Behavior Research Methods
2008, 40 (2), 422-427
doi: 10.3758/BRM.40.2.422

Computerizing the Mental Rotations Test: Are gender differences maintained?

John S. Monahan, Maureen A. Harke, and Jonathon R. Shelley
Central Michigan University, Mount Pleasant, Michigan

A computerized version of the Revised Mental Rotations Test using touch-screen technology was tested and compared with the paper-and-pencil version. One hundred ninety-two participants—133 women and 59 men—took the paper-and-pencil version; 91 participants—47 women and 44 men—took the touch-screen version. Standard scores × test half and the entire test, proportion of errors × type, and proportion correct × alternative type were calculated. The expected gender differences occurred in standard scores and proportion correct × alternative type for both test versions. Men performed better than women in all instances, but gender difference effect sizes (measured by Cohen's d) were reduced from *large* for the paper-and-pencil version to *medium* for the computerized version. Scores declined at least nominally from the first to the second half of both versions, and significantly for women taking the paper-and-pencil version.

Linn and Petersen (1985) separated spatial ability tests into three categories based on effect size. One category of ability was *spatial visualization*, the ability to manipulate complex spatial information when several stages are needed to produce the correct solution. Gender differences in spatial visualization had a small effect size and were statistically nonsignificant. Another category of ability was *spatial perception*, the ability to determine spatial relations despite irrelevant information. Gender differences in spatial perception had a medium effect size and were statistically significant. The third category of ability was *mental rotation*, the ability to rotate two- or three-dimensional figures quickly and accurately in imagination and to compare them with other similar figures. Mental rotation was the only spatial ability category to yield a large gender difference effect size and was also statistically significant.

More recently, Voyer, Voyer, and Bryden (1995) conducted a meta-analysis of the three categories of spatial abilities and found similar results. Spatial visualization tasks showed significant gender differences only for participants who were over 18 years old; for younger participants, there were no significant gender differences. Spatial perception tasks showed a significant gender difference for participants who were 13 and older. Mental rotation tasks showed gender differences for participants of all ages. For all three spatial ability categories, there was a significant linear increase in effect size with increasing age, possibly indicating that sexual differentiation is important in gender differences in spatial ability. They further found that the largest effect was found for the Mental Rotations Test (MRT; Vandenberg & Kuse, 1978).

Wraga, Duncan, Jacobs, Helt, and Church (2006) report that, although traditional gender gaps in cognitive performance have diminished over many years, mental rotation tasks have consistently yielded large and reliable gender differences of about 1 standard deviation with no significant reduction. At least one study, however, found that, under certain circumstances, gender differences in the MRT did not hold.

Goldstein, Haldane, and Mitchell (1990) administered the MRT under two different sets of instructions. In Experiment 1, participants were allowed the standard 3 min of time for each 10-item half of the test. The standard method of scoring the MRT is to count correct only those items for which both correct alternatives are marked (maximum possible score of 20). Using the standard scoring method, the probability of scoring a point by chance is .16. Instead of using the standard method of scoring, Goldstein et al. calculated scores in two alternate ways. One way was to count the number of correct alternatives chosen (maximum possible score of 40). With this method, the probability of scoring a point by chance is .5 for the first guess of an item. Goldstein et al.'s second scoring method was to derive the ratio of the number of correct alternatives chosen to the number of alternatives attempted. Men chose significantly more correct alternatives than women did, but did not have a significantly higher ratio score, possibly indicating that men are more likely to guess than women are. In Experiment 2, participants were first allowed 3 min per half, but they indicated the last item completed in each half; subsequently, they were given unlimited time to complete the unattempted items. The standard score and a ratio score for the timed part of the test were calculated, as were the numbers of correct alternatives chosen for the timed and untimed portions. A gender difference was found only with the standard score in the timed portion of the test.

J. S. Monahan, monah1js@mail.cmich.edu

Thus, Goldstein et al. argued that male advantage in the MRT occurs only when the test is taken under time pressure. Of course, removing time pressure from the MRT changes it from a speed test to a power test.

