

Sex differences in visuospatial ability: Do performance factors play such an important role?

ANA R. DELGADO and GERARDO PRIETO
Universidad de Salamanca, Salamanca, Spain

This study was designed to analyze some performance factors as a possible source of sex-related bias in psychometric tests of visuospatial aptitude. Goldstein, Haldane, and Mitchell (1990) explored the effect of two response styles—slowness of performance and reluctance to guess—by using a 3-D mental rotation test (the task showing the largest cognitive sex difference) and found that time limits and raw scores contributed substantially to the male advantage. We applied two tests in the *speed–power* continuum to a representative sample of 621 males and 821 females in their last year of high school in a 2×2 (gender \times time) full factorial design. Reluctance to guess was similar for males and females. Males obtained more correct responses on both tests, and for both time conditions, than did females. These results are not only statistically significant but also are of substantial practical consequence.

Although a large proportion of psychological research during the late 1960s and early 1970s focused on age-related differences, from the late 1970s on this interest was supplanted by a fascination with sex-related differences, and especially those related to spatial abilities, as Eliot (1987) has noted. Sex-related differences have been a matter of controversy ever since Maccoby and Jacklin (1974) concluded in their review that there was a difference favoring females in verbal ability and differences favoring males in quantitative as well as spatial abilities. When meta-analytic reviews started to replace narrative surveys of the literature, Hyde (1981) reanalyzed the data reviewed by Maccoby and Jacklin and calculated the median effect d for verbal, mathematical, and spatial abilities; the results, in the expected directions, were .24, .43, and .45, respectively. Even though Hyde reported that the effect size for spatial abilities was “only” .45, accounting for *only* 4.3% of the variance, it appears that this difference can be of substantial importance in practical realms, as Rosenthal and Rubin (1982) and Burnett (1986) have cogently illustrated.

In particular, males outperform females when spatial ability is measured by means of tasks that require subjects to mentally rotate 3-D figures; Sanders, Soares, and D’Aquila (1982) reported an effect size of .80—a non-

trivial difference, as they called it. Moreover, this difference has also been reported when the task is solved under unspeeded conditions (Resnick, 1993) and seems to have remained stable over time, as meta-analytic evidence from 14 studies published between 1975 and 1992 shows (Masters & Sanders, 1993). The pragmatic significance of these results has been emphasized by Casey, Nuttall, Pezaris, and Benbow (1995), who found mental rotation skills to be mediators for sex-related differences in math aptitude among high-ability college and college-bound students.

Different theories have been postulated to account for the differences in spatial aptitude. Hypothesized causes range from sociological factors—traditional gender roles, educational practices, differential course taking—to biological factors, such as X-linked genetic transmission or a sex difference in brain organization reflecting different hormonal influences, to mention just a few. The nature/nurture argument promises to be long-range given that there is clear evidence for both sides. Spatial aptitude has been shown to be a nonlinear function of testosterone levels (Kimura, 1992; Shute, Pellegrino, Hubert, & Reynolds, 1983); on the other hand, the fact that spatial aptitude can be dramatically improved in females—but only slightly in males—by means of specific training suggests the importance of the role of experiential factors (Connor, Schackman, & Serbin, 1978; Regian & Shute, 1993).

The main criticism made of any explanation is that reported sex differences in spatial aptitude are inconsistent, and so we could be trying to find the origin of something that perhaps does not actually exist. Caplan, MacPherson, and Tobin (1985) expressed concern with respect to the weak evidence of construct validity of the so-called spatial ability, concluding that it was inappropriate to ask about sex differences until spatial abilities were adequately divided up and labeled. This supposed lack of con-

The authors wish to thank Ronald K. Hambleton and colleagues of the School of Education at the University of Massachusetts at Amherst for their helpful advice. Kathryn T. Spoehr, Diane F. Halpern, and an anonymous reviewer provided constructive criticism of an earlier version of this article. This research was partially supported by a grant from the Spanish Instituto de la Mujer (Ministerio de Asuntos Sociales) to A.R.D. and G.P. Correspondence should be addressed to A. R. Delgado, Facultad de Psicología, Avda. de la Merced 109-135, 37005 Salamanca, Spain (e-mail: adelgado@gugu.usal.es).

struct validity can be true for a large portion of the measures used in research on sex-related differences, but, as Burnett (1986) pointed out, some of the main shortcomings in the Caplan et al. analysis were the failure to recognize which studies adequately reflected the spatial ability construct and which studies were trying to measure it by means of tasks in which the psychometric characteristics were at best unknown.

