Psicothema
Psicothema 2013, Vol. 25, No. 1, 115-122
doi: 10.7334/psicothema2011.393
ISSN 0214 - 9915 CODEN PSOTEG

www.psicothema.com

The use of the bi-factor model to test the uni-dimensionality of a battery of reasoning tests

Ricardo Primi[1], Marjorie Cristina Rocha da Silva[1], Priscila Rodrigues[1], Monalisa Muniz[2] and Leandro S. Almeida[3]
[1] University of San Francisco (Brazil), [2] Universidade do Vale do Sapucaí and [3] University of Minho (Portugal)

Abstract

Background: The Battery of Reasoning Tests 5 (BPR-5) aims to assess the reasoning ability of individuals, using sub-tests with different formats and contents that require basic processes of inductive and deductive reasoning for their resolution. The BPR has three sequential forms: BPR-5i (for children from first to fifth grade), BPR-5 – Form A (for children from sixth to eighth grade) and BPR-5 – form B (for high school and undergraduate students). **Method:** The present study analysed 412 questionnaires concerning BPR-5i, 603 questionnaires concerning BPR-5 – Form A and 1748 questionnaires concerning BPR-5 – Form B. The main goal was to test the uni-dimensionality of the battery and its tests in relation to items using the bi-factor model. **Results:** Results suggest that the g factor loadings (extracted by the uni-dimensional model) do not change when the data is adjusted for a more flexible multi-factor model (bi-factor model). **Conclusions:** A general reasoning factor underlying different contents items is supported.

Keywords: battery of reasoning tests, factorial validity, item response theory, bi-factor model.

Resumen

La utilización del modelo bifactorial para testar la unidimensionalidad de una batería de pruebas de raciocinio. **Antecedentes:** la Batería de Pruebas de Raciocinio (BPR-5) tiene como objetivo evaluar la capacidad de razonamiento de las personas utilizando pruebas menores con diferentes ítems y contenidos, pero que presentan relaciones en lo referente a la inducción y la deducción que intervienen en su resolución de la tarea. La BPR tiene una organización secuencial: BPR-5i (para niños de 1º a 5º grado), BPR-5 versión A (del 6º al 8º grado) y BPR-5 versión B (Enseñanza Secundaria y Terciaria). **Método:** el presente estudio evaluó los datos de 289 protocolos de la BPR-5i, 603 de la BPR-5 versión A y 1.748 de la BPR-5 versión B. El objetivo principal fue poner a prueba la unidimensionalidad de la batería y de las pruebas que la componen. **Resultados:** los resultados sugieren que las cargas factoriales g (extraído por el modelo uni-dimensional) no cambian cuando los datos se ajustan según un modelo multidimensional (bi-factorial). **Conclusiones:** un factor de razonamiento general que subyace a los ítems de diferentes contenidos ha sido confirmado.

Palabras clave: batería de pruebas de raciocinio, validez factorial, desarrollo cognitivo, teoría de respuesta al ítem, calibración de la prueba.

Human intelligence is a hot topic in psychological research. Historically many theoretical models have been introduced and coexisted in an attempt to define this construct. The debate within the psychometric tradition has been whether intelligence has a simple or multi-faceted structure. The models oscillate between two extremes with one side conceiving intelligence as a single cognitive ability (g factor) and the other side viewing it as a group of cognitive factors with different levels of independence. This issue has been surrounded by some controversy and disagreement but in recent decades, some reconciliation that integrates both positions into a hierarchical model with different levels of specificity has been developed (McGrew, 2009). Therefore, it is possible to focus on the specific cognitive abilities attached to specific tasks or—at a broader level—on the general reasoning processes of problem solving.

This conciliatory position was derived from the models of Cattell (1963), Horn (1991), Horn and Cattell (1966) and Carroll (1993, 1997) and is called CHC theory with reference to the contribution of these authors (Cattell-Horn-Carroll). The CHC theory proposes that a first stratum is composed of several narrow factors associated with different tests, which are correlated and organised into at least ten broad ability factors (fluid reasoning, *Gf*; comprehension-knowledge, *Gc*; short-term memory, *Gsm*; visual processing, *Gv*; auditory processing, *Ga*; long-term storage and retrieval, *Glr*; processing speed, *Gs*; reaction and decision speed, *Gt*; reading and writing, *Grw*; and quantitative knowledge, *Gq*). These broad factors are associated with the cognitive functions involved (perception, memory, reasoning, creativity, etc.) or the content of the processed information (auditory-verbal, numerical-quantitative, spatial-figurative, etc.). Then, the correlation between these broad domain factors leads to a third stratum represented by a more general factor, which is interpreted as the *g factor* of Spearman or the fluid intelligence (*Gf*) of Cattell, which is assumed to be very similar to *g* (Carroll, 1993, 1997; McGrew, 2009; Woodcock, 1990).