Subsequently, a number of researchers have found results different from those found by Goldstein et al. (1990). Stumpf (1993) had participants complete a number of different spatial tests, one of which was the MRT. He found that, although ratio scoring reduced the gender difference, it remained large and significant. Resnick (1993) tested participants on the MRT without time limit and found a significant gender difference, despite comparable scores on a vocabulary test. Delgado and Prieto (1996) had participants complete two spatial tests, one of which was the MRT. The tests were given under both speed and power instructions. Significant gender differences in MRT scores were found with each type of instruction. Masters (1998) tested men and women on the MRT under timed and untimed conditions and scored the tests under both the standard procedure and the alternate procedure of using the number of correct alternatives chosen. A ratio score was not calculated. She found no effect of timing condition on the magnitude of the gender difference.

More recently, Terlecki and Newcombe (2005) have argued that gender differences in spatial experiences with computers and video games may account for part of the differences in spatial ability scores. Using a spatial activities survey, they found that women generally had less computer experience and greater variability in experience. They found that computer experience mediates the gender difference in MRT scores and that experience is more important for women than for men. They concluded that lack of computer and spatial experience prevents women from achieving their full potential in spatial ability tests. Quaiser-Pohl and Lehmann (2002) found similar results.

Peters et al. (1995) noted that the original Vandenberg and Kuse (1978) MRT had deteriorated due to repeated copying, thus requiring that the stimulus images be touched up by researchers prior to use. As a result, they argued that the tests lacked uniformity between studies. Peters et al. used a computer-assisted drawing program to redraw the MRT and made two changes. They included four additional items and enclosed the figures in rectangular, rather than circular, frames. Twenty-four items were included in the Revised MRT in two 12-item blocks. As in the original MRT, each item consisted of a target figure and four response options, of which two were correct rotations of the target figure and two were distractors. Participants must choose both correct rotations in order to get the item correct. The Revised MRT was found to be comparable to the original; men performed significantly better than women did, with a large effect size for gender difference. Peters et al. also compared men's and women's scores relative to the degree program in which they were enrolled (i.e., BSc or BA). Both men and women in the BSc program performed better than did their counterparts in the BA program. Men in each program performed better than did women in the same program, with a large effect size. They also found that the gender differences remained significant when using Goldstein et al.'s (1990) ratio scoring method.

Voyer and his colleagues tested the timing and guessing hypotheses as the sources of gender differences using Peters et al.'s (1995) Revised MRT. To test the timing hypothesis, Voyer, Rodgers, and McCormick (2004) presented items one at a time for specific intervals, instead of the standard method of presenting half of the items for a specific interval. They allowed times from 15 to 40 sec per item and found a significant gender difference in the number of correct items at each interval. They also found that scores for both men and women increased with each increase in interval length. In a second experiment, they allowed unlimited time per item and still found gender differences in scores. To test the guessing hypothesis, they divided the incorrect items into five categories according to the alternatives selected. In order of likelihood, the categories are *correct–wrong* (CW), *correct–blank* (CB), *wrong–wrong* (WW), *wrong–blank* (WB), and *blank–blank* (BB). The proportion of each type of error did not differ by gender, and Voyer et al. (2004) concluded that men did not guess at a greater rate than women did.

Voyer and Hou (2006) looked at gender differences on the MRT related to the type of item presented: ones in which distractors were mirror images of the target, ones with distractors that were structurally different from the target, and ones that had occluded parts resulting from rotation on one alternative. They gave participants unlimited time to complete the Revised MRT and found that men's scores were significantly better than women's scores across all three categories, the greatest difference being on items in which there were occluded parts. These findings supported their hypothesis that greater gender differences on items with occluded items may be due to the three-dimensional nature of the MRT. Gender differences were statistically significant, even though participants were given unlimited time to complete the test. Voyer and Hou also found that scores for both men and women were nominally lower on items on the second half of the test than on items on the first half.

