Coll. Antropol. **34** (2010) 3: 1015–1026
Original scientific paper

Latent Structure of Raven's Colored Progressive Matrices

Stanislav Fajgelj[1], Gustav Bala[2] and Ratko Katić[3]

[1] Faculty of Law and Business Studies, Novi Sad, Serbia
[2] Faculty of Sport and Physical Education, Novi Sad University, Novi Sad, Serbia
[3] Faculty of Kinesiology, Split University, Split, Croatia

A B S T R A C T

The main goal of the study was to determine the constructive validity of Raven's Colored Progressive Matrices by means of item factor analysis. The most important topic within this objective is to determine the test dimensionality, since many authors report on finding several significant primary factors. The study included 2334 children aged 4 to 11. Several types of factor analysis were used in order to obviate the influence of technique on the results. Our data suggested 3 or 4 first order factors. Based on the usual McDonald definition, the primary factors obtained could be considered difficulty factors in the majority of cases. The necessary number of factors on age subsamples, extracted by parallel analysis, was between 3 and 5. Factor structure on age subsamples indicated the youngest ages, 4 and 5, to be essentially different from the older ones. This difference was identified as underdevelopment of the goal management mechanisms.

Key words: *intelligence, Raven's Colored Progressive Matrices, preschool children, primary school children, g-factor*

Introduction

Definition and assessment of the anthropologic status dimensions provide a basis for anthropologic studies of the psychosomatic development of children and adolescents. In line with the theory of integrated development (Bala and Katić, 2009)[1], relations between morphological characteristics and motor abilities (Bala et al., 2009)[2], relations between morphological-motor development and cognitive abilities (Katić, 1977; Bala et al., 2009)[3,4], and relations between conative characteristics and motor abilities (Katić, 1977)[5] are analyzed. In the present study, metric characteristics of Raven's Colored Progressive Matrices, a measuring instrument for assessment of cognitive abilities were analyzed.

Raven first published his progressive matrices test in 1938, with a subtitle Perceptual Intelligence Test. The revised version was published in 1956[6] and this test is known today as Standard Progressive Matrices (SPM). The test is non-verbal, intended for measuring g-factor according to the classic Spearman terminology (Spearman, 1927, 1950)[7,8]. Since the test is non-verbal, it is considered to be more culture fair, and therefore it is often found in the batteries of cross-cultural tests. SPM are intended for general, average population (»aged 6 to 80«), have five sets of 12 items each; sets A and B have 6, and sets C to E 8 solutions offered. Sets within the test and items within the sets are arranged according to difficulty.

Later, Raven published two more main versions of his progressive matrices, which followed the concept of standard matrices in all respects. The Colored Progressive Matrices (CPM) test was constructed by Raven in 1947 as an alternative to SPM and was intended for children aged 5 to 11, special populations, and those who do not speak English. Advanced Progressive Matrices (APM) are intended for the above-average population above the age of 12. The current version of APM was published in 1962. It is considerably more difficult than Standard Matrices and its goal was, among other things, to compensate for the influence of Flynn effect, but also to create a test suitable for triage of 20% of those who rank highest on cognitive achievement (Cotton, Kiely, Crewther, Thomson, Laycock, & Crewther, 2005; Lynn, Allik, & Irwing, 2004; Bors & Vigneau, 2003)[9–11].

All progressive matrices, due to the large number of their applications and published papers, served in the debate about Flynn effect (Neisser, Boodoo, Bouchard, Boykin, Brody, Ceci, et al., 1996)[12]. Pind, Gunnarsdottir, & Johannesson (2004)[13] mention the results of two stan-

Received for publication May 25, 2010

dardizations of SPM published in 1942 and 1992. Comparing different cohorts of included participants, the authors discovered an increase in median from 25 to as much as 55 test scores (and maximal score is 60). The highly above-average participants from 1942 would nowadays be classified as mentally dull. Recently, a standardization of CPM has been conducted on children in Voivodina (Serbia) (Fajgelj, Bala, & Tubić, 2007)[14] and, in comparison to the norms from test manual, an increase in percentile scores from 1 to 6 points at younger and middle ages was determined, at various cognitive levels. Because of the ceiling effect, the progress was reduced at higher cognitive levels and in older ages. Also, the progress was smaller at lower cognitive levels.

Test Reliability and Validity

Low reliabilities at younger ages, both of the type of internal consistency, and test-retest, recurred in all standardizations familiar to us, including the paper of Fajgelj et al. (2007)[14]. Raven (n.d.) determined test-retest reliability of 0.80 for CPM at the age of 9.5, and 0.60 at the age of 6.5 (however, on a rather small sample). Rushton, Skuy, & Bons (2004)[15] state that for APM typical alpha coefficients range from 0.75 to 0.86. On SPM, the test-retest reliability obtained was as high as 0.96 (Williams & McCord, 2006)[16]. Cotton et al. (2005)[9] think that the average reliability of CPM, reported in papers, equals 0.80. In our research (Fajgelj et al., 2007)[14], a stable coefficient alpha of 0.91 was obtained in a sample of 2300 children aged 4 to 11.

In his manual, Raven states the correlation of CPM with Terman-Merill L scale of 0.66, and with Crichton Vocabulary Test of 0.65 (at the age of 9). Pind et al. (2004)[13] state that Raven established the correlation of SPM and tests of academic achievement between 0.20 and 0.60, where higher correlations were obtained in the field of mathematics and natural sciences, and lower in the field of languages. In a population of pupils from Iceland, the authors established correlations between SPM and mother tongue: 0.38 (10-year-old), 0.64 (13-year-old) and 0.53 (16-year-old), and between SPM and mathematics 0.50, 0.75 and 0.64. These correlations are higher than in other studies with SPM, and are similar to correlations obtained between SPM and Wechsler's test WISC-III.

Rushton et al. (2004)[15] state that prediction of school achievement and success at work, based on APM scores, is between 0.20 and 0.50. Williams and McCord (2006)[16] mention Burk's study, which determined the correlation of SPM with WAIS of 0.57. However, the authors also mention the study by McLaurin and Farrara, who did not establish a significant correlation between SPM and school achievement.

Other Findings about the Test

In test manual, Raven does not mention any differences between boys and girls. In Australia, Cotton et al. (2005)[9] obtained a significant difference in favor of boys only at the age of 6, but the size of groups was 25. On Iceland, Pind et al. (2003)[13] did not obtain a significant effect of sex on SPM in school children aged 6–16, or significant interaction of age and sex. In the study by Fajgelj et al. (2007)[14] on 2300 children aged 4–11, there was no significant sex effect or significant interaction of sex and age. Mackintosh and Bennet (2005)[17] found significant sex differences in some of Raven's items (males were better at items dominated by the rule of addition and subtraction, Carpenter, Just, & Shell, 1990)[18], while in other items the differences were not significant. Generally, it can be considered that there are no significant sex differences in the level, but differences can be expected in parts of the test and latent structure.

Bors and Vigneau (2003)[11] found a statistically significant effect of practice (training) on APM, i.e. of about one point after each repetition, accompanied by an increase in reliability (all test-retest reliabilities were above 0.85). That is in keeping with the rare previous studies of practice effect on SPM, which ranged from 0.4 points for the young, 1.2 for the middle-aged to 2.9 for the old, but these differences were not statistically significant.

