Spatial Ability Evaluated by a Mental Rotations Test and Its Relations to Undergraduate Graphics Education

（メンタル・ローテーション・テストで評価される空間認識力及びその図学教育との関係に関する研究）

Kumiko SHIINA

1995

PhD Thesis

Spatial Ability Evaluated by a Mental Rotations Test and Its Relations to Undergraduate Graphics Education

Kumiko SHIINA

General Systems Studies
Graduate Division of International and Interdisciplinary Studies
University of Tokyo

Contents

List of Figures

List of Tables

Chapter 1

INTRODUCTION

1.1 Evaluation of Spatial Ability

In many fields of engineering, it is very important and necessary to communicate through drawings. Engineers are required to connect two-dimensional drawings and three-dimensional real objects and to "think in three dimensions" through drawings. One of the important objectives in early undergraduate graphics curricula is not only acquisition of technical skill of drawings but also enhancement of students' ability to think in three dimensions. In graphics education, "spatial ability" is used as a term that means the ability to think in three dimensions.

Recently, some researchers have made attempts to evaluate students' spatial abilities not only by performance on term-end tests but by some other indices. In their quasi-experimental research, psychometrical "spatial tests" are used as such indices.

In psychometric studies, many investigators have been interested in the ability of "spatial domain" as a component of human intelligence. A large number of "spatial" tests were independently constructed by different researchers in order to estimate the ability of spatial domain. They are called "spatial tests" mainly because they consist of two- or three-dimensional figures and because don't contain verbal problems. The relation between the ability evaluated by such psychometric "spatial tests" and "spatial ability" required in early undergraduate graphics curricula is not exactly clear at present.

1.2 Objectives of Research

A Mental Rotations Test (hereafter "MRT") developed by Vandenberg and Kuse[1] has been used as a measure of spatial ability in psychometric studies. Because of its high reliability, the MRT is widely used in recent quasi-experimental research whose objective is to evaluate students' spatial ability from the viewpoint of how it is related to graphics curricula.

Although Vandenberg et al.[1] reported high correlation between the MRT and other spatial tests, it is not clear which aspects of spatial ability are evaluated by the MRT. The objectives of this paper are to get some insight into the following points:

- which aspects of the spatial ability are evaluated by the MRT,
- and how the MRT is related to undergraduate graphic curricula.

1.3 Construction

This dissertation is constructed from 5 chapters.

In Chapter 2, the author will review studies on spatial ability from the following points of view:

- spatial ability evaluated by so-called spatial tests,
- the individual differences in the spatial ability, and
- the relationship between the spatial ability and undergraduate graphics education.

In Chapter 3, the solving process of the MRT will be analyzed by using the following methods:

- error analysis of the MRT in the paper-and-pencil version,
- direct monitor of eye fixation during solving each question and strategy reports, and
- analysis of response time as a function of an angular disparity.

By combination of these methods, the ability reflected by the score in the MRT will be discussed.

In Chapter 4, the relation between the MRT and performance in descriptive geometry will be discussed from the following viewpoints:

- whether the ability reflected by the MRT has correlation with the ability acquired through education of descriptive geometry, and
- whether the ability evaluated by the MRT can be enhanced through undergraduate graphics curricula.

In Chapter 5, the results of this dissertation will be summarized.

Chapter 2

REVIEW OF STUDIES ON SPATIAL ABILITY

Many investigators have shown interest in the human ability of "spatial domain". A large number of "spatial" tests were independently constructed by different researchers in order to estimate the abilities of spatial domain. Recently, some researchers have been paying attention to the relation between students' spatial abilities and graphics curricula in universities. In this chapter, the author will review the spatial ability from the following points of view:

1. the spatial ability evaluated by the so-called spatial tests,
2. the individual differences in the spatial ability, and
3. the relationship between the spatial ability and graphics education.

2.1 What is "Spatial Ability"?

2.1.1 Spatial ability in intelligence and in graphics education

In psychometrical studies, "spatial ability" means the total ability in the spatial domain as a component of human intelligence. "Spatial ability" is considered to be reflected by scores in "spatial tests". Tests, which consist of two- or three-dimensional figures and which do not contain verbal problems, are called spatial tests. According to Zimovski and Wothke[2], numerous "spatial" tests share one characteristic: "all required the processing visuospatial stimuli (p.5)", although the difficulty or complexity of spatial tests varies one to another. Some researchers[3][4][5] classified spatial ability into two or three sub-components based on factor analysis, however, their interpretations were seen to differ.

