

Multiple solutions test Part II: Evidence on construct and predictive validity

Marko Živanović¹, Jovana Bjekić², and Goran Opačić¹

¹*Department of Psychology, Faculty of Philosophy, University of Belgrade, Serbia*

²*Institute for Medical Research, University of Belgrade, Serbia*

In this study, we complement data on Multiple solutions test (MST) by examining its construct and predictive validity. Unlike conventional matrices where a single solution is required, MST sets three types of problems before the participants, requiring them to solve matrices for *the best*, *the second-best*, and *the least accurate solution*. A total sample of 235 individuals (age $M = 22.65$, $SD = 3.33$, 199 females) participated in the study. Construct validity of each task within MST was tested in relation to the KOG9 battery of intellectual abilities ($N = 156$), while predictive value of individual tasks and full-scale performance was tested in relation to scholastic achievement measured by GPA ($N = 235$). The results have shown high between-task correlations, but also pointed to the specificities of each of them. Additionally, differential difficulties among the tasks were found with *the least accurate* task being the most difficult, followed by *the second-best*, and *the best* one. The test showed satisfactory convergent validity in relation to Gf/Gv test markers within KOG9 battery. Furthermore, MST has shown predictive validity, along with the incremental value of alternative tasks above the standard one (when the scholastic achievement was taken as a criterion), as well as incremental validity in predicting GPA above the KOG9 battery. In general, MST has shown to be a valid instrument for an intelligence assessment, and its alternative tasks have a potential to be a useful addition to standard matrices with one type of solution.

Key words: Multiple solutions test (MST), figural reasoning, fluid intelligence (Gf), construct validity, predictive validity

Highlights:

- Multiple solutions test (MST) demonstrated satisfactory convergent validity in relation to Gf/Gv test markers within KOG9 battery.
- Alternative tasks within MST showed an incremental value above the standard task in the prediction of GPA.
- MST demonstrated incremental value in prediction of GPA above KOG9 battery of intelligence tests.

Corresponding author: marko.zivanovic@f.bg.ac.rs

Acknowledgement. The study was supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia (project number 179018); The second author was supported by project number 175012.

Conventional, highly *G*-saturated tests, i.e., intelligence tests that are the most central proxies of general intellectual ability, are the best independent predictors of real-life success. These tests successfully predict a variety of relevant real-life criteria, more than any other psychological construct (Gottfredson, 1997; Jensen, 1998; Kuncel, Hezlett, & Ones, 2004; Salgado et al. 2003; Schmidt & Hunter, 1998). Yet, traditional intelligence tests are often criticized by means of their “distance” from the real-life (Ceci, 1990; Gardner, 1993; Sternberg, 1985; Sternberg, 1999; Sternberg & Wagner, 1986). The rationale for such criticism can be summarized by the notion that standard intelligence tests capture a very narrow segment of a person’s intellectual functioning, thus failing to capture a wide range of abilities which are important for real-life problem solving (Ceci, 1990; Gardner, 1993; Sternberg, 1985; Sternberg et al., 2000). Nevertheless, most of the authors questioning the validity of conventional tests and measures of intelligence do not deny their importance in the assessment of one’s abilities (e.g., Gardner, 1993; Sternberg, 2003).

A number of context-based theories of intelligence have offered alternative conceptualizations of intelligence, as well as means for its measurement. Some authors advocated different types of intelligence, i.e., intelligence domains, which are mutually independent (e.g., Gardner, 1993; Sternberg, 1999) and suggested alternative intelligence tests, which could, in terms of their requirements and formal characteristics, be closer to real-life problem solving and thus more comprehensively capture and assess intelligence in comparison to conventional intelligence tests (Sternberg et al., 2000). Furthermore, some authors believe that other forms of intelligence which are not measured by standard tests may offer a better prediction of various external criteria (e.g., Sternberg, 1999). However, these tests have proven to be too domain-specific, relating to a very narrow range of abilities, and capturing limited aspects of intellectual functioning, more specifically, expertise in very specific areas, thus falling prey to the same criticism. Additionally, there is no evidence that these tests measure abilities fundamentally different from those assessed by standard intelligence tests (Brody, 2003; McDaniel & Nguyen, 2001; McDaniel & Whetzel, 2003). Finally, the validity coefficients of these tests are often lower than those of highly *G*-saturated tests (Gottfredson, 2003), despite the fact that the most of those tests are developed for the purpose of solving the issue of the insufficient validity of conventional intelligence tests.

A somewhat different approach, aiming to bring cognitive assessment to more ecologically valid context, tried to tap a variety of intellectual processes involved in complex problem solving by using laboratory computer-based simulations of real-life problems (see Frensch & Funke, 1995; Funke, 2010; Gonzalez, Vanyukov, & Martin, 2005). However, the validity of these tasks in relation to intelligence measured by conventional psychometric tests seems to be limited (see Frensch & Funke, 1995; Kretzschmar, Neubert, Wüstenberg, & Greiff, 2016).

Besides from insufficient comprehensiveness, an additional aspect of criticism frequently leveled at intelligence tests regarding their remoteness from the real-life context is their lack of flexibility. Conventional intelligence tests lack flexibility in that they present a person with a problem that has limited options to which only one solution is adequate (Sternberg & Wagner, 1993), making all other options equally “wrong”, therefore displacing a problem from real-life context where a person would be able to express its abilities in a more flexible manner through perceiving and choosing between options of varying accuracy. In that way, intelligence tests disable one to deal with a problem in an alternative way, expressing his/her abilities to the full extent.

Multiple solutions test (MST) is based on the older ‘Multiple solutions test’ (original name: Test višestrukih rešenja, TVR) developed by Bujas and colleagues in mid-sixties (Bujas, 1966; Bujas, Bartolović, & Vodanović, 1967). TVR provides an empirical framework for making typical test-markers of intelligence more flexible, more complex, and closer to real-life problem-solving. In this test, a person is not faced with a problem to which only one solution can be applied, but rather has an opportunity to express its capacity in broader and more flexible manner, namely to deal with a problem from different perspectives, simultaneously comparing and facing alternatives, and choosing between options of varying degrees of accuracy (for further information consult the Part I of this study – see Živanović, Bjekić, & Opačić, 2018).

