

## Gender differences on tests of crystallized intelligence

Solange Muglia Wechsler<sup>1</sup>, Tatiana de Cassia Nakano<sup>1</sup>, Simone Ferreira da Silva Domingues<sup>2</sup>, Helena Rinaldi Rosa<sup>3</sup>, Roselaine Berenice Ferreira da Silva<sup>4</sup>, José Humberto da Silva-Filho<sup>5</sup> and Carla Alexandra da Silva Moita Minervino<sup>6</sup>  
<sup>1</sup>Pontifical Catholic University of Campinas; <sup>2</sup>Cruzeiro do Sul University; <sup>3</sup>University of Sao Paulo; <sup>4</sup>University of Santa Cruz do Sul; <sup>5</sup>Federal University of Amazonas; <sup>6</sup>Federal University of Paraiba (Brazil)

This study aimed to determine whether performance on tests of crystallized intelligence is affected by gender and to ascertain whether differential item parameters could account for the gender disparities. The sample comprised 1.191 individuals (55% women) between the ages of 16 and 77 years old ( $M=22$ ;  $SD=9.5$ ). The participants were primarily college students (58.3%) living in four Brazilian states. Four verbal tests measuring crystallized intelligence (vocabulary, synonyms, antonyms and verbal analogies) were constructed and administered in a group setting. An analysis of variance revealed no significant differences in the overall performance between men and women. However, a differential item functioning analysis indicated significant differences on 8.7% of the items, which indicates the existence of gender bias. Because bias can limit women's access to social opportunities, the results obtained indicate the importance of reducing item bias in cognitive measures to ensure the accuracy of test results.

*Keywords:* Intelligence, tests, cognition, gender, culture.

*Diferencias de género en pruebas de inteligencia cristalizada.* Este estudio tuvo como objetivo determinar si el desempeño en las pruebas de inteligencia cristalizada es afectado por el género y para determinar si los parámetros diferenciales de los ítems podrían explicar las disparidades de género en los resultados. La muestra fue de 1.191 personas (55% mujeres) de edades comprendidas entre 16 y 77 años ( $M=22$ ,  $SD=9.5$ ). Los participantes eran principalmente estudiantes universitarios (58.3%) de cuatro estados brasileños. Cuatro pruebas verbales para medir la inteligencia cristalizada (vocabulario, sinónimos, antónimos y analogías verbales) fueron construidas y administradas en un ambiente de grupo. El análisis de varianza no reveló diferencias significativas en el rendimiento general entre hombres y mujeres. Sin embargo, el análisis de funcionamiento diferencial de los ítems indicó diferencias significativas de 8.7% de los ítems, lo que significa la existencia de sesgo para género. Debido al hecho que la existencia de sesgo puede limitar el acceso de las mujeres a las oportunidades sociales, los resultados obtenidos indican la importancia de tentar reducirlo en medidas cognitivas para garantizar la exactitud de los resultados.

*Palabras clave:* Inteligencia, pruebas, cognición, género, cultura.

The effect of gender differences on cognitive abilities has been under continual scrutiny by researchers. Studies have attempted to explain such differences by investigating genetic and hormonal factors, neuroanatomical functions and environmental influences (education and social class). However, the results have been mixed, indicating the need for a psychobiosocial approach to understanding gender differences in intellectual ability (Doherty, Kovas & Plomin, 2011; Nisbett *et al.*, 2012; van der Sluis *et al.*, 2008). If these differences do exist, then it is unclear when they emerged and how they developed. Confirmation of the existence or absence of gender differences in cognitive abilities is crucial to political decision making and has a wide range of implications for public policy (Halpern & LaMay, 2000). If intellectual gender differences are confirmed, then a full array of psychological and educational tests must be revised to avoid bias in specific items or activities to ensure the fair application of test results (van de Vjiver & Leung, 2000).

Cognitive abilities can be defined in several ways. Cattell (1971) proposed two major types of cognitive abilities: Gf (fluid intelligence) and Gc (crystallized intelligence). Gf includes deductive and inductive reasoning, and Gc comprises abilities that reflect the influence of culture and education on verbal knowledge (Flanagan, McGrew & Ortiz, 2000). Subsequently, Horn (1985) and Carroll (1993) differentiated among general, broad and specific cognitive abilities in their Cattell-Horn-Carroll model (CHC), which was expanded by McGrew (2005). Gc has often been measured through tests involving information, synonyms, antonyms and verbal analogies and plays an important role in the majority of intelligence tests, primarily reflecting the systematic role of language and acculturation in determining an individual's intellectual potential (Woodcock, McGrew & Mather, 2001).

