Intelligence 41 (2013) 11–18

Contents lists available at SciVerse ScienceDirect

Intelligence

Sex differences on g and non-g intellectual performance reveal potential sources of STEM discrepancies

Gina C. Lemos [a,*], Francisco J. Abad [b], Leandro S. Almeida [c], Roberto Colom [b]

[a] Universidade de Évora, Portugal
[b] Universidad Autónoma de Madrid, Spain
[c] Universidade do Minho, Portugal

ARTICLE INFO

Article history:
Received 11 March 2011
Received in revised form 17 July 2012
Accepted 23 October 2012
Available online 14 November 2012

Keywords:
Sex differences
General intelligence
Cognitive abilities
STEM (science, technology, engineering and math)

ABSTRACT

The analysis of sex differences in cognitive abilities is largely confusing because these differences are masked by the pervasive influence of the general factor of intelligence (*g*). In this study a battery of five reasoning tests (abstract [AR], numerical [NR], verbal [VR], mechanical [MR], and spatial [SR]) was completed by a sample of 3233 young and old adolescents representative of the population. Using a latent variable approach, mean differences on the general factor were estimated after examining measurement invariance. Results show that the difference, favoring boys in latent *g* increases with age from two to four IQ points. Further, boys outperform girls in all the subtests and the observed differences were generally explained by *g*. However, mechanical reasoning is a systematic and strong exception to this finding. For the young adolescents, the observed difference in MR is equivalent to 10 IQ points, and this difference increases to 13 IQ points for the old adolescents. Only 1 (young) or 2 (old) IQ points of the sex difference in MR can be accounted for by *g*. The findings suggest that the persistent – and usually neglected average large advantage of boys in mechanical reasoning (MR) — orthogonal to *g* – might be behind their higher presence in STEM (science, technology, engineering, and math) disciplines. A new look at this relevant social issue is proposed in this study.

© 2012 Elsevier Inc. All rights reserved.

1. Introduction

Intelligence differences have often captured scientific and public interest (Deary, Penke, & Johnson, 2010; Jensen, 1998; Mackintosh, 1996, 1998). As underscored by Hunt and Wittmann (2008), this core psychological trait has broad implications for varied everyday life behaviors. In this regard, presumed sex differences in general intelligence (*g*) and cognitive abilities (verbal, numerical, spatial, and so forth) are of central interest (Halpern, 2000; Johnson, Carothers, & Deary, 2008; Lohman & Lakin, 2009). Setting the 20th century new "habits of mind" aside, and rejecting "to classify things as a prerequisite to understanding" (Flynn, 2010, p. 364), the approach of differentiating things to capitalize on their differential utility is worthy for society. Therefore, studies on intelligence tests, and also on sex differences, will continue to gather an understanding with educational, political, social, and ethical values (Rindermann, 2007).

Since Terman's (1916) finding of irrelevant sex differences in IQ on a sample of nearly one thousand American 4–16 year olds, the conclusion has been repeatedly asserted in terms of a negligible better performance of girls in IQ. Further research showed no significant sex differences on either IQ or *g* and in two main general cognitive abilities (fluid-abstract and crystallized-verbal intelligence) (Brody, 1992; Colom, García, Juan-Espinosa, & Abad, 2002; Colom, Juan-Espinosa, Abad, & Garcia, 2000; Deary, Irwing, Der, & Bates, 2007; Deary, Thorpe, Wilson, Starr, & Whalley, 2003; Halpern, 2000; Jensen, 1998; Lubinski, 2000; Mackintosh, 1996; Strand, Deary, & Smith, 2006).

* Corresponding author at: Departamento de Psicologia, Universidade de Évora, Apartado 94, 7002-554 Évora, Portugal. Tel.: +351 912299077(voice).
E-mail address: ginalemos@uevora.pt (G.C. Lemos).

0160-2896/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.intell.2012.10.009

This general picture changed when specific cognitive abilities were considered (Reynolds, Keith, Ridley, & Patel, 2008). The literature yielded some consensus regarding significant sex differences on three core cognitive abilities, designed by Hyde (1990) as the "Holy Trinity", namely: verbal, visuospatial, and quantitative abilities (e.g. Carroll, 1992; Maccoby & Jacklin, 1974; Snow & Lohman, 1989). Halpern et al. (2007) reported a comprehensive review of the available scientific evidence to conclude that: (a) by the end of grade school and beyond, females tend to excel in verbal abilities, especially when assessment includes writing and language-usage items (e.g. word fluency, read, speed articulation, verbal analogies); (b) males outperform females on most measures of visuospatial abilities (e.g. mental rotation, spatial perception) and in quantitative abilities (e.g. geometry, problem-solving); and (c) sex differences in visuospatial and quantitative abilities are smaller for the mid-range of the ability distribution than for those with the highest levels of ability, because of the higher variability of male performance in visuospatial and quantitative abilities. This larger variability of male performance on cognitive tests, as well as a large percentage of male adolescents among high-scoring individuals, was found by Hedges and Nowell (1995) when reanalyzing data from six studies done between 1960 and 1992 with national representative samples in the United States. These male differences in variability and frequency in the upper tail were not found in reading comprehension, perceptual speed, and associative memory only. Moreover, all of the reported effect sizes were small except for abilities associated with vocational aptitude scales (e.g. mechanical reasoning), in which the average and top 10% of boys had much better performance than the average and top 10% of girls (Hedges & Nowell, 1995). Findings from the Lubinski and Benbow (2006) 20-year follow-up of longitudinal studies pointed out boys' overrepresentation in the upper tails, namely in math–science abilities and STEM fields.

