

Time limits and gender differences on paper-and-pencil tests of mental rotation: a meta-analysis

Daniel Voyer

Published online: 17 November 2010
© Psychonomic Society, Inc. 2010

Abstract The purpose of this study was to summarize research relevant to the influence of time limits on gender differences in paper-and-pencil tasks of mental rotation by means of a meta-analysis. Thirty-six effect sizes from published and unpublished studies examining the influence of time constraints were retrieved. Results showed that gender differences in mental rotation are significantly larger when the task is administered with time constraints compared to when such constraints are absent. In addition, the magnitude of gender differences was linearly related to the amount of time available for test completion. These findings were not related to the age or the year of birth of participants in the retrieved studies. The results are discussed with regard to their implications for explanations of gender differences in mental rotation and cognitive abilities in general.

Keywords Gender differences · Mental rotation · Meta analysis · Spatial cognition · Time pressure

Gender differences in favor of men in paper-and-pencil tests of mental rotation are well documented (Hedges & Nowell, 1995; Linn & Petersen, 1985, 1986; Voyer, Voyer, & Bryden, 1995). Furthermore, this type of task produces some of the largest cognitive gender differences according to meta-analytic findings (Linn & Petersen, 1985; Voyer et al., 1995).

Considering their prevalence, many explanations have been offered to account for these gender differences (see Halpern, 2000). However, the present review is only

concerned with one of the possible factors relevant to gender differences in mental rotation. Specifically, in their discussion of the role of performance factors in mental rotation, Goldstein, Haldane, and Mitchell (1990) stated that women work slowly and cautiously when completing a mental rotation task, whereas men are more likely to guess and work faster. Following this reasoning, Goldstein et al. (1990) claimed that these performance factors accounted for the observed gender differences in mental rotations. Among others, this statement led Goldstein et al. (1990) to hypothesize that gender differences in mental rotation should be more pronounced when time constraints are present compared to when participants are given an unlimited amount of time to complete the task. The experiments presented by these authors supported this hypothesis as gender differences became non-significant on the Vandenberg and Kuse Mental Rotations Test (MRT: Vandenberg & Kuse, 1978) when time limits were removed.

The two experiments conducted by Goldstein et al. (1990) were the basis for the suggestion that men and women approach mental rotation tasks differently, and that such performance factors would account for gender differences on the MRT. However, other researchers reported evidence contrary to the Goldstein et al. (1990) hypothesis. For instance, Masters (1998) used a larger number of participants than in the Goldstein et al. (1990) studies, thus reducing the risk of a Type II error. She also implemented an independent groups design in her manipulation of time limits (as opposed to the within-subject design used by Goldstein et al.). The results of Masters' study showed that the magnitude of gender differences actually increased slightly but not significantly when the MRT was administered without time limits.

The contradictory results reported by Goldstein et al. (1990) and Masters (1998) both find echoes in the

D. Voyer (✉)
Department of Psychology, University of New Brunswick,
PO Box 4400, Fredericton, New Brunswick E3B 5A3, Canada
e-mail: voyer@unb.ca

literature. For example, Voyer and Sullivan (2003) reported a reduction in the magnitude of gender differences when time constraints were removed on the MRT. In contrast, Delgado and Prieto (1996) found a slight increase in the magnitude of gender differences when time constraints were relaxed on a test of mental rotation. These contradictory findings emphasize the variability of the results obtained when one examines the influence of such performance factors in tests of mental rotation. Considering the apparently irreconcilable difference between the results of individual studies, it is fruitful to combine the data available in a meta-analysis. In fact, a meta-analysis is likely to shed light on the state of affairs concerning the influence of time limits on mental rotation test performance. In doing so, it should allow clearer conclusions to be drawn concerning the role of this factor on the magnitude of gender differences in these tests.

Meta-analysis permits the combination and comparison of a set of studies relevant to an area of research (Rosenthal, 1991). Using this approach to quantify the influence of time constraints on gender differences in tests of mental rotation should therefore allow more definite conclusions. Accordingly, the purpose of this study was to examine the hypothesis that time constraints affect the magnitude of gender differences in tests of mental rotation by means of a meta-analysis of retrievable data.

Method

Selection criteria for inclusion in the meta-analysis

The initial goal of this meta-analysis was to assess the influence of timing manipulations on the magnitude of gender differences in all cognitive tests where such differences have been observed. However, in retrieving literature, it quickly became clear that such manipulations have been implemented only for paper-and-pencil tests of mental rotation and that considering other tests or formats would not provide a sufficient number of effect sizes to make their inclusion meaningful. Accordingly, this meta-analysis includes published as well as unpublished studies presenting results obtained with paper-and-pencil tests of mental rotation. Following the definition provided by Linn and Petersen (1985), a test of mental rotation was defined as a test measuring the ability to rotate quickly and accurately two- or three-dimensional figures, in imagination. Of course, there are many published studies in which the authors administered tests of mental rotation exclusively under the time limits recommended by the test designers as they followed the standard administration procedure. However, as the purpose of the present analysis was to

determine whether time pressures affect the magnitude of gender difference, a deviation from the standard timing procedure was necessary for inclusion of a specific study as this was critical to an examination of the question of interest. As such, studies that relied exclusively on a standard administration of mental rotation tests were excluded. If they were included, the numerous studies relying on a standard administration of paper-and-pencil mental rotation tests would produce a sample of effect sizes where such studies would overwhelm the limited sample of studies with non-standard administration. Heterogeneity in the resulting sample of effect sizes would be potentially due in a large part to differences in sampling and procedural factors within the pool of studies that used exclusively a standard time limit. This would ultimately shift the research question and defeat the purpose of conducting the present analysis. Inclusion of only those studies that had at least one condition with non-standard timing likely minimizes this extraneous variance and isolates better the variable of interest. Accordingly, the sample was limited either to studies where timing conditions were manipulated, studies where only an unconstrained timing condition was used, or studies where a particularly long time limit was implemented.