Actually, from the late seventies on, a great effort has been expended in reducing confusion in the field of spatial aptitude concerning both its psychometric and cognitive aspects. Multivariate studies have identified three major factors (Lohman, 1979): *spatial relations*, the most clearly defined factor, which requires speeded rotation or reflection of figures (although it does not represent speed of mental rotation, but rather the capacity to solve this sort of problem quickly, whatever the means); *spatial visualization*, the ability to deal with complex visual problems by imagining the relative movements of internal parts of an image; and *spatial orientation*, the capacity to imagine how a stimulus would be seen from a different point of view. In practice, it is often difficult to distinguish the third factor from the first, so we could schematically represent spatial aptitude by two factors ordered in a *speed–power* continuum and a *simple–complex* processing dimension: (1) Spatial relations tests are classified as speeded and simple, with 3-D mental rotation as the least speeded and least simple sort of test in this half of the continuum; and (2) spatial visualization tests occupy the other extreme, with surface development tests as the most powerful and complex (Pellegrino, Alderton, & Shute, 1984). A reference volume of spatial tests (Eliot & Smith, 1983), handbooks and chapters on the topic (see, e.g., Eliot, 1987; Pellegrino & Hunt, 1991), and several rigorous empirical studies on sex-related differences in spatial aptitude are now available. Feingold (1988), for example, determined the effect size of sex-related differences using the norms from the four standardizations of the Differential Aptitude Test (DAT) conducted between 1947 and 1980, and found that the differences usually encountered were decreasing. With respect to spatial aptitude, the *d* for the Space Relations subtest (DAT-SR) dropped from .37 to .15 over the period from 1947 to 1980. This approach to the assessing of sex-related differences has a double advantage: Samples are large and representative and tasks are typically administered in a public setting under timed conditions, similar to what occurs in real-life academic and professional selection processes or settings. On the other hand, Blough and Slavin (1987) failed to find a sex-related difference in an experimental task based on forms of the DAT-SR, an inconsistency that they properly attributed to the fact that few and relatively easy trials had been applied in a private context under untimed conditions. Psychometricians have long called that effect the *ceiling effect*. When psychologists design experimental tasks to explore cognitive processes, they are mostly interested in similarities between people's ways of thinking, not in

differences, so it is not unusual for them not to find such differences, since that is not their main objective. Psychometric tests, although not perfect, have been designed to reliably assess differences, so why not give tests some credit?

Apart from other considerations (see comments on Caplan et al., 1985, by Burnett, 1986; Eliot, 1986; Halpern, 1986; Hiscock, 1986, etc.), findings of sex differences in spatial aptitude—especially when operationalized by 3-D mental rotation tests—are the largest and most consistently found of the cognitive sex differences (Halpern, 1992; Masters & Sanders, 1993). But the classic report by Maccoby and Jacklin (1974) contained a less publicized observation on the tendency of females to perform tasks slowly and carefully, whereas males tended to perform tasks more quickly. In addition, other authors have indicated that females may have lower levels of confidence in their ability to perform various tasks (Parsons, Adler, & Kaczala, 1982) and consequently might be more reluctant to guess in multiple-choice tests if they are uncertain of the correct answer. These response tendencies were assessed by Goldstein et al. (1990)—who prefer to refer to them as *performance factors*—in a 3-D mental rotation test; it was found that time limits and the use of raw scores were contributing substantially to the male advantage. The suggested solution would be to apply tests in free-time conditions, or, in time-limit conditions, to correct the raw score by the attempted items. This ratio score, representing the percentage of correct alternatives chosen from among the items attempted, should eliminate the biasing effect of the time limit on performance. In our opinion, the effect of time limit on performance must be assessed by comparing experimental conditions (as in Goldstein et al.'s, 1990, second study) instead of using the ratio score; in practice, test takers would be informed of the scoring procedure and they could be granted a perfect score by answering (correctly) just one question (which could be the easiest). From a psychometric point of view, the transformation of the scale of measurement is not admissible either, given that raw scores are not divided by a constant but rather by a variable, and so we do not know the consequences on the score distribution, let alone on test reliability. However, we have used it in our research for comparative purposes.