Studies related to gender differences in specific abilities have yielded inconsistent results. In meta-analytic investigations of studies performed in the 1980s and 1990s (Hyde, 1981; Hyde, 2005; Hyde & Linn, 1988), the magnitude of gender differences in verbal abilities was considered to be small. Nevertheless, current studies continue to demonstrate the existence of gender differences in different tests designed to measure specific intellectual abilities. For instance, in studies of sex differences in latent cognitive abilities in US samples using the Woodcock-Johnson III Battery (WJ III), men have demonstrated a small but consistent advantage with respect to crystallized ability (measured by comprehension/knowledge tasks), whereas women have outperformed men on processing speed tests (Camarata & Woodcock, 2006; Keith, Reynolds, Patel & Ridley, 2007). In a study using the Kaufman Assessment Battery for Children (KABC-II) with US children and youth, Reynolds, Zeith, Ridley and Patel (2008) demonstrated that boys have a significant mean advantage in their visual spatial ability and crystallized intelligence factors, whereas girls scored higher on latent higher-order *g* factors at two age levels.

Findings from other countries also reveal gender differences in specific abilities. In research investigating sex differences on the Wechsler Intelligence Scale for Children-Revised (WISC-R) with Dutch and Belgian children, boys outperformed girls on the verbal subtests of information and arithmetic, although no significant differences in general intelligence were found (van der Sluis *et al.*, 2008). In another study, Goldbeck, Daseking, Hellwig-Brida and Pertermann (2010) used the WISC-IV with German children and adolescents and reported gender effects favoring boys on the verbal comprehension index, although no significant differences on the full scale IQ were found. An investigation of sex differences among Chinese adults using the Wechsler Adult Intelligence Scale-Revised (WAIS-R) demonstrated that men scored reliably higher than women on the verbal, performance and full scale IQ indicators (Dai, Ryan, Paolo & Harrington, 1991). A recent study comparing Brazilian youth and adults on Raven Progressive Matrices and on two Brazilian cognitive tests (AC and BPR5) indicated that the overall scores of women were superior to those of men in attention ability, whereas men had superior total scores on verbal analogies as well as on mechanical, numerical and spatial reasoning tests (Flores-Mendoza *et al.*, 2013). However, another study with Spanish children did not find gender differences neither on results of the Raven Progressive Matrices Test nor in their mathematical abilities (Aragón, Delgado, Aguilar, Araujo & Navarro, 2013).

Gender differences in verbal abilities have manifested at least one paradox. Although boys tend to outperform girls on intelligence subtests that require crystallized intelligence or verbal ability, they tend to have significantly lower scores when this ability is measured by educational achievement tests (Greene & Winters, 2006). Consideration of education and socioeconomic status (SES) plays a major role in understanding gender differences and human abilities. Indeed, in a large-scale study conducted in 70 countries through the Program for International Student Assessment (PISA), researchers found that girls outperformed boys in reading in all countries. However, their mathematics and science scores surpassed the scores of the boys only in some nations. Socio economic variables had a significant influence on these results: in 21 countries, superior scores were obtained by students attending private schools compared with those attending public schools (OECD, 2010). Family environment has also an important impact on reading abilities and school achievement (Andres, Urquijo, Navarro & Garcia-Sedeño, 2010).

Gender bias in job access, termed the glass ceiling in the literature, is noted in many countries even today, as observed by Tjardens and van Klaveren (2012) in their comparison of data from 43 countries. Because intelligence tests are often used for competency assessments in different job situations, determining what is being measured and how this measurement is performed are of great concern (Hyde & Kling, 2001). Modern test analysis using item response theory (IRT) is recommended for

understanding item discrimination based on the respondent's gender rather than the classical analytic strategy of using overall scores (Linacre, 2004; Wright, 1977). Recommendations are also extended to those who construct cognitive tests that the gender roles in different cultures be noted and taken into consideration in item scoring (Sireci & Allalouf, 2003; van de Vijver and Leung, 2000).

Most studies examining gender and intellectual performance among adults have been conducted in North American, European and Asian countries. Few data describing gender differences in the cognitive abilities of South American adults currently exist. In Brazil, for instance, most studies have focused on children and adolescents (Almeida & Primi, 2004; Wechsler *et al.*, 2010). However, as van de Vijver and Portinga (1997) have noted, gender-biased tasks in cognitive tests must be identified because their influence on the lives of adolescents and adults –in areas such as vocational guidance, job selection and future decisions– may be significant. Because language reflects acculturation and has a major influence on most tests that measure intellectual performance, it is necessary to determine whether there are gender differences in crystallized intelligence or verbal abilities among Brazilians adults.