In addition, a further distinction was required between time limits as there was some variability in the actual time limit used in studies where time constraints were applied. Specifically, in the final sample, the time limit varied from 5 to 20 min for test completion, and this varied both within and across the actual tests administered. For example, Delgado and Prieto (1996) administered a Spanish adaptation of the Rotation of Solid Figures test with a limit of either 5 min (speed condition) or 15 min (power condition), whereas the MRT is typically administered with a limit of 6 min (e.g., Peters, 2005), although a limit of 10 min has also been used (e.g., Voyer and Sullivan, 2003). Accordingly, rather than using a simple timed /untimed distinction, studies were coded as having a time limit of short (2 to 6 min) or long (10 to 20 min) duration, or no time limit at all. This distinction allows a finer grained analysis of the influence of time limits on gender differences in mental rotation tests.

PsycInfo searches were conducted in an attempt to retrieve research published in all media contained in the PsycInfo database (peer-reviewed or not, dissertations, conference proceedings, etc.) with the search terms “cognitive tests” or “intelligence” along with “gender” or “sex” as well as with “timing” “timed,” “time limits,” or “timing condition” at first. Then simply “mental rotation” coupled with “gender” or “sex” as well as with “timing” “timed,” “time limits,” or “timing condition” when it became clear that timing had been manipulated

only for tests of mental rotation. In addition, the reference list of papers obtained through this search was examined closely for relevant studies. In an attempt to gather more studies, an e-mail message requesting published or unpublished data relevant to the purpose of the present study was sent to several researchers interested in gender differences on mental rotation tests whose research was retrieved in the PsycInfo search and for whom a current e-mail address could be obtained. This request was sent to 27 researchers and received a 51.9% response rate (14/27), although only four unpublished data sets were retrieved in this manner. Presentation of the preliminary results from the present analysis at the Spatial Learning Conference at Harvard University in May 2010 also resulted in two additional sets of relevant unpublished data.

The studies selection procedure resulted in the sampling of 36 effect sizes drawn from 26 separate studies, 6 of which (23.1%) were unpublished. Note that papers presented at professional meetings were counted as unpublished because they were not published in a peer-reviewed journal. The effect sizes entered in the analysis are presented in Table 1, and the relevant studies are marked with an asterisk in the reference list. It is important to note that a few studies were relevant to a manipulation of time limits, but did not present the methodological or statistical information required for inclusion in the present analysis. In such cases, an e-mail requesting the missing information was sent to the authors, resulting in a 100% response rate, although for one 23-year-old study, the information was no longer retrievable. In any case, supplemental information obtained directly from the authors accounts for deviations between the information presented in the published research and that presented for individual studies in Table 1.

It should also be noted that, for studies that included a training or practice component, only pre-training data were considered for effect sizes calculations. Finally, when the research involved a manipulation that purported to affect the magnitude of gender differences (e.g., the manipulation of speed/accuracy emphasis in Scali, Brownlow, & Hicks, 2000), only the overall main effect of gender, collapsed across such conditions, was considered.

The present meta-analysis is presumed to provide an exhaustive review of the published literature on the influence of time constraints on gender differences in tests of mental rotation. Considering the small number of published studies, the inclusion of unpublished data sets should increase the generalizability of the present analysis in view of the reduced likelihood of publication when negative findings are obtained (Rosenthal, 1979). Nevertheless, as the present analysis included mostly published studies, this leaves open the possibility that it relies on a biased sample of the existing studies, based on

the assumption that only experiments with significant results are published (Rosenthal, 1979). This “file drawer problem” (Rosenthal, 1979) is likely to produce an overestimation of the effect sizes. In this situation, the number of studies averaging null results necessary to offset the significance of the findings at the .05 level (fail-safe number) is typically computed. This value was therefore calculated in the present analysis to estimate the resistance of the meta-analytic results to the file drawer problem. The larger the fail-safe value, the more confidence one can have in the obtained results. As a rule of thumb, Rosenthal (1991) suggested that we should reject the hypothesis that significant results are due to the file drawer problem when the fail-safe value exceeds a criterion of 5 times the number of sampled studies +10 ($5k + 10$). This criterion should be kept in mind when results are presented.

Analysis procedure

Cohen’s d was used as the measure of effect size (Cohen, 1977). This index represents the standardized difference between the mean of the groups under study (women and men in the present analysis). Effect sizes were computed using the formula presented by Cohen (1977) when means and standard deviations were available, or using the formulae presented by Wolf (1986) when only the t , p , or F statistic was available. In the present analysis, effects were computed in such a way that a positive effect size reflected a difference in favor of men. However, as d is considered a biased estimate of effect sizes, it was corrected based on the approach presented by Hedges and Becker (1986) to obtain an unbiased estimate for use in the analysis.

The meta-analysis followed the procedure presented by Hedges and Becker (1986). These authors developed meta-analytic techniques based on a fixed effects model designed for the assessment of cognitive gender differences and for the evaluation of the homogeneity of effect sizes. Homogeneity of effect sizes allows for the conclusion that the studies included in a specific meta-analysis can be considered replications of each other and that a pooled estimate of effect size provides a valid summary of the results from the sample of studies. However, when heterogeneity is detected, it is likely that the pooled estimate is not representative of the state of affairs in a sample. When this is the case, the effect sizes have to be partitioned further to achieve homogeneous groupings.

