BRAIN AND COGNITION 28, 39–58 (1995)

A Redrawn Vandenberg and Kuse Mental Rotations Test: Different Versions and Factors That Affect Performance

MICHAEL PETERS, BRUNO LAENG, KERRY LATHAM, MARLA JACKSON, RAGHAD ZAIYOUNA, AND CHRIS RICHARDSON

University of Guelph, Guelph, Ontario, Canada

The available versions of the Vendenberg and Kuse (1978) Mental Rotations Test (MRT) have physically deteriorated because only copies of copies are available. We report results from a redrawn version of the MRT and for alternate versions of the test. Males perform better than females, and students drawn from the physical sciences perform better than students drawn from the social sciences and humanities, confirming other reports with the original version of the MRT. Subjects find it very hard to perform the MRT when stimuli require rotation along both the top/bottom axis *and* the left/right axis. The magnitude of effect sizes for sex (which account, on average, for some 20% of the variance) does not increase with increasing difficulty of the task. Minimal strategy effects were observed and females did not perform differently during the menstrual period as opposed to the days between the menstrual periods. Practice effects are dramatic, confirming other reports with the original MRT, and can also be shown to be powerful in a transfer for practice paradigm, where test and retest involve different versions of the MRT. Main effects of handedness on MRT performance were not found.

INTRODUCTION

Even strong advocates of biological causes of sex differences in spatial abilities (Kimura, 1992) point out that sex differences are by no means seen in all spatial tasks, and often tasks that yield sex differences do not always do so consistently (cf. Alyman & Peters, 1993). The question is not: why are males better than females on spatial tasks, but rather, what tasks yield sex differences and why? In the pursuit of this question, the

This research was supported by National Sciences and Engineering Research Council of Canada Grant A7054. Address correspondence and reprint requests to Michael Peters, Ph.D., Dept. Psychology, University of Guelph, Guelph, Ontario, Canada N1G 2W1. E-mail: mpeters@uoguelph.ca.

0278-2626/95 $6.00

tests which provide reliable sex differences become a focus of interest. The most important of these is Vandenberg & Kuse's (1978) mental rotations test (MRT). It is important for three reasons. First, there are no difficulties in identifying this test as a test of spatial abilities. Second, and unlike most other paper and pencil spatial tests, this test favors male subjects in practically all published studies, across cultural boundaries (Jahoda, 1980; Oosthuizen, 1991; Stumpf & Klieme, 1989) and for different age cohorts (cf. present results with original Vandenberg & Kuse study). Third, the effect sizes are appreciable. Law, Pellegrino, and Hunt (1993) have recently made the point that dynamic spatial tasks yield more robust sex differences than static paper and pencil spatial tasks. However, even in their dynamic task on relative velocity judgements, sex accounted for only 6% of the variance in sensitivity of velocity judgements while in the mental rotation task a considerably larger proportion of the variance in scores is accounted for, as seen in this study.

The MRT, then, provides an ideal point of departure in the examination of what sex differences in spatial performance mean. Unfortunately, the available copies of the original Vandenberg & Kuse test have deteriorated through successive copying of the originals, so that many researchers are forced to touch up the figures and tests are used which vary in the quality of the stimulus drawings. The present note describes a refurbished version of the Vandenberg & Kuse test which was redrawn with help of a computer-assisted drawing program.[1] The primary purpose of this note is to provide test results for this version that are based on a large sample of college students.

However, in addition to describing the "behavior" of the test for a large sample of subjects, a number of factors which might be of practical significance in testing will be considered.

First, there is the element of practice. In keeping with the intent of this article, there will be no extensive discussion of practice effects on mental rotation performance in general. Instead, we asked the question: what changes can be expected due to repeat performance of the test? Casey and Brabeck (1989) found that a single repeat of the test leads to a rather marked increase in performance. No further information was given in terms of whether the sex differences remained stable for the second administration but it is clear that this factor needs to be considered, both from a practical and a theoretical point of view. An associated question is whether the practice effect is specific to a particular set of items or more general in nature.

Second, there is the factor of "preselection". When populations of

[1] The original Vandenberg and Kuse stimulus figures were redrawn with the AUTOCAD program by Diane Duncan, School of Engineering, University of Guelph. In keeping with the philosophy of Vandenberg and Kuse, we are happy to provide interested researchers with copies of the tests described here at cost.

college students are tested, individuals from different academic programmes differ in their MRT performances (Casey and Brabeck, 1990, 1989). Casey and Brabeck (1989) found that, in general, math–science majors performed better than non-math-science majors. This suggests that caution has to be used in between-subject group designs, so that results do not spuriously arise from different cohort membership. Here, a similar comparison is made in order to see if the effect of academic specialization has some generality across college populations. It should be noted that in comparisons of this kind, differences might be confounded with other, uncontrolled factors. For instance, preselection for degree majors in different universities involves different cut-off criteria for admission.

Third, there is a possibility that spatial performance tests might be sensitive to the phase of the menstrual cycle in females. There are a number of reports in the literature that concern differences in the performance of spatial and manual tasks relative to the phase of the menstrual cycle (Gouchie & Kimura, 1991; Hampson, 1990a,b; Hampson & Kimura, 1988, 1992; Silverman & Phillips, 1993). Unfortunately, Hampson and Kimura did not use the MRT, but relied on a test that is similar to the Picture folding test, the Rod and Frame Test, and a Hidden Figures test. However, their argument is that males excel on tests of spatial ability and that females perform such tests better when their estrogen levels are lowest, i.e., during the menstrual phase of their cycle. It should be noted that the Picture Folding Test used in this study, even though the means favored males, did not produce an overall significant sex effect, and embedded figures tasks do not always produce significant sex differences (Alyman & Peters, 1993; Kimura, 1994; Lee et al., 1978).

