What does the Mental Rotation Test Measure? An Analysis of Item Difficulty and Item Characteristics

André F. Caissie[1], François Vigneau[2,*] and Douglas A. Bors[3]

[1]*Centre de recherche sur la cognition et l'apprentissage, Université de Poitiers, Poitiers, France*

[2]*École de psychologie, Université de Moncton, Moncton, Canada*

[3]*Department of Psychology, University of Toronto at Scarborough, Scarborough, Canada*

Abstract: The present study examined the contributions of various item characteristics to the difficulty of the individual items on the Mental Rotation Test (MRT). Analyses of item difficulties from a large data set of university students were conducted to assess the role of time limitation, distractor type, occlusion, configuration type, and the degree of angular disparity. Results replicated in large part previous findings that indicated that occluded items were significantly more difficult than non-occluded and that mirror items were more difficult than structural items. An item characteristic not previously examined in the literature, configuration type (homogeneous versus heterogeneous), also was found to be associated with item difficulty. Interestingly, no significant association was found between angular disparity and difficulty. Multiple regression analysis revealed that a model consisting of occlusion and configuration type alone was sufficient for explaining 53 percent of the variance in item difficulty. No interaction between these two factors was found. It is suggested, based on overall results, that basic figure perception, identification and comparison, but not necessarily mental rotation, account for much of the variance in item difficulty on the MRT.

Keywords: Mental Rotation Test, spatial ability, mental rotation, item types, item characteristics.

1. INTRODUCTION

Vandenberg and Kuse's [1] Mental Rotation Test (MRT) is one of the most commonly used measures of spatial ability. Performance variability on this test is frequently substantial, even in groups of high-ability subjects such as university students. Yet, the factors responsible for the performance variability on the MRT and for the difficulty of particular items still have not been isolated and understood. It has been suggested in the literature that presence or absence of certain physical characteristics are associated with item difficulty [2].

The MRT is comprised of 24 items in which (two-dimensional) drawings of three-dimensional geometrical figures are to be compared. These figures were adapted by Vandenberg and Kuse [1] from similar figures used by Shepard and Metzler [3] in a series of now classic studies of the concept and definition of mental rotation.

Each MRT item consists of a row of five line drawings, including a geometrical target figure in the left most position followed by four response-choice figures: two rotated reproductions of the target and two distractors (see Fig. **1**). The subject's task is to indicate which two of the four response-choice figures are rotated reproductions of the target figure. In the version of the MRT presently examined [4], the 120 figures (targets and response choices) used in the 24 items are based on a total of 10 distinct reference figures (see Fig. **2**).

Various factors potentially responsible for item difficulty on the MRT have been suggested. In the present article, these factors (response rate, distrator type, and occlusion) will be examined. Two additional factors previously unaddressed in the literature on the MRT will also be explored: configuration type and angular disparity

Some studies (e.g., [5]) have indicated that item difficulty could be primarily a product of the stringent time limit use in the standard administration of the MRT and other procedural factors such as the scoring scheme associated with the test. The timed condition has also been associated with more response omissions and unanswered test items [6]. Generally speaking, low scoring subjects attempt fewer items. As Goldstein *et al.* [5] hypothesized, it is possible that lower scoring subjects approach the items with more caution and perform at a slower pace than higher scoring subjects. Therefore, items placed nearer to the end of the test are more likely to be left unanswered and thus appear more difficult than they would if they were placed nearer the beginning of the test. Eliminating the effect of procedural artefacts (e.g., time factor) could help clarify the relation between item difficulty and other factors such as spatial abilities [7] and item characteristics [2].

Some approaches to overall performance on the MRT have focused on specific item characteristics. Voyer, Rodgers, and McCormick [6] have suggested an MRT item categorization scheme based on differences in the nature of the item's distractors. According to this scheme, mirror items are contrasted with structural items. Mirror items are items whose distractors have the same relations between their configuration's segments as the target figure, but are mirror im-

