Cross-battery factor analysis of the Battery of Reasoning Abilities (BPR-5) and Woodcock-Johnson Tests of Cognitive Ability (WJ-III)

Ricardo Primi
University of San Francisco – Brazil

Tatiana de Cássia Nakano
Solange Muglia Wechsler
Pontifical Catholic University of Campinas – Brazil

Abstract

This study performed a cross battery confirmatory factor analysis of BPR-5 and WJ-III in order to investigate which latent constructs are being measured by the subtests of both batteries. The sample was composed of 90 Psychology undergraduate students (68% women), ages ranging from 20 to 46 (M=26.49, SD=7.16). These students answered eleven subtests (5 from BPR-5 and 6 from WJ-III) as a part of their assessment course. Results supported a model of three correlated factors comprised of crystallized intelligence-Gc verbal reasoning (vocabulary, synonyms, antonyms, analonyms, analogyms, analogyms), fluid intelligence-Gf (abstract reasoning, concept formation and spatial relations) and visual processing-Gv (spatial reasoning, mechanical reasoning and numerical reasoning). Highly intelligent subjects also demonstrated an imbalanced profile of Gv over Gc. In conclusion, this study demonstrated the construct validity of these test batteries and confirmed the Cattell-Horn-Carroll (CHC) broad factors model for understanding and measuring intellectual differences.

Keywords: Cattell–Horn–Carroll theory, Cognitive abilities, Validity, Intelligence, Psychometrics.

Análise fatorial inter-baterias: Bateria de Habilidades de Raciocínio (BPR-5) e Bateria de Habilidades Cognitivas Woodcock-Johnson III

Resumo

Este estudo realizou a análise fatorial inter-baterias dos instrumentos BPR-5 e WJ-III com a finalidade de investigar quais construtos latentes estão sendo medidos pelos subtestes de ambas as baterias. A amostra foi composta por 90 estudantes de graduação em Psicologia (68% mulheres), com idades variando entre 20 e 46 anos (M=26,49, DP=7,16). Os participantes responderam a onze subtestes (sendo 5 da BPR-5 e 6 da WJ- III) como parte da disciplina de avaliação psicológica. Os resultados apontaram um modelo de três fatores correlacionados, sendo inteligência cristalizada - Gc (composto pelos subtestes de raciocínio verbal, vocabulário, sinônimos, antônimos, analogias), inteligência fluida - Gf (composto pelos subtestes de raciocínio abstrato, formação de conceito e relações espaciais) e processamento visual Gv (composto pelos subtestes de raciocínio espacial, raciocínio mecânico e raciocínio numérico). Estudantes com alta habilidade apresentaram um perfil desbalanceado, no qual Gv mostrou-se mais alto que Gc. Em conclusão, este estudo demonstrou a validade de construto destas baterias de testes, confirmando o modelo de fatores amplos proposto por Cattell-Horn-Carroll (CHC) para compreender e medir as diferenças intelectuais.

Palavras-chave: Teoria Cattell–Horn–Carroll (CHC), Habilidades cognitivas, Validade, Inteligência, Psicometria.

Análisis factorial intra-baterías: Batería de Habilidades de Raciocinio (BPR-5) y Batería de Habilidades Cognitivas Woodcock-Johnson III

Resumen

Este estudio realizó un análisis confirmatorio de intra-batería de los instrumentos BPR-5 y WJ-III con el propósito de investigar qué constructos latentes están siendo medidos por las sub-pruebas. La muestra estuvo compuesta por 90 estudiantes de psicología (mujeres 68%), con edades que oscilaban entre los 20 y 46 años (M = 26,49, SD = 7.16). Estos estudiantes respondieron once sub-pruebas (5 de BPR-5 y 6 de WJ-III) como parte de su curso de evaluación. Los resultados apuntaron un modelo de tres factores correlacionados: inteligencia cristalizada Gc (raciocinio verbal, vocabulario, sinónimos, antónimos, analogías), inteligencia fluida Gf (raciocinio abstracto, formación de conceptos y relaciones espaciales) y el procesamiento visual GV (raciocinio espacial, raciocinio mecánico y raciocinio numérico). Estudiantes con alta habilidad presentaron un perfil desequilibrado, en el cual el GV se mostró más alto que el GC. En conclusión, este estudio demostró la validez de constructo de estas baterías de pruebas y confirmó el modelo de factores amplios Cattell-Horn-Carroll (CHC) para comprender y medir las diferencias intelectuales.

Palabras clave: Teoría Cattell-Horn-Carroll, Habilidades cognitivas, Validez, Inteligencia, Psicometría.

The identification of general intelligence through psychometric studies has been established since the later 1800, thus stimulating a debate on the proper way to assess this construct (Flanagan & Harrison, 2005; Roberts, Zeidner, & Matthews, 2001). The models proposed for intellectual assessment are remarkably similar in structure and organization, although notable differences are present. According to each theory, the test batteries vary in item content and task demands. For this reason, it is important to investigate how persons perform in intelligence tests according to the type of task demanded, which have been named the cross-battery assessment, initially proposed by McGrew and Flanagan (1998) as a way of assessing the total range of intellectual abilities.

This type of approach is based on the hypothesis that a combination or logical selection of psychological tests could better identify a construct, thus being able to measure a selective range of specific abilities, in a valid way, with depth and more adherence to empirical evidence than would be any single test battery (Flanagan & Ortiz, 2001; Schretlen, Van Gorp, Wilkins, & Bobholz, 1992). The extensive study carried on by Flanagan, McGrew, and Ortiz (2000) indicated that the Wechsler Intelligence Scales (WISC-III and WAIS-III), for example, were only measuring specific areas of intellectual functioning, therefore indicating that they should be combined with other measures to provide a more complete basis for intellectual assessment.