Given the importance of understanding gender differences in specific intellectual abilities, the purpose of this study was to verify whether gender bias exists in tests that have been created in Brazil to measure crystallized intelligence. The focus of this investigation was to identify item bias and to analyze whether possible disparities in item parameters and contents could explain the gender differences in the test results.

## METHOD

### *Participants*

The sample included 1191 individuals (528 men and 663 women) between the ages of 16 and 77 years old (Mean=22.0; *SD*=9.52). The participants lived in four Brazilian states (Amazonas, Paraiba, Rio Grande do Sul and Sao Paulo) and were members of the middle and upper socioeconomic classes. The majority of the sample was composed of college students in various majors (58.3%), but 19.8% of the participants had received only a secondary education, whereas 21.9% were graduate students. The participants were recruited by teachers or friends and were invited to participate in the study according to their availability.

### *Instruments*

*Tests of Crystallized Intelligence:* Four verbal tests assessing crystallized intelligence were constructed for this study: vocabulary, synonyms, antonyms and verbal analogies. The items for these tests were retrieved from the Houaiss Electronic Dictionary of the Portuguese Language (Houaiss, 2007). A list of 500 nouns and

adjectives was selected from this dictionary and administered to samples of college students majoring in different fields to verify the comprehensibility of the content and the initial difficulty level. Extremely easy items (90% or more correct responses) and extremely difficult items (10% or fewer correct responses) were eliminated from the test. From this list, 69 nouns were selected because they could be rendered pictorially and used to illustrate the vocabulary items. In addition, 57 words were chosen for the synonym test, 57 for the antonym test and 114 for the analogy test. These items were again presented to the college students and eliminated according to their comprehensibility and difficulty levels to obtain the final version, which could be administered in a group setting and completed in half an hour.

*Table 1.* Item descriptions for each crystallized intelligence test

Item	Vocabulary	Synonyms	Antonyms	Verbal Analogies
1	drumstick	help	hollow	nose is to smell as hand is to...
2	xylophone	hide	truthful	glass is to metal as bottle is to...
3	ripple	reprimand	generous	hand is to neck as ring is to...
4	eclipse	enhance	attentive	paper is to wood as tire is to...
5	flamingo	perspiration	opaque	foot is to finish as head is to...
6	boiler	vaporous	facultative	group is to people as harbor is to...
7	crotch	devour	conciliatory	chalk is to blackboard as brush is to...
8	telescope	pacify	listener	diamond is to jewel as button is to...
9	palette	visualize	zealous	shoe is to boot as car is to...
10	zeppelin	oppress	noble	rat is to rabbit as cheese is to...
11	gladiator	pagan	credulous	boy is to man as calf is to...
12	sluice	tolerate	negligent	coat is to sun as winter is to...
13	missile	babble	universal	jump is to air as escalade is to...
14	fetter	slender	skeptic	watch is to lantern as time is to...
15	totem	xenophobia	bellicose	stomach is to lung as digestion is to...
16	dome	vehement	obtuse	jail is to library as convict is to...
17	coat of arms	tormentor	opponent	fire is to oxygen as men is to...
18	quadricycle	permeate	gallant	wagon is to horse as lamp is to...
19	hurdy-gurdy	sophism	luxuriant	brick is to house as step is to...
20	monocle	shaman	eradicate	plant is to cultivate as egg is to...
21	scaffold	crave	laconic	branch is to root as to sow is to...
22	magician	mangle	fortuitous	chair is to bed as to seat is to...
23	dagger	perseverance	latent	shelf is to book as safe is to...
24	mosque	worship	eagerness	bottle is to eat as drink is to...
25	sheave	vanguard	vulgar	shield is to sword as defense is to...
26	saddle	hesitate	abstention	dive is to fly as ocean is to...
27	shrine	exalt	synopsis	
28	catapult	infringe	altruism	
29	saber	flow	scarceness	
30	abacus	opulent	veridical	
31	handbarrow	mockery		
32	turnstile			
33	scaffold			
34	banister			
35	oboe			
36	knocker			
37	blowgun			
38	mandolin			

The final version of the vocabulary test comprised 38 pictures that the participants were instructed to name. The synonym test presented 31 words as stimuli,

the antonym test contained 30 adjectives as stimuli, and the verbal analogy test was composed of 26 pairs of words. Table 1 presents each item on these specific tests.

### *Procedure*

The participants responded to the verbal tests in a two-hour group session either during or after school or work. As compensation for their participation, the individuals received a graph in the mail showing their percentage of correct answers on each test two months after the survey.