We developed an instrument for intelligence assessment which could address some of the aforementioned issues of traditional psychometrical intelligence tests, namely lack of comprehensive assessment and flexibility. But at the same time, the test is designed so it can preserve the advantages of the highly *G*-saturated tests, as well as the advantages of the standard test method, i.e., objectivity and efficiency of assessment. The idea behind the test is to present one with a typical *G*-saturated task but to impose additional requirements in order to presumably grasp flexibility of reasoning and a variety of executive processes involved in successful information management. Executive functions are of great relevance in everyday functioning and predictive for a variety of relevant real-life criteria (see Diamond, 2013). Some of the measures of these processes are incorporated in traditional intelligence tests, i.e., working memory and updating, and undoubtedly play a role in higher-order abilities measured (Ackerman, Beier, & Boyle, 2005; Chuderski, Taraday, Nęcka, & Smoleń, 2012; Colom, 2004; Colom, Abad, Rebollo, & Shih, 2005; Conway, Cowan, Bunting, Theriault, & Minkoff, 2002; Engle, Tuholski, Laughlin, & Conway, 1999; Friedman et al., 2006; Kane, Hambrick, & Conway, 2005; Kane et al., 2004; Kyllonen & Christal, 1990; Martínez et al., 2011; Oberauer, Süß, Wilhelm, & Wittmann, 2008). However, conventional measures of intelligence seem to be missing some of the fundamental supervisory functions (see Friedman et al., 2006). Therefore, through introducing additional tasks to the well-proven intelligence test we

wanted to give more load to a variety of executive control processes involved and therefore test requirements presumably bring more closely to real-life problem-solving through potentially more comprehensive assessment of core intellectual processes and abilities.

The test consists of a number of matrix problems, each of which faces a person with three tasks to be solved (see Figure 1): to find *the best* solution for the matrix problem (identical to standard matrices), to detect *the second-best* option – the solution which would be the correct one, if the “ideal” one wouldn’t be present among the options offered, and to solve for *the least accurate* option – to realize which option is the most distant from the “ideal” one. “Correctness” of three types of adequate solutions, within each item, is *a priori* quantitatively operationalized as the number of deviations¹⁰ of the figure from the “ideal”, i.e., correct solution, when following the rules that apply to a given problem.

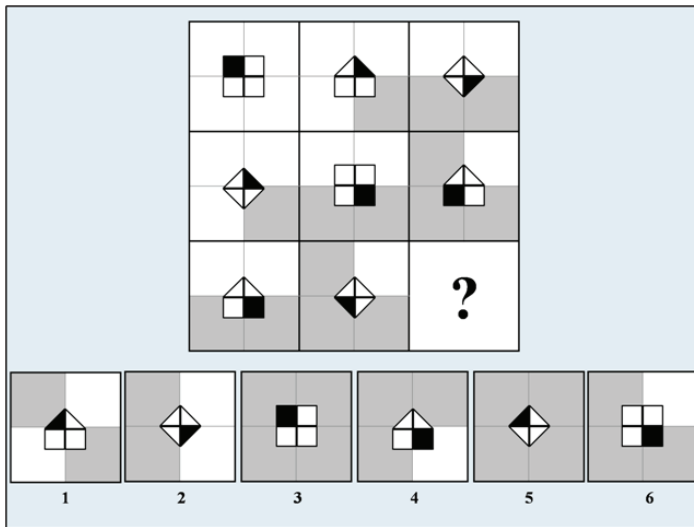


Figure 1. Example of the test's item.

Each of the tasks within the test, as well, as the test as a whole, demonstrated good psychometric properties, with internal consistency parameters lying in the range of widely used matrices, i.e., above .85 (Živanović et al., 2018). It seems that this test format provides reliable but at the same time more flexible

¹⁰ *The best* solution is a figure which follows all the rules that apply to a given matrix (Figure 3); *the second-best* solution is the figure which deviates from the matrix rules the least in comparison to all other options (except from *the best* one (Figure 5); *the least accurate* solution is the figure which, in comparison to all other options available deviate from the rules of the matrix the most (Figure 2). For details see Živanović et al., 2018.

and potentially broader measure of intelligence than standard matrices. In other words, the novelty of this approach is primarily in the alternative tasks that offer the possibility of measuring one's abilities to detect varying degrees of accuracy of different options that impose a load on coordination between rules and their transfer to the next task within the same item in order to solve a problem fully.

The present study

The present study supplements Part I (Živanović et al., 2018) and aims to provide the evidence on construct, predictive, and concurrent validity of MST as a whole, as well as the validity of each of its tasks. Construct validity of all three tasks within MST is tested in relation to the KOG9 battery of ability tests. Based on the Cybernetic model of intellectual abilities (Momirović, Bosnar, & Horga, 1982; Wolf, Momirović, & Džamonja, 1992). KOG9 represents a compilation of reliable and valid intelligence tests that originate from widely used batteries for intelligence assessment such as General Aptitude Test Battery (GATB), Army Alpha, and others (for details see Lazarević & Knežević, 2008; Wolf et al., 1992). This battery assesses the efficiency of three mental processors: perceptual, serial, and parallel. These processors, by means of their operationalizations, closely correspond to the second-order factors within Cattell-Horn-Carroll's model of intelligence (Carroll, 1993, 1997, 2005; McGrew & Wendling, 2010): *Gs*, *Gc*, and the amalgam of *Gf* and *Gv*, respectively. Since MST is designed for measuring fluid reasoning, it could be expected to manifest a number of positive relations with parallel processor tests, while achieving somewhat lower correlations with tests measuring the efficiency of serial and perceptual processor. The predictive value of the instrument is tested in relation to the scholastic achievement as measured by GPA (grade-point average), which was frequently used in previous studies for pointing to the practical validity of intelligence tests (see Jensen, 1998). Namely, studies show that intelligence is the best single predictor of scholastic achievement (e.g., Kunzel et al., 2004). Although the average correlation of IQ score or other composite measures of intelligence with scholastic achievement is estimated at approximately .50 (Gustafsson & Undheim, 1996) (the correlation between individual tests and scholastic achievement is of course somewhat lower), there are significant variations between studies. One of the reasons for such variations is the attenuation of correlation on the higher educational levels (typically .30–.40) in comparison to lower educational levels (typically between .50 and .70), due to the restriction in range in abilities at higher educational levels (Jensen, 1998).

Since MST is aiming to provide a wider coverage of core features of intelligence that are closer to real-life success we wanted to test whether alternative tasks introduced by this instrument – *the second-best* and *the least accurate* solutions, provide any additional information over the standard one, by testing their incremental value in predicting GPA, over *the best* one. Finally, in order to examine whether MST adds any practical value to currently used instruments, incremental validity of MST in predicting GPA, over and above KOG9 was tested.

Method

Participants

A total of 235 psychology students and fresh psychology graduates from the University of Belgrade and the State University of Novi Pazar (age $M = 22.65$, $SD = 3.33$, 36 men and 199 women) participated in the study. Predictive validity of the instrument was tested on the whole sample, while the construct validity was tested on the subsample of 156 students from the University of Belgrade (age $M = 22.17$, $SD = 2.21$; 22 males and 134 females) who completed KOG9 battery. All participants volunteered to participate in the study.

Instruments and measures

Multiple solutions test. Multiple solutions test consists of 40 items with three tasks. Within each item, participants are to perform three separate tasks: 1) to identify *the best* solution, i.e., the one which makes the most sense to them and completes the matrix best, 2) to identify *the second-best* solution, i.e., the one they would pick if the “right” option was not present, 3) to identify *the least accurate* option, i.e., the one which makes the least sense to them, deviating from the inherent rules of the matrix the most. Within every item, participants could perform these three tasks in no particular order. Scoring is performed for the whole test (i.e., *total score*; with the theoretical range of 0–120), and for each task separately (i.e., *the best score*, *the second-best score*, *the least accurate score*; each with a theoretical range of 0–40).