In Brazil and Portugal, Primi and Almeida (2000a, 2000b) have developed the Battery of Reasoning Test 5 (BPR-5), which

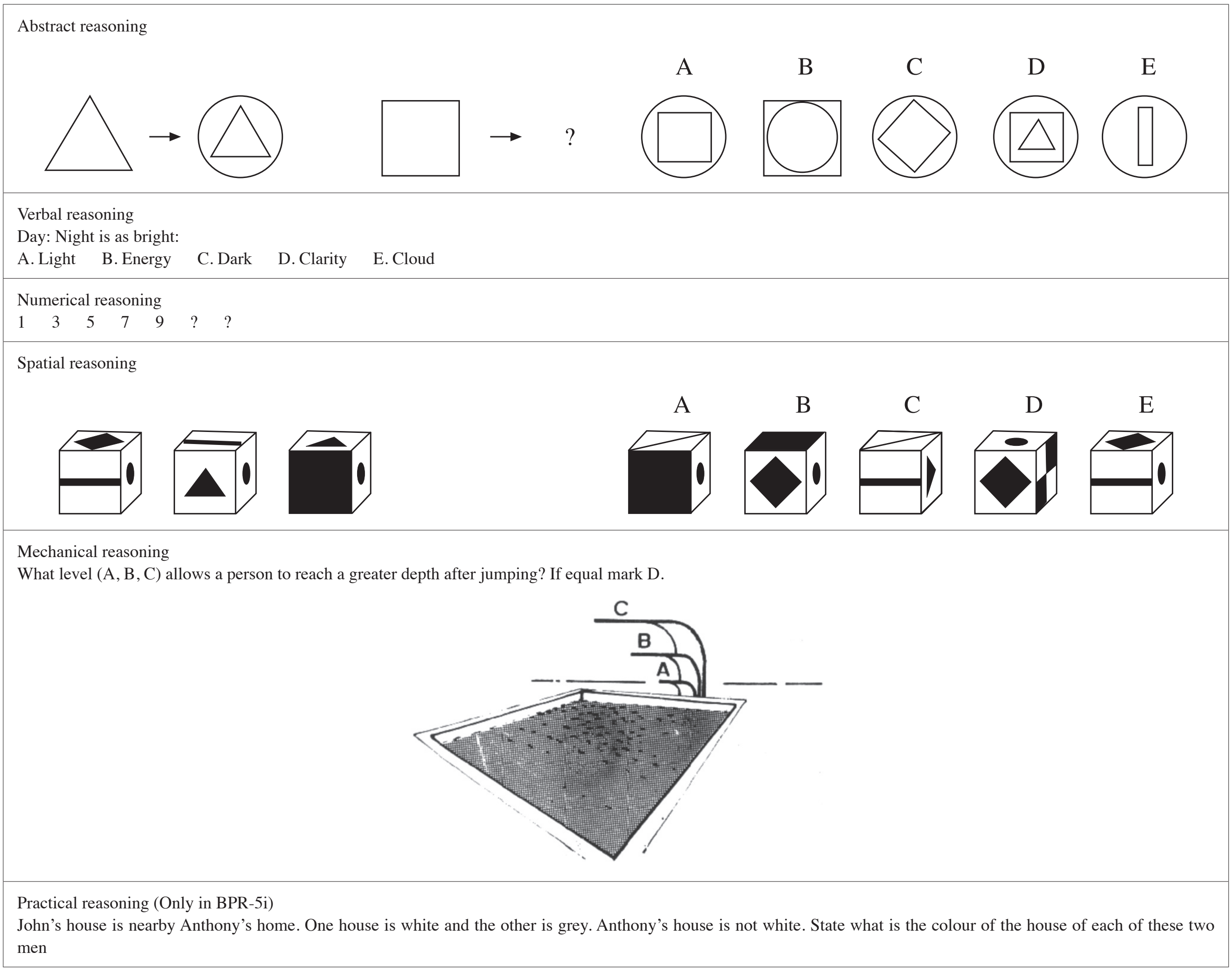
---

Received: November 25, 2011 • Accepted: May 14, 2012
Corresponding author: Ricardo Primi
Laboratório de Avaliação Psicológica e Educacional
University of San Francisco (Brazil)
CEP 13251-900 Itatiba São Paulo
e-mail: rprimi@mac.com

is composed of sub-tests sampling different contents (verbal, numerical, spatial, abstract and mechanical) and tasks (series, analogies, practical figure-text problems and syllogistic reasoning). The BPR-5 is composed of inductive and deductive reasoning tasks that require a person to analyse a relationship among a set of stimuli in order to find their organising rules and then apply it in a similar situation to find the right answer to a problem. The most typical formats for accessing inductive reasoning are analogies and series. Item models that compose the six tasks present in BPR sub-tests are exemplified in Figure 1.

The authors have used the CHC model, as well as cognitive psychology (Primi, 2002; Sternberg, 1977), to interpret what BPR-5 is measuring. They argue that the test assesses fluid intelligence, which is strongly associated with *g factor*, as well as some specific factors at the sub-test content level. The first sub-test—abstract reasoning (AR)—is supposed to be associated with fluid intelligence, which is defined as the capacity of reasoning in new situations in order to create concepts and understand implications. Verbal reasoning (VR) is supposed to be associated with *Gf* and also with comprehension-knowledge (*Gc*), which is defined as the extension and depth of verbal and vocabulary knowledge. Numerical reasoning (NR) is supposed to measure *Gf* but also quantitative reasoning (*RQ*) and (*Gq*) knowledge that can be defined as the understanding of the basic quantitative concepts, such as addition, subtraction, multiplication and division, as well as implying manipulation of numerical symbols. Spatial reasoning (SR) is supposed to measure visual processing (*Gv*), which is the ability to represent and manipulate mental images, but it is also associated with *Gf*. Mechanical reasoning (MR)— least homogeneous sub-test—is supposed to be related to *Gf*, to mechanical knowledge (*MK* within the domain-specific knowledge broad factor *Gkn*), to visual processing (*Gv*) and to reading comprehension (*RC*), because problems are presented in terms of visual schemes and explanatory text. Practical reasoning is only present in BPR-5i and accesses *Gf* deductive reasoning (RG), as well as comprehension-knowledge (*Gc*) and reading comprehension (*RC*), as the problems are presented verbally.

