Analysis of Psychometric Properties of the Cube Test to Evaluate Visuospatial Reasoning

Análise das Propriedades Psicométricas do Teste dos Cubos: Para a Avaliação do Raciocínio Visuoespacial

Monalisa Muniz*, [a], Fabián Javier Marín Rueda[b] & Juliana Carnevalli Siqueira Nery[b]
[a]*Universidade Federal de São Carlos, São Carlos, São Paulo, Brasil*
& [b]*Universidade São Francisco, Itatiba, São Paulo, Brasil*

Resumo
O objetivo da pesquisa foi contribuir com novas informações acerca das evidências de validade do Teste dos Cubos. Mais especificamente, analisaram-se as propriedades psicométricas dos itens por meio da Teoria de Resposta ao Item, e a estrutura do teste utilizando a Análise Fatorial Confirmatória. A amostra foi composta por 684 sujeitos, com idade média de 21,3 anos (*DP*=5,94), sendo 50,3% feminino. Nos resultados, as análises realizadas pela Teoria de Resposta ao Item corroboraram o ajuste dos itens ao modelo, enquanto a Análise Fatorial Confirmatória apontou o modelo com dois fatores como sendo o mais adequado. Os dados encontrados, além de oferecerem resultados favoráveis sobre o instrumento, sustentam as evidências de validade contidas no manual do Teste dos Cubos.
Palavras-chave: Raciocínio, processamento visual, análise fatorial confirmatória, teoria de resposta ao item, teste psicológico.

Abstract
The objective of the present research was to contribute with new information regarding the validity evidence of the Cube Test. More specifically, psychometric properties of the items were analyzed through the Item Response Theory (IRT) and the test structure was analyzed using Confirmatory Factor Analysis (CFA). The sample was composed of 684 subjects, with a mean age of 21.3 years (*SD* = 5.94) and 50.3% were female. Results showed that IRT analyses confirm the adjustment of the items to the model, whereas the CFA suggested that a two-factor model was the most adequate. The findings, besides offering favorable results about the instrument, supported the validity evidence present in the Cube Test manual.
Keywords: Reasoning, visual processing, confirmatory factor analysis, item response theory, psychological test.

The construction of psychological tests is one area of psychological assessment that is increasing in Brazil. In recent years, practitioners have shown a greater awareness to the importance of how the tests are constructed and supported by the characteristics of the population that will be submitted to these tools. As noted by Urbina (2006), to construct a test, one must first have an objective – a theoretical foundation that supports the construct which will be assessed – as well as knowing which population the instrument will aim to evaluate. Further, the existence of a variety of psychological tests to assess the same construct facilitates the work of the psychologist, because the professional will be able to verify which test is most appropriate for a given subject, population and purpose.

* Mailing Address: Universidade Federal de São Carlos, Departamento de Psicologia, Rod. Washington Luis, km 235, São Carlos, SP. Caixa Postal 676, CEP: 13565-905, telefone (16) 3351-8361. Email: mo_nascimento@yahoo.com.br

In an inquiry conducted by the Psychological Testing Evaluation System (Satepsi) in 2011, no instrument was found with the assent of the Federal Council of Psychology that could specifically measure visuospatial reasoning. The measure of this construct is found in a battery of tests, such as Reasoning Battery Tests – BPR-5 (Primi & Almeida, 2000), applied together with other subtests that assess different types of reasoning. However, it is intriguing that the psychologist could also have the option of a test that addresses only the visuospatial reasoning, as well as a manual that contemplates information and more detailed studies of this construct.

Accordingly, the Cube Test was approved in November 2011, to assess visuospatial reasoning (Rueda & Muniz, 2012), which proposes to measure just this type of reasoning. The construction of the test and the items that comprise it were rooted in more modern and widespread psychometric theory, which is the Cattell-Horn-Carroll model (CHC model) of cognitive abilities (Carroll, 1997).

This model conceives intelligence as a multidimensional construct that encompasses various capacities, which in turn are all linked to a general factor.

Initially, ten abilities integrated the CHC model, namely, fluid intelligence (*Gf*), crystallized intelligence (*Gc*), quantitative reasoning/knowledge (*Gq*), reading-writing ability (*Grw*), short-term memory (*Gsm*), visual processing/intelligence (*Gv*), auditory processing/intelligence (*Ga*), long-term storage and retrieval (*Glr*), processing speed (*Gs*), and decision speed/reaction time (*Gt*). Since 2009 six other abilities were gradually incorporated into the model: those were general (domain-specific) knowledge (*Gkn*), psychomotor speed (*Gps*), the psychomotor abilities (*Gp*), olfactory abilities (*Go*), tactile abilities (*Gh*) and kinesthetic abilities (*Gk*; McGrew, 2009). With regard to the last six, though some are not yet well established, the chapter described by Schneider and McGrew (2012) notes that research is being developed and, besides approaching updates to the CHC theory, reaffirms the existence of the abilities.

Considering the Cube Test, visuospatial reasoning contained in it is mainly grounded in the construct *Gv*. Schneider and McGrew (2012) describe that *Gv* can be better understood by the skills of Visualization (*Vz*), Speeded Rotation/Spatial Relations (*SR*), Closure Speed (*CS*), Flexibility of Closure (*CF*), Visual Memory (*MV*), Spatial Scanning (*SS*), Serial Perceptual Integration (*PI*), Length Estimation (*LE*), Perceptual Illusions (*IL*), Perceptual Alternations (*PN*) and Imagery (*IM*).