Two analyses were performed: an analysis of variance (ANOVA) and a one-parameter logistic IRT model (Rasch model). These analyses were performed to determine whether the crystallized intelligence tests revealed any differences attributable to gender. The ANOVA was used to investigate differences in the crystallized intelligence construct, and the Rasch model was used to identify differential item functioning (DIF) between genders on each subtest. Item functioning differences shape the item characteristic curve (ICC), which is estimated using the number of correct versus incorrect item responses (Linacre, 2004; Hunter and Schmidt, 2000; Rasch, 1966). This curve (or nonlinear regression) is described by curve parameters that are estimated statistically. Differences in the estimated ICC parameter values between groups indicate item bias; that is, if an item's ICC differs between the two groups, then that item is biased against one of the groups.

Therefore, the objective of the DIF analysis that was conducted in this study was to identify interactions between genders and item difficulty parameters, as such interactions indicate whether some items are relatively more or less difficult for a particular gender (Smith, 2004). The WINSTEPS program was used to calculate the DIF (Linacre, 2010) and to compare the performance of men and women on each test. Three criteria commonly found in the literature (Elder, McNamara & Congdon, 2004) were used to evaluate the presence of DIF: a separate calibration t-test, a comparison of the item difficulty parameters for the two groups and probability. Only critical t-test values of 2.4 or higher were considered significant. In addition, only contrasts higher than .42 or probabilities of less than .05 were considered significant. Thus, positive or negative values corresponded to an item's DIF and indicated whether the item was biased toward a specific gender.

## **RESULTS**

The first analysis assessed the influence of the gender variable on the constructs that were assessed in each of the four verbal subtests. The results from the ANOVA indicated that gender did not exercise a significant influence ( $p \leq .05$ ) on any of the crystallized intelligence tests, as shown in table 2.

Table 2. ANOVA of the gender differences for each crystallized intelligence test

Test	Gender	<i>M</i>	<i>SE</i>	<i>SSQ</i>	<i>MS</i>	<i>df</i>	<i>F</i>	<i>Sig</i>	<i>n</i> <sup>2</sup>
Vocabulary	1	10.570	.240	10.762	10.762	1	.281	.596	.000
	2	10.379	.269						
Synonyms	1	6.549	.192	6.674	6.674	1	.272	.602	.000
	2	6.700	.216						
Antonyms	1	8.916	.220	34.889	34.889	1	1.085	.298	.001
	2	9.260	.247						
V. Analogy	1	13.092	.194	10.102	10.102	1	.459	.499	.000
	2	12.894	.218						

Note: Gender 1=Women; 2=Men.

The second analysis, which focused on the influence of gender on the test items, was then conducted to determine the DIF. According to the unidimensionality requirement of Item Response Theory's (IRT), four analyses were conducted (one for each subtest). An earlier analysis was conducted to estimate the reliability and fit indices for each subtest. The results yielded adequate Cronbach's alpha reliability values for all of the verbal tests (vocabulary=.83; antonyms=.82; synonyms=.76; verbal analogies=.82). Infit and outfit indices were used to analyze the model fitness. Based on the recommendations in the literature, an infit value above 1.3 indicates model misfit (Smith, 2004; Hambleton & Swaminathan, 2010). Only one item (item 27 on the vocabulary subtest: sanctuary) yielded a value that was slightly higher than expected (1.33). All other items demonstrated adequate fit indices.

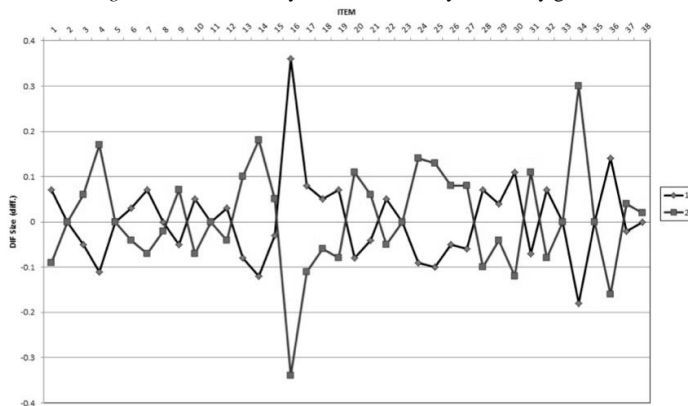
For all of the analyses, the value of difficulty (b parameter) for both groups was estimated. As explained previously, to determine the presence of DIF, three criteria were used: t-value, DIF contrast and Mantel-Haenszel probability. The item parameter was considered with respect to the value of theta (the latent variable). The probability of success is fixed at .50, a value to which the theta value is compared. For example, an item with a theta value of -1.8 is easier than an item with a theta value of .60. Therefore, as ability increases, the probability of correctly answering the item also increases.