Evidence based on factor analysis of the internal structure in the level of sub-tests supports the existence of a general factor explaining the common variance between the scales (Primi & Almeida, 2000b). Moreover, correlation coefficients with other tests measuring similar or related constructs (Cruz, 2008; Santos et al., 2000) and criterion measures, such as school achievement, age, and job performance (Almeida, Lemos, & Primi, 2011; Baumgartl



Verbal reasoning
Day: Night is as bright:
A. Light B. Energy C. Dark D. Clarity E. Cloud

Numerical reasoning
1 3 5 7 9 ? ?

Mechanical reasoning
What level (A, B, C) allows a person to reach a greater depth after jumping? If equal mark D.

Practical reasoning (Only in BPR-5i)
John's house is nearby Anthony's home. One house is white and the other is grey. Anthony's house is not white. State what is the colour of the house of each of these two men

Figure 1. *Item examples of BPR sub-tests*

& Primi, 2006; Almeida, Guisande, Primi, & Lemos, 2008; Lemos, Almeida, Guisande, & Primi, 2008; Primi et al., 2012; Primi, Ferrão, & Almeida, 2010), support the interpretation of general and specific factors associated with each sub-test.

The BPR has three test forms: BPR-5i for children from second to sixth (ages typically from 7 to 11), BPR-5 A for students from seventh to ninth grade (ages typically from 12 to 14) and BPR-5 B for high school and undergraduate students (ages from 15 to adulthood). These forms were planned to have common items, so that their scores could be vertically equated using the Item Response Theory (IRT, Hambleton, Swaminathan, & Rogers, 1991; Muñiz, 1997; Pasquali & Primi, 2003; Prieto & Delgado, 2003). Test equating makes it possible to obtain common metric and comparable scores for subjects that have been tested with two tests that measure the same construct but are composed of different items (Hambleton & Swaminathan, 1986). In order to apply IRT models, two assumptions should be observed: unidimensionality and local independence. Unidimensionality requires that the response to test items depends mainly on a dominant factor, that is, items should be measuring the same construct, which should explain most of their common variance (Hambleton & Swaminathan, 1986; Hambleton, Swaminathan, & Rogers, 1991). Local independence is another way of expressing the same principle of unidimensionality. It requires that the correlation of item responses only depends on the subject's ability. When this information is statistically removed, item responses remain uncorrelated. When the unidimensionality condition is assumed, one can estimate the subject's capacity apart from the test form used (Muñiz, 1997). Therefore, the first step regarding the equating of BPR test forms is to verify if the criteria of unidimensionality is satisfied.

As was argued, BPR sub-tests measure a general factor of reasoning but also specific factors associated with the different contents present in the forms (verbal, abstract, numerical, spatial and practical). With regard to the assumption of unidimensionality, a general factor has been consistently found in the level of sub-scales. Moreover, there is also evidence for the specific factors found in studies that investigate the correlations of sub-tests with external variables (Almeida et al., 2008, 2011; Primi et al., 2012; Primi, Nakano, & Wechsler, 2012). Therefore, an important question that is still open to inquiry concerns to what extent there is a dominant factor in the level of item responses. It is clear that, even if this condition is not satisfied, each sub-test can be analysed separately. However, if there is a main factor (which is believed to be inductive reasoning) underlying most items of different sub-tests and this factor is dominant over the specific ones, then, it would be possible to equate the three forms in order to estimate a general *Gf* score across sub-tests, even using different items. This way, the main purpose of this paper is to test the unidimensionality assumption of the BPR at the level of item responses.

Taking into consideration the recommendations of the literature about methods for testing unidimensionality (Hattie, 1985; Vitoria, Almeida, & Primi, 2006), two important peculiarities of the present data should be considered. First, one is conducting factor analysis on the level of item responses instead of test scores. The decision criteria for the number of factors to be extracted, when considering tests scores, is usually not directly applicable to item-factor analysis (Reise, Waller, & Comrey, 2000). The second point refers to the latent hierarchical structure that is supposed to exist, that is, the items are organised in sub-scales that also measure a general factor. As recommended and exemplified by Reise and Haviland (2005) and Reise, Morizot and Hayes (2007), these situations are adequately handled by bi-factor analysis. Bi-factor analysis fits a model with a general factor and, simultaneously, with group factors that capture specific remaining common variance across items uncorrelated with the general factor. In this sense, it accommodates the hierarchical structure that may exist in the data and allows the estimation of the relative magnitudes of general versus specific factors. Reise, Morizot and Hayes (2007) recommend that a traditional unidimensional model (one-factor solution) is used and then, a bi-factor model before item-factor loadings on a general factor of these two solutions are compared. If the results are similar, it is supposed that the specific factors that may exist do not distort the general factor interpretations, thus allowing the application of unidimensional IRT models to the data.

In synthesis, the purpose of this paper is to use bi-factor and full information factor analysis to test the fit of items of a unidimensional model on BPR. In doing so, this paper is an attempt to contribute to the establishment of a foundation for the application of procedures to equate BPR test forms that serve practical purposes. This study also provides information about the BPR construct validity, as it tests the assumption of an underlying general factor across forms implicit to most of their items, which is believed to be *Gf*. This paper also illustrates the application of bi-factor analysis to test the unidimensionality of the items of tests. Although highly recommended, it has not been so frequently applied.