Of these eleven skills, the items built into the Cube Test relate mainly to the skills *Vz*, *SR* and *SS*. The ability *Vz* requires that the individual can imagine, manipulate or transform objects or visual patterns mentally, independent of the processing speed of the response. In addition, subjects also must provide, visualizing mentally, how a particular object would be after a few alterations, for example, when parts of an object are rearranged, removed or relocated to other parts of that object. In turn, *SR* refers to the ability to perceive and manipulate objects or visual patterns, also mentally, but quickly, by mental rotation or transformation. Thus, in this ability it is important to be able to identify objects at different angles and positions with agility, because the fluency of identification is required. It is worth highlighting that identification is the main difference between *Vz* and *SR*. Ability *SS* requires the subject to explore a broad or visually confusing field quickly and accurately to identify objects or images. Therefore, in the implementation of the Cube Test, the successful individual will be able to imagine the suggested rotation, mentally manipulate the cube quickly and accurately, without being confused by other stimuli and, at the end of this activity, temporarily hold the image obtained, comparing it with stimuli suggested in the alternatives to find the correct item.

Reasoning is a crucial activity of thought and is present in its whole structure (Wilhelm, 2005). As described by Vinod Goel (2005), reasoning is a cognitive activity that allows derivation from information provided, involving deductions based on one or more premises that provide some information to accept another proposition, which is the conclusion. Goel's explanation refers to the ability of fluid intelligence (*Gf*), which is the mental operations of reasoning that the individual performs when faced with new situations that cannot be automatically performed, and which depend minimally on acquired knowledge. For Schneider and McGrew (2012), these operations involve relating ideas, inducing abstract concepts and solving problems, using mainly inductive and deductive reasoning. Some psychological tests are proposed for evaluating a particular type of reasoning, as in the case of the Cube Test in assessing the visuospatial reasoning. However, these same instruments can encompass other types of reasoning, as well as other abilities beyond the visuospatial. This is consistent with Schneider and McGrew (2012), who highlight that fluid intelligence covers inductive and deductive reasoning.

In the Cube Test, visuospatial reasoning can also be related to the ability of fluid intelligence, as in task resolution, wherein the subject will deal with new situations, minimally dependent on prior knowledge and seeking relationships within the contained information. In turn, inductive reasoning is a factor of fluid intelligence and refers to the ability to analyze a set of information and establish relationships within the set, creating new ideas and concepts, systematically organizing the information (Primi, 2002). Within the items of the Cube Test, it is necessary to discover the relationship between colors and position changes, in order to reach the correct answer, resulting in inductive reasoning.

Although deductive reasoning can also be a factor of fluid intelligence, is not present in the Cube Test. To Holyoak and Morrison (2005), the differences between the two types of reasoning are subtle and emphasize conclusions from some initial assumptions. For these authors, an inference is deductive if the truth of the premise guarantees the truth of the conclusion by virtue of the form of the argument. But if the true premise does not guarantee reliable conclusions, the inference is called inductive. In the Cube Test, the premise is related to mental rotations, that is, it informs the subject that he needs to figure out how the cube will appear after one, two or three rotations. The correct response, which is part of a set of alternatives already stipulated, is not the only possible configuration of the cube after the rotations, but on the test it is, because the other alternatives are incorrect.

The study by Almeida et al. (2010) is an example of scientific work showing the association between fluid intelligence, more specifically the inductive reasoning factor that composes this intelligence, with visuospatial reasoning. The researchers found a low and positive correlation between proof of Spatial Reasoning from the BPR-5 and Raven's General and G-36 Tests, which assess fluid intelligence and inductive reasoning. This is because fluid intelligence is the ability that is closer to general intelligence

(Flanagan & Ortiz, 2001) and is present in other abilities. Thus, to more specifically evaluate a certain ability, one can resort to the content, format, and task instructions that outline the type of reasoning to be employed in resolving them. To assess visuospatial reasoning through the Cube Test, items were created for which concrete images must be mentally manipulated to find the correct response. This content (images), format (cubes), and instruction (to mentally manipulate) to the test items requires the evaluated subject to predominantly utilize visuospatial reasoning. In the work of Primi, Cruz, Nascimento and Petrini (2006), we see the construction of a fluid intelligence test that sought to specifically assess the simultaneous coordination of tasks and selective attention/abstraction. For this, geometric shaped items were systematically constructed seeking to control the amount of irrelevant information and perceptual organization, that elements that are used to assess, respectively, the cited factors coordination and attention. To solve the tasks the subject must discover the underlying rule in a set of geometric figures.

Rueda and Muniz (2012) have generated two factors for the Cube Test, understood as one factor containing easier items, and a second factor with more difficult items. Despite being a test of visuospatial reasoning, other abilities that tend to be a subtle part of reasoning may have contributed to the study of internal structure contained in the test manual. The complexity of the items, which gradually increases from Item 1 through Item 15, further endorses the emergence of the factors, such as the increasing demand for the use of cognitive resources and, consequently, the visuospatial reasoning is not identical in the resolution of easier items and more difficult items. It is emphasized that the manual presents studies of its internal structure through the exploratory factor analysis by the Classical Test Theory, and by analyzing *full information factor analysis* (FIFA) based on Item Response Theory.

Considering the above so far, and especially the fact that the test was designed to measure visuospatial reasoning, finding unidimensionality would be expected. However, as already noted, other types of reasoning and abilities may be present, which makes it more difficult to achieve unidimensionality. Even the Raven's Coloured Progressive Matrices (CPM), which measures the eduction of relations (similar to fluid intelligence), has always been understood as unidimensional, though there are no studies that suggest unidimensionality. Two studies conducted in Brazil showed that the CPM tended to measure more than one dimension. The work of Pasquali, Wechsler and Bensusan (2002), found four factors, but at such a secondary level of as to indicate unidimensionality. Likewise, the study of Sisto, Rueda and Bartholomeu (2006), did not detect unidimensionality. Thus, it is important to do more studies seeking to investigate this question of structure.

Within this context, and seeking to further investigate the Cube Test, the objective of this study was to contribute new information about the validity evidence of the instrument. The main analysis is on test structure, seeking to verify whether the structure of the two factors found by the test manual authors stands in light of new data from a sample differing from the normative. Additionally, the adjustment and the difficulty of items has been studied, the item-theta and item-total correlation, and the accuracy, in order to compare data obtained from this research with the studies described in the test manual.

