Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Intelligence 32 (2004) 25–32
INTELLIGENCE

Cognitive mediators and sex-related differences in mathematics

Ana R. Delgado*, Gerardo Prieto

Universidad de Salamanca, Salamanca, Spain

Received 1 July 2002; received in revised form 25 May 2003; accepted 15 June 2003

Abstract

Sex-related differential studies on mathematical abilities have hardly taken into account the mediator role of the verbal factor, which contrasts with the interest shown in the mediator role of visuospatial aptitude. We predicted that if sex-related differences were found, mental rotation would mediate mathematical abilities typically favoring males (geometry and verbal problems) and lexical access would mediate the one favoring females (arithmetic). Data from 455 participants with a median age of 13 were analyzed, showing that sex-related effect sizes in mathematical criteria were small, ranging from $d = -.16$ to .18 (corrected by attenuation), as expected for unselected samples. Lexical access scores were consistent predictors for every mathematical subdomain, and mental rotation added to the prediction only for geometry and word problems.

Keywords: Cognitive mediators; Sex-related differences; Mathematics

1. Introduction

Research conclusions on sex-related differences in cognition are now becoming popular. A number of books and chapters thoroughly reviewing the scientific literature on that matter have been recently released (e.g., Geary, 1998; Halpern, 2000; Kimura, 1999; a good review can also be found in the chapter on group differences by Macintosh, 1998). Books on mathematical cognition have also dealt with the gender question (e.g., Butterworth, 1999; Dehaene, 1997; Geary, 1994). They all conclude similarly with respect to mathematical cognition in general populations: Differences are nonexistent or very small in size favoring males on average, however, not in every content domain (Beller & Gafni, 1996; Hedges & Nowell, 1995; Hyde, Fennema, & Lamon, 1990). Data from the

* Corresponding author. Departamento de Psicología Básica, Psicobiología y Metodología. Avda. de la Merced 109-131 37005, Salamanca, Spain.

E-mail address: adelgado@usal.es (A.R. Delgado).

0160-2896/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/S0160-2896(03)00061-8

1991 International Assessment of Educational Progress in Mathematics and Sciences show that sex-related effect sizes of mathematics for the 13-year olds are small ($d < .20$) but consistently positive, i.e., favoring boys. In addition, there is variability among countries: e.g., while in Spain the effect size was .18, in the United States it was a nonsignificant .04 (Beller & Gafni, 1996). It is well known that among those of above-average mathematical aptitude, there is an obvious preponderance of males (e.g., Benbow & Stanley, 1980). In addition, the sex difference in variance in tests of quantitative ability (ranging from 5% to 25%) indicates that there are more males than females also among the lowest scoring individuals (Geary, 1998; Hedges & Nowell, 1995; Wilingham & Cole, 1997).

Even though mathematics may be a useful label for functional objectives, different domains must be considered when searching for psychological explanations. When content domain is taken into account, the pattern of sex-related differences in mathematics would be males outperform females in geometry and solving of word problems; sometimes females outperform males in complex arithmetic (Geary, 1996).

The source of sex-related differences in mathematics has been traditionally located either in cognitive variables (e.g., McGuinness, 1993) or in attitudinal variables (e.g., Parsons, Adler, & Kaczala, 1982). Casey, Nuttall, and Pezaris (1997) compared the relative importance of the two kinds of factors, finding out that mental rotation—a visuospatial aptitude—was the most relevant mediator. Up to now, mental rotation has been the aptitude in which the largest and most consistent sex-related differences, favoring males, can be found (Masters & Sanders, 1993; Voyer, Voyer, & Bryden, 1995). Moreover, it has been shown that this difference cannot be attributed to performance factors such as differential guessing or time limit (Delgado & Prieto, 1996; Masters, 1998).