In addition, the meta-analysis followed the hierarchical approach outlined by Hedges and Becker (1986). Thus, an overall analysis examining the magnitude and the homogeneity of gender differences was first conducted, followed by partitioning into homogeneous clusters. However, some

Table 1 Studies on the influence of time limits on gender differences in paper-and-pencil tests of mental rotation

Studies	nM	nW	Test	Age	Design	Limit	<i>d</i>
Brownlow et al. (2003)	51	78	VRT	21	B/S	None	0.67
Brownlow and Miderski (2002)	20	24	MRT	21	B/S	None	0.44
Carr et al. (2007)	91	103	MRT	9	B/S	None	0.51
Cooke-Simpson and Voyer (2007)	40	40	MRT	20	B/S	None	0.66
Delgado and Prieto (1996)	309	390	RFM	18	B/S	5	0.84
	312	431	RFM	18	B/S	15	0.95
Dollinger and Patrick (1993)	244	287	MRT	29	B/S	None	0.60
Gallagher (1989)	118	121	Cubes	14	W/S	2	1.50
	118	121	Cubes	14	W/S	None	0.36
Gallagher and Johnson (1992)	114	117	Cubes	14	W/S	2	1.40
	114	117	Cubes	14	W/S	None	0.33
Glück and Fabrizii (2010)	62	83	MRT	16	W/S	6	1.02
	62	83	MRT	16	W/S	None	0.59
Goldstein et al (1990; Exp. 2)	23	23	MRT	23	W/S	6	0.83
	23	23	MRT	23	W/S	None	0.60
Heil and Jansen-Osmann (2008)	62	47	Other	8	B/S	None	0.50
Janas et al. (2006)	33	32	VRT	21	B/S	None	0.48
Kanoy et al. (2009)	33	33	VRT	21	B/S	None	0.40
MacLeod (2000)	32	53	MRT	20	B/S	10	0.79
	25	37	MRT	20	B/S	None	1.05
Masters (1998)	35	35	MRT	21	B/S	6	0.86
	35	35	MRT	21	B/S	None	0.96
Miller and Halpern (2010)	48	29	MRT	18	B/S	20	0.74
Ortner and Sieverding (2008)	87	74	Cubes	32	B/S	None	0.23
Peters (2005)	45	61	MRT	20	B/S	6	0.87
	43	63	MRT	20	B/S	12	0.68
Resnick (1993)	88	94	MRT	21	B/S	None	0.75
Scali and Brownlow (2001)	48	48	MRT	21	B/S	None	0.32
Scali et al. (2000)	58	58	MRT	21	B/S	None	0.40
Titze et al. (2008)	150	150	MRT	24	B/S	None	0.42
Voyer (1997)	97	117	MRT	20	B/S	10	0.67
	58	90	MRT	20	B/S	None	0.29
Voyer and Doyle (2010)	86	109	MRT	20	B/S	None	0.73
Voyer and Hou (2006)	100	103	MRT	20	B/S	None	0.61
Voyer and Sullivan (2003)	106	153	MRT	20	B/S	10	0.85
	111	132	MRT	20	B/S	None	0.46

nM = number of men; nW = number of women. Test: Cubes = Cubes Test; MRT = Mental Rotations Test; RFM = Spanish adaptation of the Rotation of Solid Figures test; VRT = Purdue Visualization of Rotation Test. Age: Mean age of the sample in years. Limit: Time limit in minutes. Design: B/S = Between-subjects; W/S = Within-subject

authors had the same group of participants performing the task with time limits and then complete remaining items without time limit. Such within-subject designs, denoted with “W/S” in Table 1, violate the assumption of independence of the effect sizes in the approach used here (see Rosenthal, 1991). Non-independent effect sizes should be analyzed differently (Hedges & Olkin, 1985), but this approach would require the correlation between the two measures, and this statistic was unavailable in all retrieved studies of relevance. Despite this problem, the assumption of non-independence was only violated for the overall analysis, similar to what occurred in the Linn and Petersen (1985) and Voyer et al.

(1995) meta-analyses. Accordingly, the influence of this factor was unlikely to affect the observed results in a meaningful way when effects sizes were partitioned as a function of timing conditions or other variables.

The Hedges and Becker (1986) approach also allows one to examine whether a given variable has a significant effect on the magnitude of effect sizes. Specifically, at each step, a test is calculated to determine whether the partitioning applied to the data had a significant effect on the magnitude of effect sizes. This test examines whether the difference between the heterogeneity for the whole sample (total heterogeneity) and that for the sum of the partitions (within-

group heterogeneity) results in a significant amount of between-group heterogeneity. This approach can thus be interpreted as a test of whether the specific variable used in partitioning produced significant between-group heterogeneity. This is essentially the same as determining whether a factor produces significant group differences in the context of analysis of variance (Hedges & Becker, 1986).

Results

Overall analysis

The analysis of the 36 effect sizes obtained on mental rotation tasks revealed a mean weighted d of 0.70 ($z = 27.28$, $p < .01$), demonstrating that, overall, gender differences in mental rotation tasks favoring men are large and significant in the studies retrieved here. The fail-safe analysis indicated that 7,609 studies with non-significant or contrary results would be needed to offset the significance of the mean effect size at the .05 level. The findings are thus resistant to the file drawer problem. As an additional assessment of a possible publication bias, the correlation between the number of participants and the effect size was calculated. The use of this index is based on the notion that studies that have a small effect size obtained with a small sample size are less likely to get published. Accordingly, a significant negative correlation would suggest that the sample is biased as it would likely be missing some studies with a small effect size and sample size. The present sample showed a correlation of .159 ($p > .35$), suggesting no influence of a publication bias from this perspective.

However, the effect sizes were not homogeneous, $\chi^2(35) = 126.46$, $p < .01$. This suggests that the studies included in the present analysis are not all drawn from the same population and that the pooled estimate of effect size does not provide a representative summary of the sample of effect sizes. Thus, while the gender differences are significant, they are also heterogeneous. Partitioning of the effect sizes into homogeneous clusters was therefore required.