Williams and McCord (2006)[16] compared the paper and computerized versions of SPM and concluded that both versions provided equivalent results and that anxiety had the same influence on scores in both.

What Do Raven's Progressive Matrices Measure?

Raven intended to create a test that measures the »pure g«, and when the test yielded its first results, Spearman accepted it as such as well. Since then, the interpretation of intelligence has undergone a long and intensive development, during which not only new interpretations of that ability arose, but also qualitatively different approaches to its studying have emerged. The consequence is that many papers dedicated to analysis of latent structure of Raven's test use different theories about the structure of intelligence, and accordingly, name factors differently. For example, in the domain of theories accepting the existence of general intelligence, Raven's tests are marked in different ways – for instance, as tests of g-factor or as test of fluid intelligence, according to Cattell-Horn terminology (Cattell, 1963, 1971; Horn, 1979)[19–21].

Raven attributed large importance to perception, as an ability to »create order from chaos«. In keeping with that commitment of the experienced Raven, a large number of papers confirmed that one perceptive factor does occur. It is sometimes named gestalt factor, sometimes »the speed of completing the whole«, and sometimes visual or visual-spatial factor. Besides this factor, the factor of reasoning according to analogy, that is, inductive inference is usually also singled out [I or R, according to Cattell, according to Ivić, Milinković, Rosandić, & Smiljanić (1978)[22]]. DeShon et al. (1995)[23] and Lynn et al. (2004)[10] also derived the factor of verbal reasoning (in more difficult items in APM and SPM).

Van der Ven and Ellis (2000)[24] used Rasch's model for assessment of SPM dimensionality and obtained two dimensions, gestalt continuity and analogous reasoning in set B, and analogous reasoning and coping in set E. Lynn et al. (2004)[10] have established that Van der Ven's and Ellis's factor of analogous reasoning is divided into two factors. More difficult items from B and all items except for the most difficult ones from C and D belong to one factor, while the last items from C and D and all items from E belong to the other. Mackintosh and Bennett (2005)[17] provide an overview of the analyses performed so far and claim that in SPM set A and the first items in B measure the perceptual factor, while the remainder measures analytical processes, except for E, which also measures something else. In some studies, it seems as if the fission of some of these factors into two has occurred, and some other factor has appeared in the others.

What are the opinions on the unidimensionality of Raven's tests? Van der Ven and Ellis (2000)[24] find sets A, C and D in SPM to be unidimensional, and sets B and E are not. Lynn et al. (2004)[10] conclude that, although SPM yields three factors in the first order, g-factor is obtained in the second order. Vigneau and Bors (2005)[25] confirm that APM are unidimensional, especially on the basis of Rasch's analysis. Mackintosh and Bennet, on the basis of the sex differences obtained, think that SPM and APM measure several different mental processes. DeShon, Chan, & Weissbein (1995)[23] argue that APM measure two different processes: visual-spatial and verbal-analytical. In our previous analysis, we concluded that the test was monofactorial in the second order, while in the first order it yielded several specific factors (Fajgelj et al., 2007)[14].

Mostly, one gains the impression that findings and opinions about the dimensionality of Raven's Progressive Matrices do not converge, on the contrary. It is noticeable that there is rarely an author who omits to say that Raven's tests are a »typical« or even »the best« measure for *g*, but data more often bring about doubts than confirmation. Meta-analysis of papers about the dimensionality of Raven's matrices would definitely be very useful, but in the lack of a good indicator of effect size, it must be based on some kind of qualitative integration and triangulation of different data.

Such a situation makes the job difficult for every researcher of dimensionality of Raven's matrices. If he does not want his findings to be just one of many, he must make a new breakthrough in theory of intelligence. Of course, such a result is difficult to achieve, but it seems very probable that without it, that is, without the solution to the dilemma of *g*-factor, it will not be discovered whether any intelligence test measures *g*, including Raven's CPM as well.

There are some aspects of the discussion about whether *g* exists and what it consists of, which we deemed to be of importance for the purpose of this study. One aspect attracts a lot of attention, and it refers to possible constructive elements that form *g*, or, more precisely, mental processes that participants use during solving intelligence tests. It is usually spoken about cognitive processes, memory (primarily working), learning, perception, etc. Some authors do not see the logic in splitting psychometric *g* into its integral parts, when it was already formed by joining the elements (for example, Carroll, 1991)[26]. Other authors think that it is natural for *g* to consist of some elementary cognitive processes (for example, Kranzler, 1993, and Detterman, 2002)[27,28].

Since many people primarily see the process of inductive generalization in solving the tasks in the tests such as CPM, it would be good if we knew how inductive generalization functions in a cognitive way, and not in terms of formal logic. Based on the papers of Sloutsky and Fisher (2005)[29] and Heit and Hayes (2005)[30], one gains the impression that it is not a unitary cognitive operation, but probably starts with perception, and only in the end there is induction, that is, generalization.

Finally, there have been a number of studies of processes, rules or strategies necessary for solving Raven's tests. Carpenter, Just and Schell (1990)[18] determined 5 basic rules for solving APM, where most of the items can be classified according to the main rule they are based on. DeShon, Chan and Weissbein (1995)[23] thought that tasks in APM can basically be divided into visual-spatial and verbal-analytical, and therefore they defined 6 visual-spatial rules for solving the tasks and 4 verbal-analytical. These and other similar studies did, indeed, search for mental operations, but they ended up in a typology of items from which we conclude that they were more of the analysis of stimuli than the analysis of (brain) cognitive functions.

In that respect, the research of Freund, Hofer and Holling (2008)[31] is completely clear. Based on the analysis of the tasks of the type of progressive matrices, they suggest a list of rules for the construction of such items: complete addition, addition of one element, addition of two elements, progression of position, and progression of form. Freund et al. claim that by combining these rules, varying their number, the number of elements in the matrix, the nature of elements and sequence of rules, the items of good and well-known psychometric properties can be obtained.

When the research of Freund, Hofer, & Holling (2008)[31] is compared to the work of Carpenter et al. (1990)[18] and other similar papers, it becomes pertinent to wonder how much of what those authors name cognitive processes really takes place in the participants' heads. From the taxonomy of items according to formal rules, it does not follow that the participants use the same rules in solving – from the viewpoint of cognitive neuroscience, cognitive psychology and metacognition. For example, it seems to us that the list of rules for solving of CPM (and even SPM, that is, all the tests which do not contain the instruction, an example and practice) should also include »establishment of rules in first items«, as well as the rule of establishing the hypothesis based on matrix and testing the hypothesis on the basis of the solutions offered. Anyway, what would be more useful to us is the taxonomy of participants with regard to the mental processes engaged than taxonomy of items.

From the viewpoint of cognitive neuroscience, the most interesting currently is the localization of cognitive functions in the brain. It reveals to us that even some old and well-known cognitive operations such as inductive reasoning do not have the same definition (or content) for neuroscientists. If you base your observations on PET and fMRI scans, cognitive operations look different than when you analyze these operations logically or psychologically.