The purpose of the present study is to test a computerized version of the Revised MRT by comparing its results to those of the paper-and-pencil version. A number of recent studies of conversion of paper-and-pencil instruments to computerized form have been conducted, the scores from the latter being compared with those from the original instrument. For example, Pettit (2002) compared versions of the Computer Attitude Scale (Loyd & Gressard, 1984), the Perfectionist Self-Presentation Scale (Hewitt & Flett, 1995), and the Marlowe–Crowne Social Desirability Scale (Crowne & Marlowe, 1960). She found that scores and response patterns for the two presentation modes were very similar. Hewson and Charlton (2005) found similar results with original and computerized versions of the Multidimensional Health Locus of Control Scale (Wallston & Wallston, 1981). Meyerson and Tryon (2003) measured the equivalence of versions of the Sexual Boredom Scale (Farmer & Sundberg, 1986) with five additional scales used to validate the Sexual Boredom Scale. They found the versions psychometrically equivalent.

Not all scales, however, translate accurately from paper-and-pencil to computer presentation. Buchanan et al.

(2005) found that the Prospective Memory Questionnaire (Hannon, Adams, Harrington, Fries-Dias, & Gipson, 1995) has four factor-analyzed subscales as a paper-and-pencil test, but only two factor-analyzed subscales as a computerized World Wide Web-delivered test. They caution that one cannot be sure, without testing for equivalence, whether a computerized test administered via the World Wide Web, or, presumably, via a local computer, actually measures the intended concept.

If it is a comparable variant, computerized version of the Revised MRT should show most of the characteristics of the paper-and-pencil version. There should be a significant gender difference in standard scores and in the scores of items with mirrored or structurally different foils, or occluded alternatives, as described in Voyer and Hou (2006). Furthermore, there should be a slight reduction in scores from the first to the second half of the test. Of course, at best, the scores on the computerized version should match the scores on the paper-and-pencil version. Additionally, one might expect little or no gender difference in error types. As with any measure of mental rotation, gender difference effect sizes should be large.

METHOD

Paper-and-Pencil Version

Participants. The participants were 192 undergraduate students—133 women and 59 men—recruited from introductory psychology courses at Central Michigan University. Participation was voluntary, and students received course credit for participating.

Apparatus and Materials. The Peters et al. (1995) Revised MRT was used with separate instructions (Peters, 1995).

Procedure. Participants were tested in relatively large groups (20–40 persons) in their own classrooms. The participants' task was to choose the two alternative figures that were rotations of the target. The instructions were read to the participants, who were then provided three practice problems to which they were given the correct answers in order to become familiar with responding to the items. Students were instructed to place a large "X" through the box surrounding the alternative to signify their choice. The test was separated into two blocks with 12 items in each. Participants were given 3 min to complete each block, with a 2-min interval between blocks. Participants were debriefed.

Touch-Screen Version

Participants. Participants were 91 undergraduate students—47 women and 44 men—recruited from introductory psychology courses at Central Michigan University. Participation was voluntary, and students received course credit for participating. All participants were tested during the fall semester of 2006.

Apparatus and Materials. The Revised MRT was reproduced using E-Prime software (Schneider, Eschman, & Zuccolotto, 2002) for use with touch-screen computer displays. The goal was to recreate conditions similar to those of the paper-and-pencil version. Items were presented one at a time, and the interval between blocks was self-paced.

Procedure. Testing was conducted in small groups of 2 to 6 participants. The gender makeup of the groups was not controlled. Standardized instructions were read to the group, after which the participants were told to begin. There were three practice items with correct answers given, allowing the participants to become familiar with responding on the touch screen. When participants touched an alternative to signify their answer, a red "X" was shown below that figure. To change their answer, they simply touched the figure again, and the "X" was removed. They could then choose another alternative. Only two items could be chosen at a time. When the participants were satisfied with their answers, they pressed the "Next" button and advanced to the next item. They were unable to return to an earlier item.