Method

Participants

The study includes 684 participants, high school and college students from both public and private institutions in the state of Minas Gerais. Of the total, 344 were female (50.3%). The mean age was 21.3 years (*SD* = 5.94), with a minimum age of 15 and maximum of 57. Of the total, 58.5% have incomplete higher education, and majors included Engineering (13.3%), Digital Game Design (11.5%), Psychology (13.2%), Life Sciences (11.8%) and Languages (8.6%); while 41.5% had incomplete secondary education, with 29.1% in the 2nd year and 12.4% in the 3rd (and final) year. The study sample was by convenience. It is emphasized that the Cube Test manual sample was also by convenience, but prioritized participants who attended college and majored in topics such as engineering, architecture and aerospace, which tend to demand a greater capacity for visuospatial reasoning. The present study, although with the participation of subjects in engineering and digital game design courses, had no such concern because the goal was not to verify differences between types of professions.

Instrument

Cube Test: To Assess Visuospatial Reasoning (Rueda & Muniz, 2012). The objective is to assess visuospatial reasoning in subjects between ages 18 and 66 years, of both sexes, conforming to the standardization in the manual. The test consists of 15 tasks: Items 1 through 8 have three choice alternatives, and Items 9 through 15 have four alternatives. Each task includes a cube design, like a "Rubik's Cube", and the subject must ascertain how the cube would appear after one, two or three rotations. The subject selects one option from the alternatives, each with its own cube design, with only one being the correct answer. Figure 1 is an example of the test.

Based on the example, the examinee must analyze the different faces of the cube and be able to form mental-visual representations, applying the number of transformations to find the correct answer from the options provided. For each correct answer 1 point is awarded, while the errors are not scored. The Cube Test can be administered individually or collectively, in groups of up to 20 people. The time limit for administration is 30 minutes and the test should be completed in a single session.

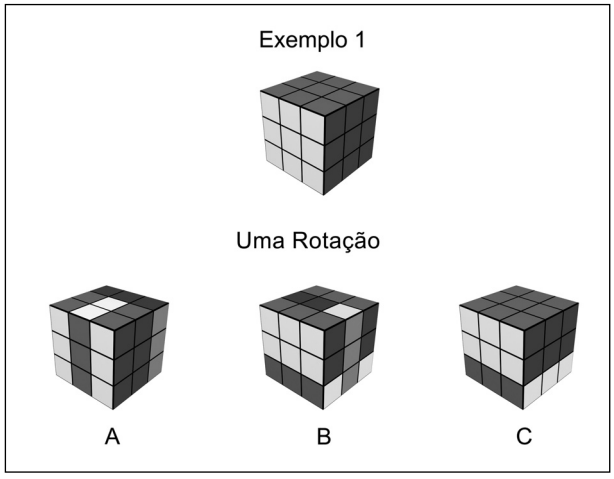

Figure 1. Test example of cubes taken from the manual with permission of the authors of the test (Rueda & Muniz, 2012).

The test manual testing contains the analyses of psychometric properties effected through Item Response Theory and demonstrates satisfactory fit indices and item difficulty of the Cube Test. Validity evidence of the internal structure is also presented, suggesting two moderately correlated factors and interpreted as assessing the same construct, which is visuospatial reasoning. Also related to the internal structure, there is information for Differential Item Functioning, DIF, suggesting that the items have a similar statistical behavior depending on the gender of the participants.

Validity evidence based on relations with external variables are also contemplated. Age and gender were external criteria and a decline in visuospatial reasoning was observed as age increased, as well as better performance by males. The university major was also an important criterion and students of Engineering and Architecture performed better. Finally, the validity evidence with tests that measure similar constructs was found with the Spatial Reasoning Test of the Battery of Reasoning Tests – BPR-5 (Primi & Almeida, 2000). All results of validity evidence are consistent with those indicated in the literature on the visuospatial reasoning construct.

With regard to accuracy, a coefficient of .80 was obtained by means of Cronbach's alpha, .84 by Spearman-Brown, and .83 by Guttman. All indices were considered highly satisfactory.

Procedure

Initially, contact with educational institutions was made, to explanation the research and request permission to collect data. After institutional authorization, the project was submitted to and approved by the Ethics in Research Committee (CEP).

After approval, again the institutions were contacted to schedule data collection. Before collection, high school students were given the Terms of Free and Clear Consent (TCLE) to take their guardians to read and, if they agreed, to sign. For higher education students, everyone being 18 years of age or older, they were given the TCLE consent form to read and sign on the day of data collection. Only students with signed consent forms participated in the survey. Data collection was conducted in group sessions of about thirty minutes on the premises of each educational institution.

Data Analysis

The psychometric properties of the test items were analyzed by Item Response Theory, through the Rasch model, using the program Winsteps (Linacre, 2005), with the purpose of verifying the accuracy of the instrument, the item-theta correlation, the item difficulty, and the adjustment of the items in relation to the scale. Regarding the adjustment, the outfit and infit were evaluated. The outfit refers to the average of the settings of the items: this parameter is more sensitive to outliers, wherein the misalignment of the item or discrepancy occurs in contradiction to skill level. That is, the cases of outfits refer to situations in which the subject does not ascertain items that fall within their ability, or conversely, ascertains correct answers for items outside the subject's ability. In turn, the infit is also a measure of item misfit, being related to discrepancies near the skill level of the subject, that is, the subject does not ascertain the items near his/her skill level (Wright & Stone, 2004). With regard to the expected values, it is desirable that the accuracy is at least .60, the item-theta correlation is at least .30, and the adjustments at most 1.2 – the latter according to Wright and Stone (2004), though Linacre (2002) contends adjustments to 1.5 are acceptable. As for the difficulty indices, values near zero are considered average difficulty; much below are considered very easy, and high above are very hard. It is emphasized that the Cube Test manual used the Item Response Theory through the Rasch model to perform the psychometric analyzes of the test items. As the present study also aims to compare the data contained in the manual with the present research, it was deemed appropriate to maintain the same type of analysis.