The analysis of the vocabulary subtest indicated that none of the items exhibited DIF according to the first criterion (t-value higher than 2.4). When the second criterion was used (DIF contrast higher than .42), item 16 (dome) exhibited DIF. Because the difficulty level was 2.94 for women and 2.24 for men, item 16 appears to be easier for men. According to the third criterion, which is a Mantel-Haenszel probability of less than .05, two items exhibited DIF: item 4 (eclipse), which yielded a difficulty level of .30 for women and .58 for men and item 34 (banister), which yielded a difficulty level of 1.33 for women and 1.81 for men. Thus, both items appear to be easier for women.

Figure 1 illustrates the effect of DIF on the vocabulary test. The gender differences found for items 4 (eclipse), 16 (dome) and 34 (banister) are larger than the differences observed for the other items. The gender differences were verified following an inspection of the types of answers given by men and women for these items. On item 16, men tended to give the specific and correct answer (dome), whereas women tended

to give a more general name for this part of a building (ceiling). On item 4, women tended to give the correct answer (eclipse); whereas men gave answers related to the earth's spatial positioning (the earth's rotation). Finally, item 34 (banister) was easier for women to identify as part of a house but more difficult for men.

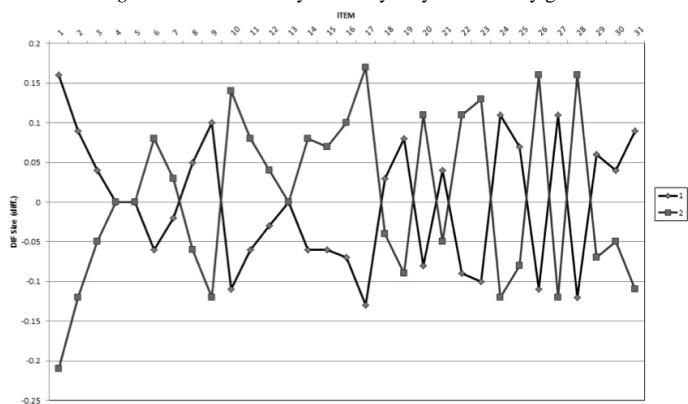
Figure 1. Item difficulty on the vocabulary subtest by gender



Note: Series 1=Women; 2=Men.

On the synonym test, further gender-based DIF patterns were observed. One item exhibited DIF (item 1: to help) according to two criteria (t-value higher than 2.4 and probability less than .05). This item was easier for men ( $b_{men} = -2.76$ ,  $b_{women} = -2.39$ ). Another item (item 17: tormentor), which exhibited DIF according to the probability criterion, was easier for women ( $b_{women} = .81$ ,  $b_{men} = 1.11$ ). No DIF was detected under the DIF contrast criterion.

Figure 2. Item difficulty on the synonym subtest by gender



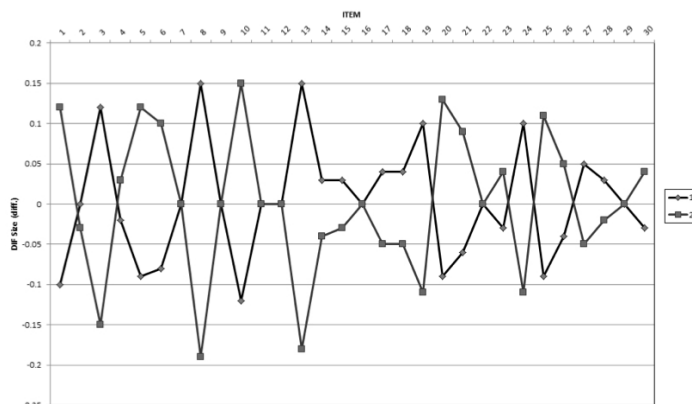
Note: Series 1=Women; 2=Men.



Figure 2 shows the DIF effects of the synonym test. The differences between the genders on items 1 and 17 were larger than on the other items. Men answered item 1 (to help) correctly (to offer assistance), whereas women tended to answer in a more general manner (to keep company, to stand beside). Item 17 (tormentor) was answered correctly by most women as a persecutor or chaser, but men tended to give answers such as boring or dull. Although there were also differences in the difficulty levels of other items, these differences were not sufficiently large to be significant, as observed for items 10, 26 and 28.

