Title
Cognitive sex differences in reasoning tasks: Evidence from Brazilian samples of educational settings

Journal
Intelligence, 41(1)

ISSN
0160-2896

Authors
Flores-Mendoza, Carmen
Widaman, Keith F
Rindermann, Heiner
et al.

Publication Date
2013

DOI
10.1016/j.intell.2012.11.002

Peer reviewed

Intelligence 41 (2013) 70–84

Contents lists available at SciVerse ScienceDirect

Intelligence

Cognitive sex differences in reasoning tasks: Evidence from Brazilian samples of educational settings

Carmen Flores-Mendoza [a,*], Keith F. Widaman [b], Heiner Rindermann [c], Ricardo Primi [d], Marcela Mansur-Alves [a], Carla Couto Pena [a]

[a] Department of Psychology, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil
[b] Department of Psychology, University of California, Davis, USA
[c] Department of Psychology, Technical University Chemnitz, Germany
[d] Department of Psychology, Universidade São Francisco, Itatiba, São Paulo, Brazil

ARTICLE INFO

Article history:
Received 7 July 2010
Received in revised form 7 November 2012
Accepted 10 November 2012
Available online 9 December 2012

Keywords:
Sex differences
Intelligence
Reasoning tasks
Brazilian people
g factor

ABSTRACT

Sex differences on the Attention Test (AC), the Raven's Standard Progressive Matrices (SPM), and the Brazilian Cognitive Battery (BPR5), were investigated using four large samples (total $N = 6780$), residing in the states of Minas Gerais and São Paulo. The majority of samples used, which were obtained from educational settings, could be considered a nonprobability sampling. Females outperformed males on the AC (by 2 IQ points), whereas males slightly outperformed females on the SPM (by 1.5 IQ points). On the BPR5, sex differences favoring males were statistically significant (on average 6.2 IQ points). The largest difference was in Mechanical Reasoning (13 IQ points), and the smallest was in Spatial Reasoning (5 IQ points). In addition, two methods were adopted for determining whether sex differences existed at the level of general intelligence. First, a *g* factor score was estimated after principal axis factoring of test scores. Men had an advantage of 3.8 IQ points (statistically significant) on the *g* score, which was reduced to 2.7 IQ points (not significant), when the *g* score was estimated without including Mechanical Reasoning. Second, a confirmatory factor analysis approach was conducted that allowed testing of mean differences at the latent variable level. Again, sex differences favoring males were found (0.23 or 3.44 IQ points). Regarding educational and SES variables, some sex differences favoring males were found in the SPM and in the BPR5. In general, our results agree with studies that identify small, but consistent cognitive sex differences in reasoning tasks. Societal implications are discussed.

1. Introduction

Some topics in the social and behavioral sciences are difficult and controversial due to the political and societal consequences and implications of the empirical results obtained. Research on cognitive sex differences is one such topic. Discussions regarding sex differences began in the early 20th century, when the first investigations were recorded negatively in books such as, "The mental inferiority of woman", written by the German physiologist Moebius (Andrés-Pueyo & Zaro, 1998). Since then, a myriad of investigations, including more technical and less prejudiced and political, have highlighted whether cognitive sex differences really exist. Empirical results accumulated to date suggest that the final answer is still far from being achieved.

Scientific uncertainty regarding cognitive sex differences arises from the vast accumulation of evidence that men and women may be reliably different on specific dimensions of ability. However, whether the sex differences can reliably be identified at the level of general intelligence is a matter that has

* Corresponding author at: Av. Antonio Carlos, 6627, Fafich, Depto. de Psicologia, Gab. 4042, Universidade Federal de Minas Gerais, CEP 31330-030, Brazil.

E-mail addresses: carmencita@fafich.ufmg.br, carmenflor@uol.com.br (C. Flores-Mendoza), kfwidaman@ucdavis.edu (K.F. Widaman), heiner.rindermann@psychologie.tu-chemnitz.de (H. Rindermann), ricardo.primi@saofrancisco.edu.br (R. Primi), marcelamansuralves@yahoo.com.br (M. Mansur-Alves), carlacouto@magnum.com.br (C.C. Pena).

0160-2896/$ – see front matter
http://dx.doi.org/10.1016/j.intell.2012.11.002

not yet been resolved. Sex differences are often reported as standardized effect sizes based on mean differences between groups. These effects are usually calculated using Cohen's d, calculated as $d=(M_m-M_f)/S$, where M_m is the mean of the male sample, M_f is the mean for the female sample, and S is the pooled data within-group standard deviation. Thus, Cohen's d indicates the number of SD units of difference in mean performance between males and females. With regard to specific ability dimensions, reported d values (where negative values indicate females scored higher than males), vary considerably across studies and domains of mental ability. For instance, Hyde (1981) and Hyde and Linn (1988) found $d=-0.24$ and $d=-0.15$ respectively in verbal tasks, thus differences favoring females. Jensen (1998) asserted that women outperform men in perceptual speed and short term memory (d values -0.20 and -0.30, respectively), whereas Colom, Quiroga, and Juan-Espinosa (1999), using data from the *Differential Aptitude Tests* (DAT), found women were slightly better in speed ($d=-0.08$) and accuracy ($d=-0.05$). In addition, Colom et al. found higher female scores in vocabulary ($d=-0.03$), verbal fluency ($d=-0.01$), and inductive reasoning ($d=-0.038$), using 1995 normative data of *Primary Mental Abilities* (PMA). On the other hand, Linn and Petersen (1985) found $d=0.73$ in rotation tasks and $d=0.44$ in spatial relations tasks, thus favoring males. Voyer, Voyer, and Bryden (1995) reported similar results. Colom et al. (1999) found higher scores for males in Verbal Reasoning ($d=0.31$), Numerical Ability ($d=0.59$), Abstract Reasoning ($d=0.39$), Spatial Relationships ($d=0.39$), Mechanical Reasoning ($d=1.14$), calculation ($d=.57$), and Mental Rotation ($d=0.36$).

More recently, Geiser, Lehmann, and Eid (2008), reported evidence of sex differences favoring males in mental rotation at different ages (ranging from 9 to 23 years), with d values varying from 0.52 to 1.49. In another study, van der Sluis et al. (2008), analyzing sex differences in the WISC-R in Dutch children (ages 11 to 13 years) and Belgian children (ages 9.5 to 13 years), found that girls outperformed boys on the Coding subtest ($d=-0.53$ and $d=-0.53$, for Dutch and Belgians, respectively), whereas boys outperformed girls on the Information ($d=0.52$ and $d=0.37$, for Dutch and Belgians, respectively) and Arithmetic subtests ($d=0.31$ and $d=0.19$, for Dutch and Belgians, respectively). Also, Liu and Lynn (2011) analyzed samples of children (ages 5 to 6 years) from China, Japan, and the USA, assessed with the Wechsler Preschool and Primary Scale and found significant sex differences favoring boys on Information ($d=0.19$, China), Vocabulary ($d=0.20$, China), Arithmetic ($d=0.14$, China), Comprehension ($d=0.23$, China), Picture Completion ($d=0.11$ for China and $d=0.21$ for Japan), and Mazes ($d=0.35$ for China and $d=0.33$ for Japan). The only sub-test in which girls significantly outperformed boys was Animal House (a non-verbal test) for Japanese ($d=-0.36$) and USA ($d=-0.31$) samples.

Therefore, considering published research, little doubt remains that men and women, in fact, differ in their specific cognitive abilities. Moreover, some abilities appear to differentiate males and females in a consistent way. For instance, considering mean differences on a distribution across the sexes, women outperform men in tasks that require semantic processing, perceptual speed, and verbal memory (Codorniu-Raga & Vigil-Colet, 2003; Halpern, 1997; Johnson & Bouchard, 2007; Lubinski, 2004; Lynn, Raine, Venables, Mednick, & Irwing, 2005; Maccoby & Jacklin, 1974); whereas men demonstrate superior performance to that of women in visual–Spatial tasks, abstract reasoning, and numerical reasoning (Colom et al., 1999; Hyde, Fennema, & Lamon, 1990; Hyde & Linn, 1988; Linn & Petersen, 1985; Voyer et al., 1995). In general, similar results have been observed, with few exceptions, in children and adult samples and in populations from developed or from developing countries (Dai & Lynn, 2001; Dai, Ryan, & Harrington, 1991; Echavarri, Godoy, & Olaz, 2007; Flores-Mendoza, Mansur-Alves, Lelé, & Bandeira, 2007).

In contrast, this "relatively stable" pattern of differential sex differences, particularly in cognitive abilities, has not been observed in investigations regarding general intelligence. Conceptually, general intelligence (also called "g") is defined as a broad mental ability for reasoning, planning, resolving problems, thinking abstractly, and learning from experiences (Gottfredson, 1997; Hunt, 2011; Lubinski, 2004; Nisbett et al., 2012). Thus, g would be an important psychological attribute for human survival and adaptation to any context, as well as for maintaining personal success across a lifetime. Specific cognitive abilities, especially those with lower loading of g, would not have the same degree of general importance in survival and adaptation due to the narrow range of application of the specific abilities. Therefore, a consistent finding in differences among human groups at the level of g would imply that one group has a greater likelihood of successful adaptation than another group in its capacity for dealing, in a global way, with the challenges present in their physical and social contexts. Because of the perceived importance of g for successful adaptation and success in life, laypersons and experts pay special attention to the results of investigations of sex differences in the general intelligence.

Studies of sex differences in general intelligence are aligned within two academic camps. The first, asserts a male advantage in general intelligence. Studies by Lynn and his colleagues (Lynn, 1999, 2002; Lynn, Allik, Pullmann, & Laidra, 2004; Lynn, Backhoff, & Contreras-Niño, 2004; Lynn, Fergusson, & Horwood, 2005) clearly support this position, which confers a male advantage of 0.26 SD units, which corresponds to a sex difference of approximately 4 points on the well-known IQ metric. Other researchers (Jackson & Rushton, 2006; Nyborg, 2003, 2005) have arrived at similar results. According to Lynn, biological bases might explain these differences insofar as developmental sex differences are observed. Females do better at younger ages, but males obtain higher scores after they achieve biological maturation, around 16 years of age (Lynn, 1999), which is later than females. This idea is not new and can be tracked across the first studies of mental abilities published during the past century (e.g., Conrad, Jones, & Hsiao, 1933).

The second position, in contrast to the first, maintains that no sex difference in general intelligence exists (Aluja-Fabregat, Colom, Abad, & Juan-Espinosa, 2000; Codorniu-Raga & Vigil-Colet, 2003; Colom, Juan-Espinosa, Abad, & García, 2000; Dolan et al., 2006; Johnson & Bouchard, 2007; Mackintosh, 1996; van der Sluis et al., 2006, 2008). According to this position, sex differences found in some studies reflect, in fact, differences in "intelligence in general" (which refers to a combination of g + a mixture of specific cognitive abilities) instead of differences in "general intelligence" (or g). A major summary commissioned by the American Psychological Association endorsed this position (Neisser et al., 1996), as did a recent review of that summary (Nisbett et al., 2012).