The MST shows good internal consistency ($\alpha = .94$), and all three tasks have satisfactory psychometric properties ($\alpha_{\text{the best}} = .91$, $\alpha_{\text{the second-best}} = .90$, $\alpha_{\text{the least accurate}} = .86$). Within every item, there are 6 available options, three of which are correct (one per task). The test was developed to have a liberal time constraint, but the completion of all items takes up to 60 minutes for most people. MST is intended to be used as a whole and to calculate only the total score, but since the main focus of this paper is on its alternative tasks, we will also use scores for each of the tasks in order to provide additional information about them and their relations.

KOG9. KOG9 (Wolf et al., 1992) is a battery of intelligence tests for adults. It consists of nine tests that assess the effectiveness of three postulated processors, each of which is assessed by three tests: *perceptual processor* (Identical figures test (it1), Hidden figures test (cf2), and Form perception test (gt7)), *serial processor* (Synonyms-antonyms test (al4), Analogies (alf7), and Synonyms test (gsn)), and *parallel processor* (Domino test (d48), Test of spatial abilities (it2), and Test of visual spatialization (s1)). Each of the subtests has strict time constraints. Battery manifests postulated three-factor structure and good psychometric properties (Lazarević & Knežević, 2008).

Scholastic achievement. Scholastic achievement is operationalized as a students' overall GPA during their university studies (range 6–10). GPA is calculated as the average grade at the moment of testing for those students that have not yet completed their studies.

Procedure

MST was individually administrated to the participants using a computer in a suitable environment (testing or computer room). After the general instructions, through short practice section participants were familiarized with items that are to be solved and the tasks they would be required to perform. After the practice round completion, participants were asked to start the test. Within every item, participants were answering by indicating the number of the figure which they consider to be *the best*, *the second-best*, and *the least accurate* solution.

Additionally, participants were asked to provide their GPA. Data on the performance of the participants on the KOG9 were obtained in a standardized group paper-pencil testing.

Results

Relations between tasks

Table 1 displays descriptive statistics for the MST's total score and its individual tasks. *The best solution* and *the second-best solution* task have shown elevated scores as indicated by asymmetry coefficients. Analysis of differences in the performance between *the best*, *the second-best*, and *the least accurate solution* have indicated differential difficulties for all three tasks, $F(2, 468) = 472.81$, $p < .001$, $\eta^2 = .669$. Namely, *the least accurate* task ($p < .001$) has shown to be the most difficult for the participants, followed by *the second-best* ($p < .001$), while *the best solution* task ($p < .001$) has shown to be easier than both alternative tasks ($p < .001$).

Table 1
Descriptive statistic for individual tasks of Multiple solutions test and its total score ($N = 235$)

	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>	<i>Sk</i>	<i>Ku</i>	<i>zSk</i>	<i>zKu</i>	<i>K-S</i>
<i>the best</i>	28.77	7.67	7 (0)	40 (40)	-0.78	-0.30	-4.89**	-0.95	2.05**
<i>the second-best</i>	22.41	8.20	5 (0)	38 (40)	-0.33	-0.85	-2.06*	-2.69**	1.52*
<i>the least accurate</i>	21.40	7.38	4 (0)	39 (40)	-0.04	-0.73	-0.26	-2.33*	0.90
MST full-scale score	72.58	22.22	17 (0)	114 (120)	-0.41	-0.75	-2.56*	-2.37*	1.32

Note. *M* – mean; *SD* – standard deviation; *Min* – minimum; *Max* – maximum; numbers in the brackets – theoretical range; *Sk* – skewness; *Ku* – kurtosis; *zSk* – standardized skewness; *zKu* – standardized kurtosis; *K-S* – Kolmogorov-Smirnov test of normality of distribution of scores; ** $p < .01$; * $p < .05$.

The correlations between scores on the three tasks have shown to be high (*the best* and *the second-best* $r = .894$, $p < .001$, *the best* and *the least accurate* $r = .838$, $p < .001$; *the second-best* and *the least accurate* $r = .857$, $p < .001$). However, when partialled-out for the performance on *the best solution* the alternative tasks (*the second-best* and *the least accurate*) have shown to correlate moderately ($r = .441$, $p < .001$). In addition, out of those who did not solve *the best* task correctly, on average, 13.5% solved *the second-best* task correctly, and 28.9% manage to solve *the least accurate* task.

Construct validity

Table 2 displays descriptive statistics for subtests within KOG9 battery and three tasks within MST, and its total score obtained on the subsample of participants. Due to characteristics of the sample elevated scores are not unexpected.

Table 2
 Descriptive statistic for individual tasks of Multiple solutions test and its total score, and KOG9 subtests (N = 156)

	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>	<i>Sk</i>	<i>Ku</i>	<i>zSk</i>	<i>zKu</i>	<i>K-S</i>
<i>the best</i>	31.10	5.64	9 (0)	40 (40)	-1.16	1.62	-5.99**	4.20**	1.59*
<i>the second-best</i>	24.65	6.32	5 (0)	36 (40)	-0.53	-0.04	-2.76**	-0.09	1.25
<i>the least accurate</i>	23.33	6.17	8 (0)	36 (40)	-0.06	-0.33	-0.32	-0.85	0.69
MST full-scale score	79.09	16.88	25 (0)	114 (120)	-0.60	0.21	-3.09**	0.54	0.96
Synonyms-antonyms (al4)	19.77	2.13	7 (0)	21 (40)	-3.24	12.94	-16.69**	33.53**	3.90**
Analogies (alf7)	23.56	3.23	12 (0)	30 (39)	-0.50	0.48	-2.59**	1.25	1.22
Test of synonyms (gsn)	25.00	4.16	7 (0)	36 (39)	-0.53	2.35	-2.76**	6.08**	1.34
Test of spatial abilities (it2)	25.17	5.76	8 (0)	38 (39)	-0.39	0.10	-2.02*	0.27	1.22
Test of form perception (gt7)	38.06	8.93	15 (0)	60 (60)	0.06	-0.34	0.31	-0.89	0.75
Visual spatialization test (s1)	23.41	4.75	3 (0)	30 (30)	-1.52	2.70	-7.81**	6.99**	2.04**
Identical figures test (it1)	27.11	4.73	7 (0)	37 (39)	-0.56	1.42	-2.89**	3.67**	0.99
Domino test (d48)	25.79	5.76	5 (0)	39 (40)	-0.52	0.49	-2.70**	1.26	1.07
Hidden figures test (cf2)	48.24	15.01	2 (0)	87 (97)	-0.25	0.48	1.29	1.25	0.61

Note. *M* – mean; *SD* – standard deviation; *Min* – minimum; *Max* – maximum; numbers in the brackets – theoretical range; *Sk* – skewness; *Ku* – kurtosis; *zSk* – standardized skewness; *zKu* – standardized kurtosis; *K-S* – Kolmogorov-Smimov test of normality of distribution of scores; ** *p* < .01; * *p* < .05.