A possible explanation of the role of mental rotation in mathematical scores is related to problem-solving strategies. Many mathematical problems can be solved by using visualization as well as analytical strategies, and it has been empirically shown that there are two such different styles: one is based on algorithm memorization and automatic application, the other on visuospatial representation of the problems. The first one is considered as generally less efficient and it has been hypothesized that it could be more frequent among female students, which would help to explain the differences (McGuinness, 1993). This is consistent with the fact that most reviews of the literature have found that males are more variable in their visuospatial performance than females (e.g., Hedges & Nowell, 1995; Wilingham & Cole, 1997). This finding has theoretical importance because one hypothesis about the cause of sex-related differences is that many females do not use a spatial strategy to solve problems that are of a spatial nature, such as visualization tests (Delgado & Prieto, 1997), geometry tasks, or word problems (Geary, 1996). However, Kimura (1999) has pointed out that there is little factual support for the hypothesis on the criticality of the effect of three-dimensional rotational ability on math reasoning tasks, given that, in her own laboratory studies, the relation between math and rotation is no higher than the relation between math and vocabulary, thus suggesting that three-dimensional rotational ability might not be contributing any more to math ability than general intelligence does. Friedman (1995) has also called our attention to the relevance of the verbal factor, whose mediator influence has hardly been studied, which contrasts with the interest shown by the research community in the mediator role of visuospatial aptitude.

In recent years, the importance of some verbal components has also been indicated from studies of mathematical cognition (Butterworth, 1999; Cipolotti & Butterworth, 1995; Dehaene, Spelke, Pinel, Stanescu, & Tsivkin, 1999). Moreover, differential studies in mathematical competence are beginning to explore the question of the supporting cognitive systems (Geary, 1994, 1996; Hoard, Geary, & Hamson, 1999). In this context, there has been a change in jargon too: It is recognized that the division of abilities into verbal, visuospatial, and quantitative has been useful, but there are more fruitful ways of examining cognition. In addition, doing it in terms of processes seems closer to scientific psychology. Halpern (2000) has classified tasks in which males and females excel according to the underlying cognitive processes: Skilled performance in the tasks favoring females requires rapid access to and retrieval of information that is stored in memory, e.g., speed of lexical access; in contrast, tasks in which males tend to excel require the ability to maintain and manipulate mental representations, i.e., visuospatial working memory.

Mental rotation tasks have a high load in visuospatial working memory. By definition, lexical access requires accessing long-term memory, although accessing meaning is not needed (recognizing a word qua word is enough). Therefore, our hypotheses were that if sex-related differences were found in mathematical domains, mental rotation would mediate those typically favoring males (i.e., geometry and verbal problems) and lexical access would mediate the one favoring females (i.e., arithmetic). Rosenthal (1988) has proposed a statistical procedure in order to consider a variable as a possible mediator for sex differences in mathematics. It is necessary to have three nonzero correlations: sex and mathematics score, sex and mediator, and mediator and mathematics score. The effect of sex on the score is examined after partialling out the effect of the mediator: If the partial r approaches zero, then the mediator could be considered as a plausible explanatory variable (Rosenthal, 1988). However, when working in a multivariate context, the pattern of results could be different—e.g., owing to the existence of suppressor effects (Velicer, 1978), covariating out mental rotation could have the effect of incrementing (instead of decreasing) sex-related differences in arithmetic. Thus, we had to rule out this possibility before putting to the test the mediator role of lexical access and mental rotation.

2. Methods

2.1. Participants

Data from 455 subjects, 241 males, and 214 females were analyzed. This sample size allows enough power to detect small-sized differences between two independent sample means and small-to-medium sized effects in multivariate analyses (Cohen, 1988). Sampling was not at random; rather, every student in the second year of "Enseñanza Secundaria Obligatoria" (ESO) or Compulsory Secondary Education in 11 schools was tested, which warrants coverage of the whole continuum of abilities. At this point in their education, all the students have the same background in language, math, and science, given that the Spanish curriculum is common to every school. Although there seems to be a disproportion in the number of male participants, the actual percentage of males in the studied cohort is close to 54%, and therefore no a priori sample bias should be suspected. Data from students who showed lack of understanding of the instructions (e.g., writing down operations when instructions dictated that all calculations had to be carried out mentally) were discarded. The median age was 13, although some students were already 14 years old (about 25%, as was to be expected for the first term of the year—students have to be 13 when the year begins in order to study the second year of ESO).

2.2. Instruments

The questionnaire was composed of three multiple-choice tests measuring Lexical Access (LAT), Mental Rotations (MRT), and Mathematics (MT).