Very interesting are Goel's findings (2005)[32] on fMRI scans of participants while solving syllogisms, i.e. reasoning deductively. The author worked with two types of syllogisms: those with content (all dogs are pets / all poodles are dogs / all poodles are pets) and without content (all P are B / all C are P / all C are B). Reasoning on the material with content activated the zones in the left middle and upper temporal lobe (BA 21/22), left temporal lobe (BA 21/38) and left lower frontal lobe (BA 47). These are essential systems for speech and memory. Reasoning on the material without content activated on both sides occipital (BA 19), on both sides upper and lower parietal (BA 7) and bilaterally dorsal (BA 6) and lower (BA 44) frontal lobe. Such a pattern of activation usually occurs during internal representation and manipulation of spatial information, inferring about geometric shapes and certain kinds of mathematical reasoning, which involves the approximation of numerical values.

Mackintosh and Bennet (2005)[17] performed brain scanning during solving SPM and found that easier items activated the right hemisphere, especially the right frontal lobe, and more difficult (analytical) the left, that is, the left frontal lobe. There is a considerable amount of similar findings that the tasks whose differences seem minor to us activate completely different areas of cortex. Is that the end of *g*-factor? At first glance, it seems so, because specific and different brain structures participate in reasoning (and reasoning lies behind solving the tasks in intelligence tests). In other words, does biological perspective indicate the existence of special »intelligences«?

The belief of DeShon et al. (1995)[23] that the basic processes in APM can be divided into visual-spatial (perceptive) and verbal-analytical is shared by a large number of authors and supported by a great deal of findings. These two processes strongly remind of division into the dorsal and ventral visual system, i.e. »what« and »where/how« system (Ungerleider and Mishkin, 1982, Milner and Goodale, 1995, Glover, 2002, 2004)[33,36].

Chabris, Jerde, Woolley, Hackman, & Kosslyn (2006)[37] found that men, students of natural sciences and persons who like computer games prefer spatial visualization (visual-spatial intelligence), and women, students of humanities and artists prefer object visualization (verbal intelligence). Persons with spatial style are better in mental rotation and tasks with labyrinth, and persons with object style are better in complex recognition of objects. Based on higher correlations, the authors conclude that spatial visualization is probably a more unitary and homogeneous ability.

Basically, the old psychological division into verbal and non-verbal intelligence seems to have a neuro-physiological basis. Under the condition that non-verbal intelligence is called »visual-spatial« and by that we bear in mind the connection of visual-spatial cognition with planning and control of motor skills – perception demands action (according to Gallese, 2007)[38].

From the aspect of discussion about the nature of *g*-factor, the key question is the cooperation of these two cognitive systems. The findings so far regarding their cooperation are contradictory. For example, it was established that visual tasks can be solved by using visual-spatial strategy, verbal-analytical strategy or both, depending on the task and the participant. On the other hand, studies indicate that if visual-spatial tasks are solved competitively, visually and verbally, then solving process is aggravated (for instance, the studies with the so-called overshadowing of DeShon et al., 1995, Schooler and Engstler-Schooler, 1990)[23,39]. In this contradiction, working memory, the way we understand it according to Baddeley's model, definitely plays an important role, where the way of representation that uses working memory more makes the other one less efficient.

Do these two cognitive systems cooperate and how does our cognition act as a whole at all? If we remind ourselves of the findings of Goel (2005)[32], we see that the participants in the tunnel of PET or fMRI scanner gave correct answers both to the tasks »with content« and to the tasks »without content«. It is correct that anatomically different structures were activated in the brain, but all of them had as their *goal* the correct answer. Deductive reasoning, studied by Goel *via* different syllogisms, is something we perform successfully, activating different neurological mechanisms. Whether correct answers are arrived at by the brain correctly arranging the tasks, or by integrating the results of processing from different zones, or in both of these ways, is not known for now. However, such a »controller system« must exist. From the viewpoint of cognitive neuroscience, it could contain the »centre for *g*«.

We thought that very important for this paper is the question of planning and coordination of mental activities. According to cybernetic models of intelligence, one of the components of central information processing is the component of planning, decision-making and goal management (Embretson, 1999; Zarevski, 2000)[40,41]. The process of planning and decision-making definitely involves working memory and is burdened by task complexity, that is, the number of relations the item contains. Also, it is probably connected with (self)criticism, already introduced by Terman in the definition of intelligence (according to Zarevski, 2000)[41], maybe also with test wiseness, training, that is, practice in solving tests, reacting to time limitations in test administration, etc.

A large number of authors (and even those mentioned here) use the terms such as »goal management«, »metacognition«, »control«, »planning«, etc. Observed either from psychological or neuro-physiological point of view, it is clear that some integrative mechanisms or central controllers must exist. Neuroscientists seem to be inclined to locate these advanced mechanisms in frontal, or, more precisely, in prefrontal lobes. Maybe metacognition (understood as management) is that process or part of cognition which is responsible for covariance of intelligence tests, that is, for psychometric *g*.

Finally, if we stick to the standard methodology of confirmatory research, our task could be understood as testing the hypothesis of unidimensionality of CPM. Unfortunately, we think that unidimensionality, that is, congenerity, that is, monofactorial structure is difficult to determine, even when it exists. Determining unidimensionality is, really, essentially contradictory to the idea of factor analysis, and even to any other multivariate analysis. Multivariate analysis is a powerful tool when we want to perceive the separate clouds of dots in the space of variables or cases. Multivariate analysis does not provide good answers to the questions about the relation between the sample and population, either because sampling distribution is not developed, or because »structural hypotheses« are different from statistical. For this reason, structural answers to structural questions are often based on subjective estimation. For example, if you perceive that there are two clouds of dots, that is, two factors, which are in addition in keeping with theoretical expectations, your task is accomplished. Based on such a perception you will theorize, set further hypotheses, or make structural conclusions. If you perceive your variables as being grouped in two or more wholes, that is a valuable finding both heuristically and epistemologically.

On the other hand, if you want to prove that there is no more than one cloud of dots in your space, it is a completely different way of looking at things. The situation is very unfavorable. In research practice, it never happens that all your variables/items form one and only one whole. In that case, if you want to prove unidimensionality, you must prove that some separate clouds of dots are artifacts or a coincidence, and that only one cloud is the »main« one. To put it simply, you must prove that something does not exist. That is no longer a structural question (because you reject the existence of a structure, everything is one), but the question of quantification of dimensionality. There are different attempts at quantification of unidimensionality, but their successfulness is insufficient. After all, if we had a good quantification of unidimensionality, we would deal with the dilemma of existence of *g*-factor to a considerably smaller degree.

Method

Subjects

The sample of participants comprised 2334 children (1252 boys and 1082 girls), aged 4 to 11, who attended preschool institutions and primary schools (the first four grades) in towns in Voivodina (a province in the north of Serbia). Testing was performed during April and May 2006 in the premises of the aforementioned institutions. Testing was performed under the surveillance of psychologists, and it took 20 to 45 minutes to solve the test tasks, depending on age and individual characteristics of participants. Since the sample of participants included children below the age of five, and even four, the help of examiner was necessary in writing the answers to the items, by which test administration assumed the character of individual testing. Other participants had group testing.

Research was conducted within a wider project, in which numerous measurements on small children were performed, from psychological, over motor, to anthropometric. For this reason, the institutions selected in the sample had to have a gym in their facility. The choice of schools was convenient, which means that the sample comprised better equipped schools from urban areas. In order to check whether the choice of »better« school and preschool institutions could have jeopardized the representativeness of the sample, we used the questionnaire about the socioeconomic status, which was completed by parents and contained different questions about the education of extended family and closest friends, then about the professional and financial status of the family. Out of 20 questions we formed the index of socioeconomic status in two ways: as a simple summation score and as the score on the first principal component. The correlations of the total score on CPM and these two indices were 0.046 and 0.030, respectively (neither was significant). Based on that, we think that bias in choosing schools did not jeopardize the representativeness of the sample of participants with respect to intelligence.