Likewise, a wide body of research reviewed by Willingham and Cole (1997), including 24 large data sets, led to the conclusion that sex differences are small in elementary school grades ($d>.2$ favoring females at 4th grade in writing, language use, and reading). These differences become larger by the end of high school, and again girls outperform boys (e.g. writing, d between .5 and .6; language usage, d between .4 and .5). The comprehensive meta-analysis by Hedges and Nowell (1995) underscored males' disadvantage in verbal abilities by the middle to end of secondary school.

With respect to visuospatial ability, several meta-analyses (e.g. Linn & Petersen, 1985; Masters & Sanders, 1993; Voyer, Voyer, & Bryden, 1995) reported the largest and most consistent sex difference in the cognitive literature, although the effect size differed depending on the task. For example, chronometric mental rotation tasks showed effect sizes around .37; two-dimensional mental rotation tasks revealed effect sizes between .31 and .44; and complex three-dimensional mental rotation tasks found effect sizes between .70 and .95, always favoring males.

Several meta-analytic studies (e.g. Geary, 1996; Hyde, Fennema, & Lamon, 1990; Willingham & Cole, 1997) reported small sex differences in quantitative abilities in elementary school (until 4th grade) where girls outperform boys. This difference is almost zero in the remaining primary-school grades. Afterwards, a small male advantage through higher secondary-school grades is detected (d between .1 and .2). The nature of the task involved is again crucial for understanding these sex differences: in early elementary school years, quantitative ability is measured mainly through computational tasks, at which girls outperform boys; as we go through higher secondary-school grades, mathematical concepts require more reasoning abilities and become more spatial in nature (e.g. problem solving in geometry and calculus), which favors boys.

Various explanations for sex differences in cognitive abilities have been delivered, ranging from biological factors (e.g. Benbow, 1988; Hooven, Chabris, Ellison, & Kosslyn, 2004; Kimura, 2002; Lynn, 1994, 1999, 2001; Lynn, Allik, & Must, 2000) to those more socially rooted (e.g. Baenninger & Newcombe, 1995; Guiso, Monte, Sapienza, & Zingales, 2008; Hyde & Linn, 2006; Quaiser-Pohl, Geiser, & Lehmann, 2006) but these are mainly interactive (e.g. Ceci & Williams, 2010; Halpern et al., 2007). Regardless of the ultimate cause, theoretical accounts stressing the nature of the task, along with the involved cognitive processes, deserve close attention. In this respect, it becomes pertinent to note that observed sex differences may be masked by the variance explained by the general factor of intelligence or *g* (Johnson & Bouchard, 2007), or even by choices in curriculum made by boys and girls (Calvin, Fernandes, Smith, Visscher, & Deary, 2010). Interesting methodological issues have been discussed regarding different approaches (e.g. sum scores, factor score estimates, latent variables) for understanding results derived from studies investigating sex differences in intelligence (Steinmayr, Beauducel, & Spinath, 2010). Sex differences in STEM fields are also discussed regarding non-cognitive variables. Life values, personality dimensions, and vocational interests have been considered in the last decades for framing global sex differences and also for explaining why males are overrepresented on STEM's graduation and careers (Del Giudice, Booth, & Irwing, 2012; Ferriman, Lubinski, & Benbow, 2009; Su, Rounds, & Armstrong, 2009). These non-cognitive variables might play a role in adolescence when students start to explore and make vocational choices. These can also help to explain why sex differences on verbal, spatial and, quantitative abilities become more evident at junior and senior high school.

In this study we analyze two large, representative samples of young and old adolescents from Portugal totalling to 3233 participants. These samples completed an intelligence battery comprising five reasoning tests (abstract, numerical, verbal, mechanical, and spatial reasoning). Because of the pervasive influence of *g*, average sex differences are computed including and excluding *g* variance from the five tests. This would provide straightforward findings regarding the presence or absence of average sex differences in specific cognitive abilities.