The LAT was constructed for this study in order to translate the typical experimental task designed to find out differences in speed of lexical access into a psychometric test measuring this cognitive ability with high reliability. Here, it is considered a means of assessing how easy it is to retrieve information, such as arithmetic facts, from long-term memory (Gathercole & Adams, 1994). The LAT is composed of 80 items showing three two-syllable pseudowords and only one two-syllable word, which must be crossed for the item to be awarded as correct. The location of the correct option in the item (first, second, third, or fourth) was randomly assigned, with the constriction that the correct option could be found 20 times in each location. Three pseudowords were constructed from each word, as homophones (when there were any of them), or as pronounceable letter sequences in which one or two letters had been substituted into the original word. For example, for the word "verde" (green) there is one homophone "berde" and two pronounceable letter sequences: "velde" and "verda"; for the word "cabra" (goat) there is one homophone, "cavra" and two pronounceable letter sequences: "cabla" and "vabra." Homophones differed only in spelling (Spanish has a quite transparent spelling, and sources of homophony are scarce. Some consonants sound or can sound equal when followed by certain vowels: b and v, ll and y, g and j, c and z, and c and k. The letter h does not sound. In standard Spanish, there are clear rules concerning how letters must be pronounced when in a word.) Pseudowords were randomly assigned to the items. The LAT is a speeded test, and therefore test–retest or alternate form reliabilities should be estimated; however, the only estimate available for this study was the Spearman–Brown corrected odd–even reliability coefficient (r=.98). This coefficient overestimates test reliability when tests are speeded, and therefore when it is used to disattenuate a correlation coefficient, the corrected value will be an underestimation.

The MRT is the redrawn version of the original Vandenberg and Kuse test, based on figures provided by Shepard (Peters, 1995). It is a spatial test composed of 24 items showing a target figure (an angular composition of 10 cubes) and four options from which only two are vertically rotated versions of the target. Following the recommended procedure, one point per item is awarded if and only if the two correct options have been identified. Therefore, the maximum score attainable in the MRT is 24. Although the MRT is a speeded test, the only available reliability estimate was the Spearman–Brown corrected odd–even reliability coefficient; it was .87 for this study.

The MT consists of 30 items arranged in three blocks of 10 questions corresponding to arithmetic (ARIT), geometry (GEOM), and word problems (PROB). Each question consists of a stem and four options of which only one is right (for example, item 23 reads, There are 116 candies for the guests at a birthday party. Half of them are mint. Of the rest, half are strawberry and half are blackberry. How many blackberry candies are there? Response options are 23, 24, 29+, 58. Item 24 reads, The distance between Springfield and Paul's town is 100 miles. Halfway between them there is a hill. Once past the hill, 15 miles towards Paul's town there is a country church. How far is this church from Springfield? Response options are 35, 45, 55, 65+.) The MT is an achievement test constructed for this study in order to obtain (1) specific, even if brief, measures on these three content domains as specified by the Spanish curriculum for the previous academic year, and (2) a global mathematics measure. It is a power test. Internal consistency α was .72 for the MT (.50, .51, and .46 for arithmetic, geometry, and word problems, respectively).

2.3. Procedure

Test application followed ethical standards. Data were collected by only one person, instructed by us, but blind to the hypotheses of the study. Questionnaires were applied in the classroom, during the first hours of the morning class, to natural groups of about 25 students per group. The order of application was constant: LAT, MRT, and MT. Standardized instructions for each test included practice items. It was clearly indicated that every question should be mentally solved, insisting on the fact that the ballpen should be used only to mark the answers. Application time was controlled by means of a chronometer: LAT = 3 minutes and 30 seconds; MRT = 8 minutes; MT = 25 minutes.

3. Results

Table 1 shows sex-related effect sizes d for every test, as well as d' (effect sizes corrected by attenuation, i.e., divided by test reliability index) and male/female variance ratios.

It can be seen that sex-related effect sizes for ARIT, GEOM, and PROB were small, as scientific literature on general populations of this age would lead us to expect. However, at a descriptive level, both signs as well as variance ratios were consistent with the bibliography for samples of 13-year olds. As regards possible mediators, a very small-sized (nonsignificant) effect, favoring females, was found for the LAT; as expected, the effect size for MRT was large.

Except for the MRT, sex-related effect sizes d did not reach the .20 conventional minimum. A one-factor (sex) MANOVA with two covariables (LAT and MRT) on three dependent variables (ARIT, GEOM, and PROB) was carried out in order to simultaneously test our hypotheses without incrementing the overall type I error probability. Multivariate F tests for sex did not reach statistical significance [Wilks' lambda exact $F(3,449) = 1.41$; $P > .05$]. Multivariate sex-related effect size was less than 1% of the variance, consistent with the descriptive univariate effect sizes. No evidence of suppressor effects was found.