The confirmatory factor analyses were conducted in the Mplus program (Muthén & Muthén, 2010a), through a tetrachoric matrix, using the estimation method of Weighted Least Squares Mean and Variance-Adjusted (WLSMV), specific to dichotomous data (Muthén & Muthén, 2010b). Two adjustment models were tested, the first following the division of two latent factors found through exploratory factor analysis. This model was assumed in the test manual and covers a factor with the easiest items (item 1-9), and another factor with the most difficult items (item 10-15), but assessing the same construct. The second model tested a general latent factor with the measures of visuospatial reasoning, where the 15 items carried in a single dimension, expecting the existence of unidimensionality, since theoretically the Cube Test evaluates only a construct.

To check which model fit better, we used the following fit indices: Ratio chi-square by degrees of freedom ($\chi 2/df$);

Root mean square error of approximation (RMSEA), Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI). For $\chi 2/df$, values less than 3 indicate models adjusted to the sample; for the RMSEA values between .06 and .08 (with confidence interval below .10) indicate low levels of residuals; and for the CFI and TLI, values above .90 indicate fit (Brown, 2006; Byrne, 2010; Schreiber, Stage, King, Nora, & Barlow, 2006).

In addition, the reliability of the factorial structures was evaluated by the Composite Reliability method, which evaluates the reliability of indicators of the determined construct through the specific variance of items in relation to the measurement error (Fornell & Larcker 1981). Reliability values above .60, and preferably greater than .70, suggest adequacy of the model (Fornell & Larcker, 1981; Marôco, 2010).

Other data obtained was the variance of the factors assessed by the average variance extracted test (AVE; Fornell & Larcker, 1981; Marôco, 2010), which is the variance of indicators explained by the latent factor in relation to the variance due to measurement error (Marôco, 2010). Factors with greater than .50 AVE are adequate, since the latent construct explains more than half of the variance of the observed variables (Dillon & Goldstein, 1984; Fornell & Larcker, 1981; Hair, Anderson, Tatham, & Black, 1998; Marôco 2010).

Unlike the initial analysis of the psychometric properties of the items, which maintained the same type of analysis, for the study of the test structure it was decided to test the model found by the manual's authors, using the estimation method of Weighted Least Squares Mean and Variance-Adjusted (WLSMV), specific for dichotomous data (Muthén & Muthén, 2010b). In the Cube Test manual, there are two types of exploratory factor analysis, one conducted by using SPSS 11.5 (Statistical Package for the Social Sciences, 2002) based on Classical Item Theory and the other by Testfact based on TRI. In this case, the comparison between the data in the test manual and the present study is limited, but there is the possibility to test the proposed structure and verify that it maintains, even with other forms of analysis.

Results

Seeking to verify if the psychometric parameters of the Cube Test remain acceptable, an analysis was done of the items by the Rasch model. Table 1 presents the information related to the difficulty of the items, *infit* and *outfit* values, the item-theta and item-total correlations. For comparison purposes, there are two columns for each psychometric property – the column in plain typeface presents the data in the manual, while the column with variables highlighted in bold refers to the data from current research.

Table 1
Psychometric Properties of the Items by the Rasch Model

Item	Difficulty	**Difficulty**	*Infit*	***Infit***	*Outfit*	***Outfit***	Item-total	**Item-total**	**Item-theta**
15	2.09	2.77	1.07	1.00	1.38	1.70	.24	.24	.47
14	1.59	2.50	1.04	.88	1.19	1.12	.20	.28	.54
13	1.55	2.35	.96	.89	1.05	.90	.29	.34	.56
11	1.27	1.85	1.01	.81	1.07	.83	.27	.32	.61
12	.97	1.80	1.00	.98	1.09	1.03	.37	.32	.53
10	.95	1.34	1.04	.92	1.07	.95	.37	.30	.57
9	.92	.97	.96	1.00	1.08	1.14	.44	.37	.53
8	.34	-.12	1.03	.95	1.07	1.23	.30	.30	.52
6	-.34	-.84	.93	1.03	1.00	1.62	.37	.37	.44
5	-.90	-1.51	1.01	.91	1.09	.94	.30	.37	.48
7	-1.02	-1.12	.82	.75	.71	.63	.32	.44	.59
4	-1.11	-1.87	1.03	1.07	.95	1.35	.32	.27	.35
3	-1.32	-2.00	.98	1.16	.90	1.72	.34	.29	.29
2	-2.40	-2.74	.94	1.17	.97	2.06	.28	.20	.21
1	-2.58	-3.39	.96	1.12	1.07	1.53	.24	.24	.20

Note. Variables highlighted in bold refer to the data of current research.

The results showed that four items presented unacceptable values ranges relative to *outfit*, which according Linacre (2002) is between .70 and 1.50. The remaining items proved fit, as in the study contained in the manual. In relation to *infit*, all items show acceptable value ranges, both those in the research and those in the manual. With regards to item difficulty, it is noted that the degree of difficulty sequence is similar to that identified in the manual. However, in the manual, item 5 tested with a higher level than 7. An interesting observation in this study sample is that, comparing the level of difficulty of each item, up to item 9 the difficulty levels are higher than the manual study, yet, from item 8, are smaller. Finally, with regard to the item-total correlation, it is observed that, as in test manual, all values were greater than .20. By comparing the data, it can also be observed that there was a balance between the two studies, that is, some items had higher values in the test manual, while others showed higher correlations in the present study. In Figure 2 the item representation can be seen on a scale from the most difficult to the easiest.

Figure 2 shows that the average ability of the subjects (letter 'M' on the left side of the figure) was slightly higher than the average difficulty of the items (letter 'M' on the right side of the figure). This result was also found in the analysis conducted for the manual. Even item-theta correlations – except for items 1, 2 and 3 – were adequate. In the test manual information was used through the TCT of item-total correlation, similar to the item-theta correlation. The data contained in the manual were also adequate, having as parameter the values highlighted by the American Institutes for Research that establishes the value of .20 as a good indicator of item discrimination. Finally, the accuracy obtained through the Rasch was .72, through the TCT was .77, both deemed adequate, as well as the accuracy in the manual, which obtained .80 through Cronbach's alpha.