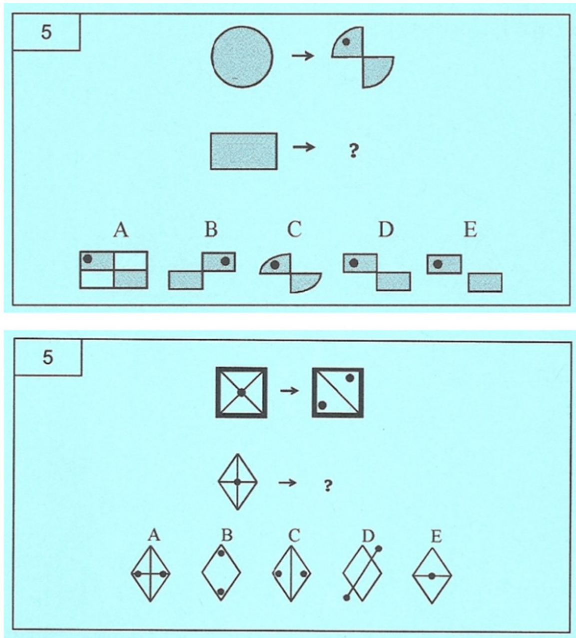
2. Method

2.1. Participants

Two independent samples comprising 3233 students (1564 boys and 1669 girls) were considered. The first sample included 1714 students in the third cycle of elementary school (young adolescents; mean age = 13.5, SD = .96, range 12–15 years), of whom 828 were boys and 886 were girls. The second sample included 1519 secondary schools students (old adolescents; mean age = 16.9, SD = .88, range 16–19 years), of whom 736 were boys and 783 were girls.

The participants drawn were nationally (Portuguese) representative and randomly selected for the Reasoning Test Battery (RTB; Almeida, 2003; Almeida & Lemos, 2007). According to the annual school census of the Department of Assessment and Foresight and Planning – Ministry of Education – the samples gathered 6% of the Portuguese student population in the considered school levels. Schools were selected based on these criteria and samples were stratified across regions in the country, school grade, and sex

Example of Abstract Reasoning items from the Junior and Senior High School Battery, respectively



Example of Numerical Reasoning items from the Junior and Senior High School Battery, respectively

10. 26 27 29 30 33 34 38 39 ? ?

10. 26 27 29 30 33 34 38 39 ? ?

Example of Verbal Reasoning items from the Junior and Senior High School Battery, respectively

11. Wing is to air as wheel is to...

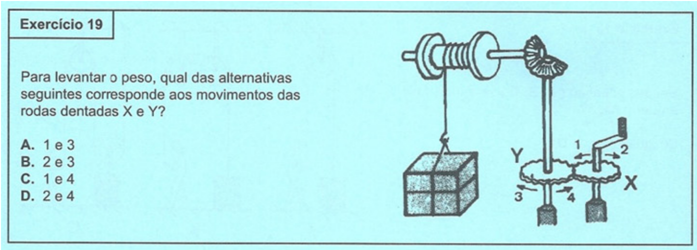
A. Tire B. Car C. Axis D. Road E. Speed

11. Door is to home as title is to...

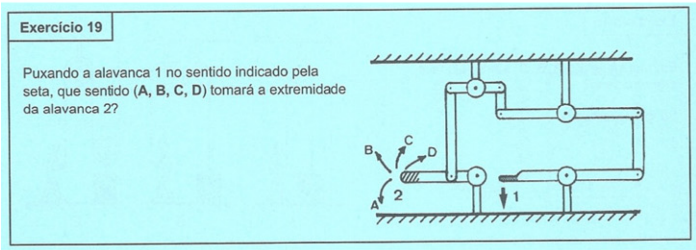
A. Entry B. Book C. Chapter D. Preface E. Cover

Fig. 1. Example items from the administered intelligence battery.

Example of Mechanical Reasoning items from the Junior and Senior High School Battery, respectively

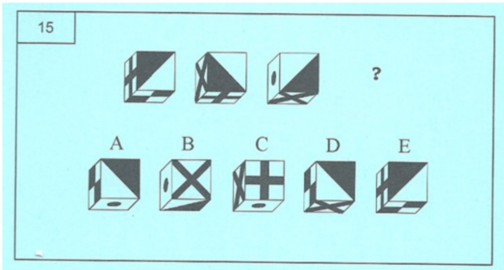


To lift the weight, which of the following matches the movement of sprockets X and Y?
A. 1 and 3 B. 2 and 3 C. 1 and 4 D. 2 and 4



When pulling the lever 1 in the direction indicated by the arrow, which way (A, B, C, D) will the end of the lever 2 take?