In graphics education, "spatial ability" is used as a term that refers vaguely to the ability to think in three dimensions. Miller, Wiley and Bertoline[6] used "visualization" which corresponds to spatial ability and defined it as "the ability to mentally create and edit visual information". Spatial ability in graphics education is more restrictive than that in psychometrical studies and emphasizes the ability to connect two-dimensional drawings and three-dimensional real objects.

2.1.2 Components of spatial domain

Factor analytic research distinguished "spatial domain" from human intelligence. Snow[7] showed a structural model based on the intercorrelation between 34 mental tests and affixed labels to the clusters of tests. As factor labels, he used G_c, G_f, G_v, terms which stand for "crystallized ability", "fluid ability", "visualization". In his analysis, the distinction between verbal (G_c) and spatial factors (G_f, G_v) was clear, although it was difficult to distinguish between G_f and G_v. As for figural and spatial relations tests, some were attributed to G_f and others to G_v.

Several investigators conducted factor analytic research to gain insight to the structure of spatial ability. Two reviews of spatial ability[3][4] claimed the existence of two or three factors.

Based on his review of the spatial abilities, McGee[3] proposed two distinct spatial abilities as follows:

1. *Spatial orientation.* This factor involves "the comprehension of the arrangement of elements within a visual stimulus pattern and the aptitude to remain unconfused by the changing orientations in which a spatial configuration may be presented (p.893)".

2. *Spatial visualization.* This factor involves "the ability to mentally rotate, manipulate, and twist two- and three-dimensional stimulus objects (p.896)".

According to McGee's formulation, *spatial orientation* emphasizes the "invariable arrangement" of elements in spite of changing orientations. On the other hand, *spatial visualization* is characterized by various "mental manipulations".

Lohman and Kyllonen[4] proposed the following three factors: Spatial Orientations, Spatial Relations, and Visualization. Lohman et al.[4] and McGee[3] agreed on the names, but not the definition, of two of the three factors.

1. *Spatial Orientation.* The factor involves "the ability to imagine how a stimulus array will appear from another perspective (p.111)".

2. *Spatial Relations.* "Although mental rotations is the common element, the factor probably does not represent speed of mental rotation; rather, it represents the ability to solve such problems quickly, by whatever means (p.111)".

3. *Visualization.* The tests that load on this factor are "all administered under relatively unspeeded conditions, and most are much more complex (p.111)" than tests that load on other factors.

The definition of spatial orientation by Lohman et al.[4] agreed with McGee's[3] that both involved a stimulus seen from a different position, though McGee[3] further required "comprehension" of the invariable arrangement of elements.

As for spatial visualization, the definitions disagree with each other. Lohman et al.[4] added "spatial relations". In Lohman et al.'s criteria, speededness and complexity play an important role in distinguishing the factors. Complexity and unspeeded condition characterize visualization and simplicity and speededness characterize spatial relations. The spatial relations factor is called "speeded rotation" afterwards by Lohman[8] (cited in Juhel[5]). Lohman et al.[4] distinguished speeded mental rotation from more complex manipulation and classified them as spatial relations, although McGee[3] described all tests involving mental rotation of two- and three-dimensional objects as spatial visualization.

Juhel[5] shares his distinction in spatial ability with Lohman et al.[4]. Juhel classified spatial ability into three groups: *spatial orientation, spatial visualization* and *speeded rotation.* He described "*speeded rotation*" as follows : "It is primarily defined by simple tests setting mental rotations of forms or objects in action. Because of their simplicity, the speed of rotation is of great importance in these tests (p.118)." According to Juhel's definition, simple task with high accuracy belongs to speeded rotation factor. Conversely, a more complex task, which requires more time to get the correct answer, belongs to the spatial visualization factor. Juhel's distinction between speeded rotation and spatial visualization is similar to that of Lohman et al.[4].

As above, the inconsistencies among the formulations[3] [4][5] are due to the limitations of the factor analytic approach. Factor analysis attempts to describe the process of solution through inferences concerning those features common to tests that load on the same factor and doesn't directly clarify the internal solving process.