Table 1 presents basic indicators of the total score on CPM by age groups and total. Besides the moments of distribution, the sizes of age groups are provided, as well as internal consistence reliability (coefficient α).

Test

Colored Progressive Matrices consist of three sets of 12 items each: A, Ab and B. Within each set, items are (approximately) arranged according to difficulty, and it is similar with sets – set B is the most difficult. In the upper

TABLE 1
BASIC INDICATORS OF TOTAL SCORE ON CPM ACCORDING TO AGE AND TOTAL (Fajgelj et al., 2007)

Age	$\overline{X}$	SEM	SD	N	Percent	Reliability α
4	14.88	0.33	3.61	116	5.0	0.594
5	18.28	0.24	4.57	341	14.6	0.753
6	20.79	0.25	5.75	512	21.9	0.849
7	24.38	0.28	5.79	421	18.0	0.862
8	27.33	0.36	5.53	229	9.8	0.853
9	27.94	0.37	5.89	249	10.7	0.875
10	29.30	0.34	5.41	250	10.7	0.861
11	31.27	0.30	4.52	216	9.3	0.864
Total	24.06	0.14	7.10	2334	100.0	0.908

part of each task there is a picture in which, in the lower right corner, one part is missing, and below it there are six suggested solutions, out of which only one fits exactly the missing part of the picture. In set A, the picture is a lightly colored field in which geometric figures are drawn, primarily lines and dots. In sets Ab and B, the image consists of four separated figural elements – three are shown, and the fourth is missing. Therefore, the elements in sets Ab and B are presented in the form of 2 x 2 matrix and after this all tasks of that type got the name »matrix completion tasks«. Regularities that the participant has to discover in the picture are not in any way connected with background color, but the purpose of color is motivational – to emphasize the problem and preserve the attention of the child.

Results

Number of factors

The usual Guttman-Kaiser criterion »higher than one« yielded eight factors in full correlation matrix. Cattell's »scree test« (Cattell, 1966)[42] suggested four or five factors. According to the criterion from PRELIS (which is based on RMSEA), the optimal number of factors would be three. As will be seen, two models in confirmatory factor analysis, with four and with three factors, yield approximately equal fit indicators. Finally, logical analysis of the factors obtained indicates that with the increase in the number of factors (with one to four factors) certain groupings of items are obtained that make sense intuitively, but not essentially-rationally. For this reason, it seems that the solutions with three or four factors are optimal.

Exploratory factor analysis of the test

In this study, we performed exploratory item factor analysis of Raven's Colored Progressive Matrices in several ways. Our goal was to establish the latent structure of the test, endeavoring as much as possible to eliminate the »method factor«, that is, the influence of the technique of factor analysis on the results. Table 2 presents all the configurations of factor-analytical procedure that were applied and their basic parameters. Among others, we used Testfact, the program of Bock et al. (Du Toit, 2003; Wilson, Wood, & Gibbons, 1991; Knol & Berger, 1991)[42–45]. This program is considered appropriate because it can take as an initial matrix the matrix of tetrahoric correlation coefficients, which is »smoothed« in such a way to be positively definite. In addition to this, this program uses Minres method as a technique of initial factoring, which helps us increase the variety of the factor-analytical techniques applied in order to be able to perceive the possible influence of the technique on the results. Besides Minres, for initial factoring the method of maximum likelihood (ML, within Lisrel) and the standard method of principal components were used. Besides oblique solution (promax), we also tried the orthogonal (varimax) because other authors also used varimax and because we do not reject the thesis that the orthogonal solutions are better for clear delimitation of constructs. Finally, we presented the two basic variants: with three and with four factors. The solution with three factors has an advantage in parsimony, but we think that the fourth factor illustrates well what happens when the number of factors is increased. Namely, it happens that certain groups of items form separate factors based on the similarity of graphic content. We cannot be certain whether grouping of cognitive processes lies behind it, or these items should be treated as testlets in which participants demonstrate the same solving method.

Later in the text we label the models we adopted for testing in confirmatory factor analysis as »CFA«. The model »CFA4« was formed by generalizing the results of the first four solutions of exploratory factor analysis (item B2 had to be included both in the first and the fourth factor). »CFA3-TF« is a three-factorial model derived on the basis of the results of Testfact, and »CFA3« is a three-factorial model derived from the solution VI.

Finally, Table 3 presents a short review of factoring and significant additional information important for factor interpretation. In the second row of Table 3, a short description of factors is provided. Then follow variances (lengths) of factors for each solution. The lower part of Table 3 contains intercorrelations of factors in all oblique solutions.

Confirmatory factor analysis

Since we saw that first order factor analysis yielded three to four well-defined factors, we decided to set a hierarchical model, in which the first, main and basic construct of measurement would appear on the second level.

TABLE 2
TYPES OF FACTOR ANALYSIS USED, BASIC CHARACTERISTICS AND NUMBER OF FACTORS

	Program	Matrix	Specified number of factors	Extraction	Rotation
Solution I	Lisrel	covariances	4	ML	varimax
Solution II	Lisrel	covariances	4	ML	promax
Solution III	SPSS	r	4	principal components	varimax
Solution IV	SPSS	r	4	principal components	promax
Solution V	Testfact	r_{tet} (smoothed)	3	minres	promax
Solution VI	Lisrel	covariances	3	ML	promax

TABLE 3
BASIC DESCRIPTION OF EXTRACTED FACTORS, THEIR VARIANCE AND INTERCORRELATIONS

		Factor 1	Factor 2	Factor 3	Factor 4
		Basic description of factors			
		Items of medium difficulty from all sets	The most difficult items from all sets, primarily set B	The easiest items in set A (the easiest items in the test)	Items Ab1, Ab2, Ab3, B1 (the same four figures, one is missing)
Solution I		4.99	3.41	1.66	1.76
Solution II		5.21	3.02	1.53	1.46
Solution III		5.72	4.18	2.16	2.33
Solution IV		5.85	3.76	2.03	2.15
Solution V		8.04	6.39	4.91	
Solution VI		5.24	3.21	1.76	
Factor 2	II	0.6			
	IV	0.5			
	V	0.7	1.0		
	VI	0.6			
Factor 3	II	0.4	0.2		
	IV	0.4	0.1		
	V	0.7	0.6	1.0	
	VI	0.5	0.2		
Factor 4	II	0.4	0.1	0.4	1.0
	IV	0.3	0.2	0.2	

The upper part of the table contains factor variances, and the lower part correlations

In order to do that, we formulated a structural model in Lisrel and tested it against the data obtained.

Specifically, we defined the general factor as a »ksi« variable in Lisrel terminology, that is, as an (exogenous) independent latent variable. First order factors were defined as »eta« variables, that is, as (endogenous) latent dependent variables, while the items played the role of Y-variables, that is, dependent observables. In order for the model to be identifiable, the variance of general factor was fixed on 1, in keeping with recommendations of Kline (2005)[46] and Thompson (2004)[47]. In the case of CFA4 and CFA3 it was also necessary to fix the structural coefficient from the general factor to the first primary factor at 1, while in the model based on Testfact it was not necessary to introduce this limitation. That is how we got three models: CFA4, CFA3-TF and CFA3, described in the previous chapter.