Example of Spatial Reasoning items from the Junior and Senior High School Battery, respectively



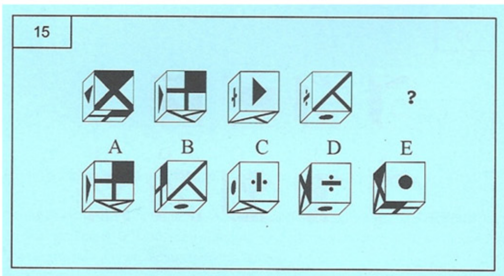


Fig. 1 (*continued*).

within the class group at each school level. Since classes are heterogeneously organized, students were selected through random selection of classes within schools.

The school system in Portugal consists of three cycles in elementary school and one cycle in secondary school. This research takes students from the 3rd cycle of elementary school, equivalent to junior high school in other countries (7th to 9th grades), and from secondary school (10th to 12th grades), when students choose among several curricular options in order to specialize in different undergraduate subjects or their chosen professionalism. The first school level corresponds to the first sample and the second level matches the second sample. The secondary school sample was also stratified across gender and curricular options in order to ensure its representativeness due to an asymmetric gender distribution, as far as the domains of sciences and of the humanities are concerned.

2.2. Measures

Intelligence was assessed through the Reasoning Test Battery (RTB). The young adolescents performed the version designed for the first level (third cycle of elementary school) and the old adolescents performed the version designed for the second level (senior high school battery). Examples are depicted in Fig. 1.

The RTB consists of five reasoning time-limited subtests: abstract reasoning (AR, 25 figural analogies and 5 min of administration time), numerical reasoning (NR, 20 numerical series and 10 min of administration time), verbal reasoning (VR, 25 verbal analogies and 4 min of administration time), mechanical reasoning (MR, 25 mechanical problem-solving items and 8 min of administration time), and spatial reasoning (SR, 20 spatial orientation and cube rotation series and 9 min of administration time).

Reliability indices were computed by test–retest and internal consistency methods. Obtained coefficients ranged from .63 (mechanical reasoning subtest) to .84 (numerical reasoning subtest). Factor analyses computed from different samples confirmed a single factor explaining between 50 and 60% of the variance (Almeida & Lemos, 2007). Further, Almeida, Guisande, Primi, and Lemos' (2008) analyses suggest that, for the old adolescents, the mechanical and spatial reasoning subtests require a visualization factor (gv) as defined by the CHC theory (Carroll, 1993; McGrew, 2009).

2.3. Procedures

A team of psychologists was trained for RTB administration. Before administration itself, strictly adhering to conditions specified in the battery's manual, participants are acquainted with the study's main ethical concerns, such as data confidentiality, as well as the relevance of their contribution for an investigation of this nature, outlining their role representing other students. The Portuguese Secretary of Education authorized this administration. The RTB was completed by participants in classes with no more than 25 students during normal teaching hours, with the agreement of teachers.

A multi-group confirmatory factor model with mean-structures was fitted for obtaining *g*-latent mean sex differences. In the model, a general factor of intelligence (*g*) predicts the five measures: abstract reasoning (AR), numerical reasoning (NR), verbal reasoning (VR), mechanical reasoning (MR), and spatial reasoning (SR). In order to probe these analyses valid, measurement invariance across sex was tested. For each sample, the same sequence of steps was followed. The first step was to establish that the same structure was present at several testing occasions (configural invariance). Then, equality of loadings, intercepts and residual variances were progressively tested. Fit of the models was assessed using the root mean square error of approximation (RMSEA) and the comparative fit index (CFI). Values close to .95 for CFI and below .06 for RMSEA suggest a good fit (Hu & Bentler, 1999). Because the progressively restrictive tests of measurement invariance create nested models, a chi-square statistic was obtained to test statistical differences between them. This statistical criterion is too sensitive to sample size and was complemented with ΔCFI as a "practical criterion", setting a critical value of .01 to reject a null hypothesis of invariance (Cheung & Rensvold, 2002). When the results suggested non-invariance, modification indices were used sequentially in order to relax the equality constraints and detect sources of invariance.

3. Results

Results for the analysis of invariance in each sample are shown in Table 1. Baseline models show a good fit and partial invariance may be supported. All the loadings and most of the intercepts and residual variances are invariant across sex. For the young sample, the exceptions are the intercept of MR and the residual variance of NR. For the old sample, the exceptions are the intercept of MR and NR and the residual variance of MR. This means that observed sex differences in these subtests are not necessarily explained by differences in the general latent factor (*g*). Thus, these parameters are let free across sex in order to obtain an unbiased measure of the *g* latent difference.

Table 1
Models tested for checking invariance of factorial structure.