The cross-battery assessment principle represents a significantly improved method of measuring cognitive abilities. This approach is theory focused and can be reached through the combination of subtests by satisfying logical criteria to optimize the measurement of intelligence founded on the best evidence available (Flanagan & Ortiz, 2001). According to Flanagan (2000), the cross-battery studies provides a set of psychometrical and theoretical principles and procedures for supplementing any intelligence battery with tests from other batteries to broaden the range and improve upon the measurement of the intelligence abilities represented in the assessment. The practical recommended procedure is to select tests based on Cattell–Horn–Carroll theory of cognitive abilities - CHC theory (McGrew, 2009), trying to combine abilities which are not measured in the original test selection. Therefore, at least two qualitative different core abilities can represent each broad category in the theoretical model and can be used for intellectual assessment (Flanagan & Ortiz, 2001).

In order to construct a test battery, according to this model, the first step is to conduct an exploratory factor study when the factorial structure is unknown and the researcher is attempting to ascertain what

structure may exist. However, the confirmatory factor analysis procedure requires prior knowledge about the expected factorial structure with the purpose to test hypotheses about a possible n exploratory model believed to underlie the data (Woodcock, 1990). According to this author, some problems may occur when only one battery is factor analyzed, due to variables restriction to tasks only represented in the battery itself, which might not include enough markers for each embedded factor (Woodcock, 1998).

An exception would be the Woodcock-Johnson III battery (Woodcock, McGrew, & Mather, 2001). This battery is a revision of previous measures, the Woodcock-Johnson Psycho-Educational Battery (Woodcock & Johnson, 1977) and the Woodcock-Johnson Psycho-Educational Battery Revised (Woodcock-Johnson, 1989), which were constructed based on Cattell-Horn's Gf-Gc theory. The crystallized intelligence (Gf), as proposed by Cattell, is defined as the ability to make relationships between stimuli and inferences and is highly associated with knowledge, thus being impacted by learning and culture (McGrew & Flanagan, 1998; Primi, 2002). On the other hand, the Gf (fluid intelligence) is associated with process and skills to perform basic activities, thus not depending on cultural experiences (Almeida, Guisande, & Ferreira, 2009; Almeida, Lemos, Guisande, & Primi, 2008). The revisions presented in the Woodcock-Johnson III battery (WJ-III), cognitive and achievement tests were based on the CHC theory (Woodcock, McGrew, & Mather, 2001; McGrew & Woodcock, 2001). The WJ-III cognitive tests, standard and extended versions, allows the measurement of narrow abilities from stratum I, seven broad intellectual abilities from stratum II as well as the general ability (g) from stratum III, according to Schrank and Flanagan (2003). The following broad abilities can be assessed through the WJ-III cognitive battery: crystallized intelligence (Gc), long term retrieval (Glr), visual spatial thinking (Gv), auditory processing (Ga), fluid reasoning (Gf), processing speed (Gs) and short term memory (Gsm), as described by Schrank, Flanagan, Woodcock, and Mascolo (2002).

According to the present federal regulations by the Brazilian Federal Council of Psychologists (Conselho Federal de Psicologia, 2003, 2010) all tests to be used in the country have to demonstrate scientific quality, which involves validity evidences and reliability as well as adequate norming to the population. These guidelines are approved by the International Testing Commission (2011), thus regulating that all foreign psychological tests to be used in another country have to ensure that the adaptation process takes full account of linguistic and cultural differences in the target population.

In order to attend to the previously mentioned requirements of test usage, a series of investigations were performed to adapt the Woodcock-Johnson III cognitive tests to Brazilian children and youth. Wechsler, Vendramini, and Schelini (2007), for example, verified the need to complement the WJ-III comprehension tests (vocabulary, synonymous, antonymous, analogies) with items drawn from Brazilian text books. In a national study with all cognitive tests from the WJ-III standard battery (Wechsler et al., 2010) the results indicated that tests measuring auditory processing as well as crystallized intelligence (verbal comprehension) had to be adapted in order to be used for Brazilians' intellectual assessment. In addition, validity evidences for the WJ-III Brazilian adapted version were obtained as gains in cognitive process were observed according to children's development (Wechsler & Schelini, 2006). Children with learning difficulties performed significantly worse in WJ-III Brazilian version when compared to those with no difficulties, thus indicating criteria validity of this battery (Mól & Wechsler, 2008). The convergent validity of the WJ-III Brazilian version was confirmed in another study due to the high correlations of their results with another validated intelligence measure in Brazil, the WISC-III, thus confirming findings obtained in the US with the original WJ-III battery (Chiodi & Wechsler, 2009).

Another test battery validated in Brazil is the Battery of Reasoning Tests-5 (Bateria de Provas de Raciocínio-5, Almeida & Primi, 1998), which is an adapted version of a Portuguese battery called Battery of Differential Reasoning Tests (BPRD). This battery is comprised of the most used psychological tests for assessing intelligence in Portugal, thus spreading its popularity to Brazil (Almeida, Lemos, & Primi, 2011). BPR-5 enables the assessment of cognitive aspects more related to the g factor as well as other

components associated with specific aptitude. According to their authors, this battery comprises a set of different tests in content, aimed to assess the capabilities of understanding of relationships between elements (inductive reasoning) and the application of these inferred relationships to new situations (deductive reasoning). The battery includes Form A (6th grade through 8th grade of elementary school) and Form B (1st, 2nd and 3rd grades of high school). The differences between forms occur in the item content for each subtest (abstract-figurative, numeric, verbal, practical, mechanical and spatial).