Most studies finding cognitive sex differences are based on comparisons of raw scores or factor scores estimated from one or a small set of measures. In contrast, most studies denying such differences are based on analyses at the latent variable level, which supposedly remove the effects of error and unique variance from estimates of the relations among constructs and, thus, permit the investigation of sex differences in a more valid way with regard to the true constructs of interest. Nevertheless, even using sophisticated multivariate analysis, such as structural equation modeling, divergent conclusions are observed.

For example, some investigators fail to find significant sex differences in *g*. As can be seen in Table 1, using multi-group mean and covariance structure analysis, Dolan et al. (2006) and van der Sluis et al. (2006, 2008), asserted that there was an absence of sex differences in the standardization samples for the WAIS-III (Spain and The Netherlands) and for the WISC-R (samples from The Netherlands and Belgium).

In contrast, Steinmayr, Beauducel, and Spinath (2010), using the same analytical method (and additional methods) for analyzing the German Intelligence-Structure-Test 2000-R, found sex differences in G*f* (reasoning), G*c* (knowledge), numerical, and figural abilities, all favoring males, although findings for verbal ability varied according the method of analysis employed. Recently, Irwing (2012) conducted a multi-group confirmatory factor analysis for two models, hierarchical and bi-factor mean structure, using the American normative data of the WAIS-III. Both models indicated sex differences in *g*, Arithmetic Information, as well Symbol Search, favoring males, although differences in Processing Speed favored females.

Moreover, a third kind of result can be found: sex differences in *g* favoring females. For example, Keith, Reynolds, Patel, and Ridley (2008) analyzed the American standardization sample of the Woodcock–Johnson III cognitive battery (WJ-III). Despite significant differences favoring males on latent factors with high loading in *g*, such as Comprehension Knowledge — G*c* (0.85), Quantitative Reasoning — RQ (0.84), and Visuo-Spatial — G*v* (0.81), and only one significant difference favoring females on the latent factor (Processing Speed — G*s*) with lowest loading (0.63) in *g*, the authors surprisingly found a difference in *g* favoring females that was equivalent to 1.2 points in IQ. In another study, Keith, Reynolds, Roberts, Winter, and Austin (2011) used the American normative dataset of the Differential Ability Scales (a test that, according to the authors, is related to the Cattell–Horn–Carroll theory) and tested developmental changes in sex differences across 5- to 17-years of age in both

Table 1
Studies of cognitive sex differences at the latent level in developed countries.

Author	Country	*N*	% females	Population representativeness	Measures of *g*	Method of analysis*	Conclusion
Colom et al. (2000)	Spain	10,475[a]	41.6	No	PMA (Vocabulary, Verbal Fluency, Spatial Rotation, and Inductive Reasoning), Monedas	MCV	No sex differences related to *g*
Aluja-Fabregat et al. (2000)	Spain	1565[a]	51	No	Scholastic's Achievement, Toulouse–Pieron test, Memory Test (MAI), and Reasoning	MCV	No sex differences related to *g*
Colom et al. (2002)	Spain	1369[a]	51.3	Yes	WAIS-III	MCV	No sex differences related to *g*
Nyborg (2005)	Denmark	62[a]	50	No	The Rod-and-Frame test, Embedded-Figures, Money Left–Right Discrimination, Tapping, Oral Fluency, and WAIS-III	MCV	Significant sex differences favoring males
Dolan et al. (2006)	Spain	588[a]	47.7	Yes	WAIS-III	MGCMS	No sex differences related to *g*
van der Sluis et al. (2006)	Netherlands	522[a]	56.3	No	WAIS-III	MGCMS	No sex differences related to *g*
van der Sluis et al. (2008)	Netherlands and Belgium	1498[a]	52	No	WISC-R	MGCMS	No sex differences related to *g*
Steinmayr et al. (2010)	Germany	977[a]	56.4	No	German Intelligence-Structure-Test 2000-R	MIMIC and MGCMS	Significant differences in G*f* and G*c* favoring males
Keith et al. (2008)	USA			Yes	Woodcock–Johnson III	MIMIC	Significant differences favoring females
Keith et al. (2011)	USA	2600[b]	50.0	Yes	Differential Ability Scales	MIMIC and MGCMS	No sex differences related to *g*, but sex differences in processing speed and visual processing favoring girls and boys respectively.
Irwing (2012)	USA	2450[a]	53.2	Yes	WAIS-III	MGCFA	Sex differences in *g*, Arithmetic, Information, and Symbol Search favoring males. Processing Speed favorable to females.

Note:
[a] Adult sample.
[b] Children/adolescent sample.
* MVC = Method of Correlated Vectors, MGCMS = Multi-Group Covariance and Means Structure Analysis, MIMIC = Multiple Indicator–Multiple Cause, and MGCFA = Multi-Group Confirmatory Factor Analysis.

broad cognitive abilities and mean levels of *g*. The multi-group structural equation modeling indicated sex differences in processing speed ability favoring females across all ages, and differences in visual processing favoring 8 year old males. But, no statistical sex differences were found at the level of the *g* factor.

The bases for these conflicting results at the *g* level across studies are not clear; however, Nyborg (2003), Steinmayr et al. (2010), and Keith et al. (2011) argued that the nature of measures, size, quality and kinds of samples, and type of analysis may account for the conflicting patterns observed in each result. In addition, the study by Madhyastha, Hunt, Deary, Gale, and Dykiert (2009) suggested that biases from sample selection may enormously affect results and thus deserves special attention in sex differences research.

Existing studies of sex differences at the *g* level (with measures loading highly on *g*, such as reasoning or *gf* tasks) have been conducted, virtually without exception, in samples from developed countries (Table 1). Whether these results can be generalized to developing countries is unclear. Empirical studies from developing countries, in which the culture is supposedly more sexist (Lehmijoki & Palokangas, 2006; Morrison & Jutting, 2005) and social inequality is more profound (United Nations — UN, 2010), are few in number. Apart from the meta-analysis of 57 studies by Lynn and Irwing (2004) in which some studies from developing countries were included, an extensive literature search regarding cognitive sex differences in databases, such as PsycInfo, Scopus, and Scielo (a Brazilian database), revealed few studies published from developing countries (Table 2). From these studies, we could conclude that: (a) in children, no clear pattern of sex differences is found; but (b) in adults, sex differences appear to favor males. This tendency lends some support to the developmental theory of Lynn (1999), which asserts that, due to females maturing at an earlier age than males, females could outperform males until 15 or 16 years of age. The male advantage would emerge after late adolescence or in early adulthood, because males complete their biological maturation at a later age. Several studies have supported this position (Colom & Lynn, 2004; Lynn, 1999, 2002; Lynn & Irwing, 2004).

However, before we accept the conclusion that genuine cognitive sex differences are also present in developing nations, we must consider the fact that virtually all studies conducted in developing countries have relied on the administration of just one test, often the Standard Progressive Matrices test (SPM). The only study (conducted in Argentina) employed a multiple-subtest cognitive battery (five measures of the DAT) when testing sex differences and reported the well-known pattern of specific cognitive differences: males were better in verbal, abstract, and numerical reasoning tasks, and women were better in verbal fluency tasks (Echavarri et al., 2007). But we could not find a single study from developing countries that investigated cognitive sex differences at the *g* level using adult samples. Moreover, variables such as SES and educational attainment were rarely or never considered in existing studies from developing countries. This is unfortunate, as Levine, Vasilyeva, Lourenco, Newcombe, and Huttenlocher (2005) showed that social variables, such as SES, could challenge the hypothesis that the male advantage reflects ability differences in the population as a whole. For instance, Levine et al., showed cognitive sex differences in children from middle- and high-SES backgrounds, but no sex differences in the low-SES group were found.

According to Colom, Garcia, Juan-Espinosa, and Abad (2002), and Dolan et al. (2006), while IQ is a significant predictor of educational attainment, they found that the direct regression of educational attainment on *g* is not significant. Rather, some first order cognitive factors seem to predict educational attainment. Additionally, Dolan et al. found no sex differences at the *g* level. These results would mean that if sex differences in educational attainment are found, they cannot be attributed to *g* because there are not sex differences at *g*, and *g* is not responsible for differences in educational attainment. As previously said, neither educational attainment nor SES were considered in studies conducted in developing countries. Thus, the impact of each educational level on the cognitive performance of sexes could be verified through the present Brazilian dataset.

The goal of the present research was to provide empirical data on key points related to the topic of cognitive sex differences in developing countries. Specifically, we sought to test for: (a) sex differences in Brazilian samples on specific cognitive ability dimensions as well as at the level of general cognitive ability, and (b) any associations with background variables, such as educational level, and socioeconomic status.

2. Method

2.1. Participants

Four datasets comprising 6780 individuals (4771 adolescent/adults, and 2009 children) were analyzed. The adolescent/adult sample ($N = 4771$) was assessed using diverse tests: 2186 people took the AC (Brazilian Attention Test; Cambraia, 2003), 2076 participants took the BPR5 (Brazilian Cognitive Reasoning Battery; Primi & Almeida, 2000), and 988 individuals took the SPM (Standard Progressive Matrices; CEPA, 2001). Only 479 participants took both the SPM and the BPR5. The sample of children ($N = 2009$) was assessed using only the SPM.

Regarding the adolescent/adult sample, the majority of the participants were from the state of Minas Gerais and were assessed for the purpose of data collection in several projects conducted by the Laboratory of Individual Differences Assessment (www.fafich.ufmg.br/~ladi). Moreover, in this sample, normative data from the state of São Paulo related to the BPR test were also obtained.

The sample of children was composed of fourth- and fifth-grade students from 28 primary public schools, randomly selected from a list of schools available from the Secretary of Education of Belo Horizonte, Minas Gerais. This representative sample was part of the first phase of the project, entitled: "Survey and identification of gifted children in primary public schools of Belo Horizonte" (www.fafich.ufmg.br/~ladi).

The states of São Paulo and Minas Gerais are the most important commercial and industrial markets in Brazil (first and third highest GDP, respectively). Table 3 shows sample size, age range, sex distribution, and year of testing for each dataset. According to this table, it is possible to note that there is an equivalent number of Brazilian males and females in the childhood period. However, after the adolescent period, there is a higher presence of females in the Brazilian population (www.ibge.gov). One of the reasons could be the higher rate of

Table 2
Studies of cognitive sex differences in developing countries.