Time limits as a partitioning factor

Effect sizes were first partitioned as a function of time limits (short, long, or none) as seen in Table 2. This partition produced significant between-group heterogeneity, $\chi^2(2) = 73.96$, $p < .01$, reflecting a significant effect of timing conditions on the magnitude of the advantage in favor of men. In addition, when the actual time limit was correlated with the unweighted effect size (including no time limits conditions coded as one unit above the longest limit observed in the retrieved studies, that is $20+1 = 21$), the obtained correlation was $-.73$ ($p < .01$, $N = 36$), although the correlation dropped to $-.40$ ($p < .018$) when weighted effect sizes were used in the correlation analysis. In addition, comparisons among clusters, computed as outlined by Hedges and Becker (1986), showed that the condition without time limits produced significantly smaller gender differences than both the other timing conditions (short vs. no time limits: $z = 8.08$, $p < .01$; long vs. no time limits: $z = 5.27$, $p < .01$). In addition, the “short time limits” partition produced significantly larger effect sizes than the “long time limits” grouping, $z = 2.32$, $p < .05$. However, partition of the effect sizes on time limits still produced heterogeneity, but only for the short time limit effect sizes (see Table 2). Further partitioning was therefore required to achieve homogeneous clusters for the short time limits grouping.

Specific test partitioning for short time limits

The short time limits cluster included results obtained with three different tests: the Spanish adaptation of the Rotation of Solid Figures test (RFM), the MRT, and the Cubes test. As only one effect size was available for the RFM, tests were grouped in two clusters: Cubes or others (that is, RFM and MRT). As seen in Table 3, partition of the effect sizes on this factor resulted in two homogeneous clusters and reflected significant between-group heterogeneity, $\chi^2(1) = 22.69$, $p < .01$. This finding indicated a significantly larger advantage in favor of men with the Cubes test ($d = 1.45$) than with other tests ($d = 0.86$) when administered with short time limits.

Table 2 Effect sizes (ES) for gender differences in tests of mental rotation as a function of time limits

Time limits	k	Weighted estimator of ES	Test of significance for ES (Z)	Fail safe	Homogeneity statistic (χ^2)
Overall	36	0.70	27.28*	7609	126.46*
Short time limits	7	1.03	18.75*	689	23.75*
Long time limits	6	0.85	15.43*	368	4.33 ^a
No time limits	23	0.51	15.12*	1717	24.42 ^a

* $p < .05$

^a Homogeneity achieved

Table 3 Effect sizes (ES) for gender differences in tests of mental rotation for short time limits

Tests	k	Weighted estimator of ES	Test of significance for ES (Z)	Fail safe	Homogeneity statistic (χ^2)
Short time limits	7	1.03	18.75*	689	23.75*
Cubes	2	1.45	13.95	117	0.23 ^a
Others (RFM/MRT)	5	0.86	13.40	234	0.83 ^a

* $p < .05$ ^a Homogeneity achieved

Supplemental analyses

Age of sample, year of publication, and time limits

Considering that the magnitude of gender differences in spatial performance has been shown to increase with age (Linn & Petersen, 1985; Voyer et al., 1995), it would be possible that age of the sample varied systematically with time limits in the studies retrieved here, and this could account for the observed findings. It is therefore important to discount the possible influence of this factor on the results. Essentially, this requires demonstrating that there is no correlation between the time limit and the age of the sample. Note that the age used for this analysis (and shown in Table 1 for each study) was based on the values reported in the articles themselves. When the mean age was not reported, it was assumed that children in Grade 1 are usually 6 year olds, whereas first-year undergraduate students are typically 19 years old, following the approach used by Voyer et al. (1995). Finally, a mean age of 21 years was assumed when a mixed undergraduate sample was used and no mean age was reported. Using this approach, the correlation between age of the sample and the time limit used (again with “no time limit” conditions coded as 21) was .18 ($p > .30$). Similarly, considering time limit as the categorical variable presented in Table 2 produced no significant effect of time limit category when age of the sample was used as the dependent variable, $F(2, 33) = 0.55$, $p > .57$, $MSe = 19.60$. It is thus unlikely that age of the sample could account for the observed influence of time limits on the magnitude of gender differences. As a matter of interest, it is also worth noting that the correlation between age and the unweighted effect size was $-.25$ ($p > .13$) in the present sample. However, one should not draw definite conclusions from the direction and lack of significance of this correlation between age and magnitude of gender differences considering the restricted range of ages in the sample (from 8 to 32).

There have been reports that year of publication could be related with a decline in the magnitude of cognitive gender differences, and this was linked to changes in the social environment (Feingold, 1988), although this claim found no support for mental rotation tasks in the Voyer et al. (1995)

meta-analysis. Regardless of the replicability of this finding on various tests, it is also simple to demonstrate that the social environment in which one was raised (measured as year of birth; see Voyer et al., 1995) had no effect on the results. Specifically, there was no significant correlation between year of birth and the time limit used (once more with “no time limit” conditions coded as 21), $r = .17$, $p > .33$. Similarly, considering time limit as a categorical variable produced no significant effect of time limit category when year of birth was used as the dependent variable, $F(2, 33) = 0.33$, $p > .71$, $MSe = 61.19$. Year of birth is thus unlikely to account for the observed influence of time limits on the magnitude of gender differences. As this might interest some readers, the correlation between year of birth and the unweighted effect size was $-.11$ ($p > .53$) in the present sample. Again, however, it is not possible to draw definite conclusions from the lack of correlation between year of birth and the magnitude of gender differences as the present study relied on a restricted range for year of birth (from 1963 to 2000).