Finally, in this study effect sizes (in terms of the percentage of variance accounted for) were calculated for the sex effects, and this allowed an evaluation of whether the size of such effects varies with the difficulty of the task. In all, three versions of the Vandenberg & Kuse task were used. These are illustrated in Fig. 1. Figure 1A is the standard Vandenberg & Kuse task, with two sets of 12 items each. Figure 1B is a set of 12 items each which is composed of the same stimulus figures, but arranged in different sets. Figure 1C is a set which was designed for greater difficulty. This set is also composed of 2 sets of 12 items. The individual items require subjects to rotate the figure around the left/right as well as the top/bottom axis.

METHOD

Subjects

In the principal study, 636 undergraduate students from different academic programs at the University of Guelph were tested. These subjects participated on a voluntary basis.

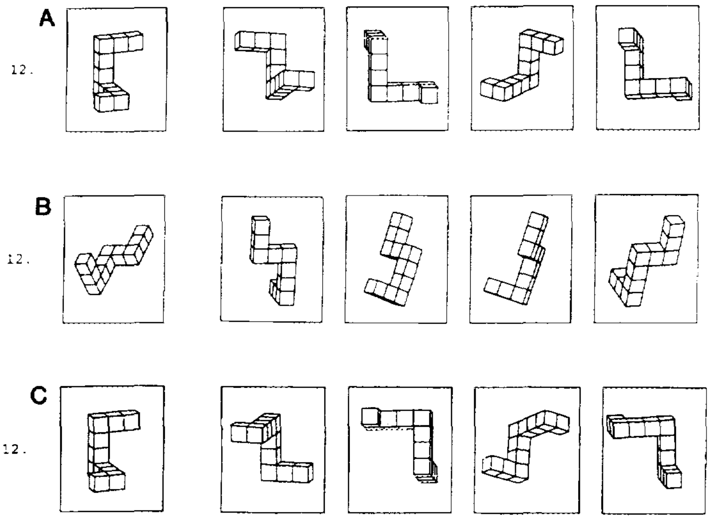

FIG. 1. Item 1 from set MRT(A) = top row, MRT(B) = middle row, and MRT(C) = bottom row. In each case, the stimulus on the left is the target. Subjects have to determine which two of the four sample stimuli on the right of the target are rotated versions of the target stimulus. There are 24 items in each test, and each item has two and only two correct matches. The fact that the targets in MRT(A) and MRT(C) are identical is coincidental; for other items all five stimuli may be different.

Subjects came from the "science" (engineering, biological, and physical sciences) and "arts" (social sciences, arts, and humanities) programs. The composition of the sample was as follows. There were 102 males and 222 females from the Arts program and 135 females and 177 males from the Science program. The average age in years was 21.3 for males and 20.5 for females. Additional subjects from the same undergraduate populations were tested for specific purposes; their numbers, sex, and academic program are provided together with the results.

Procedure

General testing procedure. Subjects were tested in groups ranging from 10 to 50 students. Before testing, each student filled out a handedness questionnaire. This consisted of 14 items (Peters & Perry, 1991), and students were given a "left/right" or "either hand" choice. For the purposes of classification in this study, the writing hand was used as criterion for right- or lefthandedness. Additional questions asked students to state whether they frequently, sometimes, or never played computer games such as "Block-out" or "Tetris" which involve manipulation of spatial images. Females were also asked to state whether they used contraceptive pills. They were also asked to state whether their menstrual cycle was regular and the day of their period (counting the first menstrual day as Day 1). Participants were given the explicit option of not answering any of these questions. Each student was given the 24-item MRT(A) set, and 3 min were given for each subset of 12 items, separated by 4 min. The version of the test was the one given in Fig. 1. Subjects had to identify *both* of the correct alternatives, and a score of "1" was given if and only if both choices were correct. Thus, the maximum score was 24. Specifics of testing procedure for the other tests will be provided in the appropriate places below.

The card rotation and the picture folding test. In order to compare the effect sizes for the MRT(A) with other tests that also involve mental rotation, a subset of subjects were also given mental rotation tests that involve mental rotation in the frontal plane only (card rotation test), and a the Picture Folding test (Ekstrom et al., 1976). A within-subjects design allowed evaluation of how sex differences for the three tests would compare for given subjects.

RESULTS

Mental Rotation Performance

Effects of Sex

The results of the overall comparison are seen in Table 1. There was a significant overall effect of sex, with males performing better than females. The effect size calculation shows that 17.7% of the variance in scores was accounted for by this variable. Figure 2 shows a histogram of the performances of males and females. To construct this histogram, standard scores for males and females within the BA and BSc program were calculated separately so that Academic Program differences would not impact on the distribution. It can be seen that there is a significant overlap, but that the curves are indicative of separate distributions. The data contained in this figure also allow an evaluation of Harris' (1978) prediction that some 25% of females should have performances in excess of the midpoint performance of the males, as based on his genetic/environmental model of sex differences in spatial performance. In order to control for the different performances of BA and BSc students, the z scores for males in the BA program were adjusted upward by the mean difference between males in the BA and BSc program, and the same was done for females in the BA program. This served the purpose to eliminate

TABLE 1
Overall MRT(A) Scores for Male and Female Students in the BA and the BSc Programs

	M	*SD*	*N*
Males			
BA	12.1	4.8	102
BSc	14.8	4.8	135
Females			
BA	8.2	3.8	222
BSc	10.4	4.2	177
Overall	10.8	5.0	636

SEX $F(1/632) = 135.75$ $p < .0001$ *ES* .177 *Power** 1.000
PROG $F(1/632) = 46.77$ $p < .0001$ *ES* .069 *Power* 1.000

* Power was determined throughout relative to an α value of .05.