The participants had 3 min to complete a block of 12 items. Time between blocks was self-paced, except that participants could not begin the second block until a 2-min period between blocks had passed. As participants finished the test, they were debriefed and allowed to leave the testing room.

RESULTS

Standard Scoring

Table 1 presents standard scores and effect sizes by test half, gender, and test version. Scores from Peters et al. (1995) for BA students are also included in Table 1. There was a main effect of gender: Men's total scores and scores by test half were greater than women's scores, regardless of test version or test half. There was also a main effect of test half: Both men and women had higher scores in the first than in the second half with either version. Second-half scores for women taking the paper-and-pencil version were significantly lower than first-half scores [$t(132) = 4.61, p < .001$], and second-half scores for men taking either version and for women taking the computerized version were nominally lower. There was no significant effect of test version, and overall test scores did not differ by test version. There were no significant interactions.

Gender difference effect sizes were medium for each half and for the total score for the computerized version, as well as for the first half of the paper-and-pencil version. Effect sizes for the paper-and-pencil version were large for the second half and for the total score. The gender difference effect size for the total score from Peters et al.'s (1995) BA students was also large. Additionally, *t* test comparisons showed that scores from the paper-and-pencil version of the present study and from Peters et al.'s BA students were not significantly different for men or for women.

Standard score by test half (2), gender (2), and test version (2), was analyzed using a partially repeated measures ANOVA. Test half was a significant factor in determining the number correct [$F(1,279) = 24.09, MS_e = 3.16, p <$

Table 1
Standard Scores (SS) by Test Half, Gender, and Test Version

	Men		Women		
Test Half	SS	*SD*	SS	*SD*	*d*
	Computerized Version				
First	6.59	2.73	5.06	2.86	0.55
Second	5.75	2.29	4.51	2.48	0.52
First and second	12.34	4.51	9.57	4.90	0.59
	Paper-and-Pencil Version				
First	6.76	2.65	4.65	2.86	0.75
Second	6.07	2.48	3.50	2.86	1.19
First and second	12.83	4.69	8.15	4.36	1.05
	Peters et al.'s BA Students				
	12.10	4.80	8.20	3.80	0.94

Note—Computerized version = 44 men, 47 women; paper-and-pencil version = 59 men, 133 women; Peters et al.'s (1995) BA students = 102 men, 222 women.

.001, Huynh–Feldt $E = 1$]. Gender was also a significant factor in determining the number correct [$F(1,279) = 39.19$, $MS_e = 20.66$, $p < .001$]. Neither test version nor any of the interactions had a significant effect on the number correct. Cohen's *d* (Cohen, 1988) was calculated to estimate gender difference effect sizes.

Types of Errors

Table 2 shows the proportion of errors and effect size by error type, gender, and test version. The two most common error types were CW and BB, which were not significantly different in proportion for any of the four gender–test version groups. CW and BB were significantly more likely than were CB, WW, and WB errors. The proportions of the latter three error types were not significantly different from each other.

The proportion of incorrect responses by error type (5), gender (2), and test version (2) was analyzed using a partially repeated measures ANOVA. Error type had a significant effect on the proportion of incorrect responses [$F(1.43, 401.50) = 276.98$, $MS_e = 0.11$, $p < .001$, Huynh–Feldt $E = 0.36$]. No other factor or interaction had a significant effect on the proportion of incorrect responses. Planned tests were conducted using Dunn's multiple comparison procedure (Dunn, 1961) and found that CW and BB errors were significantly more likely than were CB, WW, and WB errors.

Types of Alternatives

Following Voyer and Hou (2006), MRT items were split into three groups: items with mirror image incorrect alternatives, items with structurally different incorrect alternatives, and items with occluded correct or incorrect alternatives. Table 3 shows proportion correct and effect size by alternative type, gender, and test version. Men outperformed women on all three alternative types on both test versions. Gender difference effect sizes were medium for the computerized version and large for the paper-and-pencil version for each alternative type.