Lohman et al.[4] made an attempt to classify the types of mental transformations required on the complex spatial tests.

1. Mental movement. "Reflecting, rotating, folding, or simply imagining that a stimulus is moved from one position in an array to another position are all varieties of mental movement (p.112)."

2. Mental construction. There are different levels of construction, as follows, in order of complexity:

   (a) Copying. "The subject must correctly copy a stimulus design (p.112)".

   (b) Reproducing. "The design must be reproduced from memory", "not just recognized (p.112)".

   (c) Constructing. The subject must "construct a new mental image, usually by reorganizing the stimulus element in a new way (p.112)".

   (d) Deleting. "Mentally deleting parts of a stimulus (p.112)".

The classification suggests that Lohman et al.[4] were conscious of the importance of information processing during solving spatial tests.

2.1.3 From cognitive perspective

Zimovski et al.[2] and Linn and Petersen[9] emphasize the need for a classification scheme based on information processing, not based on the figural features of spatial tests. Their classification of spatial ability tests arose from paying attention to processes used to solve the tests.

Linn et al.[9] postulated three spatial ability categories based on similarities in the solving processes. Three distinctions presented below were characterized by their solving processes:

1. Spatial perception. Use of gravitational vertical or horizontal to locate the correct orientation in order to determine spatial relationships with respect to the orientation of their own bodies. " A gravitational / kinesthetic process".
2. Mental rotation. "A Gestalt-like mental rotation process analogous to physical rotation of the stimuli".
3. Spatial visualization. "Possibility of multiple solution strategies". "An analytic process".

The classification by Zimovski et al.[2] is very different from those of previous psychometric studies[3][4][5] and also that of Linn et al.[9]. They classified the so-called "spatial tests" into two types: *analog* and *nonanalog* measures. They called tests which require holistic gestalt-like processing "*analog* measures" and indicated the properties of *analog* measures as follows:

1. Tasks involving judgments about rotated stimuli.

2. Stimuli differing orientations other than 180 degrees.

3. Cases where the distractors of rotation tasks are mirror images of the reference stimuli or structurally equivalent forms.

4. Items requiring whole-whole rather than part-whole or part-part comparisons.

5. Cases where the items require the rotation of an entire object as a rigid whole rather than the rotation of only one or several pieces of the object relative to the whole.

Zimovski et al.[2] claimed that only analog measures should be termed *tests of spatial ability* in pure form and that nonanalog measures reflect both verbal and spatial abilities even if they have a visuospatial content.

In order to discuss "spatial ability", it is necessary to focus not only on the apparent features of tasks but also on information processing. Although there is little consensus of classification at present, further investigations will provide more appropriate classifications.

2.1.4 Solving process approach

As an attempt to clarify internal processes, Shepard and Metzler[10] provided an epoch-making index, i.e., reaction time. Shepard et al.[10] analyzed the relation between reaction time and task attributes and combined it with the internal processing of subjects. Their experiment was designed to measure the reaction time that subjects require to determine whether two figures were views of the same three-dimensional object or were mirror-images of each other. A sample task is shown in Fig. 2.1. The reaction time for the "Same" trials increased linearly with the angular difference in portrayed orientation. This orientation

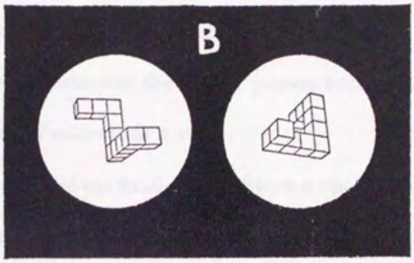


Figure 2.1: Example of Shepard-Metzler tasks (from [10])

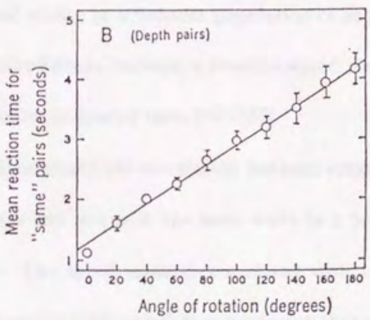


Figure 2.2: Mean reaction times for "Same" trials (from [10])

effect on reaction time shown in Fig. 2.2 is interpreted to reflect the solving process, namely "mental rotation". The slope of the reaction time function is postulated to reflect the rate of mental rotation.