Method

Participants

The enquiry used a database with a total of 2,763 students (1,748 Form B, 603 Form A and 412 Form i). These students were part of a convenience sample, but were selected to represent three different states of Brazil (São Paulo, Minas Gerais and Rio Grande do Sul), approximately equal groups, considering gender and schools from public and private educational systems. In Brazil, there is a very large socio-economic gap between the populations of the two educational establishments, thus this control helps to sample students with diverse social, economic and cultural environments. In Form i dataset, 50.9% were male and the age ranged from 7 to 14 (*M*= 9.8, *SD*= 1.5). In Form A, 51.1% were female and age ranged from 11 to 19 (*M*= 13.0, *SD*= 1.27). In Form B, 51.9% were male and age ranged from 14 to 63 (*M*= 212, *SD*= 7.7). A major part of the students were attending schools from the public system (66.9%). More details of these databases can be found in Primi and Almeida (2000a) and Cruz (2008).

Instruments

BPR-5 (Primi & Almeida, 2000a) Forms A and B consist of five sub-tests (see Figure 1): Abstract Reasoning-AR is composed of 25 items involving abstract analogies with geometrical figures (time limit: 12 minutes); Verbal Reasoning-VR is composed of 25 items involving words analogies (time limit: 10 minutes); Spatial Reasoning-SR is composed of 20 items based on a series of three dimensional cubes in motion (time limit: 18 minutes); Numerical Reasoning-NR consists of 20 items involving a series of numbers requiring completion (time limit: 18 minutes); and Mechanical Reasoning-MR is composed of 25 items regarding practical problems about physical and mechanical content (time limit: 15 minutes). Form A is designed for students from seventh to nine grade (ages typically from 12 to 14) and Form B for students from for high school and undergraduate students (ages from 15 to adulthood). BPR-5i (Almeida, Primi, & Cruz, 2007) is composed of four sub-tests. Three sub-tests have the same format and common items with BPR-5 Form A: namely, AR contains 20 items with abstract analogies; VR contains 32 items with word analogies; NR contains 30 items using number series. The fourth test Practical Reasoning-PR contains 15 verbal terms measuring syllogistic reasoning using everyday situations common to children. BPR-5i is designed for children from second to sixth (ages typically from 7 to 11). Common items were included in AR (19), VR (18), MR (19), SR (12) and NR (12) - Forms A and B - and in AR (9), VR (7) and NR (10) - Forms A, B and i - in order to equate scores across forms using IRT. In general, there is an average of 54% of common items across forms. Several studies have been published documenting evidence of validity, reliability and normative expectations, which show positive evidence of reliability and validity of its sub-tests (Almeida & Primi, 2004; Almeida et al., 2008, 2010; Cobêro, Primi, & Muniz, 2007; Lemos et al., 2008; Primi & Almeida, 2000b; Primi et al., 2012).

The BPR has three test forms: BPR-5i for children from second to sixth (ages typically from 7 to 11), BPR-5 A for students from seventh to ninth grade (ages typically from 12 to 14) and BPR-5 B for high school and undergraduate students (ages from 15 to adulthood).

Procedures

The data for the present paper came from a database concerning validity and normative studies of the BPR-5. Students answered the tests in their classrooms in two sessions with an interval of 15 minutes between sessions. All students or their parents and their school principal, as was the case, signed an informed consent for participation in the study.

Data analysis

The data analysis strategy followed the recommendation of Reise and Haviland (2005) and Reise, Morizot and Hayes (2007) and it was done using the TESTFACT programme (Wilson, Wood, & Gibbons, 1991), which performs a full factor information analysis (Bock, Gibbons, & Muraki, 1986) and a bi-factor analysis using the algorithms of the Item Response Theory. In addition, a confirmatory factor analysis was run on MPLUS using a weighted least squares estimator (WLSMV) with categorical variables (Muthén & Muthén, 2010).

The analysis was performed in two steps: first, a full information factor analysis modelling a unidimensional solution was carried out; then, a bi-factor analysis was performed considering items from each sub-test forming a group factor. Therefore, the bi-factor model accommodates the general factor simultaneously with the specific common variance that eventually is left over, due to specific content similarities among items of the same sub-test (verbal, numeric, visual, etc.). Afterwards, the g-factor loadings of the two analyses were compared to see if they changed substantially. Small and negligible differences were expected to be seen that would mean there was a small role of specific factors. Such a result would mean that specific factors do not distort the meaning of the general fluid reasoning factor that is measured generally by items of all sub-tests.

Model fit was based on three indices: Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), which calculates the relative adjustment of the model by comparing it with the null model, and the Root-Mean-Square Error of Approximation (RMSEA), which is a measure of discrepancy between modelled and observed covariate matrix that penalizes model complexity. Values for CFI and TLI above .95 and RMSEA less than .05 indicate good fit (Byrne, 2001; Schweizer, 2010). An adjusted chi-square difference between the two models was also computed using the WLSMV estimators to test whether the bi-factor model fit data better than the unidimensional model (Muthén & Muthén, 2010).