The proportion correct by alternative type (3), gender (2), and test version (2) was analyzed using a partially repeated measures ANOVA. Alternative type had a significant effect on the proportion correct [$F(2,558) = 85.80$, $MS_e = 0.026$, $p < .001$, Huynh–Feldt $E = 1$]. Gender also had a significant effect on proportion correct [$F(1,279) = 37.18$, $MS_e = 0.109$, $p < .001$].

DISCUSSION

The paper-and-pencil and computerized versions of the Revised MRT produced similar scores. There was no significant main effect of test version, nor was there a significant interaction involving test version. As expected, with both test versions, the men's scores were significantly greater than the women's scores. The main difference between the two versions was in the gender difference effect sizes. The paper-and-pencil version yielded the expected large effect sizes, but the computerized version yielded medium effect sizes. Additionally, the gender difference effect size for the paper-and-pencil test for the present study was similar to the effect size for that in Peters et al. (1995), as expected from Wraga et al. (2006), indicating little, if any, change in effect size for the Revised MRT since its introduction.

As expected from Voyer and Hou (2006), participants achieved at least nominally higher scores in the first half of the test than in the second. This effect occurred despite the difference in the timing between the present study (3 min per half) and the timing in Voyer and Hou (unlimited time to complete). The unexpected sharp decline in second-half scores for women who took the paper-and-pencil version may indicate greater fatigue, slightly more difficult test items, or greater frustration for women than for men taking that test version. Additional study of this outcome is warranted. Subsequent studies conducted in this lab have not replicated a significant drop in score from the first to the second half for either women or men. Thus, the result in the present study may be an anomaly.

Possibly, there is some advantage for women in using the computerized version, as indicated by the lower gender difference effect size. One potential explanation for the advantage of the computerized version for women is based on the difference in the size and composition of the groups during testing: large mixed sex groups for the paper-and-pencil version and small, sometimes mixed, sometimes single-sex groups for the computerized ver-

Table 2
Proportion of Total Errors (PE) by Error Type, Gender, and Test Version

	Men		Women		
Error Type	PE	*SD*	PE	*SD*	*d*
Computerized Version					
CW	.49	.24	.47	.31	0.06
BB	.38	.29	.39	.31	−0.03
CB	.06	.09	.06	.11	−0.03
WW	.05	.09	.05	.09	−0.05
WB	.02	.06	.02	.04	−0.02
Paper-and-Pencil Version					
CW	.46	.25	.40	.23	0.26
BB	.42	.29	.49	.26	−0.24
CB	.05	.13	.04	.08	0.13
WW	.04	.08	.06	.09	−0.21
WB	.02	.06	.01	.04	0.22

Note—Error types: CW = correct–wrong, BB = blank–blank, CB = correct–blank, WW = wrong–wrong, WB = wrong–blank.

Table 3
Proportion Correct (PC) by Type of Alternative, Gender, and Test Version

	Men		Women		
Alternative Type	PC	*SD*	PC	*SD*	*d*
Computerized Version					
Mirror	.56	.26	.40	.25	0.63
Structure	.59	.23	.50	.26	0.37
Occluded	.39	.23	.31	.21	0.36
Paper-and-Pencil Version					
Mirror	.59	.24	.37	.26	0.87
Structure	.61	.22	.42	.21	0.89
Occluded	.42	.23	.24	.19	0.89

sion. Data concerning the makeup of the groups are not available. If women perform better on the Revised MRT in small groups or in single-sex groups, one would expect the computerized version to lead to better performance for women.

An additional explanation might be that women see computerized instruments as being more interesting than do men. Previously, the opposite might have been the case. Popovich, Hyde, Zakrajsek, and Blumer (1987) found that women's attitudes toward computers were significantly more negative than men's on most aspects of computer usage. Terlecki and Newcombe (2005), on the other hand, report more recent Web statistics showing that women marginally surpass men in their Internet usage; but they also found that women lag behind men in using computers in spatial activities. Finally, the use of computers in society has expanded to the point where almost all students have had considerable experience with computers, presumably reducing any gender difference in computer usage. Therefore, one might expect that computerizing the Revised MRT should have only a small effect on gender differences in scores. Whether the reduction in effect sizes is caused by testing group size and composition, or the fact that computers are used, or the visual structure of the test, or some other factor, is not illuminated by the present study.