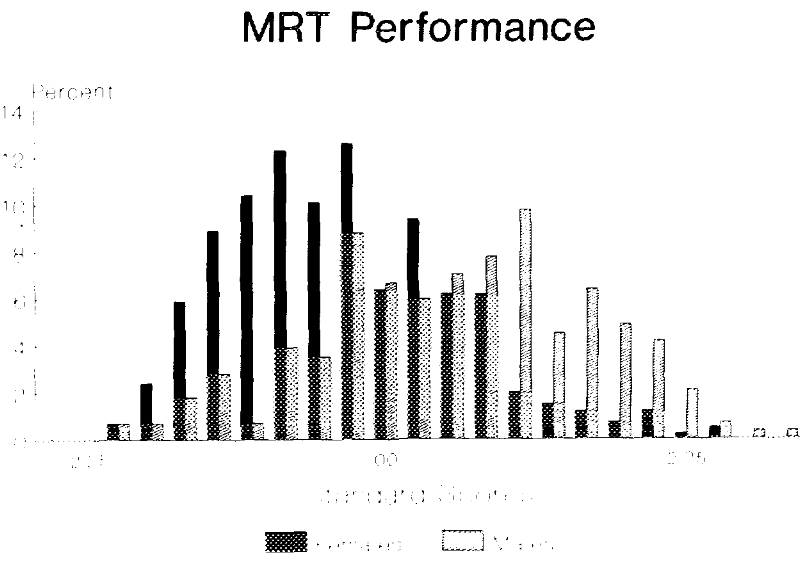

FIG. 2. Male and female performances on the MRT(A), in z scores, adjusted for program differences. 405 females, 283 males.

mean differences between males and females in the two degree programs while preserving the overall z score difference between the sexes. The results showed that only 15% of the females had z scores in excess of the mean z score of the males, and this falls short of the 25% predicted by Harris' model. Whether the discrepancy reflects an inadequacy of Harris' model or whether the 15% is a result of environment interacting with disposition to depress female scores remains to be determined. It is of interest to note that Bouchard and McGee (1977), in a direct test of the hypothesis that sex differences in spatial ability are due to an x-linked recessive gene effect, concluded that their results with the Mental Rotations Test did not support the hypothesis.

Academic Program

As in the case of the work by Casey and Brabeck (1989), individuals in the BSc program performed better than individuals in the BA program (Table 1). The effect size of Academic Program was weaker; it accounted for 6.9% of the variance in scores. It should be noted that our definition of the BSc and the BA program was very broad. Results of Scheffe posteriori tests (.05) show that some differentiation of the sex/program effects is possible.

	F BA	F BSc	F BEng	M BA	M BSc	M BEng
F BA						
F BSc						
F BEng	*					
M BA	*					
M BSc	*	*	*			
M BEng	*	*	*	*		

The Effects of Handedness

Table 2 gives the performance of lefthanders and righthanders, collapsed over sex because there were no significant interactions involving sex. In agreement with Casey, Pezaris, and Nuttall (1992), there was no overall effect of handedness. However, there was a significant interaction between Academic Program and Handedness. The means suggest that lefthanders in the BSc program performed better than righthanders and righthanders in the BA program performed better than lefthanders. Interactions among handedness, general ability level, and MRT performance have been reported by Harshman, Hampson, and Berenbaum (1987) but the minute effect size found in our interaction, which accounts for only 1.3% of the variance, suggests that the result should not be overinterpreted. In general, it appears that handedness does not have to be controlled for in studies in which lefthanders are proportionally represented. However, in studies where handedness itself is of interest and where group numbers are matched for handedness, handedness will have to be controlled for in conjunction with Academic Program.

Differences between Males and Females in Terms of Individual Questions and Strategy

In order to see whether male and female subjects approached the 24 items differently, an index of difficulty was calculated for each 24-item

TABLE 2
MRT(A) Performance of Lefthanders and Righthanders

		Mean	*SD*	*N*
Lefthanders				
BA	1	7.628	4.281	39
BSc	2	13.447	5.140	38
Righthanders				
BA	1	9.667	4.546	284
BSc	2	12.131	4.946	270
Total		10.823	4.977	631

Program $F(1/627) = 51.50$ $p < .0001$ *ES* .08 *Power* 1.000
HG × program $F(1/627) = 8.46$ $p < .004$ *ES* .01 *Power* .825

set. This index was calculated by dividing the number of subjects successfully solving a particular item into the number of subjects attempting a item. The maximum value would be 1, when all subjects attempting an item also solve it correctly. High positive correlations across all items would suggest that items found easy by one sex were also found easy by the other. The correlation obtained (Pearson's ρ) was .93, indicating that sexes responded similarly; items which proved easier for males were also found easier by females and items that proved more difficult for males were also found more difficult by females. There were no questions which favored males or females disproportionately.