With regard to the internal structure, in the Cube Test manual a two-factor structure is presented, verified both by factorial analysis of the Classical Test Theory, and by the full information factor analysis (FIFA) based on the TRI. It is understood that this division resulted from the item difficulty, since in the first factor are items with lower difficulty and featuring three response alternatives, while the second factor encompassed the more difficult items with four response choices, in addition to the item requiring three – not two – rotations of the figure to ascertain the correct response (Rueda & Muniz, 2012).

The greater complexity observed in the items of the second factor, just by increasing the number of alternatives and rotations to be considered, tends to require greater demand for cognitive resources, such as further developed working memory to store and operate the information pertaining to visuospatial reasoning with three, instead of one or two, rotations. However, all test items have the same content, figure type, and necessity of visuospatial reasoning to solve the task. Thus, independently of the two factors, the test assesses visuospatial reasoning, given that the second factor would be evaluated just as some specificity of this construct (Rueda & Muniz, 2012).

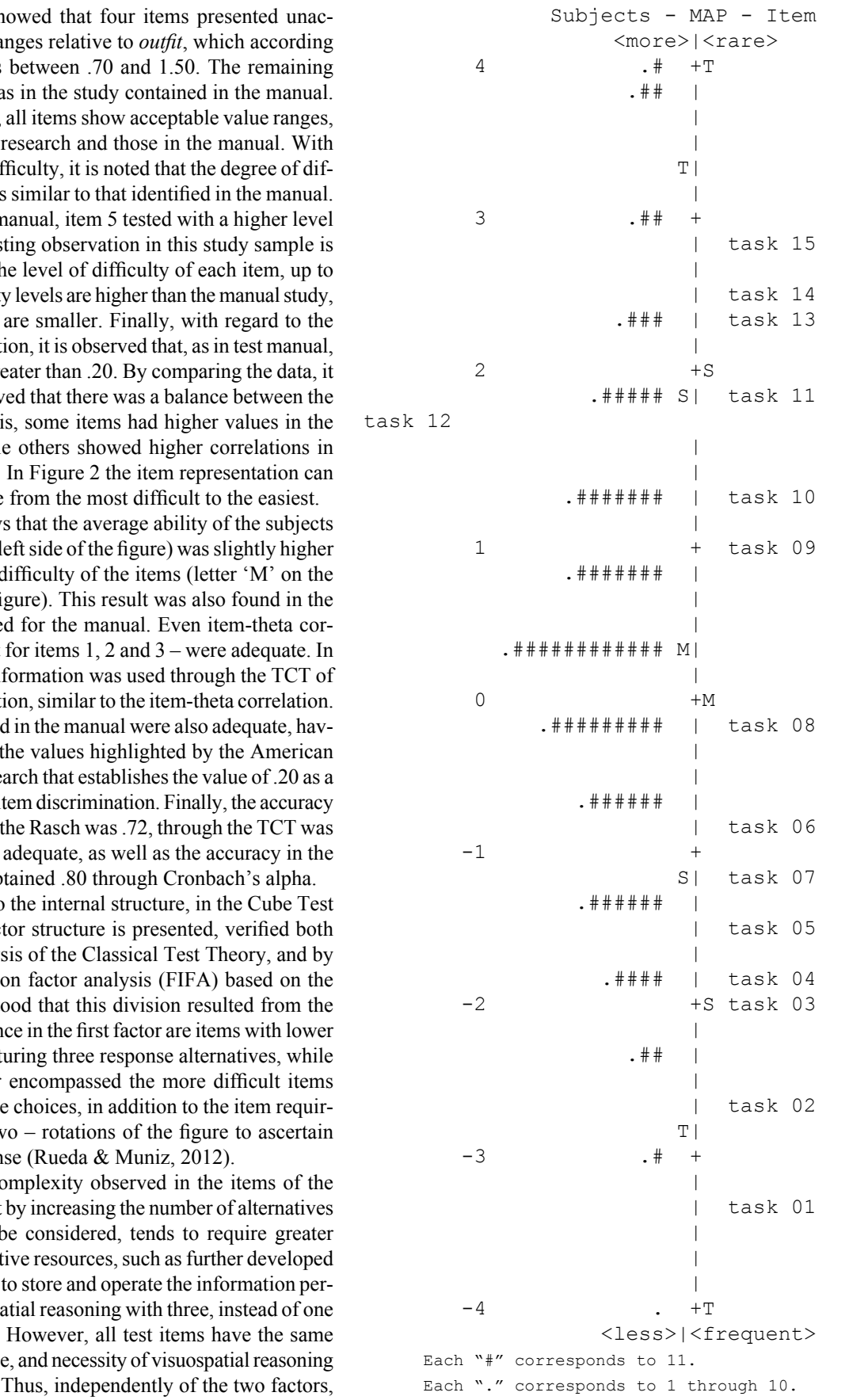

Figure 2. Representation of the items in order of difficulty and distribution of subjects by scale.

Recognizing the presence of just one construct in the two factors, the test can also be considered a measure for evaluating a general dimension, which is visuospatial reasoning. Thus, we sought to verify by means of confirmatory factor analysis, whether the one-factor model was more adequate than the two-factor model. So, initially we proceeded to the analysis with the two-factor model, as described in the manual, followed by the one-factor model, comprised of all test items. Tables 2 and 3 present the factor loadings, standard errors and significance values of the factor loadings for two-factor and one-factor models, respectively.

For Factor I a composite reliability of .445 and the average variance extracted of .339. In Factor II, the values achieved were .701 and .657, respectively. The values of composite reliability for Factor I were lower than expected, and acceptable for Factor II. The average variance extracted showed suitable only for Factor II. The correlation between the factors was -.074 ($p < .001$).

In the structure with one factor the composite reliability index showed .466 and the average variance extracted of .369. Both values are lower than expected. Regarding the fit indices for both models tested, these can be seen in Table 4.