The analysis of the antonym subtest showed that none of the items exhibited DIF according to the first criterion (t-value higher than 2.4). The same result was observed when the second criterion was applied (DIF contrast higher than .42). According to the probability criterion (less than .05), two items exhibited DIF: item 8 (listener) and item 13 (universal), both of which were easier for men. The difficulty level of item 8 was -1.21 for women and -1.54 for men. The difficulty level of item 13 was -.19 for women and -.52 for men.

Figure 3. Item difficulty on the antonym subtest by gender



Note: Series 1=Women; 2=Men.

Figure 3 shows the differences in difficulty according to gender on the antonym test. The largest gender differences occurred for items 8 and 13. Items 3 and 10 also revealed differences, but these differences were not sufficiently large to confirm DIF. Item 8 was easier for men, who provided correct antonyms for listener (lecturer or speaker), whereas women tended to give answers such as careless and neglectful. On item 13, men typically gave the correct answer for the antonym for universal (local or regional), whereas women tended to give answers such as private or not useful.

The results of the verbal analogy subtest indicated that none of the items exhibited DIF according to the first criterion. The same result was observed using the

second criterion (DIF contrast higher than .42). However, one item exhibited DIF according to the probability criterion (less than .05): item 17 (fire is to oxygen as man is to...). The difficulty level for this item was 1.51 for women and 1.17 for men. Therefore, this item appears to be easier for men.

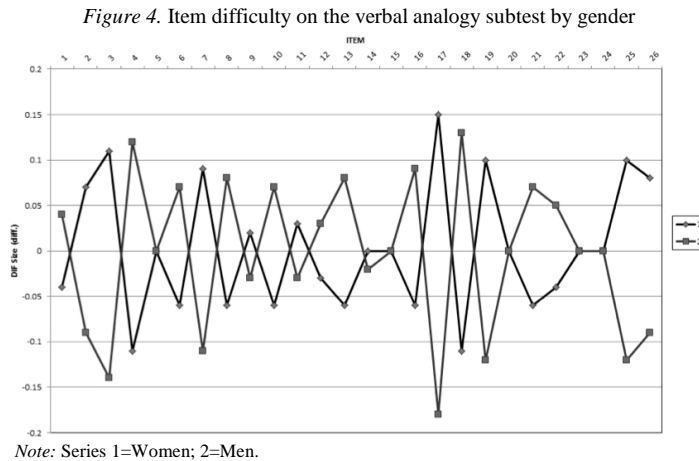


Figure 4 shows the DIF effect on the verbal analogy test. Item 17 showed the largest gender difference among all of the items. Men tended to give the word ‘energy’ to correctly complete this analogy, whereas women attempted to solve it with the word ‘food.’ Although other items also showed differences in difficulty between the genders, these differences (those for items 3, 4, 18 and 25) were not sufficiently large to be significant.

## DISCUSSION

The study aimed to investigate whether there were gender differences in crystallized intelligence that could be observed through tests of verbal ability.

The first analysis of the overall scores for both genders on each crystallized intelligence test revealed no significant differences, thus demonstrating that gender does not influence Gc, which plays a central role in intelligence measures. This finding confirms the conclusions of Hyde’s meta-analytic studies (1981, 2005), which demonstrated no significant gender differences on verbal tests, but is contradictory to several other studies that have indicated that men have superior crystallized intelligence (e.g., Dai *et al.*, 1991; Flores-Mendoza *et al.*, 2013; Keith *et al.*, 2008). Nevertheless, women continue to suffer from bias that favors the intellectual performance of men in

many countries, and the confirmation of possible gender bias in tests is important to define public policies (Acker, 2009).

Differences between women and men were primarily related to the manner in which they perceived the item content for 8.7% of the items. Women tended to find content related to their daily lives to be easier (e.g., banister), to respond more globally to the stimuli than men (e.g., eclipse) and to grasp content related to their social concerns (e.g., chaser). Items involving verbal analogies were more difficult for women (e.g., fire is to oxygen as man is to...). However, the question remains as to whether verbal analogies measure only crystallized intelligence or whether they also implicate fluid intelligence, an area in which boys tend to outperform girls, as noted by Steinmayr, Beauducel and Spinath (2010). Explanations for gender differences in perceiving and responding to test items are complex and may be related to the history, biology, traditions, cultural restrictions and educational experiences of women, thus requiring a broad psychobiosocial approach to be fully understood (Deary *et al.*, 2006; Nisbett *et al.*, 2012).