Model	χ^2	gl	CFI	RMSEA	SRMR	Nested model comparison	χ^2	gl	p	ΔCFI
Young adolescents										
M0: configural invariance	25.93	10	.993	.043	.016					
M1: equal factor loadings	29.74	14	.993	.036	.025	M1–M0	3.81	4	.432	<.01
M2: equal intercepts	232.11	18	.900	.118	.073	M2–M1	202.37	4	.000	>.01
M2*: M2 excluding MR	48.93	17	.985	.047	.027	M2*–M1	19.19	3	.000	<.01
M3: equal residual variances	97.86	22	.964	.063	.063	M3–M2*	48.93	5	.000	>.01
M3*: M3 excluding NR	67.40	21	.978	.051	.056	M3*–M2*	18.47	4	.001	<.01
Old adolescents										
M0: configural invariance	16.21	10	.996	.029	.014					
M1: equal factor loadings	23.51	14	.994	.030	.032	M1–M0	7.3	4	.121	<.01
M2: equal intercepts	243.54	18	.856	.128	.089	M2–M1	220.03	4	.000	>.01
M2*: M2 excluding MR	66.38	17	.968	.062	.037	M2*–M1	42.87	3	.000	>.01
M2**: M2 excluding MR, NR	39.28	16	.985	.044	.027	M2**–M1	15.772	2	.000	<.01
M3: equal residual variances	79.82	21	.962	.061	.080	M3–M2**	40.54	5	.000	>.01
M3*: M3 excluding MR	55.78	20	.977	.049	.048	M3*–M2**	16.50	4	.002	<.01

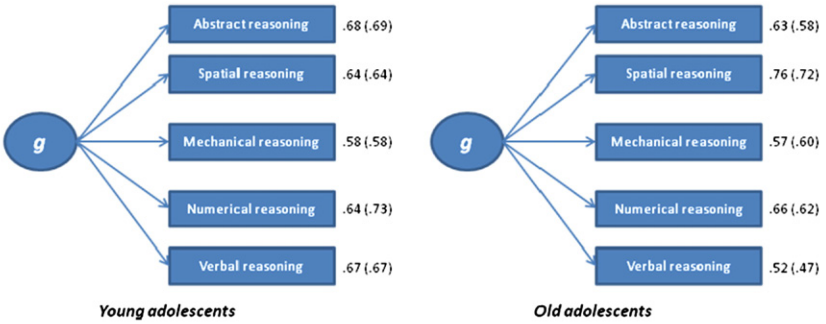


Fig. 2. Confirmatory models for the young and old adolescents. Regression weights are represented on the left (values in parenthesis are for girls).

Fig. 2 shows the within-group standardized loadings of the five subtests comprised in the battery for the final model (M3*).

Almost all the weights depicted in Fig. 2 are above .50. Nonetheless, it is worth mentioning that the factor loadings are not very large by normal standards. For the young adolescents, the largest values (>.65) are for VR, AR, and NR. NR loading was smaller for boys (as expected from the residual variance difference). For the old adolescents, NR and SR showed the highest values. Loadings are slightly higher for boys, as *g* variance was larger in this group (the exception was MR, with higher residual variance for boys). Note also that MR showed the lowest weight for the young adolescents, whereas VR showed the lowest weight for the old adolescents.

Table 2 shows observed subtest differences and *g* latent differences for both samples (as estimated by the M3* model). Significance tests were computed for testing the equality of *g* variance by comparing nested models (with and without equal variance constraint). The same strategy was used for testing the equality of *g* means (conditioned to the *g*-equality variance test result).

Equality of *g* variance can be maintained for the young group (χ^2 (1) = .13; p = .718) but not for the older group (χ^2 (1) = 6.31; p = .012) and average differences in the general factor of intelligence (*g*) are significant for young and old adolescents (p < .05; young group: χ^2 (1) = 5.26; p = .022; old group: χ^2 (1) = 20.74; p < .001). Boys outperform girls in both samples, but the difference increases with age from two to four IQ points.

Furthermore, according to the observed latent *g*-mean difference boys perform better in all the tests, but the differences are negligible for almost all the tests (and almost null for the young group). Most of the small subtest differences where due to *g*. One important exception to this pattern was mechanical reasoning (MR) for both groups, and numerical reasoning (NR) for the older group, in which boys show higher scores. These large differences were specific for these subtests and not due to *g* (see Table 2). By a large amount, the greatest difference is found for mechanical reasoning (MR).

4. Discussion

By analyzing large samples of adolescents representative of the population, we have shown that answers to the question of whether or not there are average sex differences in specific cognitive abilities are largely dependent on the variance accounted for by the general factor of intelligence (*g*). For the young adolescents, almost all the average differences favoring boys at the subtest level are accounted for by *g*. For the old adolescent this happens in three out of the five subtests. Nevertheless, mechanical reasoning (MR) is

Table 2
Model observed and latent differences for the M3* model.