As expected from Voyer et al. (2004), the proportion of errors by error type did not differ by test version or gender, only by error type. One possible conclusion is that computerizing the MRT does not cause a major alteration in the manner in which men and women perform the MRT—that is, they are not guessing more or guessing less. No major shift in strategy occurred. Although the pattern of errors in the present study is somewhat different from Voyer et al.'s, our results, like their results, show that men and women do not differ in their patterns of results. Presumably, the difference between Voyer et al.'s patterns of errors and ours has to do with timing condition differences: a set time per item for Voyer et al., and a set time overall for the present study.

As found by Voyer and Hou (2006), men in the present experiment had a greater proportion correct for mirror items, structurally different items, and occluded items. Although men scored higher than women did on all three types of alternatives on both test versions, gender difference effect sizes were only medium on the computerized version, but large on the paper-and-pencil version. Thus, the improved performance by women on the computerized version was not specific to a particular type of alternative, but carried over to all three alternative types. The improved performance by women appears to be generalized across item types.

The major difference between versions is the reduced gender difference effect size from large with the paper-and-pencil version to medium with the computerized. As Stumpf (1993) found, a performance factor can affect the level of gender difference without eliminating it. Thus, the effect seems to tap a general factor in MRT performance, rather than a specific one. The computerized version of the Revised MRT appears to be a reasonable substitute for the paper-and-pencil version. Outcomes are similar for the two versions in gender differences in overall scores and in error types. Gender differences in performance by alternative types were maintained, but with some improvement for women on the computerized version. Thus, scores on the Revised MRT seem to be affected by computerized testing conditions, but only to the extent that gender difference effect size seems to be reduced but not eliminated.

The present study tested students enrolled in psychology classes. Presumably, they are similar in spatial ability to Peters et al.'s (1995) BA students, although some may have been science students taking a psychology class as an elective. The question of whether these sorts of results would be obtained with science students, such as Peters et al.'s BSc students, is undetermined. Nevertheless, the computerized version of the Revised MRT could be useful in studies in which it is difficult to assemble all participants at the same time. Participants could start and stop with minimal disruption to others who are taking the computerized test in the same laboratory. Furthermore, although one might expect that using a mouse rather than a touch screen to respond would have little effect on scores, such an approach has not been tested.

AUTHOR NOTE

The authors thank the Central Michigan University Faculty Research and Creative Endeavors Committee for its financial support. We thank Kyunghee Han for her insightful comments on an earlier version of this article. We thank Lindsay Nelson for her initial rendering of the program. We thank Rachel Mack for conducting the pilot tests and Kayanna Sanders for her help in data collection, entry, and analysis. Correspondence concerning this article should be addressed to J. S. Monahan, Department of Psychology, Central Michigan University, Mount Pleasant, MI 48859 (e-mail: monah1js@cmich.edu).

REFERENCES

Buchanan, T., Ali, T., Heffernan, T. M., Ling, J., Parrott, A. C., Rodgers, J., & Scholey, A. B. (2005). Nonequivalence of on-line and paper-and-pencil psychological tests: The case of the prospective memory questionnaire. *Behavior Research Methods*, **37**, 148-154.

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

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, **24**, 349-354.

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

Dunn, O. J. (1961). Multiple comparisons among means. *Journal of the American Statistical Association*, **56**, 52-64.

Farmer, R., & Sundberg, N. D. (1986). Boredom proneness–The development and correlates of a new scale. *Journal of Personality Assessment*, **50**, 4-17.

Goldstein, D., Haldane, D., & Mitchell, C. (1990). Sex differences in visual–spatial ability: The role of performance factors. *Memory & Cognition*, **18**, 546-550.