The study by Shepard et al.[10] brought a new paradigm "reaction time" as the objective indicator of information processing during solving spatial tasks. However it was difficult to identify the microstructure of the solving process because reaction times include any processes occurred in subjects as a whole.

Just and Carpenter[11] analyzed eye fixations of subjects during Shepard-Metzler tasks and presented component processes in the tasks. Under the rules for classifying the locus of the eye fixation, they separated the solving performance into the three processes: (1) search, (2) transformation and comparison, and (3) confirmation.

Then, the next question is the source of individual differences. It is valuable to analyze the variation of spatial ability in a broader population of subjects. There are some studies which report the correlations between a rotation speed measure computed based on laboratory tasks and scores in spatial tests [12]–[15].

Just and Carpenter[12] analyzed the correlation between rotation speed measures in the Cube Comparison tasks and scores in the same tasks in a paper-and-pencil version (Cube Comparison test). The speed measure correlated with the psychometric score $r(21) = -.46, P < .05$. Negative high correlation suggested that the score in the paper-test reflected the rotation speed, namely the amount of time the subject took on those problems that required mental manipulation.

Studies on correlations between the slope of the reaction times in a Shepard-Metzler task and reference spatial tests other than paper-and-pencil versions of the same task have been carried out[13]–[15]. In these studies, the slope was considered to reflect the rate of rotation, namely the speed of mental rotation.

Lansman[13] reported a negative high correlation between the slope measure and 6 kinds of spatial tests which loaded on G_v (visualization in Cattel's term) factor ($r = -.50$, P$< .01$). McGue, Bouchard, Lykken and Feuer reported a negative low correlation ($r = -.13$, cited in [14]). Egan found no correlation (cited in [12]) ($r = -.06$, according to Lohman[14]) and Poltrock and Brown[15] a positive correlation $r = .20$ between the slope measure and 8 kinds of spatial tests.

The inconsistency of these studies suggests that not all subjects used the same strategies, namely, mental rotation. In other words, strategies other than "rotation" might contribute to no or positive correlation.

The solving processes were compared between subjects with high and low scores in spatial tests.

Snow[7] analyzed the individual differences in solving process of paper-folding tests from the eye tracks and introspective reports. He showed that high-ability subjects were efficient in assembling a systematic strategy for the task, their control of its application, and their flexibility in changing strategies as task difficulties demanded, whereas low-ability subjects were stuck in a principal strategy.

Just et al.[12] were also interested in the differences in the strategy of Cube Comparison tasks between experts and novices. The low-ability subjects rotated the cube around axes that were perpendicular to the faces of the cube, namely standard trajectories. The high-ability subjects were able to find axes that made the trajectory shorter than the standard ones where an alternative trajectory existed. The choice of trajectory is consistent with the flexibility of strategies shown by Snow[7].

As stated above, high- and low-ability subjects differed in their strategies in some spatial tasks. As for Shepard-Metzler tasks, some researchers[12][17] indicated individual differences within the same strategy. Differences in strategy preference were also reported[12][16].

Just et al.[12] used Shepard-Metzler tasks as a task which is less open to alternative strategies, and showed that high- and low ability subjects in Cube Comparison tasks differed in the rotation rate of Shepard-Metzler tasks. The difference in rotation rates between experts and novices was also found by Lohman[17]. Lohman[17] indicated that the estimated slope was generally steeper for low-ability subjects than high-ability ones.

Just et al.[12] also presented a strategy that requires no mental rotation. They stated that it was possible to construct orientation-free representations by taking an imaginary walk through the interior corridors, although their subjects seldom reported such a strategy.

Sayeki[18] reported that the instruction of a "body analogy" for the figures evoked a strategy other than mental rotation. The reaction time of the subject who took such a strategy was short and constant, or didn't increase with the angular difference. This suggests that strategy preference is related to a reaction time function, or performance in mental rotation.