Author	Country	Measure[a]	Sample		Age	d	t
			Male	Female			
Klingelhofer (1967)	Tanzania	SPM	1104	836	13–18	–	NS
Baraheni (1974)	Iran	SPM	191	174	9	–	2.37**
			226	141	10	–	2.07*
			158	168	11	–	2.53*
			165	172	12	–	2.82*
			282	223	13	–	2.19*
			267	384	14	–	1.65
			291	425	15	–	1.57
			185	279	16	–	2.37*
			265	261	17	–	0.32
			176	128	18	–	0.05
Sabogal, Molina, and MacVean (1979)	Guatemala	OLSAT	360	360	6–16	–	NS
Dai et al. (1991)	China	WAIS-R	703	703	>17	0.92	20.92**
Costenbader and Ngari (2001)	Kenya	CPM	583	639	6–10	0.75	5.26***
Dai and Lynn (2001)	China	WISC-R	1132	1104	6–16	0.19**	–
Wechsler and Shelini (2002)	Brazil	DMT	145	110	7-9	–	−8.37**
Lynn (2002)	South Africa	SPM – White	490	566	15–16	−0.03	−1.1
		SPM – Indians	530	533		0.25	4.25***
		SPM – Colored	386	381		0.18	2.56*
		SPM – Black	554	539		0.33	8.12***
Lynn and Tse-Chan (2003)	Hong Kong [China]	APM	903	594	15–18	1.6	5.53***
Ivanovic et al. (2004)	Chile	WAIS-R	47	49	17–19	−0.02	−.039
Lynn, Allik, Pullmann, and Laidra (2004)	Estonia	SPM	1250	1459	12–18	−0.16	−4.03***
Lynn, Backhoff, and Contreras-Niño (2004)	Mexico	SPM	472	448	7–10	0.09	1.131
Lynn, Fergusson, and Horwood (2005)	Mauritius	WISC-R	636	622	11	0.39	7.01***
Flores-Mendoza, Abad, and Lelé (2005)	Brazil	DMT	295	264	7–12	–	0.567
Abdel-Khalek and Lynn (2006)	Kuwait	SPM	3278	3251	8–15	−0.08	−3.24***
Flores-Mendoza et al. (2007)	Brazil	DMT	77	62	5–7	−0.54*	–
			83	59	8–9	−0.08	–
			51	44	10–11	−0.10	–
Flores-Mendoza et al. (2007)	Brazil	CPM	106	120	5–6	0.06	–
			103	112	7–8	0.12	–
			122	114	9–10	0.12	–
			55	47	11–12	−0.08	–
Echavarri et al. (2007)	Argentina	VR	713	816	College students	0.11	2.21*
		NR	713	816	College students	0.16	3.11**
		AR	713	816	College students	0.16	3.20**
		WF	713	816	College students	−0.53	−10.23**
		VC	713	816	College students	−0.65	−12.52**
Khaleefa and Lynn (2008a)	United Arab Emir.	CPM	183	232	6	−0.38*	–
			185	215	7	−0.24	–
			180	210	8	−0.43	–
			189	238	9	−0.26	–
			190	220	10	−0.51*	–
			197	251	11	−0.53*	–
Khaleefa and Lynn (2008b)	Syria	SPM	1739	1750	7–18	−0.06	–
Schonhaut, Maggiolo, Herrera, Acevedo, and Garcia (2008)	Chile	WPPSI	32	29	3–5	−0.25	–
Gottsfritz and Alves (2009)	Brazil	R-1	66	88	16–77	0.88	2.183*
Rosseti, Rabelo, Leme, Pacanaro, and Guntet (2009)	Brazil	APM	104	265	17–63	1.22	4.21**
Lynn, Chen, and Chen (2011)	Taiwan	CPM	794	712	7	0.03	0.58
		SPM	991	915	10	−0.10	−2.58*
		SPM	948	845	13	0.10	2.11*
		SPM	604	481	16	0.16	2.61*
Liu and Lynn (2011)	China	WPPSI	728	603	3–7	0.14*	–

Note: NS: not statistically significant; *d*: negative values indicate higher female performance. Values missing indicate that the authors did not provided information. (Countries Classification can be found in http://data.worldbank.org/about/country-classifications/country-and-lending-groups).

[a] Measures are Wechsler Adult Intelligence Scale Revised (WAIS-R), Wechsler Intelligence Scale for Children Revised (WISC-R), Differential Aptitude Tests (DAT), Verbal Reasoning (VR), Numerical Reasoning (NR), Word Fluency (WF), Verbal Comprehension, Draw-a-Man-Test (DMT), Raven's Coloured Progressive Matrices (CPM), Standard Progressive Matrices (SPM), Advance Progressive Matrices (APM), Brazilian Cognitive Non-Verbal Test (R-1), Wechsler Preschool and Primary Scale of Intelligence (WPPSI), and Otis–Lennon School Ability Test (OLSAT).

* $p<0.05$.

** $p<0.01$.

*** $p<0.001$.

male problem behavior involvement during adolescence (e.g., car accidents, criminality, and drug abuse), which increases the male mortality rate when compared to females. In this study, females represent the majority of each data set, which is to be expected according to the Brazilian census. However, the female proportion of our samples is higher to the females' proportion in the overall Brazilian population (exception is AC dataset). This is due to a well-known phenomenon: female involvement in educational settings is greater than that of males. Most participating adolescents/adults for the BPR5 and the SPM were recruited from high schools, private and public universities, preparatory courses for universities and human resource selection processes for jobs related to teaching in private schools. Therefore, the broad female presence in our study was not surprising. On the other hand, there was a high proportion of girls in the children dataset, which was not expected, according to the ratio in the Brazilian population. Despite the fact that primary schools in Brazil are compulsory for all children aged 7 to 14 years and free at all public institutions, females were predominate in our representative sample of children to which the SPM was administered, suggesting that there were more males who had dropped out of school, which is consistent with national school census data (www.inep.gov.br). For instance, in 2010 a total of 1,265,930 Brazilian students finished primary schools. From this total, 53.6% were girls. Moreover, according to the Brazilian Ministry of Education, the rate of enrollment in the first two years of primary school favors males, but from then on, the rate of school enrollment favors females (Godinho et al., 2006).

Overall, the comparison between sex ratios in our samples and the sex ratios in the Brazilian population indicates that our study used a nonprobability sampling, which strengthens the hypothesis of sample selectivity. Sample selectivity indicates that there is considerable sample selection for educational orientation and deviation from population sex ratios. This phenomenon may have affected the validation samples for the most important psychological tests created for, or adapted to, the Brazilian context, such as BPR5 (Form B), WISC-III or WAIS-III in which samples were recruited from educational settings. The same could be said of the American and Spanish WAIS III standardizations in which 53.4% and 51.4%, respectively, were females (Colom et al., 2002; Wechsler, 1997). Or the American Stanford–Binet IV standardization in which the participation by females above 17 years of age was 53.01% (the high percentage of females was more evident for ages from ages 21 to 29 years) (Roid, 2003).

Due to the difficulty in arranging cognitive testing in large samples outside of educational settings, the gathering of population-representative samples has become difficult. As shown in Table 3, the samples of our study follow this pattern. The female–male ratios in our sample had a lower male presence relative to the Brazilian census. On the contrary, for the AC test, there were more males than females relative to what was expected in the census, probably due to the fact that participants were recruited in psychological

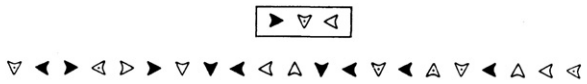

Fig. 1. Representative item from the AC test.

assessment centers authorized to assess applicants for driving licenses.

On the other hand, our adult sample included individuals who had attended only primary school (equivalent to 4th to 8th grade), high school, or university (graduation and post-graduation level degrees), and the sample contained considerable SES variance. Thus, analyzing sex differences in both variables (at educational and SES levels), was possible. In particular, the educational level of adults was scored on a three-point scale (i.e., having attended primary school, high school, or college).

2.2. Instruments

2.2.1. Cognitive measures

Three cognitive measures were used. The first cognitive measure was the *Brazilian Attention Test* [AC] (Cambraia, 2003). The AC is a Brazilian instrument for assessing attention in adolescents and adults. The test consists of several rows of figures (Fig. 1). Each figure is in a different position and is white or black, or includes a point in its center. The test taker must mark the figure equivalent to one of the three models. The total score is the number of figures correctly marked within the 5-minute time limit. According to the manual, the reliability of the AC is 0.73 (test–retest correlation). Data supporting the construct validity of the AC test were obtained by administering two additional Brazilian tests to a subsample of participants: a non-verbal reasoning test called R-1, and an attention test entitled TACOM-A. The association of the AC with the R-1 ($N = 535$) was $r = 0.483$, and with the TACOM-A ($N = 496$) was $r = 0.533$. The partial correlation, using age as the control variable, between the AC test and the R-1 and between the AC and the TACOM-A was 0.301 and 0.454, respectively. Thus, the AC test appears to share more variance with tests of attention.

The second cognitive measure was the *Standard Progressives Matrices of Raven* [SPM] (CEPA, 2001). This non-verbal test is the most frequently used test in studies of cognitive sex differences (Lynn, 1999; Mackintosh & Bennett, 2005) and is a good measure of the *g* factor (Jensen, 1998). In this study, the coefficient α (reliability) was 0.93 for the adolescent/adult sample and 0.92 for the child sample. Test takers were allowed 45 min to complete the SPM.

The third cognitive measure was the *Brazilian Cognitive Reasoning Battery* [BPR5] (Primi & Almeida, 2000). This

Table 3
Statistical description for each dataset.

Cognitive test	*N*	Year of administration	Age			Sex [%]		Male × Female[b]	
			Min–Max	Mean	SD	Male	Female	Sample	Brazil
AC	2186	2007–2009	18–78	30.05	10.9	48.5	51.4	0.94:1	0.88:1
BPR5	2076	1998–2006	13–58	20.85	8.1	45.2	54.8	0.82:1	0.94:1
SPM Raven (Adolescent/Adults)	988[a]	2006–2009	13–65	29.82	9.1	41.0	59.0	0.69:1	0.94:1
SPM Raven (Children)	2009	2009	9–11	9.94	0.7	47.5	52.5	0.90:1	1.02:1

[a] 479 participants of this sample also responded to BPR5.
[b] The male × female ratios for the Brazilian population were estimated according to age range of each studied sample.