*Address correspondence to this author at the École de psychologie, Université de Moncton, Moncton NB E1A 3E9, Canada; Tel: 506 858-4836; Fax: 506 858-4768; E-mail: f.vigneau@umoncton.ca
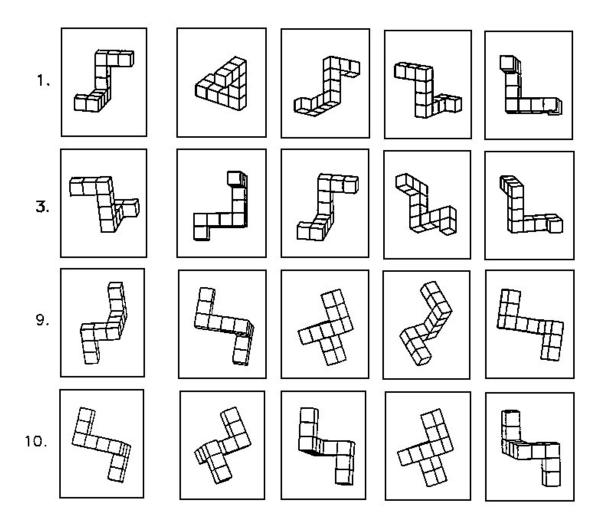
Fig. (1). Item types. Mirror items (item 1, item 9, item 10), structural item (item 3), non-occluded items (item 1, item 3), occluded items (item 9, item 10), heterogeneous items (item 1, item 3), and homogeneous items (item 9, item 10).

ages of the target figure. By contrast, in structural items, the distractors and the target figures differ in their segment relations [2]. MRT items 1, 2, 5, 6, 9, 10, 11, 12, 15, 16, 19, 20 were classified by Voyer and Hou [2] as mirror items, whereas items 3, 4, 7, 8, 13, 14, 17, 21, 22, and 24 were classified as structural. Two items, called mixed items (18 and 23), could not be classified according to this dichotomy because they each include one structural and one mirror distractor. Examples of structural and mirror items can be found in Fig. (**1**). Worth noting is the fact that item categories are not equally distributed across the test. For instance, there are more mirror items than there are structural items, and there are few mixed items.

Voyer and Hou [2] also offered an item categorization based on the presence or absence of physical occlusion in response choices. Occlusion means that significant parts of the three-dimensional figure are covered by other parts and are thus not directly visible to the observer. It is important to note that occlusion is defined in this context only with respect to correct choices, not distractor or target figures. With

occlusion so defined, Voyer and Hou [2] identified items 9, 10, 11, 14, 15, 17, 18 as occluded items, and items 1, 2, 3, 4, 5, 6, 7, 8, 12, 13, 16, 19, 20, 21, 22, 23 and 24 as non-occluded items. Examples of occluded and non-occluded items can be found in Fig. (**1**). As was the case with distractor types, occlusion item categories are somewhat clustered with respect to item position in the test. For example, no occluded items appear amongst the first eight items.

Voyer *et al.* [6] hypothesized that structural items would be easier to solve than mirror items because their distractors allow subjects to rely on simple basic object recognition processes, whereas mirror items possibly require the creation of a more integrated mental schema of the response choice figures to correctly mentally rotate and compare the distractors to the target figure. Correct performance on mirror items would imply that subjects mentally represent one of the geometrical figures in the orientation of another. This would be consistent with cognitive studies of the processes involved in judging whether figures of differing angular orientation are the same or different [3, 8]. This need to make a
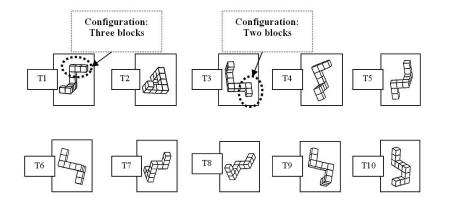
Fig. (2). Ten reference figures used to construct the MRT items. Pairs T1-T2, T3-T4, T5-T6, T7-T8, T9-T10 are mirror images. T1-T8 is a pair of structurally different figures. Figures T1, T2, T3, and T4 have peripheral segments constructed of 2 and 3 blocks (heterogeneous). Figures T5, T6, T7, T8, T9 and T10 have both peripheral segments constructed of three blocks (homogeneous).

mental representation would also relate to the idea that a mental analogue of a three-dimensional object rotation takes place during the resolution of MRT items.