Due to high non-normality of scores, nonparametric correlations were used. Correlations of KOG9 subtests within perceptual, serial and parallel processor were all positive and varied between .305–.426, .299–.346, and .342–.651, respectively. The relations between three tasks within MST, and subtests assessing the efficiency of the perceptual, serial, and parallel processor within KOG9 are provided in Table 3 while 95% confidence intervals are presented in the Appendix.

Table 3
 Correlations between individual tasks of Multiple solutions test (and total score) and subtests of KOG9 (N = 156)

	<i>perceptive processor</i>			<i>serial processor</i>			<i>parallel processor</i>		
	it1	cf2	gt7	al4	alf7	gsn	d48	it2	s1
<i>the best</i>	.16*	.13	.16*	.03	.35**	.21**	.22**	.31**	.28**
<i>the second-best</i>	.14	.06	.06	-.01	.27**	.15	.16*	.23**	.23**
<i>the least accurate</i>	.01	.03	.11	.04	.29**	.12	.09	.17*	.18*
MST full-scale score	.10	.07	.12	.01	.32**	.17*	.15	.25**	.23**

Note. it1 – Identical figures test; cf2 – Hidden figures test; gt7 – Test of form perception; al4 – Synonyms-antonyms test; alf7 – Analogies; gsn – Test of synonyms; d48 – Domino test; it2 – Test of spatial abilities; s1 – Visual spatialization test; ** *p* < .01; * *p* < .05.

*Correlation between it1 and the least accurate (the last accurate u italiku) solution is .002, but is rounded at .01 for consistency in data presentation.

Significant correlations between the performance in detecting *the best*, *second-best*, and *the least accurate* solution and subtests of KOG9 were all positive and low to moderate in size. Among speeded tests of the *perceptual processor*, the only subtests that correlated with *the best* solution were it1 and gt7. On the other hand, the remaining two tasks within MST did not correlate with any of the highly speeded subtests of KOG9 battery (perceptual processor

tests). Regarding verbal *serial processor tests*, Analogies (alf7), correlated with all three tasks of the MST. The Synonyms-antonyms (al4) showed no significant relations with MST, while the Synonyms test (gsn) correlated positively only with *the best* solution. *The best*, *the second-best*, and *the least accurate* solution all positively correlated with the subtests that assess the efficiency of the *parallel processor*. *The least accurate* task positively correlated with Spatial ability test (it2), and with the test of Visual spatialization (s1), but not with the Domino test (d48). In sum, three tasks within MST achieved the highest correlation with the verbal Analogies and a number of tests of the efficiency of the *parallel processor*, primarily tests of visualization, spatial relations. It is important to note that all the aforementioned tests represent good markers of *Gf* or/and *Gv*, while the number of relations with intelligence measures which are further from *Gf* was significantly lower or nonexistent.

Predictive validity

Descriptive statistics for GPA are presented in Table 4.

Table 4
Descriptive statistic for GPA ($N = 235$)

	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>	<i>Sk</i>	<i>Ku</i>	<i>zSk</i>	<i>zKu</i>	<i>K-S</i>
GPA	8.20	0.76	6 (6)	10 (10)	0.14	-0.13	0.88	-0.40	1.18

Note. *M* – mean; *SD* – standard deviation; *Min* – minimum; *Max* – maximum; numbers in the brackets – theoretical range; *Sk* – skewness; *Ku* – kurtosis; *zSk* – standardized skewness; *zKu* – standardized kurtosis; *K-S* – Kolmogorov-Smirnov test of normality of distribution of scores; ** $p < .01$; * $p < .05$.

GPA, as a measure of scholastic achievement, was positively correlated with the total score on the MST ($r = .437, p < .001$). Likewise, the measure of scholastic achievement manifested a positive correlation with each of the MST's task individually ($r_{\text{the best}} = .378, p < .001$; $r_{\text{the second-best}} = .431, p < .001$; $r_{\text{the least accurate}} = .444, p < .001$). With the aim of testing the incremental value of *the second-best* and *the least accurate* solution over *the best* one, in predicting GPA a hierarchical regression analysis¹¹ is performed. Since reaching *the best* solution is the initial step in solving the remaining two tasks, performance on *the best* task was entered as first, while *the second-best* and *the least accurate* task were entered in the second step of the analysis.

Results have shown that performance on *the best* solution task predicts 14.3% of the variance of a students' average grade, $R = .378, F(1, 233) = 38.928, p < .001$. *The second-best* and *the least accurate* solution accounted for the additional 6.7% of the variance of the GPA, $F(2, 231) = 9.860, p < .001$. In total, the model accounted for 21.1% of the variance of the criterion, $R = .459, F(3, 231) = 20.536, p < .001$. Additionally, upon the inclusion of the alternative tasks, the best solution lost its predictive value, which can partially be attributed to multicollinearity. The results of hierarchical regression analysis are shown in Table 5.

11 Variables which deviated from normal distribution were normalized using Rankit's formula (see Solomon & Sawilowsky, 2009).

Table 5

Hierarchical regression analysis: Incremental value of the second-best and the least accurate solution over the best one in prediction of GPA (N = 235)

model	predictors	β	ΔR^2
1	<i>the best</i>	0.38**	.14
2	<i>the best</i>	-0.15	
	<i>the second-best</i>	0.29*	.07
	<i>the least accurate</i>	0.32**	

Note. β – standardized regression coefficient; r – zero-order correlation; ΔR^2 – change in the proportion of variance accounted for (step 1 ΔR^2 presents the proportion of variance accounted for by the best score); ** $p < .01$; * $p < .05$.

Concurrent validity of MST over KOG9

In order to gain further insight into the incremental value of MST, prediction of scholastic achievement, over and above KOG9 battery was tested. In line with the factor structure of the KOG9 battery (see Lazarević & Knežević, 2008) and in order to reduce the number of predictors used in the analysis, the first principal component from the subtests within each processor has been extracted and these variables were used as predictors in a hierarchical linear regression, along with the summary score of MST. In the first step of the analysis, the measures of efficiency of the three processors were entered as predictors, while in the second, performance on MST was introduced.

The results have shown that the prediction of scholastic achievement based on all three processors of KOG9 has marginal significance. The efficiency of *perceptive*, *serial* and *parallel* processor tests in total marginally accounted for 4.7% of the variance of the criterion, $R = .217$, $F(3, 150) = 2.479$, $p = .063$, indicating that KOG9 is a poor predictor of scholastic achievement. With the inclusion of the performance on MST in the model, the percentage of the variance of criterion accounted increased by additional 13.1%, $F(1, 149) = 23.776$, $p < .001$. Variables in the final model accounted for 17.8% of variance (adj. $R^2 = .156$), of GPA in total, $R = .422$, $F(4, 149) = 8.086$, $p < .001$. By including the *MST*, the efficiency of the serial processor, the only variable that displayed predictive value became redundant. The results of hierarchical linear regression analysis are presented in Table 6.