Several cognitive abilities can be measured through the BPR-5. The abstract-figurative reasoning test evaluates Gf through analogies formed by complex graphic designs or geometric figures without any apparent relationship among them. The numeric reasoning test assesses Gf and quantitative ability through content formed by linear or alternated number sequences. The verbal reasoning test assess Gf and Gc intelligence and is formed by heterogeneous set of relationships that can be established between word analogies. The mechanical reasoning test is used to assess Gf mechanic knowledge and it is formed by problems associated with experiences of daily routine, which can be addressed or not to educational experiences involving basic knowledge of physics and mechanics. Spatial reasoning test is used to assess Gf and capacity of visual processing Gv, and is formed by a series of cubes, linear or in motion, requiring the inferences of their relative positions when changed. BPR-5 is assumed to be a battery of reasoning or fluid intelligence more than crystallized intelligence. By combining the same process of inductive reasoning with different contents, this test battery intends to evaluate simultaneously and complementarily cognitive aspects related to the more general factor of intelligence or g factor (Almeida et al., 2011).

A series of investigations performed in Brazil with the BPR-5 demonstrated positive evidences of its validity and reliability. One of the first studies (Almeida & Primi, 1998; Primi & Almeida, 2000) indicated good reliability indexes ranging from .63 to .87 for the subtests, and around .90 for the full score. These authors confirmed a single factor explaining approximately 55% of the variance

representing a compound average of fluid and crystallized intelligence, visual processing, quantitative skills and practical knowledge of mechanics. Correlations of BPR-5 with school performance were generally positive, reaching 0.54 (p <0.001). Several other studies concerning the validity of BPR-5 are summarized in Almeida et al. (2011).

Studies with the BPR-5 replicate the general factor when total score resulting from the five subtests are analyzed (Almeida et al., 2011). Although the finding of a general factor is consistent with intelligence models it has been demonstrated firstly by Woodcock (1990) that when other markers of intelligence are added in cross battery factor analysis it is possible to find a different factor structure from the original one when the only tests that are factor analyzed are the ones of only one battery. For instance, Woodcock (1990) conduced several cross battery factor analysis comparing WJ with WISC-R, WAIS, WAIS-R, K-ABC and Stanford-Binet-IV. In general these analyses support the CHC model. By comparing test batteries Woodcock found that WJ-R measures all eight of its factors (long-term retrieval, short-term memory, processing speed, auditory processing, visual processing, comprehension-knowledge, fluid reasoning and quantitative ability) with two or more clean measures. WJ measures five, the SB-IV measures four factors, K-ABC measures three factors, the WISC-R and WAIS-R measure two factors. The findings show that the combined set of subsets from the WJ-R and the other cognitive batteries load appropriately onto a set of factors defined by Gf-Gc theory. The results of these studies "demonstrate the need for factor analytic studies in which the set of variables is not constrained to the limited set of subsets that have been published as a battery. It is indicated that the set of variables to be included in a factor study must include enough breadth and depth of markers to ensure that the presence of all major effects can be identified" (Woodcock, 1990, p. 231).

The relevance of using cross-battery assessment was ratified in other studies. When comparing different test batteries, Shrank and Flanagan (2003) observed that WJ-III batteries (standard and extended versions) measured 8 factors from CHC theory (long-term retrieval, short-term memory, processing speed, auditory processing, visual processing, comprehension-knowledge, fluid reasoning and quantitative

ability) with 2 or more clean measures. However, the Stanford Binet-IV measures only four factors, the Kaufman-ABC measures three factors, the WISC-R and WAIS-R measure two factors. Another study with children using 12 subtests of WISC-III and 18 subtests of the WJ-III (Phelps, McGrew, Knopik, & Ford, 2005) demonstrated the importance of combining these subtests for intellectual assessment. Similar results were obtained by Flanagan (2000) when performing a cross battery factor analysis of WISC-R and WJ-R. These findings indicated that the combined set of subsets from the WJ-R and the other cognitive batteries is more appropriate since they allow a wide range of factors as defined by the Gf-Gc theory (Woodcock, 1990).

Although the factorial studies of BPR-5 have found a general factor that explains most of the covariance between subtests, these studies were conducted only with the battery subtests. One question that can be raised is what would happen if we conduct a factor analysis including other markers of other batteries as suggested by Woodcock (1990). Would a general factor emerge? Despite the importance and robustness of the general factor to explain the correlations among intelligence measures there is also extensive literature demonstrating the importance of broad factors of the second stratum for intellectual assessment (Ackerman, 2003; Lubinski, 2010). Thus, the purpose of this study was to conduct a confirmatory factor analysis of BPR-5 with the Brazilian version of the WJ-III to test their subtests for a cross-battery assessment, according to the recent advances on CHC literature.

Method

Participants

The sample was composed of 90 Psychology undergraduate students, 62 women (68.9%) and 28 men (31.1%) with ages ranging from 20 to 46 (M=26.49, SD=7.16). Students were from a private university from a city of the interior of the state of Sao Paulo.

Measurements

The Battery of Reasoning Tests -BPR-5 (Almeida & Primi, 1998).

BPR-5 was developed from the Differential Reasoning Tests Battery (BPRD; Almeida, 1988). It is composed of five sub-

tests: Abstract Reasoning (AR) consisting of 25 items involving analogies with geometrical figures; Verbal Reasoning (VR) consisting of 25 items involving analogies between words; Numerical Reasoning (NR) consisting of 20 items in which linear or alternating series of numbers are presented and the student must find the rules of arithmetical progression for each series in order to find the two numbers that complete the sequence; Spatial Reasoning (SR) consisting of 20 items that present sets of three-dimensional cubes in motion for the student to find the type of motion from an analysis of different faces and then choose the answer that represents the last cube in the series; and Mechanical Reasoning (MR), which is composed of 25 items containing pictures of the practical contents of physics and mechanics from which the student must choose the answer that best represents the outcome of every situation. For this study it was used the Form B for students ranging from the first to third grades of high school.