Alternatively, we also tested the models which did not contain a general second order factor, but only first order factors, among which correlations were allowed.

The third type of model was the model of measuring in which the items were dependent observables, and one latent variable (as a possible general factor) was an independent endogenous variable.

Table 4 presents basic results of confirmatory factor analysis. The selected fit indicators are presented in the first four columns. For hierarchical models, the last columns contain the correlations of primary factors with second order factors (from a completely standardized solution). It is noticeable that fit indicators for all models are very similar, except for the non-hierarchical model with a general factor which fits data worst.

For either of the models that were tested it was not possible to keep the null hypothesis according to chi-square test. All the χ^2 obtained were highly significant, and the quotient of χ^2 and the number of degrees of freedom was around 4. In the special issue of the magazine Personality and Individual Differences (Vernon & Eysenck, 2007)[48], an extensive debate has been published on whether χ^2-test must be a basic fit indicator in the SEM field or approximate fit indicators can be used equally or more justifiably. Generally speaking, a predominant attitude in the debate was that approximate indicators are useful and necessary, although insufficient for a definitive acceptance of models. In choosing fit indicators, we relied on the recommendations of Hu and Bentler (1999)[49] to use two indicators. One of those that should be definitely taken into consideration is »Standardized Root Mean Square Residual« (SRMR), which has a direct interpretation as an average correlation of residuals. As the second indicator we chose »Root Mean Squared Error of Approximation« (RMSEA), since we had a large sample. We also used »Parsimony Normed Fit Index« (PNFI), which besides evaluating the model with respect to the baseline model of uncorrelated variables, corrects the obtained value by the complexity of the model (expressed by the number of degrees of freedom).

TABLE 4
RESULTS OF CONFIRMATORY FACTOR ANALYSIS FOR MODELS WITH 4 AND 3 FACTORS – FIT INDICATORS AND CORRELATIONS WITH THE GENERAL FACTOR

	RMSEA	SRMR	PNFI	χ^2/df	Factor 1	Factor 2	Factor 3	Factor 4
			One second order factor					
CFA4	0.041	0.061	0.899	4.52	0.707	0.645	0.536	0.598
CFA3-TF	0.038	0.041	0.894	3.94	0.939	0.660	0.866	
CFA3	0.040	0.056	0.892	4.41	0.707	0.768	0.565	
			Without higher order factors					
CFA4	0.037	0.037	0.899	3.92				
CFA3-TF	0.038	0.041	0.894	4.40				
CFA3	0.039	0.042	0.893	4.09				
One first order factor	0.071	0.056	0.868	12.45				

Finally, we will mention the quotient χ^2/df as a rough overview of testing the null hypothesis of model fit. Hu and Bentler think that SRMR and RMSEA should not be bigger than 0.09 and 0.06, respectively.

We consider justified the criticism of »golden rules«, that is, firm thresholds, but we think that it does not apply only to approximate indicators, but also to the levels of significance of tests of null hypothesis. In this sense, some conventional thresholds are useful for the sake of comparability of findings, as well as for direct comparison of alternative models. Tomarken and Waller (2003)[50] recommend that several models should always be tested, or at least that such models should be explicitly discussed in the paper. Following this sound logic, besides the hierarchical model with one higher order factor (which was our basic goal), we also tested the models with only the first order factors (without a general factor), as well as the model with only one factor.

Based on χ^2-test, none of the models could be accepted, but if, with regard to sample size, we rely on two approximate indicators (SRMR and RMSEA), then all models could be accepted except for the non-hierarchical model with one factor. The differences in PNFI are neglecting (they should be at least 0.06 to 0.09), except for the model with one first order factor, where PNFI is somewhat lower than others, but not sufficiently lower to be considered different from other models.

The models with the explicitly introduced general factor do not explain data better than the model without that factor, but all models describe the data satisfactorily well. It can be seen that correlations of general factor with primary factors are always high, especially in Testfact. Generally speaking, the model which was based on solutions from Testfact defines the general factor best, partly because of the good fit, partly because it converged without additional constraints, partly because Testfact solutions yielded the highest intercorrelations of primary factors. Still, factor matrix in CFA3-TF has the same number of loadings smaller than 0.30 as the one in CFA, and that happens mostly with the easiest items in series.

CFA models we analyzed had to be modified by some additional conditions in order to achieve a satisfactory fit. The modifications regarded allowing covariance of errors between certain items, primarily the initial items in sets. In other words, there existed covariances between items that are not part of the model, more precisely, which are not explained by the set latent variables. A possible explanation for this is that the child builds a solving strategy in initial items in the set, even a wrong one. Raven initially noticed that in certain pairs of items (although not the ones from the beginning of sets) it happened that, no matter which one of pair members was placed as the first, it would turn out to be more difficult. Then, in initial items in the set, Raven noticed that if younger children started with a certain »system« of solving, even a wrong one, they tended to continue to apply it in other tasks in the set as well. In brief, our finding, as well as Raven's initial observations, leads us towards the conclusion that there are mutual influences of the initial items in the set.

Parallel analysis

We also attempted to test the unidimensionality of CPM using the so-called parallel analysis (Thompson, 2004)[47]. We used SPSS macro for parallel analysis (O'Connor, 2000)[51], with certain changes that were necessary since our matrices were singular. We selected the variant of the program that creates random variables by data permutation, which preserves the shape of distributions of raw variables/items, and therefore also the possible false correlation patterns. Parallel analysis suggested that 4 factors should be preserved on the whole sample. On age subsamples, from 4- to 11-year-old, one should keep: 4, 5, 4, 5, 4, 5, 3 and 5 factors. Figure 1 shows the basic findings, where the diagrams for ages 7 to 10 are left out, since they are essentially similar to the one for age 11.

The smaller variance of the first principal component and less steep eigenvalue curve at lower ages were at first ascribed to the lower achievement and lower reliability, that is, a small sample and measurement errors. Later, when we performed factor analysis on age subsamples, we saw that at the ages of 4 and 5, there were high negative factor loadings. That is how we concluded that the

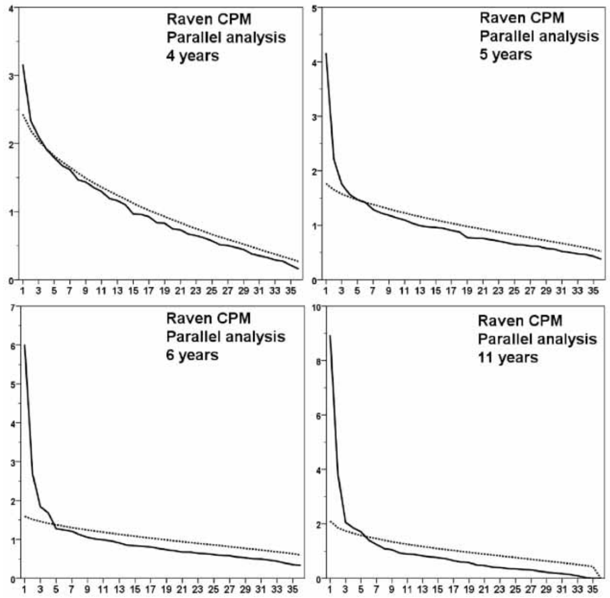

Fig. 1. Results of parallel analysis at ages 4, 5, 6 and 11.

low alpha coefficient may not be the consequence of errors (only), but also of the lack of homogeneity. Maybe at younger ages *g*, i.e. the mechanisms lying behind it, is still not sufficiently developed?