Voyer and Hou [2] argued that the occlusion of substantial visual information makes the comparison of the correct response choices with the target figure more difficult because it could cause misperceptions of the correct response figures, leading subjects to misidentify them as distractors. It could also be the case that subjects are required to do more comparative work to rule out some other alternatives in the occluded items. Thus, non-occluded correct response figures should be easier to identify than occluded ones.

Configuration type, an item characteristic not previously addressed in the MRT literature, was also examined in the present study. Each figure of the MRT has three structural angles of 90 degrees, or three vertices relating four segments. This common structure resulted in the 10 reference figures which, along with various rotations of these reference figures, constitute the entire set of target figures and response choices in the MRT (see Fig. **2**). These 10 reference figures can be reduced to two broad categories according to the number of blocks included in the peripheral segments of the figure. The figures used in items 1, 2, 3, 4, 5, 6, 7, 8, 12, and 23 are based on figure types T1, T2, T3, or T4 (see Fig. **2**). In all of these figures the two peripheral segments are constructed of two and three blocks. Because of the differing number of blocks on the peripheral segments, these figures can be called heterogeneous configurations. The figures in items 9, 10, 11, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, and 24 are based on figure types T5, T6, T7, T8, T9, and T10, all of which have both peripheral segments constructed of three blocks (homogeneous configurations). As can be observed from the order of the items in the MRT, configuration type is associated with item position. All figures used in the first eight items have heterogeneous configuration, whereas figures with both peripheral segments constructed of three blocks start on item 9. Despite this confound, it could be reasonably hypothesized that configuration type so defined could contribute to item difficulty. Heterogeneous configura-

tions are more likely to help figure identification and mental rotation by offering additional discriminative elements than are homogeneous configurations. That is, the number of blocks on a segment can simply be counted and the response choice figure retained or excluded as a possible correct match without the need of any mental rotation. Consequently, it is predicted that homogeneous configuration items will be more difficult than heterogeneous configuration items.

Finally, angular disparity was examined. Angular disparity between the target figure and the response choices figures has been considered an important feature of the MRT. Variability in angular disparity was central to Shepard and Metzler's [3] chronometric studies that evaluated subjects' reaction times in the comparison of two complex figures placed in different orientations. Results consistently showed that reaction times were directly proportional to the total degrees of angular disparity between the two figures [3, 9]. A linear relationship between angular disparity and response time was clearly demonstrated. Response latency increased monotonically from 1 degree to 180 degrees of angular disparity, and was reversed as the angular disparity became greater than 180 degrees. In the case of the MRT, these findings lead to the hypothesis that, as the degrees of angular disparity are increased, item difficulty will increase. This increase in item difficulty would then be viewed as a consequence of increased solution time.

Given the central role played by angular disparity in the experimental research on spatial cognition, it is surprising to note the lack of studies addressing the role played by this factor in item difficulty on the MRT. Although there is a lack of information as to what effect angular disparities have on subjects' performance scores, it has been commonly assumed that the MRT was based on the idea that subjects are required to mentally rotate a mental image of complex figures and then compare response choices to the model figure. For the present study, angular disparity characteristics were determined for each MRT item. They are presented in the Appendix.

Previous studies of MRT performance and item difficulty have examined factors in isolation and independently from one another. The present study does attempt to replicate those studies. Through the use of multiple regression, however, the present study goes on to test the relative importance of these factors and their possible interactions.

2. METHOD

2.1. Subjects

The data were collected through administration of the MRT to 624 undergraduate students (407 women, 217 men) from the Université de Moncton and the University of Toronto. Mean age for the group was 21 years (SD = 4.3). Subjects ranged in age from 17 to 58 years (5.4% of the subjects were older than 25 years).

2.2. Measure

The Mental Rotation Test (MRT; [1]) is comprised of 24 items, six items on four separate pages in the test booklet [4]. Each item is comprised of a row of five line drawings including a geometrical target figure in the left-most position followed by four response-choice figures: two rotated reproductions of the target and two distractors (see Fig. **1**). The subject's task is to indicate which two of the four response-choice figures are rotated reproductions of the target figure, allowing free rotation in three-dimensional space. In each item there are always two and only two correct figures and two incorrect distractor figures.