Distribution of effect sizes, outliers, and homogeneity

The present analysis has produced generally homogeneous clusters of effect sizes. However, the results presented in Table 3 suggest the possibility that the Cubes test might reflect outlying results when administered with a very short time limit. The distribution of effect sizes as a function of effect size range (in increments of 0.1) and time limit conditions presented in Fig. 1 allows a closer examination of this possibility. Figure 1 shows a broad distribution of effect sizes for no time limits conditions, although they tend to be grouped at the lower end. In contrast, short and long time limits seem to be more restricted in range, with much overlap between these two time limits categories. In addition, although no effect sizes were found in the 1.1 to 1.2 and 1.2 to 1.3 categories, the Cubes test accounts for the effect sizes obtained in the two highest categories. As was seen earlier (Table 3), the large effect sizes obtained with this test account for the heterogeneity of variance in the short time limits cluster. In fact, when the results obtained with either short or long time limits were grouped as one cluster and effect sizes obtained with the Cubes test were

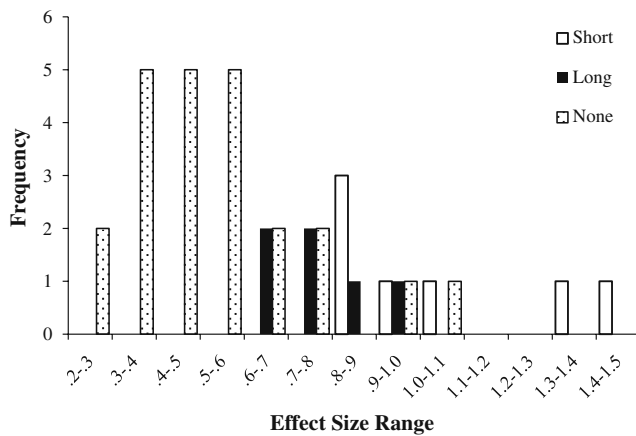


Fig. 1 Distribution of unbiased effect sizes as a function of effect size range and time limits category

excluded, results showed homogeneous effect sizes, $\chi^2(10) = 5.19$, $p > .87$. This finding indicates that the significant mean weighted effect size of 0.85 ($p < .01$) on the 11 effect sizes obtained with a time limit from 5 to 20 min is a valid summary of the state of affairs in this cluster. In addition, this effect size remained significantly larger than the value of 0.51 observed for no time limits conditions, $z = 6.32$, $p < .01$.

Discussion

This study relied on meta-analysis to examine research relevant to the influence of time constraints on the magnitude of gender differences in tests of mental rotation. This allowed the quantification of gender differences in mental rotation tests under various time limits as well as an examination of whether this factor has a significant effect on these gender differences.

The most obvious finding in the present analysis was that gender differences in favor of men were significant in all the partitions. In fact, even the smallest fail-safe value observed in Table 2 (368, for Long Time Limits) easily exceeds the criterion of $5k + 10$ (that is, $5 \times 6 + 10 = 40$) suggested by Rosenthal (1991). Therefore, the file drawer problem is unlikely to account for the significant gender differences observed here. This was expected given the finding that mental rotation tasks tend to produce the most robust gender differences in the literature (Hedges & Nowell, 1995; Linn & Petersen, 1985; Voyer et al., 1995).

The most critical meta-analytic finding was the significant between-group heterogeneity when effect sizes were partitioned as a function of time limits, reflecting that the magnitude of gender differences was larger when any time limit was used compared to when there was no time limit. Supplemental analyses also showed that this pattern of

results could not be explained by systematic variations in age of the sample or year of birth.

The finding that even using “long” time limits increased the magnitude of gender differences in tests of mental rotation when compared to the absence of time limits could have been foreseen. Indeed, Peters (2005) argued that using some time limit would be a more ecologically valid approach as, in the natural environment, perceptual speed is relevant to spatial abilities. From this perspective, the finding that any kind of time limit favors larger gender differences suggests that perceptual speed contributes to gender differences in mental rotation. This viewpoint would be in partial agreement with the perspective held by Goldstein et al. (1990) if one assumes that speed of processing, which is the factor emphasized by Goldstein et al., correlates with perceptual speed. In addition, at first glance the results presented in Table 2 suggest that a long time limit reduces the magnitude of gender differences significantly when compared to a short limit. However, the partition of effect sizes for short time limits to achieve homogeneity (Table 3) showed that the mean weighted effect size obtained for “Other” tests ($d = 0.86$) was virtually the same as that observed for long time limits ($d = 0.85$). In fact, exclusion of the data obtained under very short time limits on the Cubes test resulted in a single homogeneous cluster of effect sizes obtained with time limits varying from 5 to 20 min. This suggests that any time limit has the same effect on the magnitude of gender differences in mental rotation tests compared to no time limit. However, it is important to keep in mind that some of the results were based on a small number of effect sizes and they should be interpreted with caution.

It appears that the difference observed between short and long time limits on the magnitude of gender differences was due to the effect sizes obtained with the Cubes test. This test consists of 32 items in which a pair of cubes is presented. A different pattern is visible on each face of the cube, and participants are expected to use the spatial relation among these patterns to determine whether the two cubes are a rotated version of the same cube. A very short time limit of 2 min was used by Gallagher and colleagues (Gallagher, 1989; Gallagher & Johnson, 1992). Under this time limit, Gallagher (1989) reported that men managed to attempt an average of 69.7% of the items, whereas women attempted 46.1% of the items (a significant gender difference with $p < .0001$). The large effect size obtained under such short time limits fits with the notion that a very short time limit promotes large gender differences. Nevertheless, it is important to keep in mind the possibility that too short a time limit could result in floor effects, in which both men and women do not get enough time for accurate performance (Voyer, Rodgers, & McCormick, 2004). However, floor effects were clearly

not a factor for both genders in the studies conducted by Gallagher and colleagues, considering the large magnitude of gender differences they reported.