Table 2
Factor Loadings, Standard Errors and Significance Values for the Two-Factor Model

Variables	Factor loadings		Standard errors	*p*-value
	Factor I	Factor II		
V1	.247	-	.057	.000
V2	-.232	-	.088	.008
V3	-.292	-	.080	.000
V4	-.468	-	.070	.000
V5	-.502	-	.066	.000
V6	-.704	-	.043	.000
V7	-.634	-	.045	.000
V8	-.980	-	.034	.000
V9	-.724	-	.038	.000
V10	-	.638	.043	.000
V11	-	.728	.037	.000
V12	-	.786	.033	.000
V13	-	.918	.023	.000
V14	-	.928	.022	.000
V15	-	.828	.029	.000

Table 3
Factor Loadings, Standard Errors and Significance Values for the One-Factor Model

Variables	Factor loadings	Standard errors	*p*-value
V1	.158	.052	.002
V2	-.145	.083	.080*
V3	-.184	.071	.009
V4	-.324	.066	.000
V5	-.376	.063	.000
V6	-.579	.043	.000
V7	-.487	.046	.000
V8	-.814	.032	.000
V9	-.610	.039	.000
V10	-.596	.039	.000
V11	-.677	.035	.000
V12	-.738	.032	.000
V13	-.890	.024	.000
V14	-.910	.022	.000
V15	-.786	.030	.000

Note. *Insignificant factor load.

Table 4
Fit indices for Both Models Tested by the Cube Test

Models evaluates	Fit indices					
	χ^2	*gl*	χ^2/gl	RMSEA (90% IC)	CFI	TLI
One-factor model	632.198	90	7.02	.094 (.084 – .101)	.86	.84
Two-factor model	223.443	89	2.51	.047 (.039 – .055)	.96	.96

Note. χ^2 = chi-square; *df* = degrees of freedom; χ^2/df = ratio chi-square by degrees of freedom (χ^2/df); RMSEA = root mean square error of approximation; CI = confidence interval; CFI = comparative fit index; TLI = Tucker-Lewis index (TLI).

Considering the fit indices, the two-factor model showed acceptable values while the one-factor model showed all indices lower than expected (Brown, 2006; Byrne, 2010; Schreiber et al., 2006).

Discussion

The main objective of the study was to verify whether the internal structure with two factors of the Cube Test, presented in the manual, would remain with data analysis from another sample, using the Confirmatory Factor Analysis. In addition, we did a comparison of results obtained using the TRI between this sample and the manual data. The purpose of these objectives was to contribute new information about the validity evidence of the Cube Test, approved for use by professional psychologists.

In the analyses conducted through TRI, indices of difficulty, fit, and item-theta and item-total correlation were conducted. The difficulty of the items followed an upward pattern, with ending items having greater difficulty than the initial items, following a sequence very similar to that contained in the manual. The difference was that in this study, item 7 is easier than item 6, and in the manual it is easier than items 6 and 5. Despite this difficulty inversion among items, the results are consistent with those proposed in the construction of the Cube Test, which requires a more complex ability to use visuospatial reasoning for each test item, especially when increasing the number of rotations to be mentally visualized in representing the final object (Rueda & Muniz, 2012). For example, in items 1, 2, 3, one must imagine the cube and rotate it once to get to the correct answer. The ability of Speeded Rotation/Spatial Relations is necessary so that the subject can mentally manipulate the cube, quickly (since there is a time limit). Moreover, the mental rotation for the first three items is less complex than when this same subject must mentally manipulate the object, but doing two (items 4, 5, 6, 7, 8, 9, 10 and 11) or three rotations (items 12, 13, 14 and 15). Still, specific to the difficulty inversion of items 6 and 7, one might consider the necessity of reversing the order of these test items.

The difficulty indices, both in the manual and in the current study showed balance between easy, medium and difficult items. However, from item 10 onward, in this study there were higher levels of difficulty than in the study within the manual. This might have happened, justifiably, because the manual study was composed, in its majority, by student majoring in fields that tend to attract subjects with more developed visuospatial reasoning (architecture, engineering and aeronautics training). Thus, they tend to present a skill level closer to, or in some cases higher than, the difficulty level of the items.

Despite this fact, in the two samples it can be verified that, in general, the subject's ability was slightly greater than the item difficulty, which can be considered similar. But this is a result that shows the average difficulty and ability, and does not mean that the test is easy, since the average score was 7.96, or half the test. Analyzing the cases individually, we identified that most people tended to choose correctly on the first 8 items, but missed the others, and many fail to reach the end of the test, given the time limit of 30 minutes.

With regard to the item-theta correlations, with the exception of items 1, 2 and 3, all others were within the acceptable standard, above .30, and item 3 reached .29. The TRI data obtained in this study endorse the psychometric quality of the Cube Test items. About the accuracy of the test, it can be noted that indices remained adequate, .72 by TRI and .77 by Classical Test Theory, both compatible with the manual.

After analysis of the psychometric properties of the Cube Test items, Confirmatory Factor Analysis was applied. As can be seen, the two-factor model showed better fit quality, with all indices, as expected. The first factor is comprised of items 1, 2, 3 and 4, 5, 6, 7, 8 and 9, and the second factor covers items 10, 11, 12, 13, 14 and 15. However, the composite reliability values and average variance extracted were lower than expected for the first factor, which tends to be an indication that the factor can be improved. The factor loadings of the first three items, taken from the same factor, contribute to this statement, as they are below .30. Furthermore, the correlation between Factor I and II was significant, but low.

The first three items of the Cube Test seem to contribute to the fact that Factor I is not as robust as Factor II.

Considering that on these first three items, one arrives at the correct response after employing only one rotation of the cube as stated, the subject tends to ascertain them without much demand for visuospatial reasoning, but by simply inferring that, when compared to the stimulus cubes, the correct alternatives exhibit little color variation while the larger color variation occurs on incorrect cubes. For example, with only one rotation, the colors of the cube faces on the correct choice must vary little from the stimulus cube, so alternatives with very different color arrangement compared to the stimulus cube are automatically eliminated. In contrast, from items 4 and 12, for which the correct response requires thinking through two and three rotations, respectively, only analyzing the arrangement of colors does not help, since the cube face color changes of all alternatives are greater. This makes the visuospatial reasoning essential to seek the correct alternative.