2.3. Procedure

Standard administration instructions of the MRT were used. Accordingly, subjects were instructed to find, for each item, two response choices with figures identical to the target figure. Subjects were allotted a ten minute time limit to complete the task, and were informed when there were 5 minutes remaining and again when there were 2 minutes remaining. Instructions emphasized that subjects should refrain from guessing. Instructional items were administered followed by the 24-item MRT.

Administration sessions were carried out in groups ranging from 30 to 70 subjects. In one of the two universities, subjects were given course credits for their participation. The scoring of individual performances was according to Vandenberg and Kuse [1]: Subjects were given one point for each item on which they successfully identified both correct response (24 thus represents a perfect total score).

3. RESULTS

For the sample of 624 subjects (Sample 1), the mean total MRT score was 11.46 (SD = 6.12). The test as a whole was found to be internally consistent, Cronbach's alpha = 0.91. Percent correct and response rates for each MRT item are presented in Table **1** for Sample 1. Item difficulty (or *p*, percent correct) was associated with the items' positions in the test, the items at the beginning of the test having higher percents correct than those at the end of the test. Also as expected, response rate was associated with item position:

items at the end of the test were left unanswered more often than those at the beginning of the test. In all likelihood, this decreased response rate with increasing item number was, at least in part, a consequence of many subjects being unable to respond to all items within the time allotted to complete the test. This variability in item response rate also means that item difficulty statistics may be difficult to interpret. Specifically, the extent to which the variation in percent correct was a consequence of some item characteristics and the extent to which it was a consequence of the time limit are impossible to disentangle. To avoid such interpretational difficulties, unless indicated otherwise the subsequent analyses were conducted on a sub-sample (Subsample 1b) of 218 subjects (115 women and 103 men) comprised only of those individuals who responded to most (23 or 24) items, thus providing a form of control of the effects of time limit on response rate. Mean total MRT score for Subsample 1b was 15.00 (*SD* = 6.69). The internal reliability of the test for this sample was satisfactory (Cronbach's alpha = 0.92). As shown in Table **1**, item difficulties for these 218 subjects ranged from 31% (item 17) to 87% (item 3), and response rates ranged from 81 % (item 24) to 98 % (item 2).

Also reported in Table **1** are the corrected item-total correlations for each MRT item (Sample 1). These correlations ranged between 0.28 and 0.66 (*M* = 0.51, *SD* = 0.08), which is consistent with previous reports. In the response-rate controlled Subsample 1b, item discriminations ranged between 0.29 and 0.71 (*M* = 0.55, *SD* = 0.11). In this case of Subsample 1b, there was no significant linear relation between item discrimination and item difficulty, Sample 1: *r* = -0.33, *p* = 0.11, Subsample 1b: *r* = 0.12, *p* = 0.57.

3.1. Composite Scores Analyses

Difficulty analysis was conducted for the various categories of item characteristics. Mean scores for mirror, structural, occluded, non-occluded, homogeneous, and heterogeneous items were calculated by summing the correctly scored items in each of these six categories and dividing by the number of items in the categories, thus producing scores expressed as proportions correct. Proportion correct statistics for the various item categories are reported in Table **2** separately for Sample 1 and for Subsample 1b, along with reliability data for those composite scores. As might be expected, the mean scores for all six categories were higher in Subsample 1b than they were in Sample 1. Furthermore, and more interestingly, the pattern of differences between the two samples was stable over item characteristics.

Paired-comparison t-test conducted on Subsample 1b showed a significant difference between accuracy on mirror items and accuracy on structural items (*t* = 5.33; df = 217, *p* < 0.01). Mirror items were significantly more difficult than structural items. A paired-comparison t-test also showed a significant accuracy difference between non-occluded items and occluded items (*t* = 15.25; df = 217, *p* < 0.01), indicating that occluded items were significantly more difficult than non-occluded items. This second test represents a replication of Voyer and Hou [2]. In terms of configuration type, homogeneous configuration items were more difficult

Table 1.   Item Difficulty, Accuracy for Correct Response Figures. 1 and 2, Response Rate, and Item-Total Correlations in Sample 1 ($n = 624$) and Subsample 1b ($n = 218$)