Table 6

Hierarchical regression analysis: Prediction of GPA using KOG9 and MST (N = 154)

model	predictors	β	r	ΔR^2
1	<i>perceptive processor</i>	-0.11	.04	
	<i>serial processor</i>	0.25**	.20*	.05
	<i>parallel processor</i>	0.02	.08	
2	<i>perceptive processor</i>	-0.06	.04	
	<i>serial processor</i>	0.17	.20*	.13
	<i>parallel processor</i>	-0.09	.08	
	MST full-scale score	0.39**	.40**	

Note. β – standardized regression coefficient; r – zero-order correlation; ΔR^2 – change in the proportion of variance accounted for (step 1 ΔR^2 presents the proportion of variance accounted for by the best score); ** $p < .01$; * $p < .05$.

Discussion

Standard intelligence tests are often criticized for not measuring a wide range of skills and abilities needed for real-life problem solving (Ceci, 1990; Gardner, 1993; Sternberg, 1985; Sternberg & Wagner, 1986). Despite this criticism targeted at conventional psychometric tests, the empirical evidence supports their predictive power (Gottfredson, 1997; Jensen, 1998; Kuncel et al., 2004; Salgado et al. 2003; Schmidt & Hunter, 1998), at the same time indicating problems of construct and predictive validity of alternative tests which are trying to capture intelligence that is 'closer' to real-life (Gottfredson, 2003). Furthermore, it seems that the coverage of broad intellectual abilities is not resolved by introducing narrow domain-specific tests (Allix, 2000), nor nominating numerous abilities, reportedly independent of *G*, whose measurement-related problems are yet to be solved (Waterhouse, 2006). In fact, it seems that alternative tests do not measure anything more than manifestations of intelligence in different contexts (Brody, 2003; McDaniel & Whetzel, 2003). Finally, the validity of alternative real-life-based computer simulations for assessment of cognitive processes involved in complex problem-solving in relation to conventional psychometric intelligence tests seems to be limited (see Frensch & Funke, 1995; Kretzschmar et al., 2016).

One of the points of intersection between opposing paradigms can be found within the MST, which aims to address the lack of bandwidth and flexibility of conventional intelligence tests, by introducing alternative tasks. It could be assumed that by enabling one to deal with the problem from different perspectives, to compare and face alternatives, and choose between options of varying degree of accuracy by engaging his/her operative capacities, could bring the test more closely to real-life problem-solving. Namely, executive processes that are presumably captured by alternative tasks are only partially assessed by standard intelligence tests (see Friedman et al., 2006) but are of great relevance in everyday functioning and predictive for a variety of relevant real-life outcomes (Diamond, 2013). Therefore, it can be assumed that incorporating measures of these processes into one of the *G*-most-central tests (Carroll, 1993; Jensen, 1998; Snow et al., 1984; Spearman, 1946; Vernon & Parry, 1949) leads to the more comprehensive assessment of core intellectual abilities. Consequentially, it can be assumed that these measures could potentially provide a more valid assessment as indicated by the incremental value in prediction of relevant external criteria over conventional form of the task as well as other measures of intelligence.

The first part of the study provided evidence on good internal psychometric properties of the MST (Živanović et al., 2018). In this study, we provided preliminary data on the construct validity of MST, through examination of the relations between its individual tasks, and their relations to intellectual abilities captured by the variety of intelligence test incorporated in the KOG9 battery. Additionally, we addressed MST's predictive validity in relation to GPA, as well as its incremental value over standard-form intelligence assessment, i.e., KOG9.

Individual tasks of MST have demonstrated high between-task correlations. Bearing in mind that all three tasks share the same problem content, i.e. are performed on the same matrix one could expect them to be interrelated. Additionally, the results indicated that the performance on the alternative tasks is highly dependent on the performance on *the best solution* task. Namely, detecting the best solution can be considered as the initial step in dealing with the problem. After achieving this goal, one is enabled to set an ‘anchor’ which allows him/her to proceed with solving the next two tasks bearing in mind the rules that are to be followed and coordinated in order to successfully apply them in solving additional tasks. It has been shown that when controlling for *the best solution* the correlation between alternative tasks diminishes, highlighting that they are not entirely reducible to the abilities and processes involved in solving *the best* task and that each of them exhibits apparent specificities that are, as further results suggested, valuable for the prediction of relevant criterion.

Three tasks within MST demonstrated differential difficulties, with *the least accurate* one being the most difficult, followed by *the second-best*, and *the best* task as the easiest one. One could argue that additional tasks are more difficult than finding *the best* solution to the matrix problem because they engage additional cognitive resources and processes. It can be argued that *the best* solution could be found by eliminating those options that fail to comply with *any* of the matrix rules. On the other hand, when solving the other two tasks, one must keep in mind all the rules that apply to a given matrix along with the characteristics of the options and keep track of which one deviates from the rules and how much. In other words, alternative tasks are likely to put more load on the executive control functions, such as those systematized and elaborated by Miyake and colleagues (Miyake & Friedman, 2012; Miyake et al., 2012). *The least accurate* task is presumably more difficult than *the second-best* one because it puts even more load on the executive control. Namely, in solving this task one must keep in mind all the matrix rules and simultaneously shift attention from one to another while searching for the most adequate among available options. At the same time, one must inhibit salient deviations of distractors from the “ideal” solution but to detect aspects of stimuli which follows some of them in order to come to the most “deviant” solution. Although in solving *the second-best* task, the same processes are presumably involved, here one can alleviate executive load and “take a shortcut” by comparing relevant features of the stimuli to the ‘anchor’, i.e., *the best* solution, which is only one less deviation far from being the best one. However, since the focus of this study is more practical than theoretical one should bear in mind that untangling the processes behind solving alternative tasks needs further research on their relations with more basic cognitive functions.

Concerning the convergent/divergent validity of MST, it can be noted that the tasks for *the best* solution achieved positive, although relatively low correlation, with typically good measures of *Gf*, for example, the Domino test, as well as tests of visual spatialization and spatial abilities, that are, in addition

to being markers of G_v (Carroll, 1993, 1997, 2005), fair measures of general and fluid abilities (Bele-Potočnik, 1983; Domino, 2001; Wolf et al., 1992). Also, the task for *the best solution*, as well as two alternative tasks manifested moderate correlation with the Analogies test, which represents a good approximation of G , given that these types of tests are, depending on their content, good markers of both G_c and G_f (Carpenter, Just, & Shell, 1990; Horn, 1979). On the other hand, MST tasks have not shown substantial relations with the measures that are further away from G_f , namely measures of processing speed (Hidden figures tests, Identical figure test, Test of form perception), nor with the markedly speeded verbal test Synonyms-antonyms. One reason for the relatively low correlation that *the best* task achieved with all the subsets of the battery, in addition to restricted variability in abilities due to sample characteristics, is probably due to the difference in nature of the two tests. The battery KOG9 mainly contains tests, which are all, more or less, related to the general ability and G_f , but almost none of them represent a focal measure of fluid reasoning, as defined within CHC model (McGrew, 2009; McGrew & Wendling, 2010). Also, the majority of the subtests of the KOG9 battery are speeded tests or tests with relatively restrictive time constraints, while MST is a typical power test. Although alternative solutions have shown a similar convergent/divergent pattern of correlations with the same test-markers as *the best* solution, these correlations are somewhat lower, indicating that they capture the specific variance of ability that is not measured by KOG9. Overall, all three tasks have shown a satisfactory level of convergent/divergent validity.