Despite the observed linear relation between time limit and the magnitude of gender differences, it is worth noting that the argument, proposed by Goldstein et al. (1990) and Peters (2005) that unlimited completion time should not be expected to result in significant gender differences was not supported here. Specifically, the magnitude of the advantage in favor of men was significantly reduced, but it remained significant without time limits. This finding supports the hypothesis proposed by Lohman (1986) that gender differences in mental rotation are a matter of level of spatial ability, although the earlier discussion suggests that, contrary to what Lohman claimed, some form of speeded processing also seems to be involved.

Another argument raised by Peters (2005) is that unlimited time conditions are not truly “unlimited” and that we cannot expect participants to give a sustained effort for an extended period of time. The present findings generally disagree with the notion that any amount of extra time should be sufficient to affect mental rotation performance as the magnitude of gender differences generally remained statistically similar across the short and long timing conditions. Instead, the findings suggest the possibility that women and men do not react in the same way to the presence of time pressures. Specifically, even though women might generally work more slowly and more cautiously than men, as proposed by Goldstein et al. (1990), women might also keep their effort level higher than men when time pressures are removed. This possibility remains an empirical question. However, it is legitimate to conclude that timing conditions affect the magnitude of gender differences in tests of mental rotation.

The above discussion suggests that gender differences in a speeded component, test completion strategies, and amount of effort could potentially account for the observed influence of time pressure on mental rotation tests performance. However, it is also possible that the presence of time pressures produces more anxiety in women than in men and that this might affect women’s working memory, in a way similar to what has been hypothesized for other threatening environments (Johns, Inzlicht, & Schmader, 2008; Schmader, Forbes, Zhang, & Mendes, 2009; Schmader & Johns, 2003). Considering the demonstrated role of gender differences in working memory on mental rotation test performance (Kaufman, 2007), additional impairment of working memory for women under time pressure conditions provides another possible explanation of the present findings. However, as a meta-analysis does not allow conclusions concerning the causes of the effects of interest, the possible explanations discussed here are purely speculative and require direct empirical testing.

In considering the findings of the present analysis, it is necessary to understand that demonstrating the influence of timing conditions on gender differences in mental rotation does not necessarily require that the gender difference become non-significant without time constraints. A reduction in the magnitude of the gender difference is sufficient. However, it would appear that such a reduction often fails to produce a significant gender by timing conditions interaction in single studies. The main reason why it achieved significance here is likely due to the statistical power inherent to meta-analysis. Specifically, the present meta-analysis could be seen as reflecting a cumulation of 26 studies including a total of 2,762 men and 3,247 women. Such numbers contribute to the power of the present analysis, but they also suggest that differences in the magnitude of gender differences as a function of timing conditions are unlikely to achieve significance in the typical individual study.

Although small at the level of an individual study, the influence of time limits is meaningful when the relevant literature is considered as a whole. However, some might argue that the small number of retrievable effect sizes could result in undue influence from studies that vary in terms of scientific rigor. For example, the study by Goldstein et al. (1990) was heavily criticized by Masters (1998). It is also difficult to evaluate the quality of unpublished research. However, Linn and Petersen (1985) mentioned that homogeneity is often difficult to achieve in a meta-analysis, to the extent that they resorted to describing some clusters as “close to homogeneity” to deal with such difficulties. In contrast, homogeneous effect sizes were obtained quite easily here. It is therefore unlikely that variations in apparent rigor had a significant influence on the outcome of the meta-analysis. In fact, the consistency of the effect sizes retrieved here provides indirect support for their validity. It would also be implausible to assume that all studies sampled in a given timing condition category lacked rigor. It is therefore more parsimonious to conclude that the results obtained here reflect the state of affairs in this area of research.

The present results strongly suggest that it is worthwhile pursuing further experimentation examining the influence of time limits on the magnitude of gender differences in tests of mental rotation. In fact, this is really one of the most obvious conclusions that can be drawn from the analysis: More research is needed in this area. In particular, the role of factors such as perceptual speed, test completion strategies, amount of effort, and the influence of time pressures on working memory might offer some promising avenues for further research. In the meantime, this meta-analysis provides support for the importance of time limits on the magnitude of gender differences in mental rotation tests, suggesting that researchers should include a time limit

when their goal is to maximize gender differences. These results also suggest that any time limit should be sufficient for this purpose.

In view of the exclusion of the large body of experiments where tests of mental rotation were administered only under timed conditions, one might be tempted to argue that the present meta-analysis is not representative of the data available with this type of test. However, the overall effect size of 0.70 obtained here is comparable to the value of 0.67 reported by Voyer et al. (1995) for 35 effect sizes obtained with the Mental Rotations Test. As Voyer et al. (1995) also reported that gender differences showed a trend for an increase in magnitude with year of birth on the Mental Rotations Test, there is no reason to expect that the gender gap would have narrowed in recent years. In fact, the selected sample of studies retrieved here showed no relation between year of birth and the magnitude of gender differences, although restricted range issues temper this claim. It is thus plausible to believe that the studies sampled belong to the same population as the data available as a whole.

Another aspect of interest is the fact that, as previously mentioned, in retrieving literature for this meta-analysis, attempts were made to recover data obtained with tests of spatial and other cognitive abilities demonstrating significant gender differences and on which time constraints had been manipulated. However, not enough such studies could be retrieved to warrant their inclusion. Specifically, as far as spatial abilities are concerned, it would appear that the influence of time limits has been examined only for the Differential Aptitude Test-Spatial Relations subtest in the study conducted by Delgado and Prieto (1996), paralleling their findings of an increase in the magnitude of gender differences on the MRT under “power” rather than “speed” conditions. This suggests that more such studies are required for tests of spatial perception and spatial visualization as they would contribute to our understanding of gender differences in spatial abilities. In fact, Peters’ (2005) notion that perceptual speed is relevant to all spatial tasks would lead one to expect an influence of time limits.

