00168-9
- Reitsma, J., Scheepers, P., & Janssen, J. (2007). Convergent and discriminant validity of religiosity measures among church members and non-members. *Personality and Individual Differences, 42*, 1415 - 1426. doi: <http://dx.doi.org/10.1016/j.paid.2006.10.018>