Scholastic achievement represents a good criterion for testing the practical validity of intelligence measures (see Jensen, 1998). Previous studies (see Jensen, 1998) suggest that composite measures of intelligence typically correlate between .30 and .40 on the higher educational levels. The correlation between MST's standard matrix task obtained in this study falls in the range of aforementioned values pointing to its good predictive validity. Additionally, the correlation between alternative tasks slightly exceeded aforementioned correlations and this increase in correlations with the scholastic achievement cannot be attributed to the differences in individual tasks' reliabilities. Furthermore, it was demonstrated that the specific variance of *the second-best* and *the least accurate* solution significantly contributed to the prediction of GPA, over and above the standard task, indicating that these tasks are likely to capture abilities and processes not directly measured by standard matrices and which are an important part of a process of obtaining the real-life achievement. As shown, these two tasks demonstrated independent predictive value and suppressed the predictive value of the standard task. In sum, results underline two relevant facts: firstly, that the ability to detect *the best* solution is incorporated in the remaining two tasks, secondly, that each of the alternative tasks measures additional abilities and processes relevant to real-life achievement, that are not captured by standard matrices. Overall, obtained results support the practical value of both standard and alternative matrix tasks.

It is unlikely that these additional tasks are measures of some non-cognitive factors. Rather it's more likely that they capture the abilities and processes of flexible information management, through dealing with the problem from different perspectives, analyzing, simultaneously comparing and facing alternatives, grading their accuracy, inhibiting irrelevant aspects, and detecting crucial aspects of the problem. In other words, the ability to find *the best* solution is a standard measure of one's intellectual capacities, while the ability to solve additional two tasks is more of a measure of one's ability to adequately employ those intellectual capacities in a flexible manner.

Finally, it has been demonstrated that the performance on MST represents a better predictor of scholastic achievement than the efficiency of broad processors postulated and measured by KOG9. It seems that KOG9 has very low predictive power when it comes to scholastic achievement. This result is certainly surprising especially having in mind wide usage of its subtests and battery as a whole. However, this result can be interpreted in the light of findings that have pointed that matrices which predominantly measure *Gf* are better predictors of scholastic achievement than the measures of speeded tests or spatial abilities (see Rohde & Thompson, 2006). Additionally, it seems that an important part of the real-life achievement reflected in the abilities and processes engaged in detecting options of varying level of accuracy is not captured by this comprehensive battery.

It can be concluded that the MST has a satisfying convergent/divergent and predictive validity for all three tasks and that detecting *the second-best* and *the least accurate* solution provides relevant information on one's abilities, making them a useful addition to standard tests of this type. However, few limitations of this study should be addressed here, namely, restriction in range of abilities, and large gender imbalance of the sample used in this study, as well as the properties of the test-battery used for examination of MST's construct validity. Characteristics of the sample (student and fresh graduates) caused the restriction of range in measured abilities thus diminishing obtained correlations. Despite this fact, obtained correlations between MST and KOG9 variables demonstrated expected patterns of relations. Additionally, obtained correlations between MST's individual tasks and scholastic achievement fall in the range of those reported in previous studies which predominantly used composite intelligence scores. On the other hand, large gender imbalance of the sample used most certainly did not contribute to the strength and generalization of our findings and further studies aimed at more comprehensive examination of MST's features and validity should strive to ensure more gender-balanced and more representative sample. Finally, one may argue for a better fitting selection of criterion tests in order to provide a more convincing evidence on the convergent/divergent validity of MST since most of the KOG9 tests are predominately speeded tests, while MST is a typical power test. However, criterion tests used here are all well established and widely used both in research and in practice. Additionally, the battery is designed to cover four broad factors of intelligence (*Gf/Gv*, *Gc*, *Gs*) in an economical manner. Unfortunately, it turned out that tests within this battery do not have sufficient discriminative power for high-level educated individuals thus maybe representing "a weak rival" to MST. Therefore, further studies

should provide additional data on the construct validity of MST relating it to other well-established markers of broad factors of intelligence.

In sum, the evidence on the validity of the newly developed instrument presented here does not provide sufficient argument for all aspects of its validity. It should be noted that further validation of the instrument in relation to other relevant external criteria has to be carried out, given that this study examined its predictive validity only in relation to scholastic achievement. Additionally, further examination of the instrument's relations to more basic cognitive processes, such as executive functions is needed in order to make definite conclusions on the nature of abilities engaged in performance on alternative tasks within MST. Finally, in order to determine its divergent validity, the MST and especially its alternative tasks need to be examined in relation to more remote constructs in order to definitely discard a possibility of the impact of the noncognitive factors on performance on these tasks.

Conclusion

MST could be considered a valid measure of fluid reasoning. Additional tasks, i.e., *the second-best* and *the least accurate* solution seem to be a significant addition to the standard matrices with a single solution. Moreover, its predictive validity speaks for the practical value of MST, showing its potential for usage both in research and in practice. Although the instrument in its current form can be considered a useful and reliable tool for the assessment of intellectual abilities, its further empirical verification is needed.

References

- Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin*, *131*(1), 30–60. doi:10.1037/0033-2909.131.1.30
- Allix, N. M. (2000). The theory of multiple intelligences: A case of missing cognitive matter. *Australian Journal of Education*, *44*, 272–288. doi:10.1177/000494410004400306
- Bele-Potočnik, Ž. (1983). *Test Domino "D48" – Priručnik [Domino test "d48" – Manual]*. Ljubljana: Zavod SR Slovenije za produktivnost dela – Centar za psihodijagnostička sredstva.
- Brody, N. (2003). What Sternberg should have concluded. *Intelligence*, *31*(4), 339–342. doi:10.1016/S0160-2896(02)00190-3
- Bujas, Z., Bartolović, B., & Vodanović, M. (1967). *Test višestrukih rješenja [Multiple solution test]*. Zagreb: Odsjek za psihologiju.
- Bujas, Z. (1966). *Modifikacija Ravenovih Progresivnih Matrica [Modification of the Raven's Progressive Matrices]*. Zagreb: Odsjek za psihologiju.
- Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: a theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review*, *97*(3), 404–431.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor analytic studies*. New York: Cambridge University Press.
- Carroll, J. B. (1997). The three-stratum theory of cognitive abilities. In P. D. Flanagan, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (pp. 122–130). New York: Guilford.