With respect to other studies showing that the sex-related difference remains when the mental rotation task is solved under unspeeded conditions (Resnick, 1993), it could well be that this difference would be smaller than the difference shown when the task is solved in its standard (speeded) format; that is to say, it could well be that there was an interaction between sex and temporal conditions of application, such that the unspeeded condition was still beneficial to female performance.

Our approach to conceptualizing the guessing tendency was somewhat different from the approach taken by Goldstein et al. (1990). We think that items not answered may be considered as an indicator of reluctance

to guess *only* when they have been reached—subjects have actually attempted the item, but decided to omit the answer. Not-reached items, however—items that subjects have not even had the time to consider—are only indicators of slowness of performance, whatever the reasons: Slowness might be due to special carefulness, perhaps to lack of motivation, or even to low ability, and so on. Consequently, we make the distinction between *omissions* (items that a subject does not answer up to the last item for which an answer was given) and *not-reached items* (items that the subject did not answer, from the last item for which an answer was given to the end of the test). We believe that this distinction—usually made (Ben-Shakhar & Sinai, 1991; Budescu & Bar-Hillel, 1993, note 2), but seldom taken into account in practice—is psychologically relevant, especially when tests are speeded.

The hypotheses of this study were as follows: (1) Female performance would more closely resemble male performance under nonspeeded than under speeded conditions, and (2) omissions would be fewer for males than for females for both tests and both conditions.

Performance has been assessed in terms of right, wrong, omitted, and attempted items. Attempted items can be calculated by subtracting not-reached items from the total number of items in the test. It should be remembered that whenever the subjects have the possibility of omitting the answer to an item, or whenever there is the possibility of not answering all of the items composing the test, then the number of right items is not an inverse function of the number of wrong items, and we need to know at least three of these four variables in order to assess the subjects' performance. The fewer the attempted items, the slower the subject; the more omissions, the less the tendency of the subject to guess (note that the opposite is by no means true). But also: The more right items there are, plus the fewer wrong items, plus the fewer omitted items, and the fewer not-reached items, the more able the subject is. And we could go on with all the possible patterns. If we analyze these separate variables we will have more information on what the cause of the differences could be than by using formula scoring (whose objective is precisely the parsimonious combination of the different indicators in only one index, and for other purposes).

METHOD

Subjects

Six hundred and twenty-one males and 821 females in their last year of high school participated. Reward in terms of confidential feedback was given to those participants who asked for it. The mean age was 18.26 years ($SD = .97$, range = 17–23). It was presumed that all levels of spatial aptitude were represented, given that no other selection was carried out and most of the Spanish adolescent population is enrolled in high school. At this point in their education, all the students have a general background in math and science.

Materials

Two psychometric tests were selected from Eliot and Smith's (1983) directory: (1) the RFM test (Yela, 1968), which is the Spanish adaptation of the Rotation of Solid Figures developed by Thurstone and Thurstone in 1949, a 3-D mental rotation test, loading in the factor of spatial relations; and (2) a reduced version of the Spanish adaptation of the DAT-SR (TEA, 1990; Bennet, Seashore, & Wesman, 1974), a surface development test, loading in the factor of spatial visualization. Throughout the text these tests are referred to as MR and Vz, respectively. Items in the Spanish adaptations are identical to those from their corresponding American versions, and only minor changes were made in the standardization process.