Sample 1						Subsample 1b					
Item	p	pCR1	pCR2	RR	r_{it}	Item	p	pCR1	pCR2	RR	r_{it}
1	0.67	0.74	0.90	0.97	0.50	1	0.67	0.71	0.94	0.98	0.63
2	0.67	0.78	0.77	0.93	0.45	2	0.70	0.81	0.78	0.94	0.60
3	0.87	0.92	0.94	0.98	0.28	3	0.87	0.92	0.92	0.98	0.36
4	0.76	0.91	0.79	0.94	0.44	4	0.76	0.91	0.79	0.95	0.57
5	0.64	0.85	0.72	0.93	0.50	5	0.71	0.87	0.78	0.97	0.64
6	0.72	0.87	0.81	0.95	0.46	6	0.71	0.86	0.81	0.96	0.66
7	0.74	0.85	0.81	0.92	0.46	7	0.75	0.86	0.82	0.94	0.60
8	0.76	0.85	0.82	0.91	0.53	8	0.80	0.89	0.86	0.96	0.64
9	0.32	0.39	0.65	0.82	0.47	9	0.38	0.43	0.73	0.91	0.52
10	0.50	0.70	0.66	0.81	0.45	10	0.57	0.72	0.78	0.89	0.56
11	0.35	0.66	0.40	0.73	0.55	11	0.42	0.74	0.49	0.84	0.66
12	0.63	0.72	0.74	0.84	0.59	12	0.72	0.82	0.84	0.96	0.62
13	0.62	0.70	0.71	0.76	0.54	13	0.78	0.86	0.88	0.93	0.42
14	0.49	0.62	0.54	0.65	0.59	14	0.66	0.80	0.72	0.86	0.59
15	0.33	0.36	0.52	0.62	0.62	15	0.49	0.52	0.73	0.89	0.63
16	0.37	0.56	0.43	0.62	0.66	16	0.57	0.84	0.63	0.93	0.71
17	0.16	0.25	0.43	0.54	0.42	17	0.31	0.43	0.73	0.89	0.39
18	0.39	0.51	0.46	0.57	0.58	18	0.60	0.81	0.73	0.93	0.60
19	0.32	0.47	0.36	0.49	0.55	19	0.59	0.87	0.63	0.92	0.51
20	0.28	0.37	0.34	0.43	0.61	20	0.61	0.76	0.75	0.92	0.61
21	0.23	0.32	0.31	0.39	0.48	21	0.52	0.73	0.70	0.90	0.38
22	0.24	0.31	0.28	0.34	0.54	22	0.62	0.78	0.74	0.88	0.53
23	0.21	0.27	0.26	0.32	0.52	23	0.61	0.78	0.75	0.91	0.54
24	0.21	0.25	0.24	0.28	0.41	24	0.60	0.71	0.69	0.81	0.29

Note. p = item difficulty (percent correct); pCR1 = accuracy on correct response Fig. (**1**); pCR2 = accuracy on correct response Fig. (**2**); RR = response rate; r_{it} = (corrected) item-total correlation.

than heterogeneous configuration items ($t = 12.89$; df $= 217$, $p < 0.01$).

3.2. Item Analyses

Position in the test was explored as a potential source of item difficulty. In Sample 1, item position, indexed by item number (1 through 24), was strongly associated with item difficulty ($r = -0.86$, see Table **3**). In Subsample 1b, where response rate variations across items was controlled through subject selection, the correlation between item position and item difficulty was substantially reduced; it remained, however, significant ($r = -0.45$, $p = 0.03$).

In Subsample 1b, item difficulty was also found to be significantly associated with the presence of an occlusion ($r = -0.67$) and with configuration type ($r = -0.66$; see Table **3**). Distractor type (mirror vs. structural), surprisingly, showed no significant association with difficulty ($r = -0.26$, df $= 20$, $p = 0.28$). Thus, when the average performance on an item is used for the analysis rather than individual scores, difficulty is found to be only weakly and not significantly related to distractor type. Furthermore, controlling for configuration type attenuated the correlation between item position and item difficulty to the point of removing practically all association (partial $r = 0.01$, $p = 0.95$). This result, com-

Table 2. **Percent Correct, Skewness of Percent Correct, Response Rate, Skewness of Response Rate, and Alpha Reliability for Each Item Type Category in Sample 1 (*n* = 624) and Subsample 1b (*n* = 218)**