Multivariate F tests for the within + residual regression were statistically significant [Wilks' lambda exact $F(6,898) = 11.36$; $P < .001$]. The multivariate effect size amounted to 7% of the variance. Individual regression analyses for each DV showed that the LAT was always the best predictor. In fact, it was the only statistically significant predictor for arithmetic scores, explaining 5% of their

Table 1
Descriptive data—sex-related effect sizes d and d' (corrected by attenuation) and male/female variance ratios

Score	d	d'	VR
LAT	− .03	− .03	1.17
MRT	.77	.83	1.61
MT	.02	.02	1.10
ARIT	− .11	− .16	1.03
GEOM	.13	.18	0.99
PROB	.04	.06	1.08

Conventionally, the sign of effect sizes favoring males are positive.

Table 2
Correlation matrix—raw ($N = 455$)

	LAT	MRT	MT	ARIT	GEOM
MRT	.090				
MT	.320**	.185**			
ARIT	.240**	.060	.734**		
GEOM	.259**	.180**	.821**	.390**	
PROB	.241**	.190**	.755**	.300**	.480**

** $P < .01$.

variance. For geometry scores, the LAT explained about 6% of the variance, but the MRT was also included, explaining an additional 2%. For word problem scores, the LAT explained about 5% of the variance and the MRT an additional 3%.

Table 2 shows the correlations between the covariables LAT and MRT, the dependent variables ARIT, GEOM, and PROB, and the total MT score. Except for the MRT–LAT and MRT–ARIT correlations, they all reached statistical significance. Correlations with DVs were always larger for the LAT than for the MRT (significantly so, as contrasted with the t statistic for dependent correlation coefficients, which ranged from 16.2 to 16.9; $df = 452$; $P < .001$).

It has been shown that scores in the LAT are consistent predictors for the three mathematical subdomains, and that the MRT adds to the prediction only for geometry and word problems. Given that the mathematical subdomains correlate positively with one another and that the internal consistency of the total MT score is not low (α=.72), a multiple regression analysis was performed with the LAT and the MRT as predictors and the MT as the criterion. It was found that both variables taken together accounted for 12% of MT variance (with the LAT explaining 10%) and the MRT, an additional 2%.

Significance is not affected by the correction for attenuation. However, from a scientific point of view, disattenuation (i.e., division by the product of test reliability indexes) gives us a closer idea of how large the correlation between constructs would be if variables were measured with instruments of perfect reliability: this is relevant for theory testing on accumulated results (Schmidt & Hunter, 1996). Correlations corrected for attenuation can be seen in Table 3.

After disattenuation, both LAT and MRT correlate higher than before with the three mathematical domains, specially so for word problems whose scores were the least reliable. As to the relation between mathematical domains, it is shown that geometry and word problems are not even distinguishable for this sample after measurement errors have been corrected.

Table 3
Correlation matrix—corrected by attenuation

	LAT	MRT	ARIT	GEOM
MRT	.098			
MT	.380	.233		
ARIT	.343	.090		
GEOM	.365	.269	.768	
PROB	.360	.301	.629	.993

4. Discussion

With respect to our hypotheses, no statistically significant sex-related differences have been found in any of the tested mathematical domains, a null result that cannot be attributed to lack of power. When mathematical abilities in unselected populations are studied, sex-related effect sizes are usually very small (Hedges & Nowell, 1995; Hyde et al., 1990). Our results are also compatible with those from the 13-year-old Spanish sample in the 1991 International Assessment of Educational Progress in Mathematics and Sciences (Beller & Gafni, 1996). The striking results obtained in samples of gifted participants (e.g., Benbow & Stanley, 1980) can have social as well as scientific implications but should not be overgeneralized.

At a descriptive level, the small differences that have been found are consistent with previous findings—females are better at arithmetic, males at geometry and word problems—and the variability of male scores is slightly higher.

Although neither of the cognitive processes have qualified as a mediator for sex-related differences, lexical access has been shown to be a good predictor of arithmetic, geometry, and word problems. In addition, mental rotation has been shown to add to the prediction of geometry and verbal problems—domains favoring males in studies in which sex-related differences are found—but not to the prediction of arithmetic in which (if found) differences favor females. All these correlational effects are in the medium size range (Cohen, 1988). It seems that with the possible exception of gifted samples, the study of individual differences in mathematical achievement is scientifically more fruitful than focusing on sex-related differences.