Results

Table 1 shows the result of the full information factor analysis and a bi-factor analysis that was performed three times, one for each test form (BPR-5i, Forms A and B). It shows the proportion of explained variances for *g factor* in both models and for group factors in the bi-factor model, chi-square, reliability estimates for the general factors and eigenvalues for the first and second factors and their ratio. The results of the bi-factor model indicate that the general factor is always relatively more important than group factors with explained variances ranging between 30.72 and 35%. Group factors explained variance ranges from 1.0% to 8.61%. For the unidimensional model, the percentages due to

Table 1
Summary results of bi-factor and full information factor models

	BPR-5i	**BPR-5 A**	**BPR-5 B**
Bi-factor model			
g bi (%)	35.00	34.43	30.72
s AR (%)	8.61	2.35	2.39
s VR (%)	6.57	2.24	2.77
s MR (%)	–	3.04	4.28
s SR (%)	–	2.88	2.58
s NR (%)	6.29	3.93	4.41
s PR (%)	3.25	–	–
Uniqueness	40.25	51.13	52.83
Reliability (of g-factor)	.92	.90	.90
χ^2	27040.8 /105	63719.7 / 256	177566.0/1443
CFI	.981	.974	.951
TLI	.980	.973	.949
RMSEA	.017	.014	.017
Unidimensional model (*Full information*)			
Eig 1 / Eig 2	35.9/6.4 (5.6)	37.7/5.15 (7.32)	39.37/6.41 (6.14)
g full	44.08	37.42	40.42
Reliability (of g-factor)	.95	.95	.95
χ^2	34891.1 / 207	65848.91 /371	182457.3 /1558
CFI	.904	.924	.876
TLI	.902	.922	.873
RMSEA	.038	.024	.027
Corrected Chi-Square difference test for the weighted least squares estimator (WLSMV)			
$\Delta\chi^2$	1421.33	1540.46	3905.26
df	102	115	113
p	<.0001	<.0001	<.0001

g factor are slightly larger, as expected, due to the variance of part of the group factors variance, which, in this case, represents violations of local independence, which inflate the *g factor* variance, so that the non-modelled group factors will appear as part of the *g factor* variance. Nonetheless, the fact that the *g factor* percentages in the unidimensional model and bi-factor models do not differ substantially indicates that the unidimensional model is also quite adequate. The chi-squares difference test and model fit indexes obtained from confirmatory factor analysis in MPLUS indicates that bi-factor models fits significantly better as compared to unidimensional models. This result was expected as it better accommodates the multidimensionality on the data.

The results of the bi-factor analysis also suggest which sub-test has a more prominent presence of a specific factor. Numerical and mechanical reasoning sub-tests on Forms A and B are the ones that concentrate more specific variance. In Form I, abstract reasoning was the subtest with more specific variance. Another fact that comes to attention is that the proportion of general factor variance tends to decrease, as can be seen if these figures between test forms are compared, although Form B for older students has a relatively small magnitude.

Tables 2, 3 and 4 provide more details about item factor loading on the bi-factor model and on the unidimensional models. They present the mean, minimum, maximum and standard deviation of item factor loadings for BPR-5i, Forms A and B, respectively. The second to fifth columns present the results for the bi-factor model with summaries for *g-factor* loadings (Bi *g*) and specific group factors loadings (Bi s). The tables also show the proportion of items with loadings higher than .29. The last two columns present the same statistics for the unidimensional model.

In general, the *g-factor* loadings for items are appropriate for all forms. The mean loadings by sub-test ranged from .41 to .81. In the worst-case scenario, 72% of the items had a *g-factor* loading equal or greater than .30 (Form A, Mechanical Reasoning sub-test). Looking at the specific group factor loadings, it is possible to see that the ones that are more robust are the numerical, spatial and mechanical (and particularly the latter in Form B). But it is also interesting to note that the spatial and numerical items also have suitable *g-factor* loadings. This result is in accordance with the expectations that the measurement of g is not distorted by the presence of group factors.

Table 2
Summary of item g-factor loading on unidimensional and bi-factor model on BPR-5i

	Bi g	% >.29	Bi s	% >.29	Full g	% >.29
AR						
Mean	.730	100	.298	67	.647	97
Min	.359		.129		.181	
Max	.933		.456		.869	
SD	.148		.085		.163	
VR						
Mean	.645	93	-.019	13	.447	80
Min	.015		-.374		.034	
Max	.949		.362		.791	
SD	.224		.189		.184	
NR						
Mean	.816	100	.387	81	.806	100
Min	.599		.256		.520	
Max	.945		.513		.941	
SD	.085		.084		.091	
PR						
Mean	.650	100	.297	53	.508	100
Min	.493		-.007		.338	
Max	.854		.481		.830	
SD	.103		.132		.133	

Finally, factor loadings of both models were compared. The correlations of *g-factor* loadings between the two models (bi-factor versus unidimensional) were .89 for the form for the youngest children, .95 for Forms A and B ($p<.001$). Figure 2 shows the scatter plots comparing these loadings for the three forms separately. It can be noted that the loadings are almost the same. Therefore and according to what was expected, the use of a unidimensional model is plausible, because the presence of group factors does not distort general factor loadings. The general factor loadings or IRT discrimination parameters are relatively the same regardless of the fact of group factors being modelled or not.