There is some reason to believe that males and females differ in their strategies when approaching the MRT (Allen, 1974; Allen & Hogeland, 1978; Cochran & Wheatley, 1989; Casey, Brabeck, & Ludlow, 1986; Kail, Carter, & Pellegrino, 1979; Olson & Eliot, 1986; Schultz, 1991; Tapley & Bryden, 1977). A subset of subjects were given questions which related to introspection about processes and strategies used while performing the MRT. The questions are listed in the Appendix. A convincing overall sex difference was only found on Questions 1 and 3. In Question 2 significantly more males than females used the nonverbal strategy ($\chi^2 = 5.5$, df 1, $p < .018$). In Question 3, 37.8% of the females vs. 15.6% of the males stated that they used movements of the fingers, hand, or pencil to help with the rotation performance ($\chi^2 = 10.7$, df 1, $p < .001$). However, there were no differences in performance between females who did or did not use external help in rotation. Table 3 illustrates the relation between strategy and MRT scores for those questions where strategy mattered. Only one of the questions provided revealed noteworthy strategy differences; males who stated that they rotated parts of the figure did significantly worse than males who stated that they rotated the entire figure.

TABLE 3
Strategy and Performance on the MRT(A)

	Rotated whole figure			Rotated parts of figure		
	MRT	*SD*	*N*	*MRT*	*SD*	*N*
Males	13.6	5.0	59	10.1	5.5	13
Females	8.2	4.3	58	8.3	4.2	25

Interaction between strategy and sex: $F(1, 151) = 4.05$ $p < .05$, *ES* .03 *Power* .513

	Verbal strategy			Nonverbal strategy		
	MRT	*SD*	*N*	*MRT*	*SD*	*N*
Males	11.9	5.6	31	13.6	4.9	44
Females	7.5	3.9	52	9.5	4.2	37

Main effect of strategy: $F(1, 160) = 6.22$ $p < .01$ *ES* .04 *Power* .690

In addition, males and females who used a verbal strategy did worse than males and females who did not. In conjunction with the frequency differences of males and females choosing the verbal vs. nonverbal options for solving the item, this may be used as indication that part of the lower MRT performance in females is due to a more frequent use of the verbal strategy. However, it must be cautioned that the effect sizes were so small as to question whether they should be used to make any convincing argument that these strategy differences are an important part of the sex differences in MRT performance.

Finally, we wished to see if there was an effect of an overall approach difference for males and females. For instance, perhaps females were slower but solved a greater percentage of attempted questions than males. The comparison was based on the ratio "number of questions solved correctly/total number of questions attempted." Males solved 71.5% of the attempted questions correctly while females solved 56.3% of the attempted questions correctly ($F(1, 170) = 39.5$, $p < .0001$; ES .18, Power 1.000). Thus, the effect size for the sex difference on this accuracy score was similar to that observed for overall performance (Table 1).

We conclude that while there are strategy differences between males and females, they cannot be considered of major importance with regard to sex differences in MRT performance.

Experience with Computer Games and Liking of the Task

We also asked students whether they were frequent players of computer games (not at all, sometimes, frequent) in order to see whether this would have an impact on MRT(A) performance. Data were available for some 438 students. A significant main effect of Sex was found ($F(1, 436) = 34.4$, $p < .00001$, ES .07, Power 1.000) which was of small magnitude. Not unexpectedly, males indicated that they played computer games more frequently than females. However, there was no correlation between frequency of computer game playing and MRT(A) performance for either sex. A subsample of 136 students was asked how well they liked doing the MRT(A), on a 7-point scale which ranged from 1 = "I loathed it" to 7 = "I love doing this sort of task," with the expectation that subjects who liked the task would do better than those who did not. There was a significant sex difference, with males giving a more positive response ($F(1, 132) = 8.68$, $p < .004$, ER .06, Power .831) and a significantly more positive response by BSc students than for BA students ($F(1, 132) = 14.5$, $p < .0001$, ES .10; Power .965). Nevertheless, and surprisingly, within these subgroups there was no significant relation between the degree to which individual students liked or disliked doing the task and the performance of the MRT task.

Finally, the same 136 students were asked to indicate whether they played with Lego blocks (a lot, fairly often, sometimes, rarely, never) when they were children. We did not find overall sex differences in the indicated frequency of playing with Lego blocks. For females, there was a modest but significant correlation between frequency of playing with Lego blocks and MRT(A) performance (.21, $p < .05$). No such relationship was found for males.

Different Versions of the Mental Rotation Test

MRT(B). Version MRT(B) consists of the same items which make up the MRT(A), but the items are reshuffled in a different order. This makes MRT(B) suitable as alternate of MRT(A), without changing the level of difficulty. In this study, we tested 66 subjects on MRT(A), and 129 subjects were given MRT(B). The results can be seen in Table 4. There were no significant differences in performance on the two test forms, and the sex difference did not interact with test version. Data are collapsed over Academic Program because this factor did not interact with Test Version. The results indicate that MRT(B) can be used as an alternate if practice effects specific to item order are to be eliminated.

MRT(C). Finally, a very difficult version of the Mental Rotation Test was produced by generating a set of 24 items which required subjects to rotate the stimuli in two directions. For instance, in order to see if the target matches the choice figures, a subject will have to rotate a figure both around the top/bottom *and* the left/right axis. Because the test was so difficult, a within subjects design was preferred, in which subjects first took the MRT(B) in order to familiarize themselves with the general task, followed by MRT(C). Ninety-four subjects were tested in this way; data were collapsed across Academic Programs because there were no significant interactions involving this variable. Table 5 indicates that, even though subjects had prior experience with version MRT(B), they still found MRT(C) very much harder. This is attested to by a large main effect of Test Version. Perhaps of the greatest interest is the interaction between Test Version and Sex, which reflects the fact that the sex differ-