BPR-5 factors are interpreted from psychometric models and in cognitive psychology. Items are analogies and series of problems involving mainly inductive reasoning with different contents. Therefore, it is expected that all subtests measure a broad Gf factor. At the same time each subtest is based on a different content with the intent of measuring other specific factors associated with the content of tasks; namely, Gc in the VR and NR, Gv (visuo-spatial intelligence) in the SR, AR MR and quantitative reasoning (RQ) in NR. A number of empirical studies reported elsewhere support these interpretations (Primi et al., in press).

Woodcock-Johnson Tests of Cognitive Abilities-WJ- III (Woodcock, McGrew, & Mather, 2001)

The Brazilian adaptation of the WJ-III cognitive battery was selected for the present research. Two Brazilian versions were elaborated, one for children and another for adults, which are group administered. The original WJ-III provides similar format for both adults and children and is individually administered. The Brazilian children's version has already been validated in previous studies (Wechsler & Schelini, 2006; Wechsler, Vendramini, & Schelini, 2007; Wechsler et al., 2010).

The adult's version is comprised of 9 subtests, which are: Verbal Comprehension-Knowledge (Gc- measuring crystallized

intelligence), Visual Auditory Learning (Glr-reflecting the associative memory); Spatial Relations (Gv-ability to perform visual spatial thinking); Concept Formation (Gf- a test of categorical reasoning or fluid intelligence); Visual Matching (Gs- a test of visual processing speed) and Auditory Memory (Gsm-ability to use working memory).

The Verbal Comprehension-Knowledge subtest test is composed of 4 subtests: Vocabulary, Synonyms, Antonyms and Analogies. Items for these subtests were retrieved from the Electronic Dictionary Houaiss for Portuguese (Houaiss & Villar, 2009). A total of 370 words were first selected: 177 substantives for the vocabulary test, 88 for synonyms, 105 for antonyms. The substantives were transformed to drawings for the vocabulary subtest. The list of items was reviewed by groups of college students from different areas in order to eliminate those who could favor a specific knowledge area or those with highest difficulty (0% accuracy). The final list was reduced as following: Vocabulary (Voc-WJ, 38 items), Synonyms (Syn-WJ, 30 itens), Antonyms (Ant-WJ, 18 words). For the analogies test (Ana-WJ) 25 items were selected among the most difficult ones from the child's version (Wechsler, 2009). The Spatial Relation (SR-WJ) is composed by 32 items, where parts of a figure are presented to be assembled. The Concept Formation test (CF-WJ) is composed of 35 items representing figures in different situations requiring logical rules to organize two to three groups of figures. These were the only subtests from this battery used in this research due to time constraints.

Procedures and data analysis

After receiving the Ethics Committee's formal approval, Psychology students were administered both test batteries in group situations, divided in two sections. At first, students responded to WJ-III tests during approximately 1 hour 30 minutes. After one week, the second administration was made, also lasting 1 hour 30 minutes, when they answered to the five subtests of the BPR-5.

Confirmatory factor analysis was performed on the MPLUS program (Muthén & Muthén, 2010). Graphics were prepared on the AMOS 16 (Arbuckle, 2007). We estimated models employing the method of maximum likelihood. We tested four models: Model 1 (M1) was composed of a general factor where all tests were specified to load on the general factor. Model 2 (M2) was a three correlated factors: (a) Fluid reasoning (Gf) composed by the subtests: Abstract Reasoning, Concept Formation Test and Numeric Reasoning, (b) Crystallized intelligence (Gc) composed by Verbal Reasoning, Vocabulary, Synonyms, Antonyms and Analogies; and (c) Visual Processing (Gv) composed by the subtests Spatial Relation, Spatial Reasoning, Mechanical Reasoning and Numerical Reasoning. Numerical Reasoning was also permitted to load in Gv because most tasks interspersed two numerical series that needed to be separated so that the problem can be solved. A common strategy is to visualize the numbers set separately. Therefore one important requirement of the solution to these problems involves visual processing.

Model 3 was the same as Model 2 with a slight modification in which the only difference is that WJ-III Spatial Relations (SR_WJ) was reassigned to the fluid intelligence factor instead of visual processing. This was done after a task analysis of this subtest that suggested that it involves basic process of visual comparison among stimulus trying to identify the pieces that are parts of a target. This is coherent with analytical strategy related to inductive reasoning tests (Gf). Usually subtests of Gv require hard visual processing such as rotating three dimensional mental images or mentally seeing two numerical series visually apart while they are mixed in the front of their eyes. With this model we test the hypothesis that these subtests might not need such hard visual processing as the others Gv tasks.

Model 4 was an exploratory test suggested by the modification indexes. This model is the same as Model 3 with the only difference that Synonyms (Syn_WJ) was permitted to load on visual processing factor. This seems counter intuitive at first but was an unexpected and interesting pattern that further exploration indicated that is related to the contrasts of visual versus crystallized intelligence.