		Observed AR	NR	VR	MR	SR	Latent *g*
Young							
Boys	Mean	12.77	8.77	14.14	10.34	9.72	0.00
	SD	3.20	3.96	3.69	3.00	4.01	1.00
Girls	Mean	12.49	8.45	13.83	8.33	9.39	−0.13
	SD	3.22	3.52	3.72	3.01	4.03	1.02
d		0.09	0.09	0.09	0.67	0.08	0.13
IQ difference		1.30	1.29	1.28	10.02	1.22	1.90
Expected IQ difference due to g		1.30	1.29	1.28	1.10	1.22	
Old							
Boys	Mean	12.14	10.00	15.70	10.75	10.33	0.00
	SD	2.78	3.60	3.34	3.36	3.57	1.00
Girls	Mean	11.67	8.50	15.23	8.04	9.59	−0.27
	SD	2.66	3.43	3.24	2.83	3.34	0.89
d		0.17	0.43	0.14	0.87	0.21	0.29
IQ difference		2.59	6.41	2.12	13.11	3.19	4.29
Expected IQ difference due to g		2.59	2.76	2.12	2.50	3.19	

the important exception to this rule. The large average difference favoring the young boys (10 IQ points) cannot be explained by *g*, and the same occurs for the great average difference favoring old boys (13 IQ points). In this latter sample, the average difference favoring boys for the numerical reasoning (NR) subtests (6.4 IQ points) is only partly accounted for by *g*.

With respect to the general factor of intelligence assessed by this battery, the main finding is consistent with the developmental theory proposed by Lynn (1999). Young boys outperform girls by 2 IQ points, whereas this average difference increases to 4 IQ points for the old adolescents. The theory is based on the fact that boys and girls mature at different rates. The growth of girls accelerates at the age of about 9 years and remains in advance of boys until 14–15 years. At 15–16 years the growth of girls decelerates relatively to that of boys, who continue to grow. Colom and Lynn (2004) showed that this theory is consistent with developmental IQ data from Britain, the United States of America, and Spain. The results reported here support the theory from data of other countries. Nevertheless, non-cognitive variables, such as personality or vocational interests, might also be related to these sex differences (Del Giudice et al., 2012; Ferriman et al., 2009; Su et al., 2009).

Beyond the observed small average sex difference in the general factor of intelligence (*g*), the boys' large advantage in mechanical reasoning (MR) must be strongly underscored. This sex difference is not explained by *g*, and therefore the probable contributions of what is measured by relevant subtests such as abstract reasoning (AR) or spatial relations (SR) can be excluded. The MR difference is still present with almost the same magnitude when the general factor of intelligence (*g*) is removed. It is also noteworthy that, for the old adolescents, more than half of the variance associated with numerical reasoning (NR) cannot be attributed to *g*. Thus, we suggest that mental processes captured by these psychological measures are behind the documented male advantage in STEM disciplines (science, technology, engineering, and math).

The exhaustive review by Halpern et al. (2007) failed to find a clear-cut conclusion regarding the disproportionate presence of males on these disciplines, maybe because they did not consider this sort of mechanical measures. Colom and Lynn (2004) also found an average difference equivalent to 12 IQ points on the mechanical reasoning (MR) subtest from the DAT battery in 18 year olds, a result clearly consistent with the ones reported here. Overall, the present findings underline the relevance of this type of control as it has been discussed elsewhere (Aluja, Colom, Abad, & Juan-Espinosa, 2000; Colom, Contreras, Botella, & Santacreu, 2002; Colom et al., 2000; Dolan et al., 2006), and highlighted by Hedges and Nowell (1995): “areas not generally taught in school such as mechanical comprehension and other vocational aptitudes” (p. 45).

Whereas too much is said with respect to verbal, numerical, and spatial abilities (Colom, Contreras, Arend, García-Leal, & Santacreu, 2004; Geary, 1996; Hedges & Nowell, 1995; Hyde et al., 1990; Linn & Petersen, 1985; Masters & Sanders, 1993; Reynolds et al., 2008; Voyer et al., 1995; Willingham & Cole, 1997) it might well be the case that the explanation for the intriguing case analyzed by reports such as that by Halpern et al. (2007) lies in alternative places. The findings revealed in the present study support this possibility. The small average sex difference found for abstract reasoning (AR) or spatial relations (SR) invites a closer look at the mental processes specifically involved in mechanical reasoning (MR) measures. Whereas the general picture supports the pervasive influence of the general factor of intelligence (*g*) for explaining most of the average performance differences separating boys and girls, mental processes tapped by the considered mechanical measurement have little to do with this *g* factor.