TABLE 4
Comparison of MRT(A) and MRT(B), between Subjects Design

	MRT(A)		MRT(B)	
	M	F	M	F
M	12.2	8.5	11.9	7.6
SD	5.3	3.3	4.7	3.1
N	28	38	60	69

Sex $F(1, 191) = 46.3$ $p < .00001$ *ES* .20 *Power* 1.000

TABLE 5
Comparison of MRT(B) and MRT(C), within Subjects Design

	MRT(B)		MRT(C)	
	M	F	M	F
M	12.0	6.8	7.7	5.3
SD	5.4	3.7	3.7	2.5
n	54	40	54	40
	Sex $F(1, 92) = 27.8$ $p < .00001$ *ES* .23 *Power* 1.000		Sex $F(1, 92) = 12.2$ $p < .001$ *ES* .12 *Power* .931	

Sex	$F(1, 92) = 24.5$	$p < .00001$	.21	.998
Task	$F(1, 92) = 63.0$	$p < .00001$	.41	1.000
Sex × task	$F(1, 92) = 15.3$	$p < .0001$	.14	.971

ence is weaker in MRT(C) than in MRT(B). Separate ANOVAS for the main effect Sex for the two versions indicate that the magnitude of the effect size for MRT(B) is twice that for MRT(C). The theoretical implications of the changes in magnitude of sex differences with difficulty of tasks are rather more complex. Indeed, there are no convincing theoretical models that would allow a prediction of the expected changes of the magnitude of sex differences with changes of task difficulty, regardless of whether capacity or experience factors are invoked.

Menstrual Cycle Phase

The recent work by Kimura and colleagues (Gouchie & Kimura, 1991; Hampson, 1990a,b; Hampson & Kimura, 1988; Kimura, 1992) suggests that spatial test performance in females varies as a function of the phase of the menstrual cycle, and, by implication, as a function of hormonal state. The suggestion is that females perform better on spatial tasks during the menstrual period (when circulating estrogens are low) than during the midluteal portion of the cycle. Kimura and Hampson did not use the MRT task; instead, they used other spatial tasks, among them the PFT. However, Silverman and Phillips (1993) did use the MRT, and they used a method of comparing MRT performance during the menstrual cycle which allows a comparison of our data with theirs. Silverman and Phillips compared performance on Days 1 and 5 of the periods with other days, in a between-subjects design. Table 6A shows the relevant comparison for our data set. There was no significant effect for phase of cycle. However, there was an effect of Pill use, with females who used the contraceptive pill scoring somewhat higher than females who did not. This is in agreement with Silverman and Phillips' findings. It should be noted that the effect size was very small, accounting for only 3% of the variance in MRT scores.

TABLE 6
Mental Rotation MRT(A) in Relation to Menstrual Cycle

	Pill users			Non-pill users		
	N	Mean	*SD*	*N*	Mean	*SD*
(A)						
Days 1 to 5	37	10.3	(4.4)	44	8.2	(3.9)
All other days	100	10.0	(4.4)	110	8.7	(4.2)
Days of cycle effect n.s.						
Pill effect $F(1, 287) = 9.2$, $p < .003$, *ES* .03, *Power* .858						
Days 3 to 5	14	9.7	(4.3)	24	8.2	(4.2)
Days 1 to 2	23	10.6	(4.4)	20	8.3	(3.6)
Days of cycle effect n.s.						
(B)						
Days 1 to 5	37	10.3	(4.4)	44	8.2	(3.9)
Days 6 to 13	29	9.3	(5.6)	34	9.0	(4.2)
Days 14 to 17	26	9.8	(3.6)	13	9.4	(4.1)
Days 18 to 23	30	10.5	(3.9)	39	9.1	(4.1)
Days 24 to 30	15	10.6	(3.8)	24	7.2	(4.6)
	137			154		
Days of cycle effect n.s.						
Pill user $F(1, 181) = 8.3$ $p < .004$ *ES* .030 *Power* .815						

Hampson and Kimura only used Days 3 and 5 of the period because they wished to avoid possible confounding effects of menstrual distress. Table 6A shows comparisons that include a separation of Days 1–2 and 3–5, and Days 6–13 and 18–23 would be of interest because these are the phases where estrogen levels can be expected to be high (Table 6B). However, because these days were not as carefully determined as in Hampson and Kimura's study, the failure to find significant effects cannot be considered an effective challenge to Hampson and Kimura's findings. However, the absence of MRT performance differences on menstrual and non-menstrual days can be considered to be a direct failure to replicate Silverman and Phillips' (1993) findings. The data do support the somewhat higher performance in females who are taking the contraceptive pill observed by Silverman and Phillips. However, the effect sizes are small, and opposite claims have been made elsewhere (Genetta-Wadley & Swirsky-Sachhetti, 1990), albeit without supporting data.

Our conclusion is that in group MRT studies involving females, the failure to control menstrual phase is not likely to confound conclusions, even though precise determination for phases within each individual might well find a link between menstrual phase and MRT performance (Hampson & Kimura, 1992).

Practice Effects

Performance on the MRT and similar tasks is sensitive to practice (e.g., Baenninger & Newcombe, 1989; Casey & Braebeck, 1989; Hampson, 1990b; Kail, 1986). Here, the response to MRT(A) performance to practice was investigated. Table 7 shows the results from a sample of 27 subjects who were given the MRT once weekly for 4 weeks. There was a main effect of Sex and a main effect of Session, with no significant interactions, showing that sex difference was stable over practice, in agreement with the previous study. There was no significant effect of Academic Program, most likely due to the fact that there were too few subjects from the Science program. In a second study, 21 females from the BA and BSc programs were chosen on the basis of their initial scores. The scores had to be within a standard deviation of the average of the entire group to which they belonged. This was done in order to avoid unusually large "jumps" in scores due to subject who might have scored unusually badly on the initial test. The results are shown in Table 8. It can be seen that the practice effects are very marked. In a study reported by Peters, Chisholm, and Laeng (in press), a paradigm was used in which subjects were tested on MRT(A) first, and then on MRT(B). This study showed a significant transfer effect; subjects did significantly better on the MRT(B) than subjects who had not previously taken any MRT.