Sample 1							
	Difficulty			**Response Rate**			
Item type category	*M*	*SD*	Skew	*M*	*SD*	Skew	Cronbach's alpha
Mirror	0.48	0.29	0.14	0.76	0.32	-0.68	0.85
Structural	0.54	0.25	-0.10	0.67	0.21	-0.04	0.77
Occluded	0.36	0.30	-0.52	0.68	0.22	-0.60	0.77
Non-occluded	0.54	0.26	0.01	0.70	0.24	0.00	0.87
Heterogeneous	0.67	0.27	-0.74	0.87	0.13	-1.71	0.82
Homogeneous	0.34	0.29	0.66	0.57	0.33	-0.16	0.88
Subsample 1b							
	Difficulty			**Response Rate**			
Item type category	*M*	*SD*	Skew	*M*	*SD*	Skew	Cronbach's alpha
Mirror	0.59	0.32	-0.36	0.93	0.13	-2.21	0.89
Structural	0.67	0.25	-0.62	0.91	0.16	-2.17	0.77
Occluded	0.49	0.33	-0.02	0.89	0.19	-1.80	0.81
Non-occluded	0.68	0.28	-0.59	0.93	0.12	-2.53	0.89
Heterogeneous	0.73	0.30	-0.96	0.96	0.10	-3.18	0.87
Homogeneous	0.55	0.29	-0.11	0.89	0.17	-1.93	0.87

Table 3. **Correlations between Item Difficulty (*p*), Item Position, Item Response Rate, Distractor Type, Occlusion, and Configuration Type Computed on 24 MRT Items in Sample 1 and Subsample 1b**

	1	**2**	**3**	**4**	**5**	**6**
1. Item difficulty	1	-0.45*	0.61**	-0.26	-0.67**	-0.66**
2. Item position	-0.86**	1	-0.64**	-0.29	0.09	0.66**
3. Item response rate	0.90**	-0.97**	1	0.18	-0.48*	-0.71**
4. Distractor type	-0.06	-0.29	0.28	1	0.09	0.00
5. Occlusion	-0.35	0.09	-0.06	0.09	1	0.54**
6. Configuration type	-0.75**	0.66**	-0.64**	0.00	0.54**	1

Note. * *p* < 0.05, ** *p* < 0.01. Sample 1 (*n* = 624; 407 women and 217 men) below the main diagonal, Sample 1b (*n* = 218; 115 women and 103 men) above the main diagonal.

bined with the fact that controlling for item position attenuated only slightly the correlation between item difficulty and configuration type (partial *r* = 0.55), indicates that despite the substantial association between item position and configuration type (*r* = 0.66), configuration type, not item position, may be the main determinant of item difficulty in a response rate controlled sample.

3.3. Angular Disparity

Degrees of angular disparity relative to the target figure for each of the two correct response figures are presented in the Appendix for each item. Accuracy rates for each of the 48 correct response figures are presented in Table **1.** Degrees of angular disparity were obtained by measuring angular rotations of three-dimensional analogue figures placed to

conform to the various items of the MRT. They are expressed in terms of orthogonal axes (x, y, z) in the three-dimensional space. With respect to the present angular disparity characterization, MRT items were found to vary mainly in terms of rotation about the y axis (see Appendix).

Analyses failed to show any association between angular disparities on the y axis and accuracy for correct response figures ($r = 0.01$, $p = 0.97$). Analyses also failed to show any significant association between accuracy for correct response figures and angular disparities on the x ($r = -0.13$, $p = 0.39$) and the z axis ($r = -0.07$, $p = 0.65$). Thus, there is little evidence to suggest that degree of angular disparity per se between response figures and the target figure is a determining factor of item difficulty.

3.4. Models of Item Difficulty

Based on the 24 Vandenberg and Kuse [1] scored items, several models of item difficulty were examined. To begin with, distractor type, occlusion, and configuration type were used to predict item difficulty. In the response-rate controlled Subsample 1b, two significant predictors with independent contribution emerged: occlusion and configuration type, multiple R = 0.76 ($R^2 = 0.57$, adjusted $R^2 = 0.53$, see Table **4**). The model that included distractor type only increased the adjusted R^2 to 0.55. Further regression analyses revealed that there was no three-way interaction nor were there any two-way interactions.