Researchers should use tests constructed within the culture being studied to avoid culture bias in understanding item meanings, as recommended by Bartram (2001) and Sireci and Allalouf (2003). However, although the verbal tests in the current study were elaborated using Portuguese words, gender bias was still evident in the manner in which the respondents perceived and responded to the stimuli; thus, the results indicate that the content of some items may be easier or more difficult for either gender. Modern statistical methods, such as DIF, which was used in this study, are recommended when the objective is to investigate item bias on intelligence tests and to facilitate the fair application of test results, as suggested by Hambleton and Swaminathan (2010).

The limitations of this study can be attributed to the exclusive focus on crystallized intelligence to determine gender differences. Further research is still needed to identify whether gender bias exists in other specific abilities, such as visual-spatial ability or logical reasoning. More studies should be conducted to investigate the effect of other variables on gender differences, such as age and educational level, which were not controlled in this study. Given that intelligence tests are used worldwide for different purposes, the acquisition of additional information in other Latin American countries can be greatly beneficial in understanding women's perceptions and their responses to cognitive tests.

An important contribution of this study was the finding that gender bias can be present in test items and can affect women's performance on tests of specific abilities. Item analysis and calibration are essential to ensuring fairness in cognitive assessments, although such procedures are seldom performed. Because of the influence of test results on women's social opportunities, the examination of test items and possible sources of bias is recommended to reduce gender-based discrimination in all nations.

## REFERENCES

- Acker, J. (2009). From grass ceiling to inequality regimes. *Sociologie du Travail*, 51, 199-217.
- Almeida, L.S. & Primi, R. (2004). Perfis de capacidades cognitivas na Bateria de Provas de Raciocínio -BPR-5) [Cognitive ability profiles on the Battery of Reasoning Tests-BPR-5]. *Psicologia Escolar e Educacional*, 8(2), 135-144.
- Andrés, M.L., Urquijo, S., Navarro, J.I. & García-Sedeño, M. (2010). Contexto alfabetizador familiar: relaciones con la adquisición de habilidades prelectoras y desempeño lector. *European Journal of Education and Psychology*, 3(1), 129-140.
- Aragón, E.L., Delgado, C.I., Aguilar, M., Araujo, A. & Navarro, J.I. (2013). Estudio de la influencia de la inteligencia y el género en la evaluación matemática temprana. *European Journal of Education and Psychology*, 6, 5-18.
- Bartram, D. (2001). The Development of International Guidelines on Test Use: The International Testing Commission Project. *International Journal of Testing*, 1, 33-53.
- Camarata, S. & Woodcock, R. (2006). Sex differences in processing speed: Developmental effects on males and females. *Intelligence*, 34, 231-252.
- Carroll, J. (1993). *Human cognitive abilities: A survey of factor analytic studies*. New York: Cambridge University Press.
- Cattell, R.B. (1971). *Abilities: Their structure, growth and action*. Boston: Houghton Mifflin.
- Dai, X., Ryan, J.R., Paolo, A.M. & Harrington, R.G. (1991). Sex differences on the Wechsler Adult Scale-Revised for China. *Journal of Consulting and Clinical Psychology*, 3, 282-284.
- Deary, I.J., Strand, S., Smith, P. & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, 35, 13-21.
- Doherty, S.J., Kovas, Y. & Plomin, R. (2011). Gene-environment interaction in the etiology of mathematical ability using SNP sets. *Behavior Genetics*, 41, 141-154.
- Elder, C., McNamara, T. & Congdon, P. (2004). Rasch techniques for detecting bias in performance assessments: an example comparing the performance of native and non-native speakers on a test of academic English. In E.V. Smith Jr. & R.M. Smith (Eds.), *Introduction to Rasch measurement: theory, models and applications* (pp.419-444). Minnesota: Maple Grove.
- Flanagan, D.P., McGrew, K.S. & Ortiz, S.O. (2000). *The Wechsler Intelligence scales and Gf-Gc theory*. Needham Heights, MA: Allyn and Bacon.
- Flores-Mendoza, C., Widaman, K.F., Rindermann, H., Primi, R., Mansur-Alves, M. & Pena, C.C. (2013). Cognitive sex differences in reasoning tasks: Evidence from Brazilian samples of educational settings. *Intelligence*, 41, 70-84.
- Goldbeck, L., Daseking, M., Hellwig-Brida, S., Waldman, H.C. & Petermann, F. (2010). Sex differences on the German Wechsler Intelligence Test for Children (WISC-IV). *Journal of Individual Differences*, 31, 22-28.
- Greene, J.P. & Winters, M.A. (2006). *Leaving boys behind: Public high School graduation rates. Civic Report*. Center for Civic Innovation at Manhattan Institute, NY.
- Halpern, D.F. & LaMay, M.L. (2000). The smarter sex. A critical review of sex differences in intelligence. *Educational psychology Review*, 12, 229-246.