- Carroll, J. B. (2005). The three-stratum theory of cognitive abilities. In P. D. Flanagan, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (2nd edition., pp. 69–76). New York: Guilford.
- Ceci, S. J. (1990). *On intelligence... more or less: A bio-ecological treatise on intellectual development*. Englewood Cliffs, NJ: Prentice Hall.
- Chuderski, A., Taraday, M., Nęcka, E., & Smoleń, T. (2012). Storage capacity explains fluid intelligence but executive control does not. *Intelligence*, *40*(3), 278–295. doi:10.1016/j.intell.2012.02.010
- Colom, R. (2004). Working memory is (almost) perfectly predicted by g. *Intelligence*, *32*(3), 277–296. doi:10.1016/j.intell.2003.12.002
- Colom, R., Abad, F. J., Rebollo, I., & Shih, P. C. (2005). Memory span and general intelligence: A latent-variable approach. *Intelligence*, *33*(6), 623–642. doi:10.1016/j.intell.2005.05.006
- Conway, A. R., Cowan, N., Bunting, M. F., Theriault, D. J., & Minkoff, S. R. . (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, *30*(2), 163–183. doi:10.1016/S0160–2896(01)00096–4
- Diamond, A. (2013). Executive Functions. *Annual Review of Psychology*, *64*, 135–168. doi:10.1146/annurev-psych-113011–143750.Executive
- Domino, G. (2001). The D48 Application in Mexican-American Children of a Culture Fair Test. *School Psychology International*, *22*, 253–257.
- Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short term memory and general fluid intelligence: A latent variable approach. *Journal of Experimental Psychology: General*, *128*(3), 309–331. doi:10.1037/0096–3445.128.3.309
- Frensch, P. A., & Funke, J. (Eds.) (1995). *Complex Problem Solving: The European Perspective*. Hillsdale, NJ: Erlbaum.
- Friedman, N. P., Miyake, A., Corley, R., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological science*, *17*(2), 172–179. doi:10.1111/j.1467–9280.2006.01681.x
- Funke, J. (2010). Complex problem solving: A case for complex cognition? *Cognitive Processing*, *11*, 133–142. doi:10.1007/s10339–009–0345–0
- Gardner, H. (1993). *Frames of mind: The Theory of multiple intelligences*. New York: Basic Books.
- Gonzalez, C., Vanyukov, P., & Martin, M. K. (2005). The use of microworlds to study dynamic decision making. *Computers in Human Behavior*, *21*(2), 273–286. doi:10.1016/j.chb.2004.02.014
- Gottfredson, L. S. (1997) Why g matters? The Complexity of Everyday Life. *Intelligence*, *24*(1) 79–132. doi:10.1016/S0160–2896(97)90014–3
- Gottfredson, L. S. (2003). Dissecting practical intelligence theory. *Intelligence*, *31*(4), 343–397. doi:10.1016/S0160–2896(02)00085–5
- Gustafsson, J. E., & Undheim, J. O. (1996). Individual differences in cognitive functions. In D. C. Berliner R .C. Calfee (Eds.), *Handbook of educational psychology* (pp. 186–242). New York: Prentice Hall International.
- Horn, J. L. (1979). The rise and fall of human abilities, *Journal of Research and Development in Education*, *12*(2), 59–78.
- Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, Connecticut: Praeger Publishers.
- Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2005). Working memory capacity and fluid intelligence are strongly related constructs: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, *131*(1), 66–71. doi:10.1037/0033–2909.131.1.66
- Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal

- and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General*, 133(2), 189–217. doi:10.1037/0096-3445.133.2.189
- Kretzschmar, A., Neubert, J., Wüstenberg, S., & Greiff, S. (2016). Construct validity of complex problem solving: A comprehensive view on different facets of intelligence and school grades. *Intelligence*, 54, 55–69. doi:10.1016/j.intell.2015.11.004
- Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2004). Academic performance, career potential, creativity, and job performance: Can one construct predict them all? *Journal of Personality and Social Psychology*, 86(1), 148–161. doi:10.1037/0022-3514.86.1.148
- Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence*, 14, 389–433. doi:10.1016/S0160-2896(05)80012-1
- Lazarević, Lj., & Knežević, G. (2008). Provera faktorske strukture baterije za procenu intelektualnih sposobnosti KOG9 [Factor structure of KOG9 battery for assessment of intellectual abilities]. *Psihologija*, 41(4), 489–505. doi:10.2298/PSI0804489L
- Martínez, K., Burgaleta, M., Román, F. J., Escorial, S., Shih, P. C., Quiroga, M. Á., & Colom, R. (2011). Can fluid intelligence be reduced to “simple” short-term storage? *Intelligence*, 39(6), 473–480. doi:10.1016/j.intell.2011.09.001
- McDaniel, M. A., & Nguyen, N. T. (2001). Situational judgment tests: A review of practice and constructs assessed. *International Journal of Selection and Assessment*, 9, 103–113. doi: 10.1111/1468-2389.00167
- McDaniel, M. A., & Whetzel, D. L. (2003). The Emperor’s New Clothes: Additional Critiques of Sternberg’s Practical Intelligence Theory. Paper presented at the *Fourth Annual Conference of the International Society for Intelligence Research*, Newport Beach, CA
- McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, 37, 1–10. doi:10.1016/j.intell.2008.08.004
- McGrew, K. S., & Wendling, B. J. (2010). Cattell-Horn-Carroll cognitive-achievement relations: What we have learned from the past 20 years of research. *Psychology in the Schools*, 47(7), 651–675. doi:10.1002/pits.20497
- Miyake, A., & Friedman, N. P. (2012). The nature and organisation of individual differences in executive functions: four general conclusions. *Current Directions in Psychological Science*, 21(1), 8–14. doi:10.1177/0963721411429458
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “Frontal Lobe” tasks: a latent variable analysis. *Cognitive Psychology*, 41(1), 49–100. doi:10.1006/cogp.1999.0734
- Momirović, K., Bosnar, K., & Horga, S. (1982). Kibernetički model kognitivnog funkcionisanja: Pokušaj sinteze nekih teorija o strukturi kognitivnih sposobnosti [Cybernetic model of cognitive functioning: An attempt to synthesize some theories about structure of cognitive abilities]. *Kineziologija*, 14, 63–82.
- Oberauer, K., Süß, H.-M., Wilhelm, O., & Wittmann, W. W. (2008). Which working memory functions predict intelligence? *Intelligence*, 36(6), 641–652. doi:10.1016/j.intell.2008.01.007
- Rohde, T. E., & Thompson, L. A. (2007). Predicting academic achievement with cognitive ability. *Intelligence*, 35(1), 83–92. doi:10.1016/j.intell.2006.05.004
- Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C., de Fruyt, F., & Rolland, J. P. (2003). A Meta-Analytic Study of General Mental Ability Validity for Different Occupations in the European Community. *Journal of Applied Psychology*, 88(6), 1068–1081. doi:10.1037/0021-9010.88.6.1068
- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124(2), 262–274. doi:10.1037/0033-2909.124.2.262