The practice effects have practical and theoretical implications. The practical aspect, in terms of experimental design, is best illustrated by reference to Hampson's (1990b) observation that the expected effect of period phase on spatial performance did not materialize in a within-

TABLE 7
Practice Effects: Means and *SD*s on the MRT(A) Test for Unselected Males and Females in the BA and BSc Programs

	Session							
	1		2		3		4	
	M	F	M	F	M	F	M	F
				BA students				
M	13.3	6.9	18.0	12.8	19.2	15.0	20.8	15.1
SD	3.6	5.5	3.5	4.3	3.1	6.2	2.0	6.6
				BSc students				
M	11.5	9.9	21.0	15.6	22.0	17.6	22.5	20.1
SD	2.1	3.7	2.8	5.9	1.4	4.9	.7	3.2

Sex effect $F(1/23) = 5.4$ $p < .03$ *ES* .19 *Power* .605
Session effect $F(3, 69) = 45.8$ $p < .0001$ *ES* .67 *Power* 1.000
Program effect ns

Note. BA males, $n = 6$; BSc males, $n = 2$; BA females, $n = 8$; BSc females, $n = 11$.

TABLE 8
Practice Effects: Means and *SD*s on the MRT(A) Test for Selected Females in the BA (n = 13) and BSc (n = 8) Programs

	Session						
	1	2	3	4	5	6	7
				BA			
M	7.0	10.8	13.3	14.8	16.3	17.6	19.4
SD	3.1	4.8	5.7	5.0	4.1	3.9	3.2
				BSc			
M	9.2	12.1	17.3	19.5	20.8	21.5	21.9
SD	3.4	3.4	3.8	4.2	3.3	2.4	2.0

Program effect $F(1/20)$ = 6.4 $p < .02$ *ES* .25 *Power* .672
Session effect $F(6,114)$ = 73.9 $p < .0001$ *ES* .80 *Power* 1.000

Note. BA, n = 13; BSc, n = 8.

subject design, as opposed to a between subjects design. She attributed this to the strong practice effects for her tasks and this concern can be extended to the MRT as well. It should be pointed out that the extreme responsiveness of MRT performance to practice contrasts sharply to some other behaviors in which sex differences between males and females have been found, such as a fine motor task. Here, even very extensive practice brings only minimal improvements in performance (Peters, 1981).

The MRT in Relation to the Paper Folding Test and the Card Rotation Test

Performance data on the PFT were available for 603 subjects. Table 9 shows that there was an effects of sex for this test which approached

TABLE 9
Picture Folding Test (PFT) Performance for 603 Subjects

	Mean	*SD*	*N*
Arts			
Males	12.0	3.8	94
Females	11.4	3.3	213
Sciences			
Females	14.2	3.2	129
Males	13.7	3.0	167
Total	12.7	3.5	603

Program effect $F(1, 599)$ = 66.64 $p < .0001$ *ES* .10 *Power* 1.000
Sex effect $F(1/599)$ = 3.69 $p < .055$ *ES* .01 *Power* .482

significance, and there were significant effects of Academic Program. Data were also available for 54 males and 47 females who had all taken the MRT, the PFT, and the Card Rotation Test, and this allowed comparison of sex differences across the tasks for the same group of subjects (Table 10). The subjects performed the tests in the same order, with the Card Rotation Test first, the PFT second, and the MRT third.

Table 10 illustrates that this group did not show significant sex differences for the PFT and the Card Rotation Test, but the sex differences on the MRT were comparable to those observed for other groups of subjects. The results confirm our impression that the PFT should not be used as a test that yields reliable sex differences for populations of university students. A similar statement can be made with regard to the Card Rotation test.

GENERAL DISCUSSION

Our data show that the redrawn version of the MRT yields sizable and replicable sex differences. This makes the current version acceptable for general use. Indeed, in terms of paper and pencil tests of spatial abilities, the MRT remains the most convincing test in terms of demonstrating sex differences, as shown by the fact that the versions of MRT described here yielded consistent differences. The magnitude of the differences can be expressed in different ways. First, there is the expression of the magnitude as a ratio formed by dividing the difference between the male and female means by the common standard deviation (d'). However, because of the considerable overlap, we feel that the "percent of variance accounted for" provides a more informative measure of the effect size. Using Cohen's (1969) proposed classification of effect size (.25 = small, .25 to .50 = medium, > .50 = large), our observed effect sizes fell into the small–lower end of the medium range. There is some suggestion in the literature (Emanuelson & Svensson, 1986; Stumpf & Klieme, 1989; Richardson, 1994; Voyer, Voyer & Bryden, in press) that the magnitude

TABLE 10
Performance by the Same Set of Subjects of the Card Rotation Task, the Paper Folding Task (PFT), and the Mental Rotation Task, MRT(A)

	Card rotation		Paper folding		Mental rotation	
	M	F	M	F	M	F
n	54	47	52	47	54	47
M	108.7	107.8	11.2	11.0	12.8	8.9
SD	29.0	33.2	3.5	3.2	5.5	4.1
Sex	n.s.		n.s.		$F(1, 99) = 15.8$ $p < .0001$, *ES* .14 *Power* .976	

of effect sizes has fallen over recent years. However, based on comparisons of d′ for the differences obtained here and estimates of d′ for the original Vandenberg & Kuse study, there is no discernible reduction in effect for the values obtained in the present study.