Finally, it is relevant to note that in terms of Carroll's structure of cognitive abilities, lexical access tasks such as the LAT would load in a Stratum I factor subordinated to Stratum II factor crystallized intelligence, which is concerned with mental processes reflecting both the influence of fluid intelligence and the effects of experience, learning, and acculturation (Carroll, 1993). The predictor effect of a low-level process—speed of lexical access—in mathematical competence is promising for applied studies. Values of r ranging from .34 to .38 (disattenuated) between lexical access and mathematical abilities are larger than most correlations in this context and indicate that the role of lexical access in mathematical achievement should be thoroughly explored.

Acknowledgements

This research was supported by DGICYT grant PB98-0263.

References

Beller, M., & Gafni, N. (1996). The 1991 International assessment of educational progress in mathematics and sciences: The gender differences perspective. *Journal of Educational Psychology*, *88*, 365–377.

Benbow, C. P., & Stanley, J. C. (1980). Sex differences in mathematical ability: Fact or artifact? *Science*, *210*, 1262–1264.

Butterworth, B. (1999). *The mathematical brain*. London: Macmillan.

Carroll, J. B. (1993). *Human cognitive abilities*. New York: Cambridge University Press.

Casey, M. B., Nutall, R. L., & Pezaris, E. (1997). Mediators of gender differences in Mathematics College Entrance Test scores. *Developmental Psychology*, *33*, 669–680.

Cipolotti, L., & Butterworth, B. (1995). Toward a multiroute model of number processing: Impaired number transcoding with preserved calculations skills. *Journal of Experimental Psychology. General*, *124*, 375–390.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: LEA.

Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. New York: Oxford University Press.

Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., & Tsivkin, S. (1999). Sources of mathematical thinking: Behavioral and brain imaging evidence. *Science*, *284*, 970–974.

Delgado, A. R., & Prieto, G. (1996). Sex differences in visuospatial ability: Do performance factors play such an important role? *Memory & Cognition*, *24*, 504–510.

Delgado, A. R., & Prieto, G. (1997). Mental rotation as a mediator for sex-related differences in visualization. *Intelligence*, *24*, 405–416.

Friedman, L. (1995). The space factor in mathematics. *Review of Educational Research*, *65*, 22–50.

Gathercole, S. E., & Adams, A. (1994). Children's phonological working memory: Contributions of long-term knowledge and rehearsal. *Journal of Memory and Language*, *33*, 672–688.

Geary, D. C. (1994). *Children's mathematical development*. Washington, DC: American Psychological Association.

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

Geary, D. C. (1998). *Male, female. The evolution of human sex differences*. Washington, DC: American Psychological Association.

Halpern, D. F. (2000). *Sex differences in cognitive abilities* (3rd ed.). London: LEA.

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

Hoard, M. K., Geary, D. C., & Hamson, C. O. (1999). Numerical and arithmetical cognition: Performance of low- and average-IQ children. *Mathematical Cognition*, *5*, 65–91.

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

Kimura, D. (1999). *Sex and cognition*. Cambridge, MA: MIT Press.

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

Masters, M. S. (1998). The gender difference on the mental rotations test is not due to performance factors. *Memory & Cognition*, *26*, 444–448.

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

McGuinness, D. (1993). Sex differences in cognitive style: Implications for mathematics performance and achievement. In L. A. Penner, G. M. Batsche, H. M. Knoff, & D. L. Nelson (Eds.), *The challenge of mathematics and science education: Psychology's response* (pp. 251–274). Washington, DC: American Psychological Association.

Parsons, J. E., Adler, T. F., & Kaczala, C. M. (1982). Socialization of achievement attitudes and beliefs: Parental influences. *Child Development*, *53*, 322–339.

Peters, M. (1995). *Revised Vandenberg & Kuse Mental Rotations Test: Forms MRT-A to MRT- D* (Technical Report). Guelph, ON, Canada: University of Guelph, Department of Psychology.

Rosenthal, R. (1988). Evaluating explanations of sex differences in mathematical reasoning scores. *Behavioral and Brain Sciences*, *11*, 207–208.

Schmidt, F. L., & Hunter, J. E. (1996). Measurement error in psychological research: Lessons from 26 research scenarios. *Psychological Methods*, *1*, 199–223.

Velicer, W. F. (1978). Suppressor variables and the semipartial correlation coefficient. *Educational and Psychological Measurement*, *38*, 953–958.

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.

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