Figure 1 shows the results of parallel analysis for ages 4, 5, 6 and 11. Solid line indicates the diagram of eigenvalues of raw data, and broken line the 95th percentile of eigenvalues obtained from the set of 500 randomly generated data.

Distractor Analysis

Distribution of incorrect answers depended on the difficulty of items and on age. However, in total sample it is noticeable that in 12 items there are distractors with over 20% of answers. The strongest distractors occur with: Ab12 (4=31%), A11 (2=30%), B11 (2=27%), B9 (1=26%). At youngest ages (4 and 5), some fifteen more items have favorite distractors (>20%). At youngest ages, the most frequent distractor is A11 (2=52%) and A12 (4=41%). At older ages (>5), the most frequent distractor is Ab12 (4=34%).

From the point of view of content, the most common errors were choosing the distractor that is equal to: a) the element above the empty element, and b) the element to the left side of the empty element. However, there was also a tendency to choose the solution that is spatially the closest to the empty element in the matrix, and that is 2. In Table 5 we present the distribution of the most common distractors, where cells contain the number of items in which that distractor was chosen by more than 20% of participants. At youngest ages, distractor 2 is significantly more favorite (χ^2=21.7 and 4.6, p=0.0 and 0.03), but it is also the most common at older ages (although not significantly). »Content« and »spatial« errors often overlap precisely on the distractor number 2.

Factor analysis by age

Since the sample was very heterogeneous according to age, the question arose whether latent structure changes from age to age and in which way. We applied the standard solution, principal components plus promax, and at each age we kept the number of factors suggested by parallel analysis. At the age of 4, the factors do not resemble much the factors obtained at older ages. The first factor consists of several items of medium difficulty from all sets. The second factor consists of some easier items from set A. The third factor is defined by items A8, Ab5, Ab12, Ab8, Ab10, B11 and B12, where the first three have a positive, and the last four negative loading. The fourth factor is defined by the items A10, Ab9, B9, B10, A1 and B4, where the first four have a negative, and A1 and B4 positive loading. Factor intercorrelations are low and range from –0.03 to 0.12.

At the age of 5, a factor pattern that is later repeated at all ages is already formed: a) the factor with the easiest items, b) the factor with the most difficult items, and c) the factor mostly comprised of items of medium difficulty. However, there also are two smaller factors with negative loadings, similar to the factors at the age of 4. The fourth factor is defined by the items Ab2, Ab3, B1, B2 and Ab12, where the first four have a positive, and Ab12 negative loading. The fifth factor is defined by the items A3, A11 (negatively) and Ab10 (positively). Factor intercorrelations are somewhat higher than at the age of 4: from –0.10 to 0.24.

Negative factor loadings do not appear at older ages anymore, and factor correlations are similar to those in total sample (see Table 3) and relatively stable. We calculated a simple canonical index as a measure of similarity between the columns of factor pattern matrices. Canonical index between factors at the age of 4 and older ages was around 0.5. Between the age of 5 and older ages it was between 0.6 and 0.8, and the canonical index between the ages 6, 7 and more than 7 was above 0.9. This supports our impression that factor structure of CPM at ages 4 and 5 is not similar either mutually or with older

TABLE 5
DISTRIBUTION OF CHOICES OF THE MOST COMMON DISTRACTORS (>20%) ACCORDING TO AGE

Distractor	Age			
	4	4 and 5	>5	Total
1	6	5	1	2
2	15	9	5	5
3	4	6	3	4
4	4	6	3	3
5	2	5	3	3
6	2	2	2	2

ages, while factor structures of older ages are largely equivalent.

The question now is how to interpret the factors that have negative loadings (between –0.4 and –0.6)? We cross-tabulated the items with positive and negative loadings and obtained the significant chi-square at the age of 4 in pairs Ab5/Ab8, Ab12/B12, Ab10/Ab12 and A8/A11 (from the third factor) and A10/B4 and A1/Ab9 (from the fourth). At the age of 5, only the Ab2/Ab12 pair from the fourth factor and A3/Ab10 from the fifth factor had a significant correlation. By reviewing the observed and expected frequencies, we were not able to establish a universal system. It seems that in pairs of items with significant correlation two basic sources of negative correlation are dominant. The first one consists in children using a simple principle of gestalt completion, in such a way as to complete the empty element in the matrix by the solution presenting the neighboring element (above or below). The second is to choose the solution that is spatially the closest to the empty element, and that is number 2.

Therefore, at earlier ages some children strived towards applying one and the same strategy (rule) to all items, or at least several same strategies. We think that our bipolar factors originated because in some items that is a successful solving strategy, and in some not. At later ages, this phenomenon disappeared, and this can only be ascribed to the children not applying the same strategy blindly on the same tasks anymore, but adjusting the strategy to the task. Older participants had probably developed the mechanisms of control, coordination or cooperation of means and goals, based on which they recognized the situations when the old solving strategy did not work, but a new one should have been tried instead.

In other words, bipolar factors at the ages of 4 and 5 could be named the factors of strategy inflexibility, that is, non-adjustment of the strategy to the task. If those factors no longer exist at later ages, are they a measure of absence of Raven's *g* factor? Do our findings imply that *g*-factor is developed in the first five years?

Are primary factors difficulty factors?

All our factors analyses indicated that the results on CPM are not monofactorial at older ages as well. At older ages between 3 and 5, factors are also necessary, but they, indeed, do not have significant negative loadings and are in relatively high positive correlations. Are they the primary intelligence factors, that is, do they represent separate constructs, or are they, maybe, difficulty factors?

Short descriptions of primary factors from Table 3 suggest that we are dealing with difficulty factors, that is, artifacts. The problem of difficulty factors was perceived a long time ago and there is no simple solution for it (McDonald, 1965)[52]. Gorsuch (1974)[53] provides an illustrative example with items of the imagined Guttman scale, which is ideally unidimensional, but the items of which are equally distributed from very easy to very difficult. Factor analysis yielded four factors, where a general factor was obtained in the second order, which was strong indeed, but insufficient to explain the whole variance. McDonald calls this a Ferguson dilemma from 1941: whether to accept the multi-factor solution if factors stem from different difficulty of groups of items that have the same measuring content. McDonald operationally defines the difficulty factor as the factor whose loadings significantly correlate with difficulty of items.

In order to check this, we calculated the correlations of factor loadings and item difficulties, and found that only the first factor, the most significant in the test, which comprised items of medium difficulty from all series, did not have significant correlations with difficulty of items. All other factors significantly and highly correlated with difficulties (Table 6). This goes for all EFA solutions we applied, except for Testfact solution, whose all three factors significantly correlated with difficulties of items. The variance of factor loadings of the first factor is practically the highest in all solutions and the distribution the most symmetric compared to other factors, which indicates that insignificant correlation is not the consequence of range restriction.