Multiple regression models using the 48 correct response figure accuracies were analysed to include an assessment of angular disparity along with the other predictors used in previous regression. The significant predictors were again found to be occlusion and configuration type in Subsample 1b (see Table **5**). Again, further regressions revealed no significant interactions. Surprisingly, angular disparities accounted for little of the variance in difficulty in both these analyses. Angular disparity may be related to response latency, but it does not appear to be related to item difficulty.

4. DISCUSSION

The present study aimed at extending our understanding of the sources of item difficulty on the MRT. In this undertaking we have replicated the importance of some previous factors, failed to replicate another, and uncovered a new factor. The findings revealed a substantial simple association between MRT item difficulty and three item characteristics: distractor type, occlusion, and configuration type. Specifically, mirror items, occluded items, and homogeneous configuration items were found to be significantly more difficult than their structural, non-occluded, and heterogeneous counterparts. This was found to be the case for both the overall sample and the subsample for which response rate was controlled.

With respect to distractor type and occlusion, the results replicated and generalized the earlier findings of Voyer and Hou [2], who used an untimed administration of the MRT. The analyses of the items' configuration types were original to the present study and demonstrate a substantial effect of this additional item characteristic on item difficulty. Specifically, homogeneous figure items—items with peripheral segments of the same length—were shown to be significantly more difficult than heterogeneous items. Again, this was the case in both the overall and the response-rate controlled samples. The fact that heterogeneous items were easier to solve than were the homogeneous items may be due to the ease with which the segments can be discriminated. In this respect, heterogeneous items, with their peripheral segments of differing lengths, likely offer more salient cues as to figure orientation, thus making them easier to compare with rotated versions. This is a factor worthy of further investigation.

Somewhat surprisingly, for both samples, the degree of angular disparity between correct response figures and target figure was not found to be significantly related to response accuracy. Based on these results, it thus appears

Table 4.   Regression Model of MRT Difficulty (24 Items; Subsample 1b (*n*=218))

	Correlation Coefficients						
Final model (R^2=0.57; adjusted R^2=0.53)	**Beta**	***t***	**Sig.**	**Zero-order**	**Partial**	**Semi-partial**	**Tolerance**
Occlusion	-0.43	-2.55	0.02	-0.67	-0.49	-0.36	0.71
Configuration type	-0.43	-2.53	0.02	-0.66	-0.48	-0.36	0.71

Table 5.   Regression Model of MRT Difficulty (48 Correct Responses Figures; Subsample 1b (*n*=218))

	Correlation Coefficients						
Final model (R^2=0.36; adjusted R^2=0.33)	**Beta**	***t***	**Sig.**	**Zero-order**	**Partial**	**Semi-partial**	**Tolerance**
Occlusion	-0.34	-2.4	0.02	-0.53	-0.34	-0.29	0.71
Configuration type	-0.35	-2.4	0.02	-0.53	-0.34	-0.29	0.71

that although angular disparity may slow a subject's response rate, it is not a characteristic that, in itself, significantly affects an item's difficulty. Such a result was clearly unexpected and is inconsistent with a view of the MRT as a test of mental rotation. Potential reasons for this discrepancy may be found in the fact that most of the research reporting a relation among item difficulty, response time and mental rotation use simple two-dimensional stimuli rather that two-dimensional representation of tri-dimensional objects (see the seminal work of Cooper & Shepard [10]), or use highly practiced subjects (e.g., Shepard & Metzler [3], or Albers & Höft [11] for a recent example). In any case, it is suggested, based on overall results, that basic figure perception, identification and comparison, but not necessarily mental rotation, account for much variance in item difficulty on the MRT.

Our conclusion is supported by the multiple regression analyses. It was found that occlusion and configuration type together accounted for over 50 percent of the variance in item difficulty. The addition of the other variables (distractor type and angular disparity) that would upon inspection appear to be more related to the process of mental rotation failed to significantly increase the amount of explained variance. It should be noted that we do recognize that these findings and the relative importance of the explanatory variables may be a reflection of the university population from which we sampled. Perhaps a sample from a more general population would increase the importance of distractor type and angular disparity. The effect of practice on the relative contribution of the various difficulty factors should also be investigated in future research.