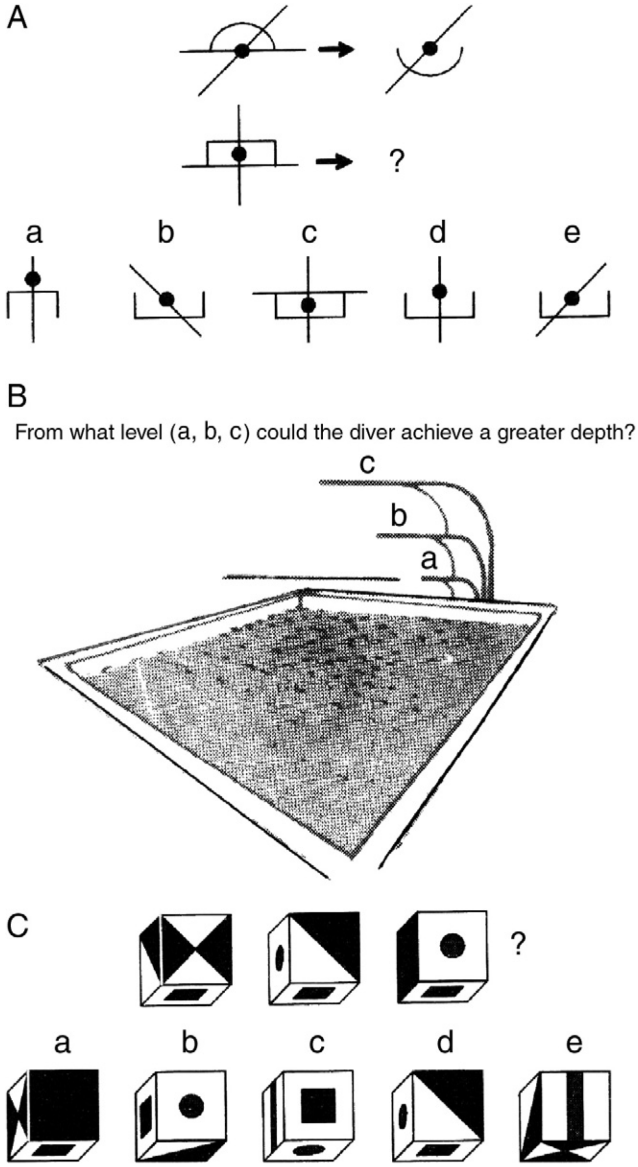

Fig. 2. a. Representative item from the Abstract Reasoning (AR) test. b. Representative item from the Mechanical Reasoning (MR) test. c. Representative item from the Spatial Reasoning (SR) test.

instrument is a Brazilian battery for cognitive assessment of adolescents and adults. The battery consists of five reasoning tests: Verbal Reasoning (VR), Numeric Reasoning (NR), Spatial Reasoning (SR), Mechanical Reasoning (MR), and Abstract Reasoning (AR). The VR takes 10 min, NR and SR 18 min each, MR 15 min, and AR 12 min. Thus, a complete and individual administration of BPR5 takes 73 min. Example items are shown in Fig. 2a, b, and c.

The BPR5 has two forms: Form A for people with only primary school education, and Form B for people with at least high school education. Form B was used in this study. Data derived from two samples: people from Minas Gerais recruited for this study, and people from São Paulo who were in the normative sample for the BPR5. Items from part of the Minas Gerais sample were available for analyzing the reliability of this battery. Coefficients α (*N*s varied from 259 to 276) were: 0.83 for VR; 0.85 for AR; 0.78 for MR; 0.82 for SR, and 0.89 for NR.

2.2.2. Socioeconomic measures

A socioeconomic questionnaire was filled out by a part of the sample from Minas Gerais. Socioeconomic status (SES) was classified by "*Criterio Brasil*", a measure based on a broad socioeconomic survey by IBOPE (*Instituto Brasileiro de Opinião Pública e Estatistica*), a multinational group that supplies broad information about the Brazilian and Latin American markets. "*Criterio Brasil*" is based on two criteria: (1) available resources at home and (2) providers' educational level (see www.abep.org.br). Points accumulated with respect to each criterion are calculated and converted into a classification scale with seven categories valid until 2007: E, D, C, B2, B1, A2, and A1. The normative sample of the São Paulo BPR-5 dataset was classified in three broad socioeconomic categories based on income and educational level of the participants[1]: high, medium and low. To connect both samples regarding this variable, the classes E, D and C were equated to low SES; classes B2 and B1 were equated to middle SES; and, finally, classes A1 and A2 were equated to high SES.

All the instruments for the MG sample were administered by the authors of the present study and by the psychologists who collaborated with the Laboratorio de Avaliação das Diferenças Individuais at the Universidade Federal de Minas Gerais.

2.3. Data analyses

Three types of analyses were performed. First, we calculated *d* indexes on scores from the four datasets (AC, BPR5, SPM adolescent/adults, and SPM children). We also estimated *d* scores for each SES and educational level of samples in the SPM and BPR5 datasets (adolescent/adults), data available only for these samples. Second, we attained data from Minas Gerais adolescents/adults ($N=479$) who answered the SPM and BPR5 tests (a total of six scores). We extracted a *g* factor using principal axis factoring (PAF) which, in contrast to principal components, permits factors to reflect the common variance of the variables, excluding unique variance. The first factor accounted for 63% of the variance. The *g* factor scores of subjects were saved and labeled $g_{1.}$ To test potential bias in the g_1 score obtained by PAF, another *g* factorial analysis excluding the Mechanical Reasoning test (a specific ability known to favor males) was performed. The variance explained by the first factor, now called $g_{2,}$ increased only slightly to 65%. For these two *g* factor scores we estimated the *d* score. In addition, because all measures employed in this study demanded cognitive abilities, all *d* scores were transformed into differences on an IQ metric in order to compare our results to published studies with similar designs.

Third, because a *g* factor obtained by PAF does not guarantee that the underlying psychological construct is the same for females and males, we use confirmatory factor analysis to test the factorial invariance of measures across groups, following procedures outlined by Widaman and Reise (1997). After establishing factorial invariance, we then evaluated mean and variance differences across groups at the latent variable level. We assumed that our sample satisfied the conditions for these analyses. These conditions include: (a) the reliability of all tests were adequate; (b) the sample employed for the present study was not selected on the basis of variables clearly related to cognitive abilities; and (c) the only test presumably biased to favor males (Mechanical Reasoning) could be accommodated in analyses.

3. Results

3.1. Sex differences on specific ability tests

Table 4 shows descriptive statistics, *d* indexes, and differences on the IQ metric between the sexes for each cognitive measure. Females outperformed males only on AC, and the difference was small in magnitude ($d=-0.13$, or 2 IQ points). Sex differences on the SPM were non-significant and varied close to zero for both child and adults samples. On the BPR5 total score, the sex difference favoring males was statistically significant ($d=0.41$, or 6.2 IQ points). When controlling for age, which affects cognitive scores (Schaie, 1994), a slightly higher difference (7.5 IQ points) was observed. The same general pattern was found on all five BPR5 subtests, with the largest differences found in Mechanical Reasoning ($d=0.89$) and Spatial Reasoning ($d=0.35$), which corresponded respectively to 13 and 5 IQ point differences favoring males.

3.2. Sex differences at the level of g

To test sex differences in *g* factor scores, we used the sample from Minas Gerais that took the SPM and the BPR5. Despite its size ($N=479$) and predominance of females (60% female), the sample was heterogeneous with regard to age (13 to 58 years), SES (E = 0.4%; D = 11.3%; C = 33.5%; B2 = 23.8%; B1 = 16.3%; A2 = 12.1%; A1 = 2.5%), race/ethnicity (White = 44.9%; Brown = 43.2%; Black = 11.9%), and educational level (primary school = 5.9%; high school completed and uncompleted = 54.3%; university completed and uncompleted = 39.8%).

Significant sex differences on the g_1 factor scores, corresponding to $d=0.25$, or 3.8 IQ points, favoring males, were found ($p<.001$). The mean sex difference decreased to a non-significant 2.7 IQ points on g_2 factor scores (excluding Mechanical Reasoning) (Table 4).

To obtain information regarding the external validity of our *g* score, we compared mean *g* scores between adult participants (≥ 18 years; $N=324$) from high and low SES, and for those with primary school and university education. We found a $d=1.38$ (21 IQ points) for high versus low SES, and a $d=-1.05$ (16 IQ points) for primary school versus university (completed and uncompleted). These results are similar to those estimated by Hart et al. (2003) and Nettle (2003). Thus, we considered the *g* scores derived from our factor analyses to have practical validity.

3.3. Sex differences at the latent variable level

Next, we used confirmatory factor analysis to evaluate sex differences at the latent variable level. To this end, we used Mplus 6.2 (Muthén & Muthén, 1998–2010). To assess overall model fit, several goodness-of-fit indices were used, none of them being unequivocally superior (Diamantopoulos & Siguaw, 2000). Models are often assessed by the χ^2/df ratio, for which there is no clear guideline about minimally acceptable values (Klen, 2000; Thompson, 2000). The thresholds for this fit index vary between 1.0 (Hair, Anderson, Tatham, & Black, 2005) and 5.0 (Kelloway, 1995; Klen, 2000), the smaller of these being the better value. The root mean square error of approximation (RMSEA) is another measure of fit, and values between 0.00 and 0.08 indicate a reasonable fit (Ackerman, 2002; Byrne, 2010; Jöreskog & Sörbom, 1993). We also used the Tucker–Lewis Index (TLI; Tucker & Lewis, 1973), the SRMR (Standardized Root Mean Squared Residual) and the CFI (Comparative Fit Index). The SRMR is sensitive to model misspecifications and is robust against violations of distributional assumptions and sample size. The TLI and CFI are measures of relative fit and penalize

[1] In the case of children samples, their parent's educational level was used. In the case of adult samples, their own educational level was used.

Table 4
d index for AC, SPM and BPR-5 (and subscales).

Measures	Male			Female			*d*	IQ
	N	Mean	SD	*N*	Mean	SD		
AC	1061	81.16	25.57	1125	84.70	26.99	−.013*	−2.0
SPM/Adult	454	49.16	9.0	534	48.25	8.98	0.10	1.5
SPM/Children	997	29.55	10.5	1102	30.32	10.0	−0.08	−1.2
BPR5	930	70.16	19.70	1146	62.43	18.11	0.41*	6.2
VR	1033	16.89	4.49	1254	16.44	4.41	0.10*	1.5
NR	1014	11.28	5.07	1215	10.32	4.86	0.19*	2.9
SR	946	12.18	4.85	1173	10.54	4.52	0.35*	5.3
AR	949	16.56	4.58	1186	15.88	4.53	0.15*	2.3
MR	944	13.51	4.89	1171	9.49	4.15	0.89*	13.4
g_1	187	0.142	0.97	292	−0.091	0.94	0.25*	3.8
g_2	187	0.104	0.96	292	−0.066	0.94	0.18	2.7

Note: AC = Brazilian Attention Test, SPM = Standard Progressive Matrices Raven, BPR5 = Brazilian Cognitive Reasoning Battery (2076 participants took all subtest, but for each subtest there was a variation between 2115 and 2287 participants), VR = Verbal Reasoning, NR = Numerical Reasoning, SR = Spatial Reasoning, AR = Abstract Reasoning, MR = Mechanical Reasoning, g_1 = g factorial score from six tests (SPM + all subtests of BPR5), g_2 = g factorial score from five tests (SPM + BPR5 without Mechanical Reasoning).
* *n.sig.* = 0.001.

model complexity (Marsh, Hau, & Wen, 2004). Smaller SRMR values and higher CFI and TLI values are better. Commonly accepted criteria for good fit are: SRMR ≤ 0.08 (Hu & Bentler, 1998, 1999) or SRMR ≤ 0.05 (Schermelleh-Engel, Moosbrugger, & Müller, 2003) and TLI and CFI > 0.95 (Hu & Bentler, 1998, 1999) or > 0.97 (Schermelleh-Engel et al., 2003).

We first fit a model to the entire sample comprising both males and females. Model 1 was a one-factor model, which fit the data fairly well (e.g., TLI and CFI > .96), but the RMSEA was above 0.08 (Table 5). We then allowed the unique factors for the SPM and the RA to covary, because these two tests share specific content, as both involve inductive reasoning with geometric figures. The resulting model, Model 2, was a more precise data fit than Model 1, $\Delta\chi^2$ (1) = 25.72, $p<.001$, and all practical fit indices were good, with an RMSEA = 0.062, and TLI and CFI > .98. The basic factor model is shown in Fig. 3, and all standardized factor loadings were very high.