Research regarding spatial dynamic tasks exemplifies the type of analyses we are suggesting and it indicates that the same pattern might emerge from mechanical reasoning tasks (Law, Pellegrino, & Hunt, 1993). ‘Dynamic’ refers to the prediction of where a moving object is going and when it will arrive at its predicted destination. Several studies analyzed sex differences in these dynamic spatial tasks, controlling various candidates for explaining the average performance difference favoring males which is persistently observed, such as educational options differentially selected by the sexes or very specific performance factors comprised in the dynamic tasks themselves, or even presumed relevant covariates, such as static spatial performance. Virtually all the potential sources of variance which were controlled failed to turn the difference to the point of non-significance (Colom, Contreras, Shih, & Santacreu, 2003; Colom, Contreras, et al., 2002; Colom et al., 2004; Contreras, Colom, Shih, Alava, & Santacreu, 2001; Contreras, Rubio, Peña, Colom, & Santacreu, 2007). These results open the door to a probable genuine sex difference in mental processes specifically related to the above mentioned STEM disciplines, beyond much more popular measures of intellectual performance.

References

Almeida, L. S. (2003). *Bateria de Provas de Raciocínio.* Braga: Universidade do Minho.

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

Almeida, L. S., & Lemos, G. C. (2007). *Bateria de Provas de Raciocínio: Manual.* Braga: Universidade do Minho.

Aluja, 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.

Baenninger, M., & Newcombe, N. (1995). Environmental input to the development of sex-related differences in spatial and mathematical ability. *Learning and Individual Differences*, *7*, 363–379.

Benbow, C. P. (1988). Sex differences in mathematical reasoning ability among the intellectually talented: Their characteristics, consequences, and possible explanations. *The Behavioral and Brain Sciences*, *11*, 169–183.

Brody, N. (1992). *Intelligence* (2nd ed.). San Diego, CA: Academic Press.

Calvin, C. M., Fernandes, C., Smith, P., Visscher, P. M., & Deary, I. J. (2010). Sex, intelligence and educational achievement in a national cohort of over 175,000 11-year-old schoolchildren in England. *Intelligence*, *38*, 424–432.

Carroll, J. B. (1992). Cognitive abilities: The state of art. *Psychological Science*, *3*, 266–270.

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

Ceci, S. J., & Williams, W. M. (2010). *The mathematics of sex: How biology and society conspire to limit talented women and girls.* New York: Oxford University Press.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*(2), 233–255.

Colom, R., Contreras, Mª. J., Arend, I., García-Leal, O., & Santacreu, J. (2004). Sex differences in verbal reasoning are mediated by sex differences in spatial ability. *Psychological Record*, *54*, 365–372.

Colom, R., Contreras, Mª. J., Botella, J., & Santacreu, J. (2002). Vehicles of spatial ability. *Personality and Individual Differences*, *32*, 903–912.

Colom, R., Contreras, Mª. J., Shih, P. C., & Santacreu, J. (2003). The assessment of spatial ability through a single computerized test. *European Journal of Psychological Assessment*, *19*, 92–100.

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

Colom, R., Juan-Espinosa, M., Abad, F., & Garcia, 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.

Contreras, Mª. J., Colom, R., Shih, P., Alava, Mª. J., & Santacreu, J. (2001). Dynamic spatial performance: Sex and educational differences. *Personality and Individual Differences*, *30*, 117–126.

Contreras, Mª. J., Rubio, V., Peña, D., Colom, R., & Santacreu, J. (2007). Sex differences in dynamic spatial ability: The unsolved question of performance factors. *Memory & Cognition*, *35*, 297–303.

Deary, I. J., Irwing, P., Der, G., & Bates, T. C. (2007). Brother–sister differences in the *g* actor in intelligence: Analysis of full, opposite-sex siblings from the NLSY1979. *Intelligence*, *35*, 451–456.

Deary, I. J., Penke, L., & Johnson, W. (2010). The neuroscience of human intelligence differences. *Nature Reviews Neuroscience*, *11*, 201–211.

Deary, I. J., Thorpe, G., Wilson, V., Starr, J. M., & Whalley, L. J. (2003). Population sex differences in IQ at age 11: The Scottish Mental Survey 1932. *Intelligence*, *31*, 533–542.

Del Giudice, M., Booth, T., & Irwing, P. (2012). The distance between Mars and Venus: Measuring global sex differences in personality. *PLoS One*, *7*(1), 1–8.

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

Ferriman, K., Lubinski, D., & Benbow, C. P. (2009). Work preferences, life values, and personal views of top math/science graduate students and the profoundly gifted: Developmental changes and sex differences during young adulthood and parenthood. *Journal of Personality and Social Psychology*, *97*, 517–532.

Flynn, J. R. (2010). The spectacles through which I see the race and IQ debate. *Intelligence*, *38*, 363–366.

Geary, D. C. (1996). Sexual selection and sex differences in mathematical abilities. *The Behavioral and Brain Sciences*, *19*, 229–284.

Guiso, L., Monte, F., Sapienza, P., & Zingales, L. (2008). Culture, gender, and math. *Science*, *302*, 1164–1165.

Halpern, D. F. (2000). *Sex differences in cognitive abilities.* Mahwah, NJ: Lawrence Erlbaum.