In terms of comparisons with two-dimensional rotation tests (Berenbaum, 1990) and the picture folding test (Gouchie & Kimura, 1991), which also yield sex differences, the effect sizes of the MRT are considerably more convincing; in the present study, the Card Rotation Test and the Picture Folding Test failed to yield statistically significant sex differences for the very subjects who showed a sex difference on the MRT(A), even though the means favored males throughout.

Part of the reason for constructing a very difficult version of the MRT test was that sex differences would emerge more clearly with greater task difficulty, as would be predicted if the sex differences were a function of capacity differences. This proved not to be the case, and the interpretation of the sex differences remains elusive; it is well to remember that some 80% of the difference in performance between individuals is due to factors other than sex. In addition, the practice effects demonstrate that females can show dramatic improvements in mental rotation performance after what would have to be considered minimal practice. The term "minimal" is used relative to the acquisition of motor skills where rather large amounts of practice are required to yield tangible improvements in performance. At the very least, the substantial practice effects have to be taken into account in repeat testing of the same subjects, and alternate versions of the test should be considered.

Because the primary purpose of this paper to provide background information for the updated version of the Vandenberg & Kuse MRT, relatively little attention was paid to substantive issues concerning the rotation of 3D images. However, even parametric information can be of some interest in this regard. Often, the claim is made that the better spatial performance by males is related to better skill in mathematical reasoning and science-related activities in general. Such claims have practical implications in terms of selection for specific occupations. However, the link between spatial ability as assessed by MRT (and there is currently no better test) and performance in mathematics and science is by no means well established. In a recent study, Peters, Chisholm, and Laeng (in press) showed that while the effect size which separates female and male engineering students is just as large as for other samples, there is no significant sex difference in academic performance in physics, chemistry, calculus, computer programming, and engineering design. Thus, if there is a relationship between MRT performance and mathematics and other science subjects, it is not obvious.

The relationship between MRT performance and targeting is of somewhat greater interest. There is a very marked sex difference in throwing at targets (Kimura, 1992; Peters, 1990; Watson & Kimura, 1992). Although evolutionary arguments have been made for the importance of throwing and targeting in their own right (Calvin, 1982), there is a possible link between targeting and MRT performance. Targeting involves action in a modelled three dimensional space, as well as the ability to anticipate the outcome of action. Pellizer and Georgopoulos (1993) have suggested that there is a direct link between mental rotation and motor mechanisms. Further inquiries may well show that the correlation between MRT performance and targeting is not spurious, but instead related to an underlying common mechanism in manipulating objects in space.

APPENDIX

1.
__ I rotated the whole figure in my mind when making the comparison
__ I rotated a section of the figure in my mind when making the comparison
__ I am not sure how I did it
__ Other (explain) ______________________________

2.
__ I thought through the steps verbally in my mind (i.e. "two cubes up and three down")
__ I relied mainly on visualizing the figures and did not talk myself through the steps
__ I am not sure

3.
__ I used movements of my finger, hand and/or pencil to help me with the task
__ I did not use movements of my finger, hand and/or pencil to help me with the task

4.
__ I scanned the options for the most likely match and then made my choices
__ I went through the options systematically, trying the first, then the second etc.
__ I went through the options in a haphazard nonsystematic way
__ Other (specify) ______________________________

5.
__ I always compared the options to the target figure
__ Once I found the match, I compared the rest of options to the match
__ I did a bit of both

6.
__ I developed a specific approach to solve the problems
__ I tried various approaches to solve the problems
__ I had no specific approach

7.
__ I was more concerned with getting the right answers than I was about the time limit
__ I was more concerned with getting all the answers completed than I was about getting the correct answers
__ I did not care how I did

8.
__ I double checked my answers before moving on to the next problem
__ I was vaguely confident of my answers before I moved to the next problem
__ I was unsure of my answers before moving on to the next problem
__ I guessed most of the time

REFERENCES

Allen, M. J. 1974. Sex differences in spatial problem solving styles. *Perceptual and Motor Skills*, **85**, 249–277.

Allen, M. J., & Hogeland, R. 1978. Spatial problem-solving-strategies as functions of sex. *Perceptual and Motor Skills*, **47**, 348–350.

Alyman, C., & Peters, M. 1993. Performance of male and female children, adolescents and adults on spatial tasks that involve everyday objects and settings. *Canadian Journal of Psychology*, **47**, 730–747.

Baenninger, M., & Newcombe, N. 1989. The role of experience in spatial test performance: A meta-analysis. *Sex Roles*, **20**, 327–344.

Berenbaum, S. A. 1990. Congenital adrenal hyperplasia: Intellectual and psychosexual functioning. In C. Holmes (Ed.), *Psychoneuroendocrinology: brain, behaviour, and hormonal interactions*. New York: Springer-Verlag.

Bouchard, T. J., & McGee, M. 1977. Sex differences in human spatial ability: not an x-linked recessive gene effect. *Social Biology*, **24**, 332–335.

Calvin, W. H. 1982. Did throwing stones shape hominid brain evolution? *Ethology and Sociobiology*, **3**, 115–124.

Casey, M. B., & Brabeck, M. M. 1989. Exceptions to the male advantage on a spatial task: Family handedness and college major as factors identifying women who excel. *Neuropsychologia*, **27**, 689–696.

Casey, M. B., & Brabeck, M. M. 1990. Women who excel on a spatial task: Proposed genetic and environmental factors. *Brain and Cognition*, **12**, 73–84.