Next, we fit a series of two-sample models to investigate factorial invariance across the male and female samples, following recommendations by Widaman and Reise (1997); fit indices for all models are shown in Table 5. All two-group models retained the covariance between unique factors for the SPM and the RA tests found in the one-group factor model. The first two-group model, Model 3, was a configurally invariant model, with minimal identification constraints. Specifically, the factor mean was fixed at 0 and the factor variance at 1.0 in the male sample. The first factor loading and the first intercept were constrained to invariance across groups, and all other one-factor estimates were made in each group (cf. Widaman & Reise, 1997). Model 3 fit the data well, with an RMSEA of 0.078 and TLI and CFI > .97. Next, we constrained the factor loadings to be invariant across groups, leading to Model 4, the weak factorial invariance model. Model 4 fit the data non-significantly worse than Model 3, $\Delta\chi^2$ (5) = 6.79, $p = .24$, and the RMSEA, TLI, and CFI, all improved.

Table 5
Summary of fit statistics for confirmatory factor models.

Model	χ^2	df	χ^2/df	RMSEA	TLI	CFI	SRMR
One-group models							
Model 1	53.39	9	5.93	0.088	0.961	0.977	0.027
Model 2	27.67	8	3.46	0.062	0.981	0.990	0.019
Two-group models							
Model 3: configural	47.19	16	2.95	0.078	0.970	0.984	0.024
Model 4: weak FI	53.98	21	2.57	0.070	0.976	0.983	0.043
Model 5: strong FI	116.78	26	4.49	0.104	0.946	0.953	0.066
Model 5a: strong FI	68.57	25	2.74	0.074	0.973	0.977	0.043
Model 6a: strict FI	84.40	32	2.64	0.072	0.975	0.973	0.056

Note. *N* = 640. Indices of fit are: χ^2 = minimum fit function chi-square, df = degrees of freedom, χ^2/df = ratio of chi-square to degrees of freedom, RMSEA = root mean square error of approximation, TLI = Tucker–Lewis Index, CFI = Comparative Fit Index, SRMR = standardized root mean square residual correlation. For Models 5a and 6a, the designation 'a' in the model description means that the model relaxed the invariance constraint on intercept for the MR manifest variable. See text for distinctions among the factorial invariance (FI) models.

The strong invariance model, Model 5, added invariance of the manifest variable intercepts. Model 5 fit the data significantly worse than did Model 4, $\Delta\chi^2$ (5) = 62.80, $p<.001$, and practical fit indices fell to unacceptable levels (e.g., RMSEA > 0.10; TLI and CFI < .96). Following suggestions by Byrne, Shavelson, and Muthén (1989) and Meredith and Horn (2001) for dealing with unique mean differences across groups, we freed the cross-group constraint on the intercept of the MR indicator, resulting in Model 5a. Model 5a fit the data much better than Model 5, $\Delta\chi^2$ (1) = 48.21, $p<.001$, and all practical fit indices returned to acceptable levels. The reason for this modification comes from past studies that indicated the MR test has the greater amount of unique variance as compared to others subtests (Primi & Almeida, 2000; Primi, Silva, Santana, Muniz, & Almeida, in press). Therefore sex differences in this particular subtest might be confounded with specific content factor, and not *g*. To make sure that this hypothesis was plausible, a confirmatory factor analysis with covariates was performed (Muthén & Muthén, 2010). We assumed the same model presented in Fig. 3, but with sex as a covariate (dummy code: 1 = men, and 0 = women). Two paths from the covariate were specified: one directed to the latent *g* factor and another directed to the MR test. By doing this, we could test the effect of sex mediated by *g* (first path) and the direct effect of sex that is not mediated by *g* (second path), that is, a measurement non-invariance due to differential item functioning. If the direct path is significant, it means that people

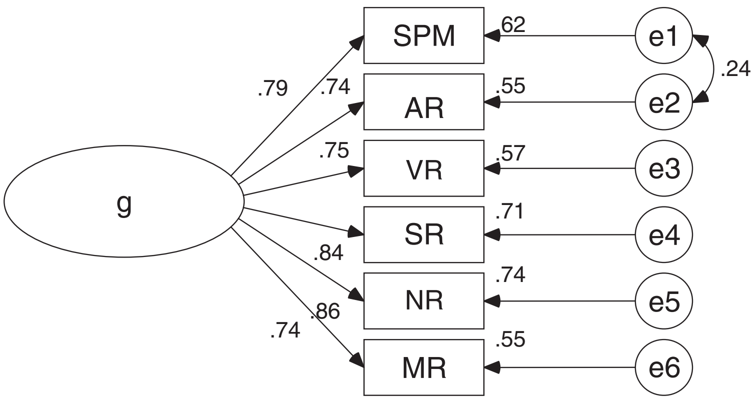

Fig. 3. CFA of *g* using reasoning measures [SPM = Standard Progressive Matrices; AR = Abstract Reasoning; VR = Verbal Reasoning; SR = Spatial Reasoning; NR = Numerical Reasoning; MR = Mechanical Reasoning].

with similar *g* latent scores differs in RM indicator due to a specific sub-dimension and not g latent factor. This model was tested six times, each one changing the direct path from sex to RM, RN, RS, RV, RA and SPM. The standardized regression weights of direct effects of sex were −.21 for MR, 0.13 for VR and 0.07 for SPM. The direct effects on the remaining subtests were not significant. This result provided strong evidence that the MR test is a complex indicator composed of a specific mechanic content factor plus *g*-spatial factor. Therefore, deriving a general factor that is more "pure" in relations to general components of reasoning, we allowed for the MR intercepts to vary across males and females in order to solve the DIF problem.

Our final model was a strict invariance model, Model 6a, that constrained all unique factor variances and covariances to invariances across groups. Although Model 6a fit the data marginally worse than Model 5a, $\Delta\chi^2$ (7) = 15.83, $p<.05$, all practical fit indices for Model 6a were approximately equal to those of Model 5a, so Model 6a was identified as the optimal model for the data.

In Model 6a, the mean difference between males and females at the latent variable level was fairly small ($d=0.23$), but significant (SE = 0.088; C.R. = 2.60; $p=0.01$) and favored males. This difference corresponds to a difference of 3.44 points in the IQ metric, which was similar to the result obtained using g_1 and g_2 factor scores.

3.4. Sex differences related to educational level and SES

Regarding educational level, differences between males and females were found on the SPM (adolescent/adults, $N=988$), but only for the group with a high school education. These differences favored males ($d=0.26$; $p=0.001$). With respect to the BPR5 ($N=2077$), significant sex differences ($p=0.001$) favoring males were found for high school ($d=0.21$) and university ($d=0.21$) graduates in total raw score without the MR subtest. Higher mean differences were found after including the MR subtest ($d=0.55$ for high school, and $d=0.46$ for University). In addition, on the AC test (attention test; $N=1061$), to which only adults had participated, significant sex differences ($p=0.001$) favoring females were found in primary school level ($d=-0.429$), as well as high school level ($d=-0.12$), but not at the university level, as these differences were significantly favorable toward males ($p<0.001$; $d=0.41$).

Regarding SES, no significant cognitive sex differences on the SPM test for any SES level were found. Conversely, on the BPR5 battery, significant differences ($p=0.001$) were found for all SES levels ($d=0.30$ for low; $d=0.21$ for middle; $d=0.22$ for high) favoring males, when the total raw score without the MR sub-test was used. But again, when the MR subtest was added, this increased the differences between males and females. Information about the participant's SES in the AC test was not available. Thus, estimation of sex differences on that variable was not possible.

4. Discussion

In the present study, *d* indexes were used to investigate cognitive sex differences in Brazilian samples. The results indicated that females outperformed males on the attention test (AC); however, males outperformed females slightly on the SPM test and moderately in all the BPR5 subtests. Considering that age effects were present for all cognitive tests, *d* indexes were calculated using *z*-scores that controlled for age. The results remained the same, with better performance by females in attention and better performance by males in all BPR subtests. The superior female performance in attention tasks is consistent with results of Anastasi (1968) and Cecilio-Fernandes and Rueda (2007).

The tendency for a slightly better performance of girls in the child sample and better performance of males in the adult sample on the SPM is consistent with Lynn's theory which proposes that females mature at an earlier age than males. However, we note that the *d* indexes for the SPM test were not statistically significant for either the adult or the child sample. If we consider the SPM test to be the strongest single-measure *g* index (Jensen, 1998), this result is inconsistent with results found by Lynn (1999, 2002) and Lynn and Irwing (2004).

On the other hand, males outperformed females on all BPR5 subtests. When these differences were converted to the IQ metric, the male advantage ranged from 1.5 to 13.4 points across subtests. The largest IQ difference in favor of males was found in Mechanical Reasoning (13.4 points). This value is quite similar to the one reported by Jensen (1998), who asserted that the difference between males and females in mechanical cognitive tasks is approximately $d=1.0$ (i.e., 15 IQ points). Regarding verbal content, Jensen (1998) and

Steinmayr et al. (2010) affirmed that females perform better in verbal fluency tasks, whereas males outperformed females in Verbal Reasoning tasks. In the present study, the male advantage on the Verbal Reasoning subtest was 1.4 IQ points. Our results showing better male performance in Numerical Reasoning and Spatial Reasoning is also consistent with other studies demonstrating that men outperform women in mathematics, spatial orientation, and visual–spatial transformations (Codorniu-Raga & Vigil-Colet, 2003; Halpern, 1997; Lubinski, 2004; Lynn, Fergusson, & Horwood, 2005). The d found in the present research for Numerical Reasoning and Spatial Reasoning, favoring men, was 0.19 (2.9 IQ points) and 0.35 (5.3 IQ points), respectively. Similar d values were identified by Colom et al. (1999), Linn and Petersen (1985), and van der Sluis et al. (2008).

When differences were estimated at the g score, we found sex differences around 3.8 IQ points, favoring males. This result is consistent with that reported by Jackson and Rushton (2006) who found, after extracting g on the Scholastic Assessment Test (SAT) in a broad sample of American students, a 3.63 IQ point difference in general intelligence in favor of males. However, in our study, the difference of 3.8 IQ points decreased to 2.7 points and was only marginally significant when Mechanical Reasoning was removed from calculation of the g factor score.

Regarding educational level, significant sex differences favoring males were found at the high school level (SPM and BPR5) and at university level (BPR5 and AC). The same pattern of results was found when the performance of the older adult sample on the SPM test and BPR5 was analyzed. Thus, in the highest educational levels, cognitive sex differences tend to favor males.

Regarding SES, Jackson and Rushton (2006) showed male–female differences in g at every level of SES favoring males. Levine et al. (2005) also found boys from middle and high-SES backgrounds outperformed girls on spatial tasks, although boys and girls from a low-SES group did not differ in their cognitive performance. In our study, significant sex differences favoring males were found for all SES levels when the comparison was carried out on the basis of summed raw scores of the BPR5, even when Mechanical Reasoning was removed. Thus, even considering apparently equal social opportunities for Brazilian males and females, cognitive sex differences favoring males at all SES levels were found.