We did not find many comments about this problem in available literature, although from factor matrices (where they are enclosed) it is clearly visible that there is space for doubt in difficulty factors, especially in the article by Lynn et al. (2004)[10], who present in parallel way their and the results of other people on factor analysis. In his paper, McDonald (1965)[52] used precisely SPM to illustrate his solution for difficulty factors. He determined that at least one of three obtained factors presented a difficulty factor. In brief, we have reason to believe that a considerable portion of specific factors (first order factors) obtained in progressive matrices can be classified as difficulty factors.

When everything is taken into account, if factor analysis of higher order, of any kind, exploratory or (espe-

TABLE 6
CORRELATIONS OF FACTOR LOADINGS OF ITEMS AND DIFFICULTY OF ITEMS (n=36)

Solution	Factor	r sig.
II	1	–0.021
II	2	–0.749**
II	3	0.585**
II	4	0.436**
IV	1	0.008
IV	2	–0.820**
IV	3	0.552**
IV	4	0.479**
V	1	0.407*
V	2	–0.875**
V	3	0.549**
VI	1	0.084
VI	2	–0.787**
VI	3	0.746**

*p<0.05, **p<0.01

cially) confirmatory, confirms the existence of a general factor, the damage done by difficulty »artefactors« is not big – unidimensionality is discovered. In this case, there remains an important question whether these difficulty factors still define some genuine differences in participants, or they are just artifacts of the method – factor analysis. In the case that unidimensionality is not confirmed, as it happened here, there remains the question whether difficulty factors could be the cause of that.

McDonald (1965)[52] and Digman (1966)[54] explain that difficulty factors have something to do with the non-linear correlation between the items themselves, or with the non-linear correlation between items and factors (characteristics). However, difficulty factor can also stem from the fact that easy and difficult items call for genuinely different mental processes. Non-linearity of correlation can, again, stem from the influence of moderator variables, such as a »development variable«. In other words, even if we establish that we obtained difficulty factor (according to McDonald's operational definition), we cannot claim that it is an artifact, at least not based on standard factor-analytical procedures. In this sense, we refrain from a definitive conclusion with respect to whether our primary factors are artifacts or not. We would only like to draw attention that, in our case, tetrahoric coefficients have not proved out to be successful means for removing difficulty factors. On the contrary, even the first factor became difficulty factor with them.

Discussion

In the total sample, 3 or 4 factors are necessary (and sufficient) in order to explain at least 30% of total variance, in order for the factors to be well-defined and for the model to have an acceptable fit. In promax solutions the first factor correlated with others between 0.30 and 0.70 (typically 0.50), while the other factors correlated lower among themselves: from 0.10 to 0.60 (typically 0.25). In confirmatory factor analysis, in models with the general factor, correlations of primary factors with a general factor were between 0.50 and 0.90. However, the models with and without a general second order factor fit data equally well.

Analysis of dimensionality through testing of models with a general factor in CFA did not give a considerable advantage to these models compared to the models without a general factor. Parallel analysis by ages also suggested 3 to 5 factors to be a natural solution.

All factors, except for the first one, show high and significant correlations with item difficulty, which defines them as difficulty factors. Standard interpretation of factors based on items that show the highest loadings on them also leaves room for a very open doubt that we are dealing with difficulty factors.

In other words, factor analysis did not prove to be a tool by which we could demonstrate the existence of a general factor, in a clear and unambiguous way. We think that such a result is, actually, a natural one, since multivariate techniques are not a good way to prove unidimensionality. However, generally observed, but also from the viewpoint of our actual findings, we cannot reject that a general factor exists in CPM. The main reasons are the following: there are some sufficiently high correlations between the factors, there is an acceptable fit of models with the general factor, and maybe the most important thing is that all primary factors (except for the first one) could be treated as difficulty factors, that is, artifacts.

Factor analysis by age indicates that factor solutions are practically stabilized after the age of five. However, at ages 4 and 5 we obtained bipolar factors (with negative loadings). We concluded that they resulted from an inflexible solving strategy, that is, poor management of goal activity such as solving tasks. In this sense, it seems to us that the development of this metacognitive process occurs in the first five years of life and that, once it gets developed, we might consider it a component *g*.

Analysis of distractors indicated that there are favorite distractors in certain items, at all ages, but that the most common distractor is 2, the one which is perceptually and motorically the closest to the empty field. The favorite distractors, and especially distractor 2, are considerably more frequent at ages 4 and 5.

While testing the structural models, we determined that fitting can be considerably improved if we allow the correlated errors between some items as a rule between those that are the easiest in the set. We think it is the consequence of the fact that children build a solving strategy on the first items in the test or set, that is, the first items serve as a practice, which introduces local dependence. This would probably be removed if examples and practice were added to the test.

Besides, initial items in sets (especially in set A) are extremely easy. This jeopardizes the entire measurement of the test, and to a considerable degree aggravates its evaluation as well. There are several solutions to this problem: items can be left out, made more difficult, replaced, or the test should not be administered to children older than 8 or 9.

Acknowledgements

This paper was done within the research projects »Anthropological status and physical activity of Vojvodina population«, co-financed by the Provincial Secretariat for Science and Technological Development, and »Integral development, aberrant behavior and physical activity of preschool children«, co-financed by the Ministry of Science and Technological Development, Republic of Serbia, which are accomplished by the Faculty of Sports and Physical Education in Novi Sad. The study was supported by grant No. 177-0000000-3410 from the Croatian Ministry of Science, Education and Sports.