Procedure

This research was conducted in Spain in accordance with prevailing ethical principles. Both tests were administered to the whole sample in a 2×2 factorial design: gender (male/female) \times time condition (speed/power).

Given the real-life conditions (1-h classes in high school), the *power* condition was not strictly untimed, but the time limit (15 min for the test on mental rotation and 25 for the visualization test) was more than double the time used in the Spanish standard application (5 min for the MR test, 12 for the Vz test), which was the time used here for the speed condition.

Students took the tests in their usual classroom and during their usual schedule. Natural groups resulted in about 30 participants per class. For roughly half of the sample (699 participants, with 309 males and 390 females), the tests were administered under the speed condition, following the standard instructions for speeded tests: "Work quickly and accurately." The rest of the sample (743 participants, with 312 males and 431 females), working under the power condition, followed what we have called "power instructions," the standard instructions accompanying nonspeeded tests: "Work accurately." The disproportion in number of males and females per cell is due to the actual disproportion of the population in schools; disproportion in the number of subjects in the speed/power condition can be attributed only to chance. Sex of the ex-

Table 1
Degree of Speededness of the Mental Rotation (MR) Test and the Visualization (Vz) Test \times Gender and Time Condition

	Speed				Power			
	Male		Female		Male		Female	
	MR	Vz	MR	Vz	MR	Vz	MR	Vz
Percentage of examinees completing test	3.9	36.6	0.8	21.5	68.6	87.2	30.4	81.0
Percentage of examinees completing 75% of test	12.9	75.5	3.1	55.5	88.7	96.5	62.4	96.0
Number of items attempted by 80% of examinees	7	22	6	18	18	30	12	30
Number of items on test	21	30	21	30	21	30	21	30

Table 2
Means and Standard Deviations × Time and Gender for Both Mental Rotation and Visualization Test Response Sum Variables

Sum Variables	Speed				Power				F Test		
	Male		Female		Male		Female		Gender	Time	Interaction
	M	SD	M	SD	M	SD	M	SD			
Mental Rotation											
Right	8.33	3.61	5.51	3.10	15.65	4.40	11.35	4.66	281.78*	941.81*	11.973*
Wrong	2.23	2.45	2.67	2.41	3.67	3.82	4.94	4.21	23.48*	116.33*	5.23
Omitted	0.11	0.48	0.11	0.71	0.12	0.46	0.24	0.72	3.92	5.28	3.78
Attempted	10.67	3.91	8.29	3.23	19.44	2.88	16.54	4.22	187.23*	1954.52*	1.86
Visualization											
Right	13.88	6.22	11.76	5.15	19.50	6.67	16.89	6.10	54.99*	283.56*	0.585
Wrong	11.31	6.76	11.08	6.16	9.29	6.43	11.49	6.16	8.93*	3.62	12.96*
Omitted	0.42	1.41	0.33	1.01	0.44	1.47	0.57	1.34	0.10	4.15	2.50
Attempted	25.62	4.61	23.17	5.18	29.23	2.54	28.94	2.68	41.33*	549.10*	26.95*

Note—*F* tests with $df = 1, 1438$ for main effects of gender and time and for gender × time interaction. * $p < .01$.

perimeter was balanced across conditions by having half of the groups for each condition tested by a female psychologist while the rest were tested by a male psychologist. Given the sample size, a stringent criterion ($p < .01$) was used throughout this study to avoid having a number of very small effect sizes interpreted.

The chosen tests are very difficult ones; for example, they are used in Spain to select aircraft pilots (Prieto et al., 1994). Nevertheless, the results were as expected when the tests were applied under speed and power instructions, as can be seen in Table 1. With respect to a possible floor effect, we will see from Table 2 that even for the group with the lowest level in MR (female/speed), the mean number of right responses was significantly above chance, provided that MR items had five response options [$t(389) = 8.19, p < .01$].