Although the present analysis focused on a limited area of spatial abilities, its findings have general applicability for most tests of cognitive abilities regardless of whether they show gender differences. Indeed, in the studies retrieved here, performance typically improved for both men and women when time limits were removed, suggesting that both men and women were affected by the time limits. Considering this finding, it is plausible to believe that the effect of time pressure on performance should extend to most test-taking situations. The present analysis suggests that this question warrants further investigation. Therefore, consideration of time limit manipulations for other cogni-

tive tests where gender differences are observed, such as mathematics (Hyde, Fennema, & Lamon, 1990) or object location memory (Voyer, Postma, Brake, & Imperato-McGinley, 2007), would also contribute to a more complete understanding of cognitive gender differences. This would be informative even if the influence of time limits turned out to be negligible as even null finding can contribute to theory elaboration (Greenwald, 1975).

The finding, mentioned above, that performance generally improved for both men and women with removal of time limits suggests the possibility that the reduction in the magnitude of gender differences without time limits could be accounted for by ceiling effects in men. Indeed, men are generally performing at a high level of proficiency in many spatial tasks even with a time limit, and this would leave less room for improvements in their scores compared to women when the time limit is removed. This possibility found some support in a study published by Glück and Fabrizii (2010). These authors followed the same approach as Goldstein et al. (1990) in that they gave their participants 6 min to complete as many items as they could on the MRT, and they then allowed them to complete the remaining items without time limits. However, the approach to scoring used by Glück and Fabrizii (2010) differed from that used by Goldstein et al. in that they did not present a composite score reflecting timed plus untimed performance.¹ Thus, their paper presented an “untimed” score that reflected only performance after completion of the timed portion and, although this score produced no significant gender differences, there was a slight advantage for women. In contrast, items completed within the 6-min time limit produced large gender differences in favor of men ($d = 0.93$). This suggests that men completed many more items successfully than women did under time limits and, having fewer items remaining, did not gain as much as women did when the time limit was removed. It thus appears that ceiling effects in men cannot be excluded as a possible explanation of the present findings. This is another avenue that warrants further investigation.

To conclude, the studies retrieved in the present meta-analysis indicate that tests of mental rotation produce significantly larger gender differences when administration involves time limits (regardless of their duration) compared to administration without such constraints. This suggests that more studies investigating the influence of time limits on gender differences in spatial tests as well as on other

¹ This composite score was obtained from J. Glück and used in meta-analytic calculations to maintain comparability with the other studies retrieved. Accordingly, the effect size presented for Glück and Fabrizii (2010) in Table 1 for their unlimited time condition was based on the same composite score as used by Goldstein et al. (1990) and others.

cognitive tests should be conducted to delimit their importance. Such studies would provide one more step toward an explanation of gender differences in spatial abilities in particular and cognitive abilities in general.

Acknowledgement This study was supported by a grant from the Natural Sciences and Engineering Research Council of Canada. The author is indebted to Susan D. Voyer for her assistance in retrieving literature and for her comments on an earlier version of this report.

References

Studies included in the meta-analysis are denoted by an asterisk (*).