In order to identify the model we set the metric of the latent variables to fixing one of the indicators variables per factor to have a regression weight of 1. We tested the model fit by looking at four indices following the recommendations of the literature (Byrne, 2001; Schweizer, 2010): (a) Chi-square that indicates the magnitude of the discrepancy

between the observed and modeled covariance matrix. High values indicate misfit, however, the chi-square is affected by sample size, so it is recommended to divide it to the degrees of freedom. Values of less than 2 are indicative of a good fit; (b) Comparative Fit Index (CFI) that calculates the relative adjustment of the model by comparing it with the null model (where the variables have zero correlation with each other). Values above .95 indicate good fit; (c) Root-Square Error of Approximation (RMSEA) is also a measure of discrepancy but that penalizes model complexity. Values less than .05 indicate a good fit; and finally, (d) Standardized Root Mean Square Residual (SMR) that reports the average standardized residual, that is, the difference between the observed and modeled correlation coefficients. Values less than .10 are indicative of good fit.

According to the research questions and hypothesis from the literature we expected that Model 1 would not present a good fit as compared with Models 2 and 3. This is assumed because a cross battery factor analysis with more pure markers of WJ-III of the three major constructs measured by the BPR-5 would help to identify the broad factors underlying the five subtests.

Results

Table 1 shows a correlation matrix among all intelligence tests of BPR-5 and WJ-III. It also shows summary information (means and standard deviation) about all the eleven tests. Expected means for this normative group are: VR=17.9, AR=18.1, MR=13.2, SR=13.9, N=13.0. Therefore this sample has a slight above average on Verbal Reasoning and Spatial Reasoning and below average in subtests Abstract Reasoning, Mechanical Reasoning and Numeric Reasoning.

The covariance matrix among eleven measures was analyzed with a confirmatory factor analysis performed in MPLUS. Model fit results of the four tested models are presented in Table 2. The final best model with estimated factor loadings is presented in Figure 1. Model 1, the general factor, produced a poor fit indicating that not all covariance among variables was accounted by a general factor. Model 2 that proposed three correlated factors (Gf measured by Abstract Reasoning, Conception Formation Test and Numeric Reasoning; Gc measured by the Verbal Reasoning, Vocabulary, Synonyms, Antonyms, Analogies and Gv measured by Spatial Relation, Mechanical Reasoning and Numeric Reasoning), showed a better fit, although not reaching acceptable levels (CFI =.84, RMSEA .10 and SMR =.11). Considering that Spatial Relation (SR_WJ) could be relatively more strongly related to fluid intelligence than visual processing, Model 3 changed the specification of the path to this indicator linking it to Gf and removing the link to Gv. With this change Model 3 reached accepted levels of fit.

Table 1 – Descriptive statistics and correlation among intelligence tests of BPR-5 and WJ-III

Subtests	M	SD	1	2	3	4	5	6	7	8	9	10
BPR-5												
1. VR	18.9	3.3										
2. AR	15.9	4.5	.254*									
3. MR	11.8	4.4	.380**	.331*								
4. SR	14.5	4.5	.341**	.108	.608**							
5. NR	11.6	4.2	.292*	.287*	.555**	.576**						
WJ-III												
6. Voc_WJ	10.7	5.5	.245*	.356**	.475**	.280*	.313*					
7. Syn_WJ	6.9	3.8	.294*	.225	.024	.059	-.021	.482**				
8. Ant_WJ	10.6	5.0	.218	.353*	.154	.288*	.216	.560**	.801**			
9. Ana_WJ	13.7	4.7	.355**	.209	.297*	.187	.268*	.348**	.552**	.496**		
10. SR_WJ	64.2	8.8	.241*	.503**	.458**	.167	.253	.256*	.223	.097	.247*	
11. CF_WJ	17.4	8.0	.302**	.359**	.241	.173	.182	.233*	.149	.147	.276**	.339**

*p≤0,05, ** p p≤0,01

Note: Abstract Reasoning (AR); Verbal Reasoning (VR); Numerical Reasoning (NR); Spatial Reasoning (SR); Mechanical Reasoning (MR); Vocabulary (Voc-WJ); Synonyms (Syn-WJ); Antonyms (Ant-WJ); Analogies (Ana-WJ); Spatial Relation (SR-WJ); Concept Formation test (CF-WJ).

Table 2 – Model fit results of Confirmatory Factor Analysis of four models

Models	X^2	df	X^2/df	RMSEA	CFI	SMR
Model 1: g-factor	117.2	44	2.66	.136	.699	.135
Model 2: Gf-Gc-Gv	77.7	40	1.94	.102	.846	.111
Model 3: Gf-Gc-Gv b	60.6	41	1.47	.073	.920	.098
Model 4: Gf-Gc Gv c	48.3	40	1.20	.048	.966	.084

Figure 1 presents the estimated standardized factor loadings of this final model composed of three factors. As can be seen, all the indicators have moderate to high loadings on their corresponding latent factors. Also, these latent factors are intercorrelated consistently with a general factor in a second level.