- Hambleton, R.K. & Swaminathan, H. (2010). *Item response theory: Principles and applications*. Boston: Kluwer-Nijhoff Publisher.
- Hyde, J.S. (1981). How large are cognitive gender differences? A meta-analysis using  $w^2$  and  $d$ . *American Psychologist*, 36(8), 892-901.
- Hyde J.S. (2005). The gender similarities hypothesis. *American Psychologist*, 60, 581-592.
- Hyde, J.S. & Linn, M. (1988). Gender differences in verbal ability: A meta-analysis study. *Psychological Bulletin*, 104, 53-69.
- Horn, J.L. (1985). Remodeling old models of intelligence. In B. Woldman (Ed.), *Handbook of intelligence* (pp. 267-300). New York: Wiley & Sons.
- Houaiss. A. (2007). *Dicionário da Língua Portuguesa*. CD-Versão eletrônica 2.0. [Portuguese Language Dictionary-CD-Electronic Version-2.0]. Rio de Janeiro, RF: Instituto Antonio Houaiss.
- Keith, T.Z., Reynolds, M.R., Patel, P.G. & Ridley, K.P. (2008). Sex differences in latent cognitive abilities ages 6 to 59. Evidence from the Woodcock-Johnson III tests of cognitive abilities. *Intelligence*, 36, 502-525.
- Linacre, J.M. (2004). Rasch model estimation: further topics. *Journal of Applied Measurement*, 5, 95-110.
- Linacre, J.M. (2010). *WINSTEPS Rasch-Model computer program*. Beaverton, OR: Winsteps.com.
- McGrew, K.S. (2005). The Cattell-Horn-Carroll theory of cognitive abilities: past, present and future. In D.P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (pp 136-181). New York: The Guilford Press.
- Nisbertt, R.E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. & Turkheimer, E. (2012). Intelligence: New findings and theoretical development. *American Psychologist*, 67, 130-159.
- Organization for Economic Cooperation and Development-OECD (2010). *PISA 2009 at a glance*. Retrived from <http://www.oecd.org/pisa/46660259.pdf> on August 27th, 2012.
- Reynolds, M.R., Keith, T.Z., Ridley, K.P. & Patel, P.G. (2008). Sex differences in latent general and broad cognitive abilities for children and youth: Evidence from higher order MG-MACS and MIMIC models. *Intelligence*, 36, 236-260.
- Sireci, S.G. & Allalouf, A. (2003). Appraising item equivalence across multiple language and cultures. *Language Testing*, 20(2), 148-167.
- Smith, R.M. (2004). Detecting item bias with the Rasch model. In E.V. Smith Jr. & R.M. Smith (Eds.), *Introduction to Rasch measurement: theory, models and applications* (pp.391-418). Minnesota: Maple Grove.
- Steinmayr, R., Beauducel, A. & Spinath, B. (2010). Do sex differences in a faceted model of fluid and crystallized intelligence depend on the method applied? *Intelligence*, 38, 101-110.
- Tijdens, K.G. & Van Klaveren, M. (2012). *Frozen in time: Gender pay gap unchanged for 10 years*. Brussels: ITUC.
- van der Sluis, S., Derom, C., Thiery, E., Bartels, M., Polderman, T., Verhulst, F.C., Jacobs, N., van Gestel, S., de Geus, E.J.C., Dolan, V.C., Boomsma, D.I. & Posthuma, D. (2008). Sex differences on the WISC-R in Belgium and Netherlands. *Intelligence*, 36, 48-67.
- van de Vijver, F.J.R. & Leung, K. (2000). Methodological issues in psychological research on culture. *Journal of Cross-Cultural Psychology*, 31, 3-5.

- van de Vijver, F.J.R. & Poortinga, Y.H. (1997). Towards an integrated analysis of bias in cross-cultural assessment. *European Journal of Psychological Assessment*, 3, 29-37.
- Wechsler, S.M., Nunes, C.S., Schelini, P.W., Pasian, S., Homsí, S.V., Moretti, L. & Anache, A.A. (2010). Brazilian adaptation of the Woodcock-Johnson III cognitive tests. *School Psychology International*, 31, 409-421.
- Woodcock, R.W., McGrew, K.S. & Mather, N. (2001). *Woodcock-Johnson III*. Itasca, IL: Riverside Publishing.

Received: August 29th, 2013

Modifications Received: January 3rd, 2014

Accepted: January 24th, 2014