RESULTS AND DISCUSSION

Table 2 summarizes results from the 2×2 (gender × time) full factorial analyses of variance (ANOVAs) on performance in both tests.

Right Items

We have found significant differences between males and females in the number of items answered correctly. Male subjects had more correct answers for both tests and under both time conditions. Effect size for gender was $d = .89$ for MR and $d = .39$ for Vz. There was also a significant interaction effect for the MR test (Figure 1) indicating that, contrary to our expectations, the sex-related difference was broader under the power condition. It seems clear that females were getting fewer correct items per unit of time than were their male counterparts, and therefore not only did the power condition turn out to be not advantageous, but it was even prejudicial for female performance.

Wrong Items

There were also significant sex-related differences in the number of wrong items. Females got more items wrong in the MR test for both time conditions. For the Vz test, there were significant main effects only for the gender factor. There was also a significant interaction effect for the Vz test (Figure 2), indicating that female sub-

jects got more items wrong than did male subjects only under the power condition; actually, simple effects of gender at time = speed were nonsignificant [$F(1, 1438) = 0.24, p > .01$].

Omissions

There were no significant differences between male and female subjects in the number of omissions. When tendency to guess is considered independently of slowness, females did not seem to be more prone to omission than were males.

Attempted

There were significant main effects of gender and time conditions on the number of attempted items in MR. There were also main effects of gender and a significant interaction effect for the Vz test (Figure 3), indicating that the difference was greater under the speed condition; simple effects of gender at time = power were nonsignificant [$F(1, 1438) = 0.95, p > .01$]. Taking into account the interactions regarding both wrong responses and at-

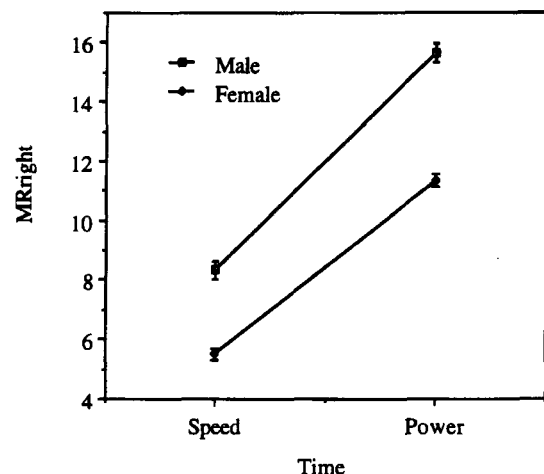


Figure 1. Interaction effects ($\pm SEM$) of gender × time on mental rotation (MR) test—right responses.

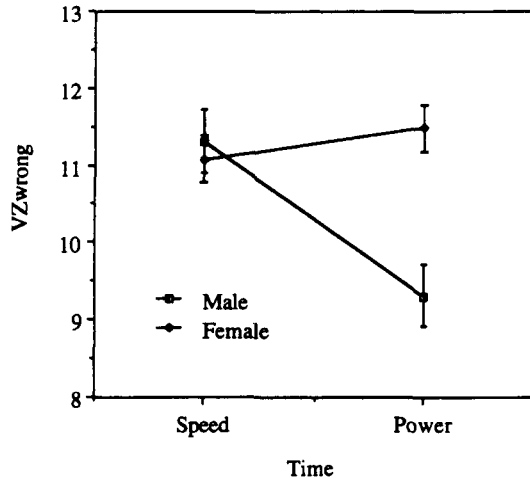


Figure 2. Interaction effects ($\pm SEM$) of gender \times time on visualization (Vz) test—wrong responses.

tempted items for the Vz test, it seems clear that females committed more errors when the time condition allowed them to answer more items (in fact, there was no difference between male and female subjects in attempted items under the power condition). Again, contrary to expectations, the power condition appears to have been prejudicial for female performance.