Discussion

This study attempted to verify the assumption of a general factor underlying all sub-tests of the BPR-5i, Forms A and B, by the means of item factor analysis. As pointed out, BPR sub-tests

Table 3
Summary of item g-factor loading on unidimensional and bi-factor model on BPR-5 Form A

	Bi g	% > .29	Bi s	% >.29	Full g	% >.29
AR						
Mean	.613	100	.283	48	.635	100
Min	.451		-.047		.420	
Max	.757		.672		.786	
SD	.086		.170		.106	
VR						
Mean	.558	100	.297	68	.540	96
Min	.297		.070		.199	
Max	.793		.568		.778	
SD	.139		.124		.148	
MR						
Mean	.408	72	.346	68	.463	76
Min	-.135		.091		-.051	
Max	.874		.610		.992	
SD	.246		.145		.260	
SR						
Mean	.536	100	.399	85	.570	100
Min	.371		.233		.376	
Max	.661		.522		.689	
SD	.088		.080		.082	
NR						
Mean	.703	95	.440	90	.726	95
Min	.268		-.005		.073	
Max	.868		.821		.861	
SD	.142		.185		.193	

Table 4
Summary of item g-factor loading on unidimensional and bi-factor model on BPR-5 Form B

	Bi g	**% > .29**	**Bi s**	**% >.29**	**Full g**	**% >.29**
AR						
Mean	.536	92	.243	48	.583	100
Min	.264		-.218		.292	
Max	.692		.630		.740	
SD	.108		.230		.114	
VR						
Mean	.442	92	.346	68	.487	96
Min	.160		.198		.208	
Max	.600		.462		.628	
SD	.104		.090		.101	
MR						
Mean	.536	96	.430	88	.639	100
Min	.289		.206		.402	
Max	.740		.595		.814	
SD	.099		.112		.096	
SR						
Mean	.644	100	.377	95	.751	100
Min	.498		.151		.609	
Max	.823		.518		.914	
SD	.102		.082		.095	
NR						
Mean	.565	95	.464	75	.683	100
Min	.208		.113		.354	
Max	.759		.684		.806	
SD	.152		.201		.128	

were supposed to measure a general factor *Gf (fluid intelligence)* but also specific factors *Gc (comprehension knowledge)*, *Gv (visual processing)*, *Gq (quantitative knowledge)*, *MK (mechanical knowledge)* and *RC (reading comprehension)* associated with the different contents present in the forms (verbal, abstract, numerical, spatial and practical). Empirical evidence for this assumption comes from studies of scale level factor analysis and correlation with external variables (Almeida et al., 2010, 2008; Lemos et al., 2008; Primi & Almeida, 2000a, 2000b) The present study adds further evidence for this assumption of internal structure of BPR by employing recent methods of item factor analysis. It tests whether a general dimension could be assumed underlying all items beyond specific factors. Based on previous research, this general dimension is interpreted as Gf factor and the secondary factors are associated with specific contents of each sub-test.

The test of unidimensionality at item level also has an important implication for future studies using IRT to equate test forms of the BPR. One of the key conditions for applying IRT models concerns the relevance of unidimensionality; in other words, that responses are determined by a major dimension and that secondary dimensions are negligible (Hambleton & Swaminathan, 1986; Hambleton et al., 1991; Muñiz, 1997). This study contributes to this topic, showing that tests of cognitive abilities measuring multiple factors can also be treated as unidimensional. This is probably related to the hierarchical organization of the intelligence constructs (Carroll, 1993, 1997; McGrew, 2009). Therefore the answers to the question of unidimensionality are not a simple yes or no because these tests have the influence of a general dimension as well as group factors. Bi-factor model is a good way to represent in the model this situation and to examine parameters associated with both components such as the proportion of variance distributed to the general and group factors.

In order to test unidimensionality, a comparison of bi-factor, which models *g* and group factors with the traditional unidimensional model, was conducted, as suggested by Reise and Haviland (2005) and Reise et al (2007). In general, results point to the dominance of a general factor. When a bi-factor model is used, it allows items to correlate within the group factors; the general factor loadings and the explained variance do not change as compared with the traditional unidimensional model that only accommodates common variance due to a general factor. This indicates that the general factor is robust, even in a more permissive situation, which tries to model the existence of group factors. If the specific factors were relatively more important, then it would have a diminished *g factor* variance and changed general factor loadings when using the bi-factor model. Results show that this does not seem to be the case. Also the reliability (internal consistency) of *g factor* scores is very similar for both models, as all are equal or greater than 0.90. Moreover, the presence of a dominant factor is shown in the ratio between the first and second eigenvalues, which was always equal to or greater than 5, thus reaching a traditional benchmark of unidimensionality (Reise et al., 2007).

These results corroborate the validity studies conducted by Primi and Almeida (2000b), Baumgartl and Primi (2006) and Cruz (2007), although these did not carry out the factor analysis of items. These results suggest that *g factor* loadings (extracted by the unidimensional model) do not change when the data is adjusted for a more flexible multi-factor model, which allows the formation of group factors, so such analysis can be taken as a validity coefficient of internal structure considering the item correlation with *g factor*. According to Reise and Haviland (2005) and Reise, Morizot and Hayes (2007), this is very strong evidence for the condition of unidimensionality for IRT.

This study also provides new and important information about which sub-tests have more specificity: numerical and mechanical reasoning. It is worth mentioning that the Numerical Reasoning-NR test is the only sub-test that does not contain multiple choice format items, that is, here, students must provide their own responses. Also NR test items that are relatively more complex combine two numerical sequences, which could require visualisation to identify both sequences. These factors could be considered as elements that contributed to the formation of a robust group factor. Mechanical reasoning requires visualisation factors, which could be an explanation of the relatively robust specific factor. These results are in accordance with previous studies showing the practical importance of these specific factors (Almeida et al., 2008; Primi et al., 2012) and they also suggest that a robust general dimension is plausible for all the sub-tests and of BPR. Finally, this study illustrates how to use modern methods of IRT to test the assumption of unidimensionality and reveal the relative importance of general and specific factors in batteries of tests that intend to measure several hierarchically related dimensions.