Modification indexes for Model 3 suggested the addition of path between Synonyms and Gv factor. Model 4 was tested with this modification and its results presented very good fit indexes, as shown in the last row of Table 1. The new path between Gv and Synonyms was estimated to be -.45. Interestingly, this result indicates an inverse relationship between one of the most valid indicators of crystallized intelligence and the latent visualization factor after the general covariance between measures are accounted for.
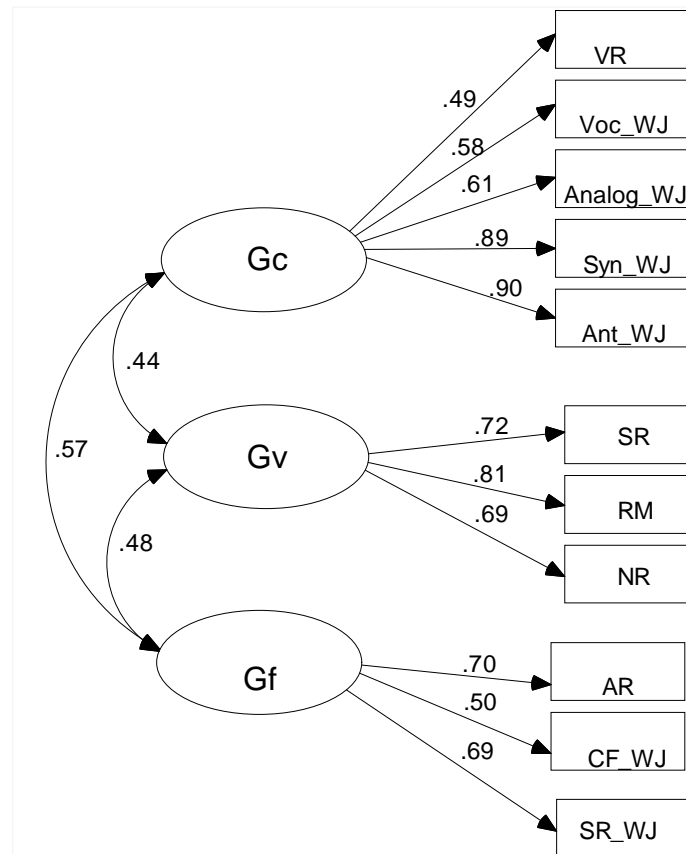

Figure 1 – Final three correlated factors (Model 3) model with estimated standardized factor loadings

Based on this result it was explored the mean of ability profiles divided by general intelligence to see if different patterns could be observed. For this analysis we first computed z scores for each observed indicator (eleven variables). Then we computed a general factor score by averaging z-scores in all indicators (g-factor), and calculated three factor scores fGv, fGf and fGv by averaging z-scores indicators of each latent factor according to the Model 3. Figure 2 summarizes the average Gv-Gg-Gc profile of four groups according to general intelligence (the groups were divided by the quartiles of g-factor z-scores. The values for percentiles 25, 50 and 75 are: .35,-.06. and .56).

It can be seen that the general level of the three factors are increasing from groups 1 to 4 (averages of -.80, -.18, .16 and .87) consistent with the g-factor model. But, interestingly, there is also an imbalance in the profile involving Gv and Gc. Group 3 has a higher Gc than Gv but group 4 has the opposite pattern. This result suggests that in the high end of general intelligence scores in addition to a global high average on all factors there is a very high Gv when compared to Gc.
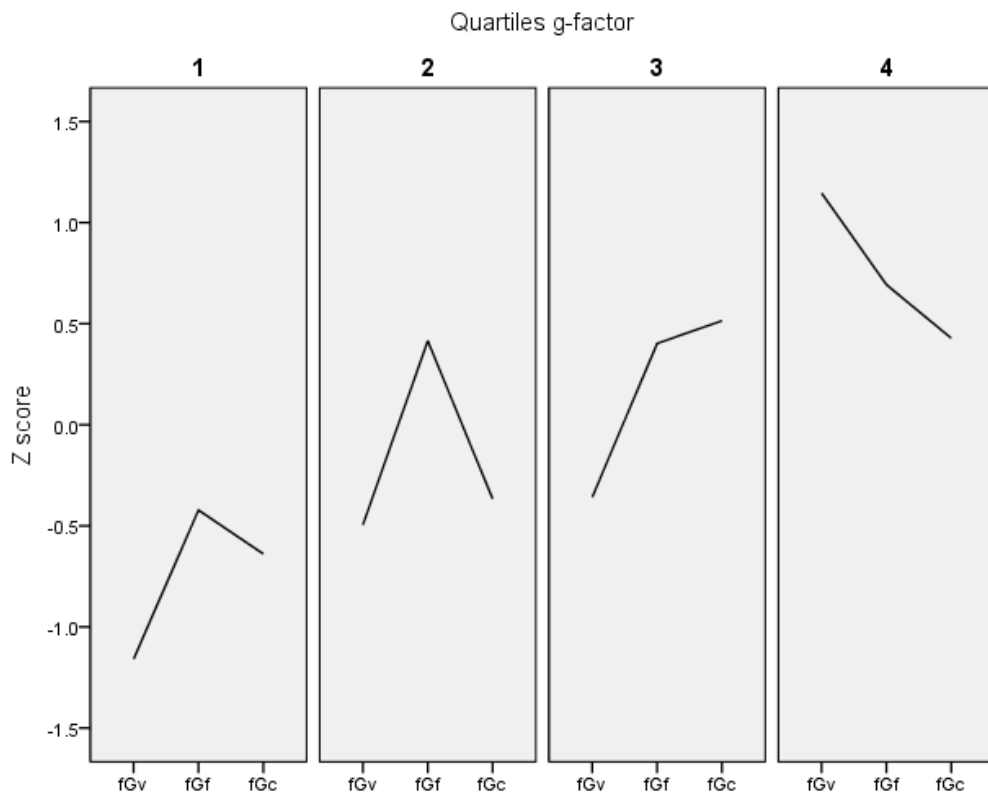

Figure 2 – Mean abilities profiles on Gv, Gf and GC by four groups of general intelligence (quartiles)

Discussion

The present study performed a cross battery confirmatory factor analysis of BPR-5 and WJ-III trying to clarify what latent constructs are being measured by both instruments. Results support a three correlated factor model Gv, Gc and Gf consistent with previous proposed conception (Almeida et al., 2011; Primi, Couto, Almeida, Guisande, & Miguel, in press), and with the CHC model (McGrew, 2009; McGrew & Flanagan, 1998). It also suggests two new clarifications on the meaning of the subtests. Numerical reasoning of BPR-5 appears to be more related to visual processing that it was originally thought. This is consistent with the task analysis that suggests that problems could be more easily solved by the use of a visual strategy. The arithmetic operations related to quantitative reasoning that is required to solve these problems might be a source of individual differences for younger

subjects. For the older subjects of the present sample these operations might be already mastered, therefore it might no longer be an essential source of individual differences.