Despite this weakness indicated by Factor I, it can be understood more specifically as a function of the items 1, 2 and 3, the model with the two-factor structure presents acceptable fit indices, confirming the results obtained by the test's authors. In the Cube Test manual, the two-factor structure was understood by considering the complexity of the items and the presence of other abilities, such as working memory, which tends to have a more effective participation in better visuospatial processing. As noted by Primi (1998), the complexity factors that form items

which assess reasoning, also define the difficulty of the task. These complexity factors that are the attributes of the items can be number of elements, number and type of transformations, and perceptual organization of the stimulus. Considering that from item 9 onward, there are four alternatives and from item 12 onward it is necessary to use three rotations, we observe that the cognitive demand for the items of Factor II is greater and may involve skills not required for the items of Factor I. One example of the constitution of a factor by the items' complexity can be seen in the work of Primi et al. (2006), which, in investigating the internal structure of a fluid intelligence assessment test through items of simultaneous task coordination and selective attention/abstraction, found three factors, one of them comprised by only the most difficult test items. These authors realized that the complexity of the items composing this factor resulted from irrelevant information contained in the items' content and tended to require more selective attention, flexibility and attention control.

The argument of the item complexity used by the authors in Cube Test manual, on the two-factor structure of the test, is consistent with the research of Primi et al. (2006). Likewise, the understanding of the Cube Test structure presented in the current work tends to reaffirm the prevalence of visuospatial reasoning as primary construct measured by the test, as indicated in the Cube Test manual. The items and the action that the subject must perform to respond to the instrument require visuospatial reasoning. All items and alternatives are cubes, the task is to encode the images of a cube (item) and find, among 3 or 4 alternatives, which is the correct image after rotating the item (Cube) once, twice or three times.

As can be seen, between one item and another the difference is the structure of the cube, represented by the array of colors of each face, besides the varying quantity of rotations that must be effected to find the correct answer, and the number of alternatives. In addition, all items were constructed based on the need to use visuospatial reasoning, more specifically, the ability of Visualization (to mentally manipulate objects and be able to predict how they appear after a reorganization), Speeded Rotation/Spatial Relations (to manipulate objects quickly and using mental rotation and to identify objects in different positions) and Spacial Scanning (scanning, exploration, or visual search of an indicated path). On the Cube Test the subject must manipulate the figure and predict how it would appear (Visualization), rapidly using mental rotation as required in the task (Speeded Rotation/Spatial Relations), given that to do this the subject scans and searches the reorganization of colors from the rotation (Visual Scanning). All items require these abilities to be solved, and as such, one can say that the instrument measures the visuospatial reasoning.

These particularities of the items require the reasoning to be, aside from visuospatial, more complex and may also require other cognitive abilities such as working memory (Rueda & Muniz, 2012). As indicated by Primi (2002), more complex items of perceptual organization require the attention control processing of selective attributes or elements, having in view the association of these selective encoding and abstraction processes, given that visual processing is also required.

In the case of the Cube Test, since there is visual processing, using some visuospatial abilities that, together with the complexity of items, also tends to overwhelm working memory, which is temporary warehouse for storage of information that can be accessed, manipulated and reorganized to be used in some task (Malloy-Diniz, Sedo, Fuentes, & Leite, 2008). Thus, in the more difficult items there is a larger quantity of information to be rearranged to arrive at the correct answer, which may be an important variable contributing to the formation of another factor.

This analysis, more cognitive of the understanding of the test items' structure, is very important, because establishing reasoning ability measures is difficult, and the construction is generally directed by the application of psychometric criteria, that is, items are constructed, applied to a sample and parameters are verified, such as correlation and factor structure, and from this the test is concluded to be "good or not," often without analyzing the psychological sense of the data. In truth, the construction should be guided by indicators strictly based on theories, derived from a cognitive model of the thinking process (Wilhelm, 2005).

The Cube Test was also cognitively grounded for the items' construction, seeking to understand which cognitive functions were present together with visuospatial reasoning. Therefore, the non-confirmation of unidimensionality is understandable, and the factors that emerged in this study and the manual mean that a more cognitive response process is very relevant to understanding the test structure. The statistical analyses and procedures amplify the scientific merit and quality of the psychological instruments, but the theoretical understanding and qualitative analysis of cognitive processes should be considered as much as the psychometric.

Both methods, psychometric and cognitive-experimental, should be used together because the benefits are mutual. Differences in the correlations between items of reasoning used in cognitive research, and the latent variables of the ability test, can reveal important differences between the experimental tasks, as well as the variability of the differences of the items in psychometric reasoning tests may possibly be explained by the application of various reasoning process theories, like the theory of mental models (Wilhelm, 2005).

An example of the importance of psychometric and cognitive models both being used, as much in construction as in the understanding of a reasoning test, can be seen in the study of Muniz, Seabra and Primi (2012). The authors constructed items of an inductive reasoning test based on the theory of Josef Klauer (1990), which identified six paradigms for inductive reasoning and specific tasks for

each of them. After the construction of test, items for each paradigm which presupposed different forms of inductive reasoning, the structure was tested, and only two factors were identified. The factors correlated moderately and the authors opted to consider the unidimensionality, because analyzing the items, all demanded inductive reasoning and utilization of the proposed paradigms.

The results presented here reinforce the quality of the Cube Test and increase the number of empirical studies concerning the validity and accuracy of the test. It also shows that good results are maintained in a different sample than the study in the manual, and contributes to confidence in the greater use of this instrument for the Brazilian population. Even so, it is recommended that more studies be conducted with other samples from different regions of the country, and to continue to seek an increasingly robust understanding of the test structure.