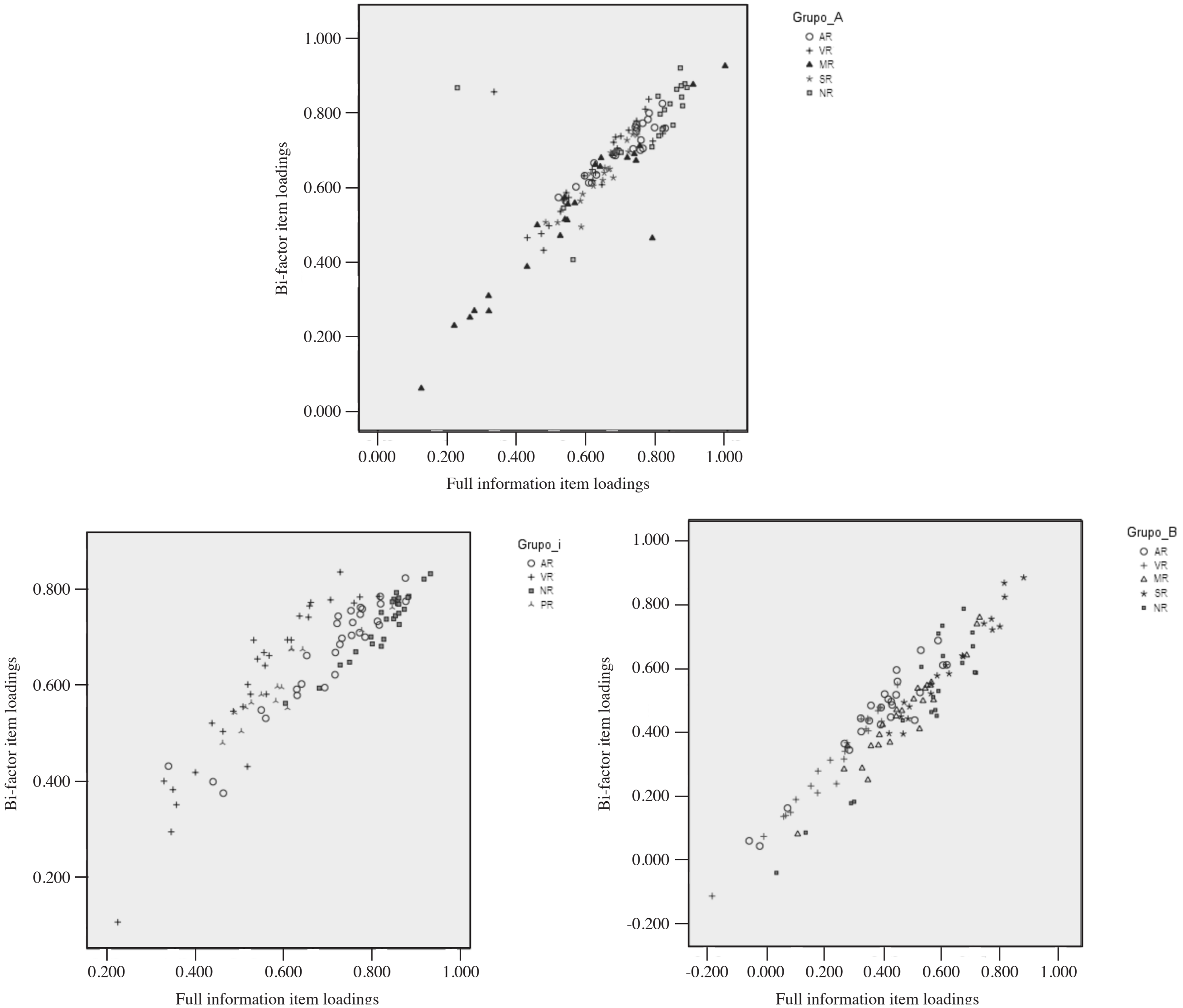


Figure 2. *Scatter plots of factor loadings in g factor with the bi-dimensional model for the youngest children (top), A (lower left) and B Forms (lower right)*

References

Almeida, L.S., & Primi, R. (2004). Perfis de capacidades cognitivas na Bateria de Provas de Raciocínio (BPR-5). *Psicologia Escolar e Educacional, 8*, 135-144.

Almeida, L.S., Primi, R., & Cruz, M.B.Z. (2007). *Bateria de Provas de Raciocínio Infantil BPR-5i*. Itatiba, SP: Universidade São Francisco, Laboratório de Avaliação Psicológica e Educacional.

Almeida, L.S., Guisande, M.A., Primi, R., & Lemos, G. (2008). Contribuciones del factor general y de los factores específicos en la relación entre inteligencia y rendimiento escolar. *European Journal of Education and Psychology*, *1*, 5-16.

Almeida, L.S., Lemos, G., & Primi, R. (2011). Recensão crítica: Bateria de Provas de Raciocínio (BPR). In Carla Machado, Miguel Gonçalves, Leandro S. Almeida and Mário R. Simões (Eds.), *Instrumentos e contextos de avaliação psicológica, Vol. 1* (pp. 285-311). Coimbra: Edições Almedina.