This study also suggests that spatial relations of WJ-III might be more related to fluid intelligence than visual processing. The WJ-III manual (McGrew & Woodcock, 2001) reports several confirmatory studies where this test loads on Gv broad factor. But an earlier study with WJ-R Woodcock (1990) indicated that spatial relations had small loading on Gv of .19 and a higher loading of .40 on Gf factor. At that time the author concluded that "spatial relation was a mixed Gv and Gf measure" (p.252). The present study supports this second interpretation.

In conclusion, this results support the construct validity of BPR-5 and the Brazilian adaptation of WJ-III. Results are in accordance with the interpretations proposed for the subtests based on task analysis cognitive psychology (Almeida & Primi, 1998). It also supports the CHC broad factor model as a way of understanding the constructs measured by cognitive test batteries.

Findings from this study support several authors' arguments (Flanagan et al., 2000; Woodcock, 1990) in favor of cross battery factor analysis as a proper way to reveal the underlying structure of cognitive batteries. Therefore, in order to clarify the constructs which are being measured by a set of cognitive tests, the inclusion of other subtests in order to obtain a broad spectrum of abilities for intellectual assessment is recommendable. The results also confirm the literature and other studies of this nature, conducted with other sets of intelligence instruments (Flanagan, 2000; Phelps et al., 2005). Only with this breadth and depth construct sample it is possible to really reveal what latent constructs each subtest is measuring.

It is interesting to note that previous studies that factor analyzed BPR-5 subtests alone have consistently found a general factor. But, in the present study that used a cross battery approach three distinct factors were obtained. Although a general factor is an undeniable phenomena in the understanding and measuring intelligence, this study also indicated that the broad factors model of CHC are also important for understanding specific abilities and patterns. This is consistent with other researchers (Ackerman, 2003) who have demonstrated the practical importance of Gf and Gc and conative factors in adult development of expertise. The relevance of Gv in predicting talent of people working on the areas of science, technology, engineering, and mathematics had already been emphasized by Lubinski (2010). In this study, the high talented students demonstrated an imbalanced profile of Gv over Gc in addition to the general high scores in all subtests (high g factor). This might indicate the practical importance of considering these broad factors in understanding and measuring individual differences in intelligence.

This study had, as a limitation, the small number of participants and the predominance of female participants, enrolled in a specific university course. As gender as well as type of course tends to impact results on Gc and Gv (Wechsler, 2011), future studies with these test batteries have to consider the impact of these variables on the explained cognitive factors. Therefore, sample diversity is recommended in order to verify if the three factors model indicated in this research can be confirmed.

References

Ackerman, P. L. (2003). Cognitive Ability and Non-Ability Trait Determinants of Expertise. *Educational Researcher*, *32*(8), 15-20.

Almeida, L. S. (1988). O impacto das experiências educativas na diferenciação cognitiva dos alunos: Análise dos resultados em provas de raciocínio diferencial [The impact of educational experiences on students' cognitive differentiation: Analysis of results in tests of differential reasoning]. *Revista Portuguesa de Psicologia*, *24*, 131-157.

Almeida, L. S., Guisande, M. A., & Ferreira, A. I. (2009). *Inteligência: perspectivas teóricas* [Intelligence: theoretical perspectives]. Coimbra: Almedina.

Almeida, L. S., Lemos, G., Guisande, M. A., & Primi, R. (2008). Contribuciones del factor general y de los factores específicos en la relación entre inteligência y rendimiento escolar [Contributions of the general factor and specific factors in the relationship between intelligence and school performance]. *European Journal of Education and Psychology*, *1*, 5-16.

Almeida, L. S., Lemos, G. C., & Primi, R. (2011). Recensão crítica: Bateria de Provas de Raciocínio (BPR) [Critical Review: Battery of Reasoning Tests]. In C. Machado, M. Gonçalves, L. S. Almeida, & M. R. Simões (Orgs.). *Instrumentos e contextos de avaliação psicológica* (pp. 285-311). Coimbra: Almedina.

Almeida, L. S., & Primi, R. (1998). *Bateria de Provas de Raciocínio (BPR-5): manual técnico* [Battery of Reasoning Tests: technical manual]. São Paulo: Casa do Psicólogo.

Arbuckle, J. L. (2007). *Amos 16.0 Users Guide*. Chicago: SPSS Inc.

Byrne, B. M. (2001). *Structural Equation Modeling with AMOS*. New Jersey: Lawrence Erlbaum.

Chiodi, M. G., & Wechsler, S. M. (2009). Inteligência: confronto entre modelos teóricos [Intelligence: confrontation of theoretical models]. *Sobredotação, 9*(1), 133-145.

Conselho Federal de Psicologia (2003). *Caderno especial de resoluções: Resolução CFP_002/2003* [Special compedium of legal decisions: Resolution CFP 002/2003]. Brasília: Conselho Federal de Psicologia.

Conselho Federal de Psicologia (2010). *Avaliação psicológica: Diretrizes na regulamentação da profissão* [Psychological assessment: guidelines for professional regulation]. Brasilia, DF: Conselho Federal de Psicologia.

Flanagan, D. P. (2000). Wechsler-based CHC cross-battery assessment and reading achievement: strengthening the validity of interpretations drawn from Wechsler test scores. *School Psychology Quarterly*, *15*(3), 295-329.