After analyzing sex differences in g at the latent variable level, we found unaltered results, where males outperformed females (standardized difference of $d = 0.23$ or 3.4 IQ points), when the intercept of Mechanical Reasoning varied across groups. With regard to model fit statistics, Table 4 shows that, in all of our models (except Model 5), the RMSEA values were a slightly higher (over 0.05), but they still fell within the range of 0.05 to 0.08, which is considered a moderately good fit. The TLI and CFI values were clearly within the acceptable range. The SRMR values were quite small, which is considered good. Further, the ratios of χ^2 to degrees of freedom tended to fall between 2 and 3, which is fairly good, especially in large samples. It is better to assess the fit of a model with several criteria, because no single fit index, by itself, is a gold standard of fit of the model to the data. In our case, the global fit indices indicated positive evidence of a similar factor structure for both sexes, sex differences in the general factor, and invariance of g. Following the assertion of Byrne (2010) that the "*judgment rests squarely on the shoulders of the researcher*" (p. 84), and considering the previous results at the raw and g score levels, we accept that our models fit the data adequately.

In previous studies with Brazilian samples (Rindermann, Flores-Mendoza & Mansur-Alves, 2010; Rindermann, Flores-Mendoza & Woodley, 2012) similar differences were found (first study: $d = 0.20$ or 2.92 IQ points, $N = 586$, mean age around 20 years, 50% men; second study: $d = 0.10$ or 1.50 IQ points, $N = 480$, mean age around 14 years, 53% men). The pattern (large differences in older age), size and the direction of sex differences in these two better balanced samples endorse our results.

Thus, the cognitive sex differences in our samples seem to be consistent. This conclusion provokes questions such as: to what extent do our samples, in fact, represent the Brazilian male and female populations? What is the nature of measures employed in the study? What possible social consequences would there be from a mean difference of 3 IQ points favoring males? Let us offer some possible explanations.

4.1. Participants

As previously exposed, the underrepresentation of males in many studies reflects a real social phenomenon that, according to Madhyastha et al. (2009), can bias study results. For instance, the high rate of male mortality and/or dropping out of school in developing countries such as Brazil, would lead not only to a disproportionally higher enrollment rate of females at all levels of education, but would also lead to sample selection differences (as more of the brightest males would survive to attend high school or university). Clearly, this would affect results in cognitive tests when mean sex differences are compared. If this argument were correct, we would expect the male population to be cognitively more homogeneous than the female population in universities, as less brilliant males are not as highly represented at these higher levels of education.

In Brazilian standardizations of cognitive tests such as the WAIS-III (Nascimento, 2000) and the BPR5 (Primi & Almeida, 2000), males are cognitively more heterogeneous (that is, they exhibit higher variability) than females in higher education samples. Such evidence appears to be inconsistent with the hypothesis of an effect of sample selection differences. But, the preceding samples for the WAIS-III and BPR5 included participants from both public universities (which have strict entrance requirements), and private universities (which have less strict standards), and this may underlie the greater variability in the male samples.

There are few studies reporting sex differences at each educational level. In our study, cognitive differences favoring males for groups with high school (SPM and BPR) and university level (BPR5 and AC) were found. These results can be an evidence of sample selection differences, insofar as males in our study were somewhat underrepresented, which as seen in previous studies referenced, is very typical. That means that the smartest males who survived and continued their education would be present in our sample to respond to the tests. But, whether underrepresentation of males is possible to overcome, at least when a study is conducted by individual researchers, or whether this underrepresentation would have affected the direction of results, is not at all clear.

Regarding this point, let us quote two studies. The first study, conducted by Lynn and Kanazawa (2011) establishing invariance of both slopes and intercepts in a measurement model, demonstrated that females cognitively outperformed males at 7 and 11 years of age, but males outperformed females by the age of 16 years on the basis of a large-scale, prospective longitudinal study. This result was interpreted as consistent with Lynn's developmental maturation model for the emergence of male superiority after reaching maturity. And, Irwing (2012) analyzing sex differences at the latent level using American normative data of the WAIS-III, in which supposedly a representative sample of the USA was employed, found results that favored males. Thus, the presumption that low-IQ men are more likely to drop out of the surveys than low-IQ women (differential attrition) can be invoked to explain the results obtained by Irwing, but cannot explain the results obtained by Lynn and Kanazawa (2011). In the latter study, a change of cognitive profile across time, favoring males, occurred in the same sample of subjects.

Nevertheless, if the presumption of an artifact of study design is correct, then selection biases may have compromised results in virtually all psychological studies published to date, including traditional standardized tests.

As indicated previously, the possibility of differential attrition by sex exists in studies done in the Brazilian context. Specifically, in our case, we tried to collect data in diverse social settings: public and private high schools, preparatory courses to enter universities, and at private and public universities. For adults above 30 years of age, we recruited participants by means of friendship networks and recruitment centers for job applicants. In addition, Brazilian people from diverse socio-demographic characteristics, such as SES, race, and educational levels are represented in our samples. Thus, we tried to sample in a way to ensure cognitive diversity that is similar to the population distribution in Brazil. However, we recognize that our sample was not randomly drawn from the population; rather, our samples were nonprobability samples. In acknowledging that our samples were not drawn randomly from the population, we are aware that our results cannot be generalized directly to the entire Brazilian population.

4.2. Regarding measures used

All six cognitive measures used to represent *g* in this study required reasoning. The percentage of variance accounted for by the first factor was 63%, a percentage characteristic of traditional cognitive batteries. Thus, our *g* score could be considered a "strong" psychometrical *g*. Nevertheless, strong *g* does not necessarily mean a "true" *g*. According to Jensen (1998), beyond sampling error and measurement error, one must consider psychometric sampling error. For example, a strong *g* factor could be extracted from a battery of eight verbal tasks, but such a *g* factor would be overly oriented toward verbal skills, not a true *g* reflecting a broad array of mental skill. In this sense, the ideal procedure for diminishing psychometric sampling error and to approximate a "true" *g* (or general intelligence) is to administer a broad battery of tests with different cognitive demands and submit them to more sophisticated multivariate analyses (Jensen, 1998). Nevertheless, as seen in Table 1, despite using similar methods of analysis, prior studies at the *g* level supported conflicting conclusions.

Could our "reasoning *g*" extracted from six cognitive measures be a "true *g*"? No easy answer can be given. Considering individual tests, the strongest sex difference found was in Mechanical Reasoning (MR) favoring males ($d = 0.89$ or 13.4 IQ points), and the MR is well known as having a specific differential sex benefit favoring males. However, on *g* factor scores as well as on *g* at the latent variable level, we found consistent sex differences, whether or not MR was included. Our battery included only reasoning tasks. Thus, if sex differences in reasoning ability are genuine, then the more reasoning tasks a battery comprises, the more sex differences favoring adult males appear, even controlling for their influence in *g* by means of sophisticated multivariate analysis. Admittedly, some studies of reasoning tasks have found no sex differences or a tendency to favor females. However, such studies frequently analyzed samples of children or adolescents whose biological maturation was ongoing and, thus, stable cognitive sex differences would not be expected. For example, Brunner, Krauss, and Kunter (2008) analyzed the performance of 29,171 German high school students (mean age = 15.8 years) in two reasoning tasks. They found that girls outperformed boys slightly by $d = -0.09$, a value almost identical to that obtained in our broad sample of children ($N = 2099$; mean age = 9.9; $d = -0.08$) using the SPM test.

Thus, the male advantage found in our study may result from one specific aspect of the six measures used, which was reasoning ability. However, our results were not unique. Our results were in line with results reported by Steinmayr et al. (2010), who conducted three methods of analysis (sum scores, *g* score, and latent variables) and found consistent sex differences in *gf* and *gc* favoring males when nine reasoning tasks were used. Thus, the ubiquitous sex differences favoring adult males could be related to the degree of demand for reasoning abilities present in measures used for comparing the sexes. However, as previously commented, the effect from sample selection differences (if that variable is possible to be controlled) cannot be ignored.

4.3. Practical effect of sex differences

The challenges of modern society are related to processing symbols, identifying relevant information from ambiguity, and independent learning. Thus, to deal with these challenges in a successful way, one necessarily must use reasoning. If *g* is an important psychological attribute for survival in any social context or epoch, then reasoning is its main characteristic (Gottfredson, 1997; Gustafsson, 1984). In this regard, Carroll (2003) wrote: "(…) *the reality of a Fluid Reasoning factor independent of g is at least questionable*…" (p. 14).

However, to judge from statistics of industrialized countries regarding the growing female participation in the labor market and their better education, a difference of 3 or 4 points in reasoning *g* does not seem sufficient to affect female social success. This situation seems paradoxical. We know, in statistical terms, that differences favoring males necessarily lead to a greater number of males at top levels of the intelligence distribution. If females have an IQ mean of 98 and males an IQ mean of 102, for example, then 12% more males than females would be found above a score of 120. In isolation, such considerations suggest that these sex differences matter. But, if we consider the impact of this difference in *g* on the criterion of job performance, the importance tends to diminish. For instance, considering *g* and job performance as two continuous *z* score variables, with a typical correlation of 0.5 between them, a difference of 4 points in *g* (standardized difference of approximately 0.27) would predict a 0.13 standardized advantage for males (or 1.8 in IQ units) on the criterion, a very small difference. Moreover, other traits beyond high IQ might be crucial for high-level jobs, such as managers or chief-executives. Such traits might include conscientiousness (Costa, Terracciano, & McCrae, 2001), social relationships, and sensitiveness (Del Giudice, Booth, & Irwing, 2012; Heckman & Rubenstein, 2001; Kelley & Caplan, 1993) in which females have higher scores, and higher levels of these traits might compensate for the small difference in *g*.

Nevertheless, the compensation for differences in *g* might not apply in specific activities such as STEM fields (Science, Technology, Engineering, and Math), in which the concentration of males is very high and personality traits such as Conscientiousness and Agreeableness may exert little influence (Charlton, 2009). This is very close to the ideas defended in studies published by Lubinski et al. (Lubinski, 2010; Wai, Lubinski, & Benbow, 2009), that report the importance of spatial reasoning ability, beyond quantitative reasoning, for STEM domains, based on the results of a major longitudinal study of mathematically precocious youth.

Some evidence from Brazil is relevant here. Data from 2001 to 2011 from the Brazilian National Research Council (CNPq) demonstrate that an approximately equal percentage of scholarships for males and females were distributed for students at Masters, Doctorate, and Post-Doctorate levels. However, the percentage of females who obtained grants varied widely across disciplines, from 30% in Engineering and Math to 65% in Public Health and Arts. Moreover, considering the five CNPq categories (categories 1A, 1B, 1C, 1D, and 2) of researchers with productivity grants in which financial rewards vary by category, 76.5% of the recipients at the top level (1A) in 2011 were male (see http://www.cnpq.br/web/guest/estatisticas1). This tendency have not changed since 2001.