Baumgartl, V.O., & Primi, R. (2006). Evidências de validade da Bateria de Provas de Raciocínio (BPR-5) para seleção de pessoal. *Psicologia: Reflexão e Crítica*, *19*, 246-251.

Bock, D., Gibbons, R., & Muraki, E. (1986). Full-information item factor analysis. *Applied Psychological Measurement*, *12*, 261-280.

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

Carroll, J.B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press.

Carroll, J.B. (1997). The three-stratum theory of cognitive abilities. In D.P. Flanagan, J.L. Genshaft and P.L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (pp. 122-130). New York: Guilford Press.

Cattell, R.B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, *54*, 1-22.

Cobêro, C., Primi, R., & Muniz, M. (2007). Inteligência emocional e desempenho no trabalho: Um estudo com MSCEIT, BPR-5 e 16 PF. *Paidéia, 16*, 337-348.

Cruz, M.B.Z. (2008). Estudo de validade e precisão da Bateria de Provas de Raciocínio Infantil - BPR-5i. *Dissertação de Mestrado*. Itatiba, SP: Universidade São Francisco.

Hambleton, R.K., & Swaminathan, H. (1986). *Item Response Theory: Principles and applications*. Norwell, MA: Kluwer Academic Publishers.

Hambleton, R.K., Swaminathan, H., & Rogers, H.J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: SAGE Publications.

Hattie, J. (1985). Methodology review: Assessing unidimensionality of tests and items. *Applied and Psychological Measurement, 9*, 139-164.

Horn, J.L. (1991). Measurement of intellectual capabilities: A review of theory. In K.S. McGrew, J.K. Werder and R.W. Woodcock (Eds.), *Woodcock-Johnson-R Technical Manual* (pp. 197-245). Chicago: Riverside.

Horn, J.L., & Cattell, R.B. (1966). Refinement and test of the theory of fluid and crystalized general intelligences. *Journal of Educational Psychology*, *57*, 253-270.

Lemos, G., Almeida, L.S., Guisande, M.A., & Primi, R. (2008). Inteligência e rendimento escolar: Análise da sua associação ao longo da escolaridade. *Revista Portuguesa de Educação*, *21*, 83-99.

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.

Muñiz, J. (1997). *Introducción a la teoría de respuesta a los ítems*. Madrid: Pirámide.

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

Pasquali, L., & Primi, R. (2003). Fundamentos da teoria da resposta ao item. *Avaliação Psicológica, 2*, 99-110.

Prieto, G., & Delgado, A.R. (2003). Análisis de un test mediante el modelo de Rasch. *Psicothema, 15,* 94-100.

Primi, R. (2002). Inteligência fluida: Definição fatorial, cognitiva e neuropsicológica. *Cadernos de Psicologia e Educação - Paidéia, 12*, 57-77.

Primi, R., & Almeida, L.S. (2000a). *Baterias de Provas de Raciocínio (BPR-5): Manual técnico*. São Paulo: Casa do Psicólogo.

Primi, R., & Almeida, L.S. (2000b). Estudo de Validação da Bateria de Provas de Raciocínio (BPR-5). *Psicologia: Teoria e Pesquisa, 16*, 165-173.

Primi, R., Couto, G., Almeida, L.S., Guisande, M.A., & Miguel, F.K. (2012). Intelligence, age and schooling: Data from the Battery of Reasoning Tests (BRT-5). *Psicologia: Reflexão e Crítica*, *25,* 79-88.

Primi, R., Ferrão, M.E., & Almeida, L.S. (2010). Fluid intelligence as a predictor of learning: A longitudinal multilevel approach applied to math. *Learning and Individual Differences, 2*, 446-451.

Primi, R., Nakano, T.C., & Wechsler, S.M. (2012). Cross-battery factor analysis of the Battery of Reasoning Abilities (BPR-5) and Woodcock-Johnson Tests of Cognitive Ability (WJ-III). *Temas em Psicologia (Ribeirão Preto), 20,* 121-132.

Reise, S.P., & Haviland, M.G. (2005). Item response theory and the measurement of clinical change. *Journal of Personality Assessment*, *84*, 228-238.

Reise, S.P., Morizot, J., & Hayes, R.D. (2007). The role of bifactor model in resolving dimensionality issues in health outcomes measures. *Quality Life Research*, *16*, 19-31.

Reise, S.P., Waller, N.G., & Comrey, A.L. (2000). Factor analysis and scale revision. *Psychological Assessment*, *12*, 287-297.

Santos, A.A.A., Primi, R., Vendramini, C.M.M., Taxa, F.O.S., Lukjanenko, M. F. S. P., Muller, F., Sampaio I., Andraus Jr, S., Kuse, F. K., & Bueno, C.H. (2000). Habilidades básica de ingressantes universitários. *Revista Avaliação*, *16*, 33-45.

Schweizer, K. (2010). Some guidelines concerning the modeling of traits and abilities in test construction. *European Journal of Psychological Assessment*, *26*(1), 1-2.

Sternberg, R.J. (1977). A component process in analogical reasoning. *Psychological Review*, *84*, 353-378.

Vitoria, F., Almeida, L.S., & Primi, R. (2006). Unidimensionalidade em testes psicológicos: Conceito, estratégias e dificuldades na sua avaliação. *Psic*, *7*, 1-7.

Wilson, D.T., Wood, R., & Gibbons, R. (1991). *TESTFACT: Test scoring, item statistics, and item factor analysis*. Chicago, IL: Scientific Software International.

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