Given the significant differences in number of attempted items, we thought that it would be pertinent to reanalyze our data considering the ratio—(right/attempted) \times 100—suggested by Goldstein et al. (1990), not only for right responses, but also for wrong and omitted items, so that the effect on the scores of slowness of performance could be eliminated.

Ratio Scores

It can be seen from Table 3 that the implications of our data remain the same, except for the interactions, which

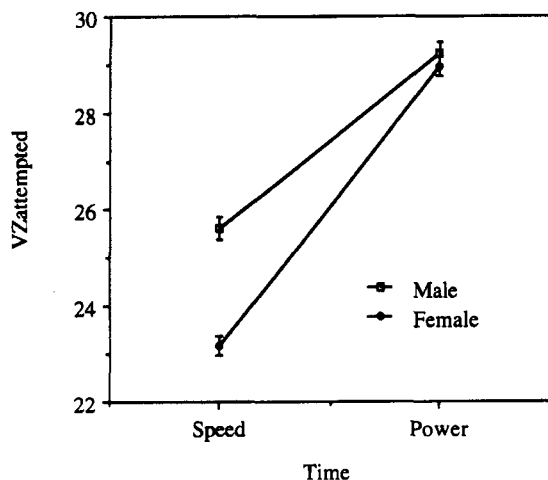


Figure 3. Interaction effects ($\pm SEM$) of gender \times time on visualization (Vz) test—attempted items.

disappear when the effect of attempted items (and, indirectly, the effect of time itself) is corrected. For the MR test, females still show lower (correct) ratio scores than do males for both conditions. Results are particularly clear with respect to wrong responses in the Vz test: After correcting for attempted items, the interaction disappears and women make more errors for both conditions. We cannot confirm Hypothesis 1. With respect to omissions, there were no significant differences even when they were corrected for attempted items, so we can conclude that our second hypothesis was not confirmed either.

It is also clear that instructions to work quickly (speed) were detrimental to both gender groups even when the effect of time itself was eliminated by using the ratio. When people work at their own pace, following instructions of accuracy, they get more responses right and fewer responses wrong. But males also get consistently more responses right and fewer responses wrong than do females. So, given that there is no interaction effect, if there is a response tendency penalizing women, it is related neither to the time itself nor to the instructions provided for this study. It could well be that with another set of instructions females would achieve better results, given that the instructions provided do not offer subjects much information about how to achieve good performance, but it is doubtful that unambiguous instructions combined with any correction formula could eliminate the individual differences resulting from differences in risk-taking tendencies (Ben-Shakhar & Sinai, 1991). In any case, this study did not manipulate formula-scoring instructions, and so no conclusions can be reached on the matter.

The fact that our results do not corroborate the conclusions reached by Goldstein et al. (1990) can be attributed to some substantial differences between the studies. First, Goldstein et al. used an easier task; second, their sample was short and selected; and third, the effects of performance factors (time limit and response style) were mixed: Not-answered items were taken as symptoms of both slowness and proneness not to guess.

Apart from the fact that the results do not confirm the hypotheses on the influence of speededness on the sex-related difference in spatial aptitude, we would like to point out that some degree of speededness is required not only in most standardized test administration, but also in some real-life settings, and it may be related to construct validity (as is true in the case of the spatial relations factor). When Goldstein et al. (1990) stated that no subject took more than 3 min longer to complete the task in the untimed portion of their study (and just 3 min should not condition opportunities in the real world), they forgot to take into consideration the fact that the timed condition was just 3 min, so, by applying a ratio, they should have noticed that this performance time was double!