Certainly, these evidences are indirect and the hypotheses are speculative while other factors, such as gender discrimination among existing science faculty members, might underlie gender disparity in academic science (Ceci & Williams, 2010; Moss-Racusin, Dovidio, Brescoll, Graham, & Handelsman, 2012), and thus deserve further investigation. In our study, the fact is that cognitive sex differences in reasoning tasks in almost all educational and SES levels were found. If these cognitive sex differences are genuine, educational programs or research of cognitive training deserve more investment, despite skeptical results (Chein & Morrison, 2010; Chooi & Thompson, 2012; Mansur-Alves et al., in press).

The present study did not intend to provide a final answer, but to open the investigation of sex differences in *g* in developing countries, like Brazil. There is still a long way to go to attain more robust results, however consistent with many prior studies on sex differences, our results support the conclusion that modest mean differences favoring males exist when reasoning tasks are used.

Finally, because our sample was non-probabilistic, and the control of the effect of differences from the sample selection was not possible, direct generalization to the Brazilian population is not possible. Therefore other studies, with more representative samples, are necessary. However, this study is the first investigation of which we are aware, into sex differences in intelligence at the *g* level in a Latin American population. We trust others will replicate and extend the current study, investigating further the conclusions presented here.

Acknowledgments

The research referred to in this article is based on datasets from diverse projects supported by grants funded by the Foundation for Research Support of the Minas Gerais state-FAPEMIG (no. PPM 00030–09), Brazilian National Research Council — CNPq (no. 310495/2009-1), and Federal Agency of Support and Evaluation of Postgraduate Education — CAPES (no. 23072.012344/2011-84). We would like to thank to professors Aljoscha Neubauer and Tara Madhyastha for their helpful comments to previous versions of this article. In addition, we are grateful to psychologist collaborators and our graduate students: Elizabeth do Nascimento, Sabrina Barroso, Mariana Telles, Claudia Terumi, Rodrigo Silva Leite, Tatiane Dias Bacelar, Álvaro José Lelé, Carlos Guilherme Schlottfeldt, Eni Ribeiro da Silva and Marco Antonio Alvarenga, for their help in the data and references collecting.

References

Abdel-Khalek, A. M., & Lynn, R. (2006). Sex differences on the Standard Progressive Matrices and in educational attainment in Kuwait. *Personality and Individual Differences*, *40*, 175–182.

Ackerman, P. L. (2002). Gender differences in intelligence and knowledge: How should we look at achievement score differences? *Issues in Education: Contributions from Educational Psychology*, *8*, 21–29.

Aluja-Fabregat, A., Colom, R., Abad, F., & Juan-Espinosa, M. (2000). Sex differences in general intelligence defined as g among young adolescents. *Personality and Individual Differences*, *28*, 813–820.

Anastasi, A. (1968). *Psicologia diferencial [Differential psychology].* São Paulo, Brazil: E.P.U.

Andrés-Pueyo, A., & Zaro, M. J. (1998). La Inteligencia de los hombres y las mujeres [Male and female intelligence]. *Mundo Científico*, *192*, 63–77.

Baraheni, M. N. (1974). Raven Progressive Matrices as applied to Iranian children. *Educational and Psychological Measurement*, *34*, 983–988.

Brunner, M., Krauss, S., & Kunter, M. (2008). Gender differences in mathematics: Does the story need to be rewritten? *Intelligence*, *36*, 403–421.

Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (2nd ed.). New York: Routledge Taylor & Francis Group.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*, 456–466.

Cambraia, S. V. (2003). *Manual do Teste de Atenção Concentrada (AC) [Brazilian Attention Test — AC].* São Paulo, Brazil: Vetor.

Carroll, J. (2003). The higher-stratum structure of cognitive abilities: Current evidence supports g and about ten broad factors. In H. Nyborg (Ed.), *The scientific study of general intelligence: Tribute to Arthur R. Jensen* (pp. 5–22). Oxford: Elsevier Science.

Ceci, S. J., & Williams, W. M. (2010). Sex differences in math-intensive fields. *Current Directions in Psychological Science*, *19*, 275–279.

Cecilio-Fernandes, D., & Rueda, F. J. M. R. (2007). Evidência de validade concorrente para o Teste de Atenção Concentrada (TEACO-FF) [Evidences of concurrent validity for the Teste de Atenção Concentrada (TEACO-FF)]. *PSIC — Revista de Psicologia da Vetor Editora*, *8*, 167–174.

Centro Editor de Psicologia Aplicada — CEPA (). *Manual das Matrizes Progressivas de Raven — Escala Geral. Séries A, B, C, D e E [Standard progressive matrices of Raven — Brazilian standardization].* Rio de Janeiro, Brazil: CEPA.

Charlton, B. G. (2009). Why are modern scientists so dull? How science selects for perseverance and sociability at the expense of intelligence and creativity. *Medical Hypothesis*, *72*, 237–243.

Chein, J., & Morrison, A. (2010). Expanding the mind's workspace: Training and transfer effects with a complex working memory span task. *Psychonomic Bulletin & Review*, *17*, 193–199.

Chooi, W., & Thompson, L. A. (2012). Working memory training does not improve intelligence in healthy young adults. *Intelligence*, *40*, 531–542.

Codorniu-Raga, M. J., & Vigil-Colet, A. (2003). Sex differences in psychometric and chronometric measures of intelligence among young adolescents. *Personality and Individual Differences*, *35*, 681–689.

Colom, R., Garcia, L. F., Juan-Espinosa, M., & Abad, F. J. (2002). Null sex differences in general intelligence: Evidence from the WAIS-III. *The Spanish Journal of Psychology*, *5*, 29–35.

Colom, R., Juan-Espinosa, M., Abad, F. J., & García, L. F. (2000). Negligible sex differences in general intelligence. *Intelligence*, *28*, 57–68.

Colom, R., & Lynn, R. (2004). Testing the developmental theory of sex differences in intelligence on 12–18 year olds. *Personality and Individual Differences*, *36*, 75–82.

Colom, R., Quiroga, M. A., & Juan-Espinosa, M. (1999). Are cognitive sex differences disappearing? Evidence from Spanish populations. *Personality and Individual Differences*, *27*, 1189–1195.

Conrad, H. S., Jones, H. E., & Hsiao, H. H. (1933). Sex differences in mental growth and decline. *Journal of Educational Psychology*, *24*, 161–169.

Costa, P. T., Jr., Terracciano, A., & McCrae, R. R. (2001). Gender differences in personality traits across cultures: Robust and surprising findings. *Journal of Personality and Social Psychology*, *81*, 322–331.

Costenbader, V., & Ngari, S. M. (2001). A Kenya standardization of the Raven's Coloured Progressive Matrices. *School Psychology International*, *22*, 258–268.

Dai, X., & Lynn, R. (2001). Gender differences in intelligence among Chinese children. *The Journal of Social Psychology*, *134*, 123–125.

Dai, X., Ryan, J. T., & Harrington, R. G. (1991). Sex differences on the Wechsler Adult Intelligence Scale — Revised for China. *Psychological Assessment*, *3*, 282–284.

Del Giudice, M., Booth, T., & Irwing, P. (2012). The distance between Mars and Venus: Measuring global sex differences in personality. *PloS One*, *7*, e29265.

Diamantopoulos, A., & Siguaw, J. (2000). *Introducing Lisrel: A guide for the uninitiated.* London: SAGE.

Dolan, C. V., Colom, R., Abad, F. J., Wicherts, J. M., Hessen, D. J., & van de Sluis, S. (2006). Multi-group covariance and mean structure modeling of the relationship between the WAIS-III common factors and sex and educational attainment in Spain. *Intelligence*, *34*, 193–210.

Echavarri, M., Godoy, J. C., & Olaz, F. (2007). Diferencias de género en habilidades cognitivas y rendimiento académico em estudiantes universitarios [Gender differences in cognitive skills and academic performance in college students]. *Universitas Psychologica*, *6*, 319–329.

Flores-Mendoza, C., Abad, F. J., & Lelé, A. J. (2005). Análise de itens do Desenho da Figura Humana: Aplicaçao de TRI [Item analysis of Human Figure Drawnings Test: Application of IRT]. *Psicologia: Teoria e Pesquisa*, *21*, 243–254.

Flores-Mendoza, C., Mansur-Alves, M. A., Lelé, J. A., & Bandeira, D. R. (2007). Inexistência de diferenças cognitivas de gênero em crianças brasileiras [There are no sex differences on g factor and specific abilities in children from two Brazilian capitals]. *Psicologia: Reflexão e Crítica*, *20*, 499–506.

Geiser, C., Lehmann, W., & Eid, M. (2008). A note on sex differences in mental rotation in different age groups. *Intelligence*, *36*, 556–563.

Godinho, T., Ristoff, D., Fontes, A., Xavier, I., Moraes, S., & Moreno, C. E. (2006). *Trajetória da mulher na educação brasileira, 1996–2003 [Women's path in Brazilian education from 1996 to 2003].* Brasília, DF, Brazil: Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP).

Gottfredson, L. S. (1997). Why *g* matters: The complexity of everyday life. *Intelligence*, *24*, 79–132.

Gottsfritz, M. O., & Alves, I. C. B. (2009). Normas do Teste de Inteligência Não Verbal R-1 para Adultos Não Alfabetizados [R-1 Non-Verbal Intelligence Test norms for illiterate adults]. *Interação em Psicologia*, *13*(59–68), 59.

Gustafsson, J. E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence*, *8*, 179–203.

Hair, J. F., Jr., Anderson, R. E., Tatham, R. L., & Black, W. C. (2005). *Análise multivariada de dados [Multivariate data analysis].* Porto Alegre, Brazil: Bookman.

Halpern, D. (1997). Sex differences in intelligence: Implications for education. *American Psychologist*, *52*, 1091–1102.

Hart, C. L., Deary, I. J., Taylor, M. D., MacKinnona, P. L., Smith, G. D., Whalleyd, L. J., et al. (2003). The Scottish Mental Survey 1932 linked to the Midspan studies: A prospective investigation of childhood intelligence and future health. *Public Health*, *117*, 187–195.

Heckman, J. J., & Rubenstein, Y. (2001). The importance of noncognitive skills: Lessons from the GED testing program. *American Economic Review*, *91*, 145–149.

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*, 424–453.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1–55.

Hunt, E. (2011). *Human intelligence.* Cambridge, UK: Cambridge Univ. Press.

Hyde, J. S. (1981). How large are cognitive gender differences? A meta-analysis using w2 and *d*. *American Psychologist*, *36*, 892–901.

Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematical performance: A meta-analysis. *Psychological Bulletin*, *107*, 139–155.

Hyde, J. S., & Linn, M. C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin*, *104*, 53–69.

Irwing, P. (2012). Sex differences in g: An analysis of the US standardization sample of the WAIS-III. *Personality and Individual Differences*, *53*, 126–131.

Ivanovic, D. M., Leiva, B. P., Castro, C. G., Olivares, M. G., Jansana, J. M. M., Castro, V. G., et al. (2004). Brain development parameters and intelligence in Chilean high school graduates. *Intelligence*, *32*, 461–479.