Flanagan, D. P., & Harrison, P. L. (2005). *Contemporary intellectual assessment: theories, tests and issues* (2nd ed.). New York, NY: The Guilford Press.

Flanagan, D. P., McGrew, K. S., & Ortiz, S. O. (2000). *The Wechsler Intelligence scales and Gf-Gc theory*. Boston: Allyn and Bacon.

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

Houaiss, A., & Villar, M. S. (2009). *Dicionário Eletrônico da Língua Portuguesa*. Rio de Janeiro, RJ: Objetiva.

International Testing Comission (2011). *ITC guidelines for quality control in scoring, test analysis, and reporting of test scores.* Retrieved October 23, 2011, from http://www.instestcom.org

Lubinski, D. (2010). Spatial ability and STEM: A sleeping giant for talent identification and development. *Personality and Individual Differences*, *49*, 344-351.

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.

McGrew, K. S., & Flanagan, D. P. (1998). *The Intelligence test desk reference (ITDR). Gf - Gc cross-battery assessment*. Boston: Allyn and Bacon.

McGrew, K. S., & Woodcock, R. W. (2001). *Technical Manual: Woodcock-Johnson III*. Itasca, IL: Riverside Publishing.

Mól, D. A. R., & Wechsler, S. M. (2008). Avaliação de crianças com indicação de dificuldades de aprendizagem pela bateria Woodcock-Johnson III [Assessing children with learning difficuties with the Woodcock-Johnson III battery]. *Psicologia Escolar e Educacional, 12*(2), 391-399.

Muthén, L. K., & Muthén, B.O. (2010). *Mplus User's Guide*. Sixth Edition. Los Angeles, CA: Muthén & Muthén.

Phelps, L., McGrew, K. S., Knopik, S. N., & Ford, L. (2005). The general (g) broad, and narrow CHC stratum characteristics of the WJ-III and Wisc-III tests: a confirmatory cross-battery investigation. *School Psychology Quarterly*, *20*(1), 66-88.

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

Primi, R., & Almeida, L. S. (2000). Estudo de Validação da Bateria de Provas de Raciocínio (BPR-5) [Validity of a Battery of Reasoning Tests]. *Psicologia: Teoria e Pesquisa*, *16*, 165-173.

Primi, R., Couto, G., Almeida, L. S., Guisande, M. A., & Miguel, F. K. (in press). Intelligence, age and schooling: Data from the Battery of Reasoning Tests (BPR-5). *Psicologia: Reflexão e Crítica*.

Roberts, R. D., Zeidner, M., & Matthews, G. (2001). Does emotional intelligence meet traditional standards for an intelligence? Some new data and conclusions. *Emotion*, *1*(3), 196-231.

Schrank, F. A., & Flanagan, D. P. (2003). *WJ-III clinical use and interpretation: Scientist-practitioner perspectives*. San Diego, CA: Elsevier Science.

Schrank, F. A., Flanagan, D. P., Woodcock, R. W., & Mascolo, J. T. (2002). *Essentials of WJ-III cognitive abilities assessment.* New York: John Wiley & Sons.

Schretlen, D., Van Gorp, W.G., Wilkins, S. S., & Bobholz, J.H. (1992). Cross-validation of a psychological test battery to detect faked insanity. *Psychological Assessment*, *4*(1), 77-83.

Schweizer, K. (2010). Some Guidelines Concerning the Modeling of Traits and Abilities in Test Construction. *European Journal of Psychological Assessment*, *26*(1), 1-2.

Wechsler, S. M. (2009). *Avaliação das habilidades cognitivas de adultos: construção e adaptação de bateria de testes psicológicos* [Assessing adults' cognitive abilities: construction and adaptation of a psychological test battery]. Technical report 2008/50252-8. São Paulo, SP: FAPESP.

Wechsler, S. M. (2011). *Compreendendo a inteligência, suas formas e impactos no desenvolvimento* [Understanding intelligence, its forms and impacts on human development]. Symposium conducted at the V Brazilian Conference of Psychological Assessment, Bento Gonçalves, RS, Brazil.

Wechsler, S. M., Nunes, C., Schelini, P. W., Pasian, S. R., Moretti, L., & Anache, A. (2010). Brazilian adaptation of the Woodcock-III tests. *School Psychology International*, *31*(4), 409-421.

Wechsler, S. M., & Schelini, P. W. (2006). Bateria de Habilidades Cognitivas Woodcock-Johnson III: validade de construto [Woodcock-Johnson III battery of cognitive abilities: construct validity]. *Psicologia, Teoria e Pesquisa*, *22*(3), 287-295.

Wechsler, S. M., Vendramini. C. M., & Schelini, P.W. (2007). Adaptação brasileira dos testes verbais da Bateria Woodcock-Johnson III [Brazilian adaptation of the Woodcock-Johnson III verbal tests] *Revista Interamericana de Psicologia*, *41*, 285-294.

Woodcock, R. W. (1990). Theoretical foundations of the WJ-R measures of cognitive ability. *Journal of Psycho-Educational Assessment*, *8*, 231-258.

Woodcock, R. W. (1998). Extending Gf-Gc theory into practice. In J. J. McArdle and R.W. Woodcock (Eds.). *Human cognitive abilities in theory and practice* (pp.137-156). New Jersey: Erlbaum.

Woodcock, R. S., & Johnson, M. B. (1977). *Woodcock-Johnson Psycho-Educational Battery*. Itasca, IL: Riverside.

Woodcock, R. S., & Johnson, M. B. (1989). *Woodcock-Johnson Psycho-Educational Battery- Revised.* Itasca, IL: Riverside.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III.* Itasca, Il: Riverside Publishing.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III tests of cognitive abilities.* Itasca, IL: Riverside Publication.