Halpern, D. F., Benbow, C., Geary, D. C., Gur, R. C., Hude, J. S., & Gernsbacher, M. A. (2007). The science of sex differences in science and mathematics. *Psychological Science in the Public Interest*, *8*, 1–51.

Hedges, L. V., & Nowell, A. (1995). Sex differences in mental scores, variability, and numbers of high-scoring individuals. *Science*, *269*, 41–45.

Hooven, C. K., Chabris, C. F., Ellison, P. T., & Kosslyn, S. M. (2004). The relationship of testosterone to components of mental rotation. *Neuropsychologia*, *42*, 782–790.

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

Hunt, E., & Wittmann, W. (2008). National intelligence and national prosperity. *Intelligence*, *36*, 1–9.

Hyde, J. S. (1990). Meta-analysis and the psychology of gender differences. *Signs*, *16*, 55–73.

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

Hyde, J. S., & Linn, M. C. (2006). Gender similarities in mathematics and science. *Science*, *314*, 599–600.

Jensen, A. R. (1998). *The g factor.* Westport, CT: Praeger.

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

Johnson, W., Carothers, A., & Deary, I. J. (2008). Sex differences in variability in general intelligence: A new look at the old question. *Perspectives on Psychological Science*, *3*, 518–531.

Kimura, D. (2002). Sex, sexual orientation and sex hormones influence human cognitive function. *Current Opinion in Neurobiology*, *6*, 259–263.

Law, D. J., Pellegrino, J. W., & Hunt, E. B. (1993). Comparing the tortoise and the hare. Gender differences and experience in dynamic spatial reasoning tasks. *Psychological Science*, *4*, 35–40.

Linn, M. C., & Petersen, A. C. (1985). Emergence and characterisation of gender differences in spatial abilities: A meta-analysis. *Child Development*, *56*, 1479–1498.

Lohman, D. F., & Lakin, J. (2009). Consistencies in sex differences on the cognitive abilities test across countries, grades, test forms, and cohorts. *British Journal of Educational Psychology*, *79*, 389–407.

Lubinski, D. (2000). Scientific and social significance of assessing individual differences. *Annual Review of Psychology*, *51*, 405–444.

Lubinski, D., & Benbow, C. P. (2006). Study of mathematically precocious youth after 35 years: Uncovering antecedents for the development of math–science expertise. *Psychological Science*, *1*(4), 316–345.

Lynn, R. (1994). Sex differences in brain size and intelligence: A paradox resolved. *Personality and Individual Differences*, *17*, 257–271.

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

Lynn, R. (2001). 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., & Must, O. (2000). Sex differences in brain size, stature and intelligence in children and adolescents: Some evidence from Estonia. *Personality and Individual Differences*, *29*, 555–560.

Maccoby, E., & Jacklin, C. (1974). *The psychology of sex differences.* Stanford, CA: Stanford University Press.

Mackintosh, N. J. (1996). Sex differences and IQ. *Journal of Biosocial Science*, *28*, 559–572.

Mackintosh, N. J. (1998). *IQ and human intelligence.* Oxford: Oxford University Press.

Masters, M. S., & Sanders, B. (1993). Is the gender difference in mental rotation disappearing? *Behavior Genetics*, *23*, 337–341.

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Studying on the shoulders of the giants of psychometric intelligence research. *Intelligence*, *37*, 1–10.

Quaiser-Pohl, C., Geiser, C., & Lehmann, W. (2006). The relationship between computer-game preference, gender, and mental rotation ability. *Personality and Individual Differences*, *40*, 609–619.

Reynolds, M. R., Keith, T. Z., Ridley, K. P., & Patel, P. G. (2008). Sex differences in latent general and broad cognitive abilities for children and youth: Evidence from higher-order MG-MACS and MIMIC models. *Intelligence*, *36*, 236–260.

Rindermann, H. (2007). Author's response the big G-factor of national cognitive ability. *European Journal of Personality*, *21*, 767–787.

Snow, R. E., & Lohman, D. F. (1989). Implications of cognitive psychology for educational measurement. In R. Linn (Ed.), *Educational measurement* (pp. 263–331). New York: Collier.

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.

Strand, S., Deary, I. J., & Smith, P. (2006). Sex differences in cognitive abilities test scores: A UK national picture. *British Journal of Educational Psychology*, *76*, 463–480.

Su, R., Rounds, J., & Armstrong, P. I. (2009). Men and things, women and people: A meta-analysis of sex differences in interests. *Psychological Bulletin*, *135*, 859–884.

Terman, L. M. (1916). *The measurement of intelligence.* Boston, MA: Houghton Mifflin.

Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin*, *117*, 250–270.

Willingham, W. W., & Cole, N. S. (1997). *Gender and fair assessment.* Mahwah, NJ: Erlbaum.