Hannon, R., Adams, P., Harrington, S., Fries-Dias, C., & Gipson, M. T. (1995). Effects of brain injury and age on prospective memory self-rating and performance. *Rehabilitation Psychology*, **40**, 289-298.

Hewitt, P., & Flett, G. (1995). *Perfectionistic Self-Presentation Scale*. Unpublished test, York University, Toronto.

Hewson, C., & Charlton, J. P. (2005). Measuring health beliefs on the Internet: A comparison of paper and Internet administrations of the Multidimensional Health Locus of Control Scale. *Behavior Research Methods*, **37**, 691-702.

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

LOYD, B. H., & GRESSARD, C. (1984). Reliability and factorial validity of computer attitude scales. *Educational & Psychological Measurement*, **44**, 501-505.

MASTERS, M. S. (1998). The gender difference on the Mental Rotations test is not due to performance factors. *Memory & Cognition*, **26**, 444-448.

MEYERSON, P., & TRYON, W. W. (2003). Validating Internet research: A test of the psychometric equivalence of Internet and in-person samples. *Behavior Research Methods, Instruments, & Computers*, **35**, 614-620.

PETERS, M. (1995). *Revised Vandenberg and Kuse Mental Rotations Tests: Forms MRT-A to MRT-D* (Tech. Rep.). Guelph, ON: University of Guelph, Department of Psychology.

PETERS, M., LAENG, B., LATHAM, K., JACKSON, M., ZAIYOUNA, R., & RICHARDSON, C. (1995). A redrawn Vandenberg and Kuse Mental Rotations Test: Different versions and factors that affect performance. *Brain & Cognition*, **28**, 39-58.

PETTIT, F. A. (2002). A comparison of World-Wide Web and paper-and-pencil personality questionnaires. *Behavior Research Methods, Instruments, & Computers*, **34**, 50-54.

POPOVICH, P. M., HYDE, K. R., ZAKRAJSEK, T., & BLUMER, C. (1987). The development of the Attitudes Toward Computer Usage Scale. *Educational & Psychological Measurement*, **47**, 261-269.

QUAISER-POHL, C., & LEHMANN, W. (2002). Girls' spatial abilities: Charting the contributions of experiences and attitudes in different academic groups. *British Journal of Educational Psychology*, **72**, 245-260.

RESNICK, S. M. (1993). Sex differences in mental rotations: An effect of time limits? *Brain & Cognition*, **21**, 71-79.

SCHNEIDER, W., ESCHMAN, A., & ZUCCOLOTTO, A. (2002). *E-Prime user's guide*. Pittsburgh: Psychology Software Tools.

STUMPF, H. (1993). Performance factors and gender-related differences in spatial ability: Another assessment. *Memory & Cognition*, **21**, 828-836.

TERLECKI, M. S., & NEWCOMBE, N. S. (2005). How important is the digital divide? The relation of computer and videogame usage to gender differences in mental rotation ability. *Sex Roles*, **53**, 433-441.

VANDENBERG, S. G., & KUSE, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual & Motor Skills*, **47**, 599-604.

VOYER, D., & HOU, J. (2006). Type of items and the magnitude of gender differences on the mental rotations test. *Canadian Journal of Experimental Psychology*, **60**, 91-100.

VOYER, D., RODGERS, M. A., & MCCORMICK, P. A. (2004). Timing conditions and the magnitude of gender differences on the Mental Rotations Test. *Memory & Cognition*, **32**, 72-82.

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.

WALLSTON, K. A., & WALLSTON, B. S. (1981). Health locus of control scales. In H. M. Lefcourt (Ed.), *Research with the locus of control construct: Vol. 1. Assessment methods* (pp. 189-243). New York: Academic Press.

WRAGA, M., DUNCAN, L., JACOBS, E. C., HELT, M., & CHURCH, J. (2006). Stereotype susceptibility narrows the gender gap in imagined self-rotation performance. *Psychonomic Bulletin & Review*, **13**, 813-819.

(Manuscript received July 20, 2007;
revision accepted for publication September 28, 2007.)