Finally, it is important to keep in mind that the list of cognitive similarities between the sexes is much greater than the list of differences (see Halpern, 1992), but this fact should not lead us to ignore the differences when these

Table 3
Means and Standard Deviations × Time and Gender for Both Mental Rotation and Visualization Test Response Ratio Variables

Ratio Variables	Speed				Power				F Test		
	Male		Female		Male		Female		Gender	Time	Interaction
	M	SD	M	SD	M	SD	M	SD			
Mental Rotation											
Right/attempted	78.42	20.97	65.59	26.20	80.57	19.41	69.41	22.60	98.53*	6.75*	0.48
Wrong/attempted	20.62	20.77	33.30	25.89	18.74	19.26	29.11	22.26	93.08*	7.28*	0.933
Omitted/attempted	0.96	3.97	1.11	5.32	0.69	2.71	1.47	4.19	4.40	0.140	1.94
Visualization											
Right/attempted	54.95	23.10	52.18	21.73	66.77	21.93	58.56	20.61	23.05*	57.90*	5.51
Wrong/attempted	43.42	23.13	46.47	21.32	31.65	21.41	39.46	20.35	23.09*	64.19*	4.35
Omitted/attempted	1.62	5.16	1.35	3.86	1.58	5.17	1.98	4.60	0.077	1.89	1.82

Note—F tests with $df = 1, 1438$ for main effects of gender and time and for gender × time interaction. * $p < .01$.

are reliably found. The underrepresentation of women in the scientific world (Steen, 1987) is a serious social problem that seems to be related to the sex difference in spatial aptitude; we do not need to know its causes (although it would obviously help) to try to attack this difference on many fronts: Parents could be taught to help female babies to actively explore their space, girls could be encouraged to pursue more math courses, and so on. A simple instructional intervention, such as placing women in discussion groups with men after practice sessions with an extremely complex laboratory spatial task, has been shown to drastically reduce sex differences (Regian & Shute, 1993). And given the strong relationship between mental rotation and math scores (Casey et al., 1995), the instructional consequences should be easily drawn. At the same time, results from research on instruction may also tell us something about causality.

What is clear is that, although it is true that carrying out effective research on the trainability of spatial aptitude implies the recognition of the differences—whatever their causes—overestimating the actual spatial aptitude of about half of the citizens in the world will not ameliorate the situation of women in mathematical and scientific careers.