APPENDIX

Voyer and Hou's[a] MRT distractor type and occlusion classifications, configuration type, and correct response Figs. (**1** and **2**) angular disparities on the y, x, and z axes.

Item	DT	Occl.	CT	CR1Y	CR1X	CR1Z	CR2Y	CR2X	CR2Z
1	M	NO	Hetero	110	17	10	170	0	10
2	M	NO	Hetero	145	20	10	80	15	10
3	S	NO	Hetero	170	0	10	60	3	0
4	S	NO	Hetero	160	15	10	85	10	10
5	M	NO	Hetero	95	8	0	165	8	0
6	M	NO	Hetero	145	1	0	80	0	0
7	S	NO	Hetero	175	0	0	68	8	0
8	S	NO	Hetero	180	0	0	90	2	0
9	M	O	Homo	120	10	0	130	15	10
10	M	O	Homo	160	15	10	70	15	10
11	M	O	Homo	130	40	0	95	2	0
12	M	NO	Hetero	90	5	10	50	7	0
13	S	NO	Homo	165	25	10	65	2	0
14	S	O	Homo	180	5	10	85	5	10
15	M	O	Homo	120	10	0	100	3	0
16	M	NO	Homo	160	55	20	160	58	80
17	S	O	Homo	180	0	0	75	2	0
18	Mixed	O	Homo	170	7	0	95	12	0
19	M	NO	Homo	110	7	0	150	20	0
20	M	NO	Homo	120	1	0	60	18	0
21	S	NO	Homo	165	14	0	45	12	0
22	S	NO	Homo	170	27	0	120	1	0
23	Mixed	NO	Hetero	155	0	0	175	0	0
24	S	NO	Homo	70	15	10	125	25	0

Note. DT = distractor type, occl. = occlusion, CT = configuration type, CR1Y: Correct response Fig. (**1**) angular disparity on the y axis; CR1X: Correct response Fig. (**1**) angular disparity on the x axis; CR1Z: Correct response Fig. (**1**) angular disparity on the z axis; CR2Y: Correct response Fig. (**2**) angular disparity on the y axis; CR2X: Correct response Fig. (**2**) angular disparity on the x axis; CR2Z: Correct response Fig. (**2**) angular disparity on the z axis. M = Mirror, S = Structural, O = Occluded, NO = Nonoccluded, Hetero = Heterogeneous, Homo = Homogeneous.
[a] [2].

REFERENCES

[1]   Vandenberg SG, Kuse AR. Mental rotation, a group test of three-dimensional spatial visualization. Percept Mot Skills 1978; 47: 599-604.

[2]   Voyer D, Hou J. Type of items and the magnitude of gender differences on the Mental Rotations Test. Can J Exp Psychol 2006; 60: 91-100.

[3]   Shepard RN, Metzler J. Mental rotation of three dimensional objects. Science 1971; 171: 701-3.

[4]   Peters M, Laeng B, Latham K, Jackson M, Zaiyouna R, Richardson C. A redrawn Vandenberg and Kuse mental rotations test: different versions and factors that affect performance. Brain Cogn 1995; 28: 39-58.

[5]   Goldstein D, Haldane D, Mitchell C. Sex differences in visual-spatial ability: the role of performance factors. Mem Cogn 1990; 18: 546-50.

[6]   Voyer D, Rodgers MA, McCormick PA. Timing conditions and magnitude of gender differences on the Mental Rotations Test. Mem Cogn 2004; 32: 72-82.

[7]   Lohman DF. The effect of speed-accuracy tradeoff on sex differences in mental rotation. Percept Psychophys 1986; 39: 427-36.

[8]   Jolicoeur P. Mental rotation and the identification of disoriented objects. Can J Psychol 1988; 42: 461-78.

[9]   Tapley SM, Bryden, MP. An investigation of sex differences in spatial ability: mental rotation of three dimensional objects. Can J Psychol 1977; 31: 122-30.

[10]   Cooper LA, Shepard RN. Chronometric studies of the rotation of mental images. In: Chase WG, Ed. Visual information processing, New York: Academic Press 1973; pp. 76-176.

[11]   Albers F, Höft S. "Do it again and again. And again?" Übungseffekte bei einem computergestützten Test zum räumlichen Vorstellungsvermögen. Diagnostica 2009; 55: 71-83.