R E F E R E N C E S

1. BALA G, KATIĆ R, Coll Antropol, 33 (2009) 353. — 2. BALA G, JAKŠIĆ D, KATIĆ R, Coll Antropol, 33 (2009) 373. — 3. KATIĆ R, Canonical relationships between psychomotor and cognitive factors. In: Proceedings. (1st EAA Congress, Zagreb, 1977). — 4. BALA G, KATIĆ R, Coll Antropol, 33 (2009) 1071. — 5. KATIĆ R, Canonical relationships between psychomotor and pathologic conative factors. In: Proceedings. (1st EAA Congress, Zagreb, 1977). — 6. RAVEN JC, Priručnik za progresivne matrice u boji – serija A, Ab i B (Revizija 1956) (Guide to Using the Colored Progressive Matrices – sets A, Ab and B (Društvo psihologa Srbije, 1956 revision, Beograd) — 7. SPEARMAN C, General intelligence, objectively determined and measured. The abilities of man. (Macmillan, London, 1927). — 8. SPEARMAN C, Human ability. (Macmillan, London, 1950). — 9. COTTON SM, KIELY PM, CREWTHER DP, THOMSON B, LAYCOCK R, CREWTHER SG, Personality and Individual Differences, 39 (2005) 647. — 10. LYNN R, ALLIK J, IRWING P, Intelligence, 32 (2004) 411. — 11. BORS DA, VIGNEAU F, Learning and Individual Differences, 13 (2003) 291. — 12. NEISSER U, BOODOO G, BOUCHARD TJ, BOYKIN AW, BRODY N, CECI SJ, American Psychologist, 51 (1996) 77. — 13. PIND J, GUNNARSDOTTIR EK, JOHANNESSON HS, Personality and Individual Differences, 34 (2004) 375. — 14. FAJGELJ S, BALA G, TUBIĆ T, Psihologija, 40 (2007) 293. — 15. RUSHTON JP, SKUY M, BONS TA, International Journal of Selection and Assessment, 12 (2004) 220. — 16. WILLIAMS JE, MCCORD DM, Computers in Human Behavior, 22 (2006) 791. — 17. MACKINTOSH NJ, BENNETT ES, Intelligence, 33 (2005) 663. — 18. CARPENTER, PA, JUST MA, SHELL P, Psychological Review, 97 (1990) 404. — 19. CATTELL RB, Journal of Educational Psychology, 54 (1963) 1. — 20. CATTELL RB, Abilities: Their structure, growth and action. (Houghton-Mifflin, Boston, 1971). — 21. HORN JL, Journal of Research and Development in Education, 12 (1979) 59. — 22. IVIĆ I, MILINKOVIĆ M, ROSANDIĆ R, SMILJANIĆ V, Razvoj i merenje inteligencije – Tom I, Inteligencija, njen razvoj i merenje (drugo izdanje) (Development and Measurement of Intelligence – Volume I, Intelligence, Its Development and Measurement (2nd edition) (Zavod za udžbenike i nastavna sredstva, Beograd, 1978). — 23. DESHON RP, CHAN D, WEISSBEIN DA, Intelligence, 21 (1995) 135. — 24. VAN DER VEN AHGS, ELLIS JL, Personality and Individual Differences, 29 (2000) 45. — 25. VIGNEAU F, BORS D, Educational and Psychological Measurement, 65 (2005) 109. — 26. CARROLL JB, Intelligence, 15 (1991) 423. — 27. KRANZLER JH, Intelligence, 17 (1993) 11. — 28. DETTERMAN DK, General Intelligence: Cognitive and Biological Explanations. In: STERNBERG RJ and GRIGORENKO EL (Eds) The General Factor of Intelligence: How General Is It? (Lawrence Erlbaum Associates, Mahwah, NJ, 2002). — 29. SLOUTSKY VM, FISHER AV, Journal of Experimental Psychology: General, 133 (2004) 166. — 30. HEIT E, HAYES BK, Journal of Experimental Psychology, 134 (2005) 596. — 31. FREUND PA, HOFER S, HOLLING H, Applied Psychological Measurement, 32 (2008) 195. — 32. GOEL V, Cognitive Neuroscience of Deductive Reasoning. In: HOLYOAK K and MORRISON R (Eds) Cambridge Handbook of Thinking & Reasoning, (Cambridge University Press, Cambridge, 2005). — 33. UNGERLEIDER LG, MISHKIN M, Two visual systems. In: INGLE DJ, GOODALE MA, MANSFIELD RJW (Eds) Analysis of visual behavior (MA: MIT Press, Cambridge, 1982). — 34. MILNER D, GOODALE MA, The Visual Brain in Action. (Oxford University Press, Oxford, 1995). — 35. GLOVER S, Trends in Cognitive Sciences, 6 (2002) 288. — 36. GLOVER S, Behavioral And Brain Sciences, 27 (2004) 3. — 37. CHABRIS CF, JERDE TL, WOOLLEY AW, HACKMAN JR, KOSSLYN SM, Spatial and object visualization cognitive styles: Validation studies in 3800 individuals. Technical Report #2, Available from: URL: http://coglab.wjh.harvard.edu/čcfc/?Chabris?2006d.pdf. — 38. GALLESE V, The »Conscious« Dorsal Stream: Embodied Simulation and its Role in Space and Action Conscious Awareness. Psyche: Available from: URL: http://psyche.cs.monash.edu.au, 13/1, April 2007. — 39. SCHOOLER JW, ENGSTLER-SCHOOLER TY, Cognitive Psychology, 22 (1990) 36. — 40. EMBRETSON SE, Psychometrika, 64 (1999) 407. — 41. ZAREVSKI P, Struktura i priroda inteligencije. (Structure and Nature of Intelligence) (Naklada Slap, Jastrebarsko, 2000). — 42. CATTELL RB (Ed) Handbook of multivariate experimental psychology (2nd ed.). (Rand McNally, Chicago, 1966). — 43. DU TOIT M, Ed. (2003). IRT From SSI: BILOG–MG, MULTILOG, PARSCALE, TESTFACT, Lincolnwood: Scientific Software International, Available from: URL: http://www.ssicentral.com. — 44. WILSON D, WOOD R, GIBBONS RD, TESTFACT, Test scoring, item statistics, and item factor analysis. (Chicago, IL: Scientific Software, Inc, 1991). — 45. KNOL DL, BERGER MPF, Multivariate Behavioral Research, 26 (1991) 457. — 46. KLINE RB, Principles and practice of structural equation modeling (2nd edition) (The Guilford Press, New York, 2005). — 47. THOMPSON B, Exploratory and comfirmatory factor analysis: Understanding concepts and applications. (American Psychological Association, Washington, DC, 2004). — 48. Personality and Individual Differences. Special issue on Structural Equation Modeling, 42 (2007) 5. — 49. HU L, BENTLER PM, Structural Equation Modeling, 6 (1999) 1. — 50. TOMARKEN AJ, WALLER NG, Journal of Abnormal Psychology, 112 (2003) 578. — 51. O'CONNOR BP, SPSS, SAS, and Matlab programs for determining the number of components using parallel analysis and Velicer's MAP test. Behavior Research Methods, Instruments, & Computers, 32 (2000) 396–402. Retrieved November 2008, Available from: URL: //flash.lakeheadu.ca/ ~boconno2/nfactors.html. — 52. MCDONALD RP, British Journal of Mathematical and Statistical Psychology, 18 (1965) 11. — 53. GORSUCH RL, Factor analysis (W. B. Saunders Company, Philadelphia, 1974). — 54. DIGMAN JM, Interaction and non-linearity in multivariate experiment. In: CATTELL RB (Ed) Handbook of multivariate experimental psychology, (Rand McNally, Chicago, 1966).

R. Katić

Faculty of Kinesiology, Split University, Teslina 6, 21000 Split, Croatia
e-mail: ratko.katic@gmail.com

LATENTNA STRUKTURA RAVENOVIH PROGRESIVNIH MATRICA U BOJI

S A Ž E T A K

Osnovni cilj ovog rada je određivanje konstruktivne validnosti Ravenovih Progresivnih Matrica u Boji putem faktorske ajtem analize. Najznačajnija tema u okviru ovog cilja je određivanje dimenzionalnosti testa, jer mnogi autori izvještavaju o tome da su našli nekoliko značajnih primarnih faktora. U istraživanju je učestvovalo 2334 djece uzrasta od 4 do 11 godina, od toga 1252 dječaka i 1082 djevojčice. Korišteno je nekoliko vrsta faktorske analize da bi se otklonio utjecaj tehnike na rezultate. U radu su izdvojena tri, odnosno četiri faktora prvog reda. Utvrđeno je da se na osnovu uobičajene McDonaldove definicije dobiveni primarni faktori u većini slučajeva mogu smatrati faktorima težine. Potreban broj faktora na uzrasnim poduzorcima, dobiven paralelnom analizom, je između 3 i 5. Faktorska struktura na uzrasnim poduzorcima pokazala je da se najmlađi uzrasti, 4 i 5 godina, suštinski razlikuju od starijih. Razliku smo identificirali kao nerazvijenost mehanizma upravljanja ciljem.