REFERENCES

- BENNET, G., SEASHORE, H., & WESMAN, A. (1974). *Differential Aptitude Test Manual* (5th ed.) New York: Psychological Corp.
- BEN-SHAKHAR, G., & SINAI, Y. (1991). Gender differences in multiple-choice tests: The role of differential guessing tendencies. *Journal of Educational Measurement*, *28*, 23-35.
- BLOUGH, P. M., & SLAVIN, L. K. (1987). Reaction time assessments of gender differences in visual-spatial performance. *Perception & Psychophysics*, *41*, 276-281.
- BUDESCU, D., & BAR-HILLEL, M. (1993). To guess or not to guess: A decision theoretic view of formula scoring. *Journal of Educational Measurement*, *30*, 277-291.
- BURNETT, S. A. (1986). Sex-related differences in spatial ability: Are they trivial? *American Psychologist*, *41*, 1012-1014.
- CAPLAN, P. J., MACPHERSON, G. M., & TOBIN, P. (1985). Do sex-related differences in spatial abilities exist? *American Psychologist*, *40*, 786-799.
- CASEY, M. B., NUTTALL, R., PEZARIS, E., & BENBOW, C. P. (1995). The influence of spatial ability on gender differences in math college entrance test scores across diverse samples. *Developmental Psychology*, *31*, 697-705.
- CONNOR, J. M., SCHACKMAN, M., & SERBIN, L. A. (1978). Sex-related differences in response to practice on a visual-spatial test and generalization to a related test. *Child Development*, *49*, 24-29.
- ELIOT, J. (1986). Comment on Caplan, MacPherson, and Tobin. *American Psychologist*, *41*, 1011.
- ELIOT, J. (1987). *Models of psychological space*. New York: Springer-Verlag.
- ELIOT, J., & SMITH, I. M. (1983). *International directory of spatial tests*. Slough, England: NFER-Nelson.
- FEINGOLD, A. (1988). Cognitive gender differences are disappearing. *American Psychologist*, *43*, 95-103.
- GOLDSTEIN, D., HALDANE, D., & MITCHELL, C. (1990). Sex differences in visual-spatial ability: The role of performance factors. *Memory & Cognition*, *18*, 546-550.
- HALPERN, D. F. (1986). A different answer to the question, "Do sex-related differences in spatial abilities exist?" *American Psychologist*, *41*, 1014-1015.
- HALPERN, D. F. (1992). *Sex differences in cognitive abilities* (2nd ed.) London: Erlbaum.
- HISCOCK, M. (1986). On sex-differences in spatial abilities. *American Psychologist*, *41*, 1011-1012.
- HYDE, J. S. (1981). How large are cognitive gender differences? A meta-analysis using ω^2 and δ . *American Psychologist*, *36*, 892-901.
- KIMURA, D. (1992, September). Sex differences in the brain. *Scientific American*, *267*, 119-125.
- LOHMAN, D. F. (1979). *Spatial ability: A review and reanalysis of the correlational literature* (Tech. Rep. No. 8). Stanford, CA: Stanford University Press.
- MACCOBY, E. E., & JACKLIN, C. N. (1974). *The psychology of sex differences*. Stanford, CA: Stanford University Press.
- MASTERS, M. S., & SANDERS, B. (1993). Is the gender difference in mental rotation disappearing? *Behavior Genetics*, *23*, 337-341.
- PARSONS, J. E., ADLER, T. F., & KACZALA, C. M. (1982). Socialization of achievement attitudes and beliefs: Parental influences. *Child Development*, *53*, 322-339.
- PELLEGRINO, J. W., ALDERTON, D. L., & SHUTE, V. J. (1984). Understanding spatial ability. *Educational Psychologist*, *19*, 239-253.
- PELLEGRINO, J. W., & HUNT, E. B. (1991). Cognitive models for understanding and assessing spatial abilities. In H. Rowe (Ed.), *Intelligence: Reconceptualization and measurement* (pp. 203-225). Hillsdale, NJ: Erlbaum.
- PRIETO, G., CARRO, J., PULIDO, R. F., PALENZUELA, D. L., ORGAZ, B., & LORO, P. (1994). Selección psicotécnica informatizada de los aspirantes a pilotos. *Revista de Aeronáutica y Astronáutica*, *633*, 429-431.
- REGIAN, J. W., & SHUTE, V. J. (1993). Basic research on the pedagogy of automated instruction. In T. de Jong, H. Spada, & D. M. Towne

- (Eds.), *The use of computer models for explication, analysis, and experiential learning* (pp. 121-132). New York: Springer-Verlag.
- RESNICK, S. M. (1993). Sex differences in mental rotations: An effect of time limits? *Brain & Cognition*, **21**, 71-79.
- ROSENTHAL, R., & RUBIN, D. B. (1982). A simple, general purpose display of magnitude of experimental effect. *Journal of Educational Psychology*, **74**, 166-169.
- SANDERS, B., SOARES, M. P., & D'AQUILA, J. M. (1982). The sex difference on one test of spatial visualization: A nontrivial difference. *Child Development*, **53**, 1106-1109.
- SHUTE, V. J., PELLEGRINO, J. W., HUBERT, L., & REYNOLDS, R. W. (1983). The relationship between androgen levels and human spatial abilities. *Bulletin of the Psychonomic Society*, **21**, 465-468.
- STEEN, L. A. (1987). Mathematics education: A predictor of scientific competitiveness. *Science*, **237**, 251-252, 302.
- TÉCNICOS ESPECIALISTAS ASOCIADOS (1990). *DAT: Tests de aptitudes diferenciales. Manual* (10th ed.) Madrid: Author.
- THURSTONE, T. G., & THURSTONE, L. L. (1949). *Mechanical aptitude II: Description of group tests* (Report No. 54). Chicago: University of Chicago, Psychometric Laboratory.
- YELA, M. (1968). *Rotación de figuras macizas. Manual*. Madrid: Técnicos Especialistas Asociados.

(Manuscript received September 3, 1994;
revision accepted for publication August 25, 1995.)