Indeed, the test structure is the part that deserves further study to understand the multidimensionality found in the manual study and this research. It is necessary to investigate which other skills are demanded together with the visuospatial reasoning, which causes the test to be more than unidimensional. The complexity of the items, the quantity of alternatives and rotations required are variables that affect the formation of the test structure, but if there is such an influence, it is because these variables also tend to require different cognitive elements beyond the visuospatial abilities reported herein. This exploration of the cognitive demands of the Cube Test is even more important to be able to identify the cognitive processes used by the subjects and, thus, also contributes to a better understanding of the strengths and weaknesses in the test, and those that the subject needs to develop to improve their performance.

References

Almeida, L. S., Nascimento, E., Lima, A. O. F., Vasconcelos, A. G., Akama, C. T., & Santos, M. T. (2010). Bateria de Provas de Raciocínio (BPR-5): Estudo exploratório em alunos universitários. *Avaliação Psicológica, 9*(2), 155-162.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: The Guilford Press.

Byrne, B. M. (2010). *Structural equation modeling with MPLUS: Basic concepts, applications, and programming*. New York: Routledge, Taylor, & Francis.

Carrol, J. B. (1997). The Three-Stratum Theory of Cognitive Abilities. In D. P. Flanagan, J. L. Gensheft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (pp. 122-130). New York: The Guilford Press.

Dillon, W., & Goldstein, M. (1984). *Multivariate analysis: Methods and applications*. New York: Wiley.

Flanagan, D. P., & Ortiz, S. O. (2001). *Essentials of cross-battery assessment*. New York: John Wiley & Sons.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing, 18*(1), 39-50. doi:10.2307/3151312

Goel, V. (2005). Cognitive neuroscienceof deductive reasoning. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 475-492). New York: Cambridge University Press.

Hair, J. F., Jr., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis* (5th ed.). Upper Saddle River, NJ: Prentice Hall.

Holyoak, K. J., & Morrison, R. G. (2005). Thinking and reasoning: A reader´s guide. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 1-9). New York: Cambridge University Press.

Klauer, J. K. (1990). A process theory of inductive reasoning tested by the teaching of domain-specific thinking strategies. *European Journal of Psychology of Education, 5*(2), 191-206.

Linacre, J. M. (2002). What do infit and outfit, mean-squared and standardized mean? *Rasch Measurement Transactions, 16*(2), 878. Retrieved April 25, 2006, from http://209.238.26.90/rmt/rmt82a.htm

Linacre, J. M. (2005). WINSTEPS Rasch measurement computer program (Version 3.60) [Computer software]. Chicago, IL: Winsteps.com.

Malloy-Diniz, L. F., Sedo, M., Fuentes, D., & Leite, W. B. (2008). Neuropsicologia das funções executivas. In D. Fuentes, L. F. Malloy-Diniz, C. H. P. Camargo, & R. M. Consenza (Eds.), *Neuropsicologia: Teoria e prática* (pp. 187-206). Porto Alegre, RS: Artes Médicas.

Marôco, J. (2010). *Análise de equações estruturais: Fundamentos teóricos, software e aplicações.* Pero Pinheiro, Portugal: Report Number.

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.

Muniz, M., Seabra, A. G., & Primi, R. (2012). Validity and reliability of the Inductive Reasoning Test for Children – IRTC. *Psicologia: Reflexão e Crítica, 25*(2), 275-285.

Muthén, L. K., & Muthén, B. O. (2010a). Mplus: Statistical analysis with latent variables (Version 6.12) [Computer software]. Los Angeles, CA: Muthén & Muthén.

Muthén, L. K., & Muthén, B. O. (2010b). *Mplus: Statistical analysis with latent variables. User's guide.* Los Angeles, CA: Muthén & Muthén.

Pasquali, L., Wechsler, S., & Bensusan, E. (2002). Matrizes Progressivas do Raven Infantil: Um estudo de validação para o Brasil. *Avaliação Psicológica, 2*, 95-110.

Primi, R. (1998). *Desenvolvimento de um instrumento informatizado para avaliação do raciocínio analítico* (Tese de doutorado, Instituto de Psicologia, Universidade de São Paulo, SP, Brasil).

Primi, R. (2002). Complexity of geometric inductive reasoning tasks: Contribution to the understanding of fluid intelligence. *Intelligence*, *30*(1), 41-70.

Primi, R., & Almeida, L. S. (2000). *Manual prático da Bateria de Provas de Raciocínio (BPR-5)*. São Paulo, SP: Casa do Psicólogo.

Primi, R., Cruz, M. B. Z., Nascimento, M. M., & Petrini, M. (2006). Validade de construto de um instrumento informatizado de avaliação dinâmica da inteligência fluida. *Psico, 37*(2), 109-122.

Rueda, F. J. M., & Muniz, M. (2012). *Cube Test para avaliação do raciocínio visuoespacial*. São Paulo, SP: Vetor.

Schneider, J., & McGrew, K. (2012). *Cattell-Horn-Carroll (CHC) Theory of cognitive abilities definitions (CHC v2.0)*. Retrieved from http://www.iapsych.com

Schreiber, J. B., Stage, F. K., King, J., Nora, A., & Barlow, E. A. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research, 99*(6), 324-337. doi:10.3200/JOER.99.6.323-338

Sisto, F. F., Rueda, F. J. M., & Bartholomeu, D. (2006). Estudo sobre a unidimensionalidade do Teste Matrizes Progressivas Coloridas de Raven. *Psicologia: Reflexão & Crítica, 19*(1), 66-73.

Statistical Package for the Social Sciences. (2002). (*Version 11.5) [Computer software]*. Chicago, IL: Author.

Urbina, S. (2006). *Fundamentos da testagem psicológica*. Porto Alegre, RS: Artes Médicas.

Wilhelm, O. (2005). Measuring reasoning ability. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence* (pp. 373-407). London: Sage.

Wright, B. D., & Stone, M. H. (2004). *Making measure*. Chicago, IL: The Phaneron Press.

Recebido: 15/05/2013
Aceite final: 27/06/2013