Jackson, D. N., & Rushton, J. P. (2006). Males have greater *g*: Sex differences in general mental ability from 100,000 17- to 18-year-olds on the Scholastic Assessment Test. *Intelligence*, *34*, 479–486.

Jensen, A. R. (1998). *The g factor: The science of mental ability*. New York: Praeger.

Johnson, W., & Bouchard, T. (2007). Sex differences in mental abilities: *g* masks the dimensions on which they lie. *Intelligence*, *35*, 23–39.

Jöreskog, K., & Sörbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language.* New York: Erlbaum.

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.

Keith, T. Z., Reynolds, M. R., Roberts, L. G., Winter, A. L., & Austin, C. A. (2011). Sex differences in latent cognitive abilities ages 5 to 17: Evidence from the Differential Ability Scales — Second edition. *Intelligence*, *39*, 389–404.

Kelley, R., & Caplan, J. (1993). How Bell Labs creates star performers. *Harvard Business Review*, *71*, 128–139.

Kelloway, E. K. (1995). Structural equation modeling in perspective. *Journal of Organizational Behavior*, *16*, 215–224.

Khaleefa, O., & Lynn, R. (2008a). A study of intelligence in the United Arab Emirates. *Mankind Quarterly*, *49*, 58–64.

Khaleefa, O., & Lynn, R. (2008b). Sex differences on the Progressive Matrices: Some data from Syria. *Mankind Quarterly*, *48*, 345–351.

Klen, L. (2000). Structural equation modeling. In L. G. Grimm, & P. R. Yarnold (Eds.), *Reading and understanding more multivariate statistics* (pp. 227–260). Washington, DC: American Psychological Association.

Klingelhofer, E. L. (1967). Performance of Tanzanian secondary school pupils on the Raven Standard Progressive Matrices test. *Journal of Social Psychology*, *72*, 205–215.

Lehmijoki, U., & Palokangas, T. (2006). Political instability, gender discrimination, and population growth in developing countries. *Journal of Population Economics*, *19*, 431–446.

Levine, S. C., Vasilyeva, M., Lourenco, S. F., Newcombe, N. S., & Huttenlocher, J. (2005). Socioeconomic status modifies the sex difference in spatial skill. *Psychological Science*, *16*, 841–845.

Linn, M., & Petersen, A. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development*, *56*, 1479–1498.

Liu, J., & Lynn, R. (2011). Factor structure and sex differences on the Wechsler Preschool and Primary Scale of Intelligence in China, Japan and United States. *Personality and Individual Differences*, *50*, 1222–1226.

Lubinski, D. (2004). Introduction to the special section on cognitive abilities: 100 years after Spearman's (1904) "'General intelligence', objectively determined and measured". *Journal of Personality and Social Psychology*, *86*, 96–111.

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

Lynn, R. (1999). Sex differences in intelligence and brain size: A developmental theory. *Intelligence*, *27*, 1–12.

Lynn, R. (2002). Sex differences on the progressive matrices among 15–16 year olds: Some data from South Africa. *Personality and Individual Differences*, *33*, 669–673.

Lynn, R., Allik, J., Pullmann, H., & Laidra, K. (2004). Sex differences on the progressive matrices among adolescents: Some data from Estonia. *Personality and Individual Differences*, *36*, 1249–1255.

Lynn, R., Backhoff, E., & Contreras-Niño, L. A. (2004). Sex differences on *g*, reasoning and visualization tested by the progressive matrices among 7–10 year olds: Some normative data for Mexico. *Personality and Individual Differences*, *36*, 779–787.

Lynn, R., Chen, H., & Chen, Y. (2011). Intelligence in Taiwan: Progressive Matrices means and sex differences in means and variances for 6- to 17-year-olds. *Journal of Biosocial Science*, *43*, 469–474.

Lynn, R., Fergusson, D. M., & Horwood, L. J. (2005). Sex differences on the WISC-R in New Zealand. *Personality and Individual Differences*, *39*, 103–114.

Lynn, R., & Irwing, P. (2004). Sex differences on the progressive matrices: A meta-analysis. *Intelligence*, *32*, 481–498.

Lynn, R., & Kanazawa, S. (2011). A longitudinal study of sex differences in intelligence at ages 7, 11, and 16 years. *Personality and Individual Differences*, *51*, 321–324.

Lynn, R., Raine, T. A., Venables, P. H., Mednick, S. A., & Irwing, P. (2005). Sex differences on the WISC-R in Mauritius. *Intelligence*, *3*, 527–533.

Lynn, R., & Tse-Chan, P. W. (2003). Sex differences on the progressive matrices: Some data from Hong Kong. *Journal of Biosocial Science*, *35*, 145–150.
Maccoby, E., & Jacklin, C. (1974). *The psychology of sex differences.* Stanford, C.A.: Stanford University Press.
Mackintosh, N. J. (1996). Sex differences and IQ. *Journal of Biosocial Science*, *28*, 559–572.
Mackintosh, N. J., & Bennett, E. S. (2005). What do Raven's Matrices measure? An analysis in terms of sex differences. *Intelligence*, *33*, 663–674.
Madhyastha, T. M., Hunt, E., Deary, I. J., Gale, C. R., & Dykiert, D. (2009). Recruitment modeling applied to longitudinal studies of group differences in intelligence. *Intelligence*, *37*, 422–427.
Mansur-Alves, M., Flores-Mendoza, C., & Tierra-Criollo, C. J. (in pres). Evidências preliminares da efetividade do treinamento cognitivo para melhorar a inteligência de crianças escolares [Preliminary evidences of cognitive training for boosting intelligence of school children]. *Psicologia: Reflexão e Crítica*.
Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis testing approaches to setting cutoff values for fit indexes and dangers in overgeneralising Hu & Bentler's (1999) findings. *Structural Equation Modeling*, *11*, 320–341.
Meredith, W., & Horn, J. (2001). The role of factorial invariance in modeling growth and change. In L. M. Collins, & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 203–240). Washington, DC: American Psychological Association.
Morrison, C., & Jutting, J. P. (2005). Women's discrimination in developing countries: A new data set for better policies. *World Development*, *33*, 1065–1081.
Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (September 17). Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences of the United States of America*, http://dx.doi.org/10.1073/pnas.1211286109.
Muthén, L. K., & Muthén, B. O. (1998–2010). *Mplus user's guide* (6th ed.). Los Angeles, CA: Muthén & Muthén.
Nascimento, E. (2000). *Adaptação e validação do teste WAIS-III para um contexto brasileiro [Adaptation and validation of WAIS III for Brazilian context].* Brasília, DF, Brazil: Tese de Doutorado não publicada, Curso de Pós-graduação em Psicologia, Universidade de Brasília.
Neisser, U., Boodoo, G., Bouchard, T., Jr., Boykin, A. W., Brody, N., Ceci, S. J., et al. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, *51*, 77–101.
Nettle, D. (2003). Intelligence and class mobility in the British population. *British Journal of Psychology*, *94*, 551–561.
Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D., et al. (2012). Intelligence: New findings and theoretical developments. *American Psychologist*, *67*, 130–159.
Nyborg, H. (2003). Sex differences in *g*. In H. Nyborg (Ed.), *The scientific study of general intelligence* (pp. 187–222). Amsterdam: Pergamon.
Nyborg, H. (2005). Sex-related differences in general intelligence *g*, brain size, and social status. *Personality and Individual Differences*, *39*, 497–509.
Primi, R., & Almeida, L. S. (2000). *Baterias de Provas de Raciocínio (BPR-5): Manual técnico.* São Paulo: Casa do Psicólogo.
Rindermann, H., Flores-Mendoza, C., & Mansur-Alves, M. (2010). Reciprocal effects between fluid and crystallized intelligence and their dependence on parents' socioeconomic status and education. *Learning and Individual Differences*, *20*, 544–548.
Rindermann, H., Flores-Mendoza, C., & Woodley, M. A. (2012). Political orientations, intelligence and education. *Intelligence*, *40*, 217–225.
Primi, R; Silva, M. C. R.; Santana, P. R.; Muniz, M. & Almeida, L. S. (in press). *The use of the bi-factor model to test the uni-dimensionality of a battery of reasoning tests*. Psicothema (Oviedo).
Roid, G. H. (2003). *Stanford–Binet Intelligence Scales — Fifth edition, technical manual.* Itasca, IL: Riverside Publishing.
Rosseti, M. O., Rabelo, I. S., Leme, I. F. A. S., Pacanaro, S. V., & Guntet, I. B. (2009). Evidências de validade das Matrizes Progressivas Avançadas de Raven em universitários [Validity evidence of Raven's Advanced Progressive Matrices in university students]. *Psico-USF*, *14*, 177–184.
Sabogal, F., Molina, B., & MacVean, R. B. (1979). Desarrollo Fisico y Cognoscitivo de Niños Guatemaltecos, en función del nivel socioeconómico y del sexo [Physical and cognitive development of Guatemalan children, depending on socioeconomic status and sex]. *Revista Latinoamericana de Psicologia*, *11*, 229–247.
Schaie, K. W. (1994). The course of adult intellectual development. *American Psychologist*, *49*, 304–313.
Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Test of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research*, *8*, 23–74.
Schonhaut, L., Maggiolo, M. L., Herrera, M. E. G., Acevedo, K., & Garcia, M. (2008). Lenguaje e inteligencia de preescolares: Análisis de su relación y factores asociados [Analysis of the relationship and associated factors among preschooler's language and intelligence]. *Revista Chilena de Pediatria*, *79*, 600–606.
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.
Thompson, B. (2000). Ten commandments of structural equation modeling. In L. G. Grimm, & P. R. Yarnold (Eds.), *Reading and understanding more multivariate statistics* (pp. 261–283). Washington, DC: American Psychological Association.
Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, *38*, 1–10.
United Nations — UN (). *Human development report 2010 — The real wealth of nations: Pathways to human development.* New York: United Nations Development Programme.
van der Sluis, S., Derom, C., Thiery, E., Bartels, M., Polderman, T. J. C., Verhulst, F. C., et al. (2008). Sex differences on the WISC-R in Belgium and The Netherlands. *Intelligence*, *36*, 48–67.
van der Sluis, S., Posthuma, D., Dolan, C. V., de Geus, E. J. C., Colom, R., & Boomsma, D. I. (2006). Sex differences on the Dutch WAIS-III. *Intelligence*, *34*, 273–289.
Voyer, D., Voyer, S., & Bryden, M. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin*, *117*, 250–270.
Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over fifty years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology*, *101*, 817–835.
Wechsler, D. (1997). *Wechsler Adult Intelligence Scale — Third edition.* San Antonio, TX: The Psychological Corporation.
Wechsler, S. M., & Shelini, P. W. (2002). Validade do Desenho da Figura Humana para Avaliação Cognitiva Infantil [Validity of Human Figure Drawing for children's cognitive assessment]. *Avaliação Psicológica*, *1*, 29–38.
Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 281–324). Washington, DC: American Psychological Association.