Casey, M. B., & Brabeck, M. M., & Ludlow, L. H. 1986. Familial handedness and its relation to spatial ability following strategy instructions. *Intelligence*, **10**, 389–406.

Casey, M. B., Pezaris, E., & Nuttall, R. L. 1992. Spatial ability as predictor of math achievement: The importance of sex and handedness patterns. *Neurospychologia*, **30**, 34–45.

Cochran, K., & Wheatley, G. H. 1989. Ability and sex-related differences in cognitive strategies on spatial tasks. *Journal of General Psychology*, **116**, 43–55.

Cohen, J. 1969. *Statistical power analysis for the behavioral sciences*. New York: Academic Press.

Ekstrom, R. B., French, J., Harman, H. H., & Dermen, D. 1976. *Kit of factor-referenced cognitive tests*. Princeton, NJ: Educational Testing Service.

Emanuelson, I., & Svensson, A. 1986. Does the level of intelligence decrease? A comparison between 13-year-olds tested in 1961, 1966 and 1980. *Scandinavian Journal of Educational Research*, **30**, 25–37.

Genetta-Wadley, A., & Swirsky-Sachhetti, T. 1990. Sex differences and handedness in hemispheric lateralization of tactile-spatial function. *Perceptual and Motor Skills*, **70**, 579–590.

Gouchie, K., & Kimura, D. 1991. The relationship between testosterone levels and cognitive ability patterns. *Psychoneuroendocrinology*, **16**, 323–334.

Hampson, E. 1990a. Estrogen-related variations in human spatial and articulatory-motor skills. *Psychoneuroendocrinology*, **15**, 97–111.

Hampson, E. 1990b. Variations in sex-related cognitive abilities across the menstrual cycle. *Brain and Cognition*, **14**, 26–43.

Hampson, E., & Kimura, D. 1992. Sex differences and hormonal influences on cognitive function in humans. In J. B. Becker, S. M. Breedlove & D. Crews (Eds.), *Behavioral Endocrinology*. Pp. 357–398.

Hampson, E., & Kimura, D. 1988. Reciprocal effects of hormonal fluctuations on human motor and perceptual-spatial skills. *Behavioral Neuroscience*, **102**, 456–459.

Harris, L. J. 1978. Sex differences in spatial ability: Possible environmental, genetic, and neurological factors. In M. Kinsbourne (Ed.), *Asymmetrical function of the brain*. New York: Cambridge University Press. Pp. 405–202.

Harshman, R. A., Hampson, E., & Berenbaum, S. A. 1987. Normal variation in human brain organization: Relation to handedness, sex and handedness differences in ability. *Canadian Journal of Psychology*, **37**, 144–192.

Jahoda, G. 1980. Sex and ethnic differences on a spatial perceptual task: Some hypotheses tested. *British Journal of Psychology*, **71**, 425–431.

Kail, R. 1986. The impact of extended practice on rate of mental rotation. *Journal of Experimental Child Psychology*, **42**, 378–391.

Kail, R., Carter, P., Pellegrino, J. 1979. The locus of sex differences in spatial ability. *Perception & Psychophysics*, **26**, 182–186.

Kimura, D. 1994. Body asymmetry and intellectual pattern. *Personality and Individual Differences*, **17**, 53–60.

Kimura, D. 1992. Sex differences in the brain. *Scientific American*, **267**(3), 118–125.

Lee, A., Fant, H., Life, M. L., Lipe, L., & Carter, J. A. 1978. Field independence and performance on ball-handling tasks. *Perceptual and Motor Skills*, **46**, 439–442.

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

Olson, D., & Eliot, J. 1986. Relationships between experiences, processing style, and sex-related differences in performance on spatial tests. *Perception and Motor Skills*, **62**, 447–460.

Oosthuizen, S. 1991. Sex-related differences in spatial ability in a group of South African students. *Perceptual and Motor Skills*, **73**, 51–54.

Pellizer, G., & Georgopoulos, A. P. 1993. Mental rotation of the intended movement. *Current Directions in Psychological Science*, **2**, 12–17.

Peters, M. 1990. Subclassification of lefthanders poses problems for theories of handedness. *Neuropsychologia*, **28**, 279–289.

Peters, M. 1981. Effects of prolonged practice on between hand performance differences. *Neuropsychologia*, **19**, 587–590.

Peters, M., & Perry, R. 1991. No link between lefthandedness and maternal age and no elevated accident rate in lefthanders. *Neuropsychologia*, **29**, 1257–1259.

Peters, M., Chisholm, P., & Laeng, B. (in press). Do engineering students show sex differences on a spatial ability test, and do such differences relate to performance in academic subjects? *Journal of Engineering Education*.

Richardson, J. T. E. 1994. Gender differences in mental rotation. *Perceptual and Motor Skills*, **78**, 435–448.

Schultz, K. 1991. The contribution of solution strategy to spatial performance. *Canadian Journal of Psychology*, **45**, 474–491.

Shepard, R. N., & Cooper, L. A. 1982. *Mental images and their transformations*. Cambridge, MA: MIT Press.

Silverman, I., & Phillips, K. 1993. Effects of estrogen changes during the menstrual cycle on spatial performance. *Ethology and Sociobiology*, **14**, 257–270.

Stumpf, H., & Klieme, E. 1989. Sex-related differences in spatial ability: More evidence for convergence. *Perceptual and Motor Skills*, **69**, 915–921.

Vandenberg, S. G., & Kuse, A. R. 1978. Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills*, **47**, 599–601.

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

Watson, N. V., & Kimura, D. 1989. Right-hand superiority for throwing but not for intercepting. *Neuropsychologia*, **27**, 1399–1989.