- *Brownlow, S., McPheron, T. K., & Acks, C. N. (2003). Science background and spatial abilities in men and women. *Journal of Science Education Technology*, *12*, 371–380.
- *Brownlow, S., & Miderski, C. A. (2002). How gender and college chemistry experience influence mental rotation ability. *Themes in Education*, *3*, 133–140.
- *Carr, M., Shing, Y. L., Janes, P., & Hettinger Steiner, H. (2007, April). Early gender differences in strategy use and fluency: Implications for the emergence of gender differences in mathematics. Poster presented at the meeting of the Society for Research in Child Development, Boston, MA.
- Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Academic Press.
- *Cooke-Simpson, A., & Voyer, D. (2007). Confidence and gender differences on the Mental Rotations Test. *Learning and Individual Differences*, *17*, 181–186. doi:10.1016/j.lindif.2007.03.009
- *Delgado, A. R. & Prieto, G. (1996). Sex differences in visuospatial ability: Do performance factors play such an important role? *Memory & Cognition*, *24*, 504–510.
- *Dollinger, S. M. C. & Patrick, C. L. (1993). *Gender and age differences in mental rotation ability: Without time constraints*. Unpublished manuscript, University of Illinois.
- Feingold, A. (1988). Cognitive gender differences are disappearing. *American Psychologist*, *43*, 95–103. doi:10.1037/0003-066X.43.2.95
- *Gallagher, S. A. (1989). Predictors of SAT mathematics scores of gifted male and gifted female adolescents. *Psychology of Women Quarterly*, *13*, 191–203. doi:10.1111/j.1471-6402.1989.tb00996.x
- *Gallagher, S. A. & Johnson, E. S. (1992). The effect of time limits on performance of mental rotations by gifted adolescents. *Gifted Child Quarterly*, *36*, 19–22. doi:10.1177/001698629203600105
- *Glück, J., & Fabrizii, C. (2010). Gender differences in the Mental Rotations Test are partly explained by response format. *Journal of Individual Differences*, *31*, 106–109. doi:10.1027/1614-0001/a000019
- *Goldstein, D., Haldane, D., & Mitchell, C. (1990). Sex differences in visual-spatial ability: The role of performance factors. *Memory & Cognition*, *18*, 546–550.
- Greenwald, A. G. (1975). Consequences of prejudice against the null hypothesis. *Psychological Bulletin*, *82*, 1–20. doi:10.1037/h0076157
- Halpern, D. F. (2000). *Sex differences in cognitive abilities* (3rd ed.). Mahwah: Erlbaum.
- Hedges, L. V., & Becker, B. J. (1986). Statistical methods in the meta-analysis of research on gender differences. In J. Hyde & M. C. Linn (Eds.), *The psychology of gender: Advance through meta-analysis* (pp. 14–50). Baltimore: John Hopkins University Press.
- Hedges, L. V., & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science*, *269*, 41–45.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. New York: Academic Press.
- *Heil, M., & Jansen-Osmann, P. (2008). Gender differences in math and mental rotation accuracy but not in mental rotation speed in 8 year old children. *European Journal of Developmental Science*, *2*, 190–196.
- Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin*, *107*, 139–155. doi:10.1037/0033-2909.107.2.139
- *Janas, A. J., Mellon, L. M., & Brownlow, S. (2006). *Are women's spatial abilities improved by successful peer models?* Poster presented at the meeting of the Southeastern Psychological Association, Atlanta, Georgia.
- Johns, M., Inzlicht, M., & Schmader, T. (2008). Stereotype threat and executive resource depletion: Examining the influence of emotion regulation. *Journal of Experimental Psychology: General*, *137*, 691–705. doi:10.1037/a0013834
- *Kanoy, A. E., Jordan, B., & Brownlow, S. (2009, February). *Can external justification for performance obviate stereotype threat effects?* Poster presented at the meeting of the Southeastern Psychological Association, New Orleans, LA.
- Kaufman, S. B. (2007). Sex differences in mental rotation and spatial visualization ability: Can they be accounted for by differences in working memory capacity? *Intelligence*, *35*, 211–223. doi:10.1016/j.intell.2006.07.009
- Linn, M. C., & Petersen, A. C. (1985). Emergence and characterisation of gender differences in spatial abilities: A meta-analysis. *Child Development*, *56*, 1479–1498. doi:10.1111/1467-8624.ep7252392
- Linn, M. C., & Petersen, A. C. (1986). Gender differences in spatial ability: Implications for mathematics and science performance. In J. Hyde & M. C. Linn (Eds.), *The psychology of gender: Advance through meta-analysis* (pp. 67–101). Baltimore: John Hopkins University Press.
- Lohman, D. F. (1986). The effect of speed-accuracy tradeoff on sex differences in mental rotation. *Perception and Psychophysics*, *39*, 427–436.
- *MacLeod, M. V. (2000). *Effect of timing conditions on the magnitude of gender differences on the Water Level Test and the Mental Rotations Test*. Unpublished Honors Thesis, St. Francis Xavier University.
- *Masters, M. (1998). The gender difference on the Mental Rotations Test is not due to performance factors. *Memory & Cognition*, *26*, 444–448.
- *Miller, D. I., & Halpern, D. F. (2010, April). Spatial training narrows gender gap in visuospatial skills. Poster presented at the meeting of the Western Psychological Association, Cancun.
- *Ortner, T. M. & Sieverding, M. (2008). Where are the gender differences? Male priming boosts spatial skills in women. *Sex Roles*, *59*, 274–281. doi:10.1007/s11199-008-9448-9
- *Peters, M. (2005). Sex differences and the factor of time in solving Vandenberg and Kuse mental rotation problems. *Brain and Cognition*, *57*, 176–184. doi:10.1016/j.bandc.2004.08.052
- *Resnick, S. M. (1993). Sex differences in mental rotation: An effect of time limits? *Brain and Cognition*, *21*, 71–79. doi:10.1006/brcg.1993.1005
- Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin*, *86*, 638–641. doi:10.1037/0033-2909.86.3.638
- Rosenthal, R. (1991). *Meta-analytic procedures for social research* (Revth ed.). Beverly Hills: Sage.
- *Scali, R. M., & Brownlow, S. (2001). Impact of instructional manipulation and stereotype activation on sex differences in

- spatial task performance. *Psi Chi Journal of Undergraduate Research*, 6, 3–13.
- *Scali, R. M., Brownlow, S., & Hicks, J. L. (2000). Gender differences in spatial task performance as a function of speed or accuracy orientation. *Sex Roles*, 43, 359–376. doi:10.1023/A:1026699310308
- Schmader, T., Forbes, C. E., Zhang, S., & Mendes, W. B. (2009). A metacognitive perspective on the cognitive deficits experienced in intellectually threatening environments. *Personality and Social Psychology Bulletin*, 35, 584–596. doi:10.1177/0146167208330450
- Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology*, 85, 440–452. doi:10.1037/0022-3514.85.3.440
- *Titze, C., Heil, M., & Jansen, P. (2008). Gender differences in the Mental Rotations Test (MRT) are not due to task complexity. *Journal of Individual Differences*, 29, 130–133. doi:10.1027/1614-0001.29.3.130
- Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotation, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills*, 47, 599–604.
- *Voyer, D. (1997). Scoring procedure, performance factors, and magnitude of sex differences in spatial performance. *American Journal of Psychology*, 110, 259–276. doi:10.2307/1423717
- *Voyer, D., & Doyle, R. A. (2010). Item type and gender differences on the Mental Rotations Test. *Learning and Individual Differences*. Published online May 17, 2010. doi:10.1016/j.lindif.2010.04.010
- *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. doi:10.1037/cjep2006010
- Voyer, D., Postma, A., Brake, B., & Imperato-McGinley, J. (2007). Gender differences in object location memory: A meta-analysis. *Psychonomic Bulletin and Review*, 14, 23–38.
- Voyer, D., Rodgers, M. A., & McCormick, P. A. (2004). Timing conditions and magnitude of gender differences on the Mental Rotations Test. *Memory and Cognition*, 32, 72–82.
- *Voyer, D. & Sullivan, A. M. (2003). The relation between spatial and mathematical abilities: Potential factors underlying suppression. *International Journal of Psychology*, 38, 11–23. doi:10.1080/00207590244000241
- 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. doi:10.1037/0033-2909.117.2.250
- Wolf, F. M. (1986). *Meta-analysis: Quantitative methods for research synthesis*. Newbury Park: Sage Publications.