- Snow, R. E., Kyllonen, C. P., & Marshalek, B. (1984). *The topography of ability and learning correlations*. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (pp.47–103). Hillsdale, NJ: Erlbaum.
- Solomon, S., R., & Sawilowsky, S. S. (2009). Impact of rank-based normalizing transformations on the accuracy of test scores. *Journal of Modern Applied Statistical Methods*, 8(2), 448–462. doi:10.22237/jmasm/1257034080
- Spearman, C. E. (1946). Theory of the general factor. *British Journal of Psychology*, 36, 117–131. doi:10.1111/j.2044-8295.1946.tb01114.x
- Sternberg, R. J. (1985). *Beyond IQ: A Triarchic Theory of Human Intelligence*. Cambridge: Cambridge University Press.
- Sternberg, R. J. (1999). The theory of successful intelligence. *Review of General Psychology*, 3, 292–316. doi:10.1037/1089-2680.3.4.292
- Sternberg, R. J. (2003). Our research program validating the triarchic theory of successful intelligence: reply to Gottfredson. *Intelligence*, 31(4), 399–413. doi:10.1016/S0160-2896(02)00143-5
- Sternberg, R. J., Forsythe, G. B., Hedlund, J., Horvath, J. A., Wagner, R. K., Williams, W. M., Snook, S. A., & Grigorenko, E. L. (2000). *Practical intelligence in everyday life*. New York: Cambridge University Press.
- Sternberg, R. J., & Wagner, R. K. (1986). (Eds.) *Practical Intelligence: Nature and origins of competence in the everyday world*. New York: Cambridge University Press.
- Sternberg, R. J., & Wagner, R. K. (1993). The G-ocentric view of intelligence and job performance is wrong. *Current Directions in Psychological Science*, 2(1), 1–5. doi:10.1111/1467-8721.ep10770441
- Vernon, P. E., & Parry, J. B. (1949). *Personnel selection in the British Forces*. London: University of London Press.
- Waterhouse, L. (2006). Multiple Intelligences, the Mozart Effect, and Emotional Intelligence: A Critical Review. *Educational Psychologist*, 41(4), 207–225. doi:10.1207/s15326985ep4104_1
- Wolf B., Momirović K., & Džamonja Z. (1992). *KOG 3 – Baterija testova inteligencije [KOG 3 – battery of intelligence tests]*. Beograd: Centar za primenjenu psihologiju.
- Živanović, M., Bjekić, J., & Opačić, G. (2018). Multiple solutions test Part I: Development and psychometric evaluation. *Psihologija*, 51(3), 351–375. <https://doi.org/10.2298/PSI161031003Z>

RECEIVED 05.02.2017.

REVISION RECEIVED 15.01.2018.

ACCEPTED 24.01.2018.

Test višestrukih rešenja Deo II: Evidencija o konstrukt i prediktivnoj validnosti

Marko Živanović¹, Jovana Bjekić² i Goran Opačić¹

¹*Odeljenje za psihologiju, Filozofski fakultet, Univerzitet u Beogradu, Srbija*

²*Institut za medicinska istraživanja, Univerzitet u Beogradu, Srbija*

U ovom istraživanju dopunjujemo podatke o Testu višestrukih rešenja (eng. Multiple solutions test – MST) ispitujući konstrukt i prediktivnu validnost instrumenta. Za razliku od konvencionalnih matrica gde se traži samo jedno rešenje, MST pred ispitanike postavlja tri tipa problema, tražeći od njih da reše matricu navodeći najbolje, drugo najbolje i najmanje tačno rešenje. U istraživanju je učestvovalo ukupno 235 ispitanika (starosti $M = 22.65$, $SD = 3.33$, 199 ženskog pola). Konstrukt validnost svakog od zadataka MST testirana je u odnosu na bateriju intelektualnih sposobnosti KOG9 ($N = 156$), dok je prediktivna validnost pojedinačnih zadataka i postignuća na testu u celini testirana u odnosu na školsko postignuće koje je operacionalizovano kao prosečna ocena ($N = 235$). Rezultati su pokazali visoke korelacije između zadataka, ali i ukazali na specifičnosti svakog od njih. Dodatno, dobijene su i razlike u težinama između zadataka, pri čemu je zadatak nalaženja najmanje tačnog rešenja bio najteži, potom zadatak nalaženja drugog najboljeg rešenja i naposljetku zadatak nalaženja najboljeg rešenja. Test je pokazao zadovoljavajuću konvergentnu validnost u odnosu na testovne markere Gf/Gv iz KOG9 baterije. MST je pokazao i prediktivnu validnost, kao i inkrementalnu vrednost alternativnih zadataka u odnosu na standardni (kada se školski uspeh uzme za kriterijum), kao i inkrementalnu validnost u predviđanju prosečne ocene u odnosu na KOG9 bateriju. U celini, MST se pokazao kao valjan instrument za procenu inteligencije, a njegovi alternativni zadaci pokazuju potencijal da budu koristan dodatak standardnim matricama sa jednom vrstom rešenja.

Ključne reči: Test višestrukih rešenja (MST), figuralno rezonovanje, fluidna inteligencija (Gf), konstrukt validnost, prediktivna validnost

© 2018 by authors



This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution ShareAlike 4.0 International license

Appendix

95% Confidence intervals (lower and upper bounds) for correlations between individual tasks of Multiple solutions test (and total score) and subtests of KOG9 ($N = 156$)

	<i>perceptive processor</i>			<i>serial processor</i>			<i>parallel processor</i>		
	it1	cf2	gt7	al4	alf7	gsn	d48	it2	s1
<i>the best</i>	.01-.31	-.02-.29	.01-.31	-.12-.19	.21-.48	.05-.35	.07-.36	.16-.45	.13-.42
<i>the second-best</i>	-.02-.29	-.01-.21	-.10-.22	-.16-.15	.12-.41	-.01-.31	.01-.31	.08-.37	.07-.37
<i>the least accurate</i>	-.15-.16	-.12-.19	-.05-.26	-.11-.20	.14-.43	-.04-.27	-.07-.24	.02-.32	.02-.33
MST full-scale score	-.06-.25	-.09-.22	-.04-.27	-.14-.17	.17-.45	.01-.32	-.01-.30	.10-.39	.08-.38

Note. it1 – Identical figures test; cf2 – Hidden figures test; gt7 – Test of form perception; al4 – Synonyms-antonyms test; alf7 – Analogies; gsn – Test of synonyms; d48 – Domino test; it2 – Test of spatial abilities; s1 – Visual spatialization test.

*Lower bounds for correlations between *the best* solution and it1 and gt7, *the second-best* solution and d48, and MST full scale and d48 are .003, .003, .005, and -.003, respectively, but are rounded at .01 for consistency in data presentation.