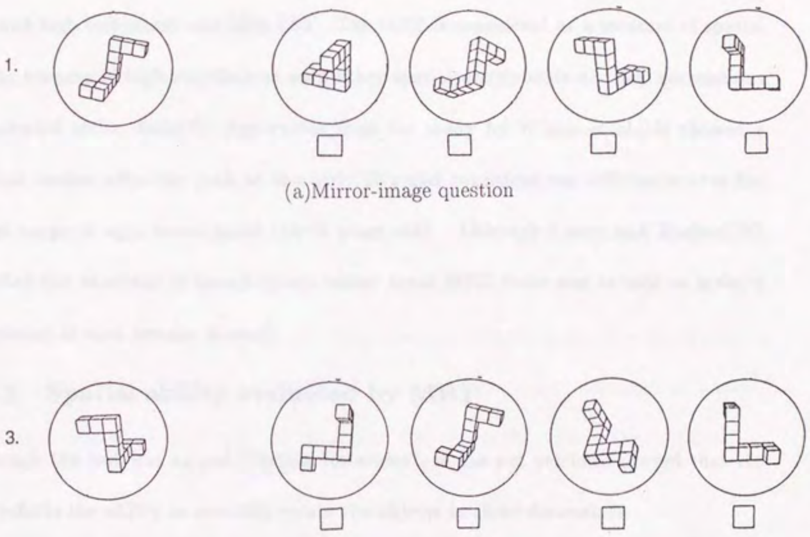
Based on post-experimental interview, Tapley and Bryden[16] classified their subjects' solving approach of Shepard-Metzler tasks into two strategies as follows: a visual-holistic strategy and a verbal-analytic strategy. The former contained a literal mental rotation; and/or a comparison of the two figures with an intuitive response. And the latter contained counting blocks; and/or isolating a particular part(s) and looking for it in the comparison figure. They reported that there were no significant effects of strategy on the slope of the reaction time function.

As discussed above, the differences both in strategies and within one strategy play a key role in the discussion of "spatial ability" in a wide range of subjects. As for Shepard-Metzler tasks, both differences have been reported[12][16]–[18]. However, it remains unclear that how strategy preference of Shepard-Metzler tasks is related to performance in mental rotation.

2.2 Mental Rotations Test

2.2.1 Vandenberg MRT

Vandenberg and Kuse[1] developed the Mental Rotations Test (hereafter MRT) based on the figures used by Shepard and Metzler[10]. Among numerous spatial tests, the MRT has been most widely used as a measure of spatial ability. Sample questions are shown in Fig. 2.3. Each question is composed of a criterion figure, two correct alternatives and two incorrect ones or "distractors". The subjects are required to find the two correct alternatives. The correct ones are always identical to the criterion in structure, but are shown in a depth-rotated position. The test contains 20 questions. For half the questions, the distractors are rotated mirror-images of the criterion, while the distractors in the other 10 questions are rotated images of one or two of the other structures. Hereafter, the former questions shall be called "mirror-image questions" and the latter "structural questions". These two kinds are arranged two by two. The test is divided into two parts. Each part contains 10 questions, and the time limit for solving each part is 3 minutes.



(a)Mirror-image question

(b)Structural question

Figure 2.3: Sample question from MRT (from [1])

The procedure of scoring is to give 2 points for a question with both choices correct, and none if one choice is correct but the other one incorrect, or if both are incorrect. If only one figure was chosen and it is correct, 1 point is given. So, a perfect score for the MRT is 40.

Vandenberg et al.[1] found substantial internal consistency (Kuder-Rechardson 20 = .88) and high test-retest reliability (.83). The MRT is considered as a measure of spatial ability because of high correlations with other spatial ability tests and low correlations with verbal ability tests[1]. Age curves from the study by Wilson et al.[19] showed a gradual decline after the peak at the early 20's and consistent sex differences over the entire range of ages investigated (14-60 years old). Although Casey and Brabeck[20] reported the existence of female group whose mean MRT score was as well as male, a population of such females is small.

2.2.2 Spatial ability evaluated by MRT

Although the test was named "Mental Rotations", it has not yet been proved that the test reflects the ability to mentally rotate the objects in three dimensions.

As for the Shepard-Metzler task, McGee[3] classified it as a test of *spatial visualization* for the reason that the task requires mental transformation. But his classification does not derive from a solving process.

Lohman[17] determined the rotation speed measures of the Shepard-Metzler tasks for each subject, then correlated these parameters with the Spatial Relations (SR) composite and the Visualization (Vz) composite. Both the SR composite and the Vz composite significantly correlated with the speed parameters. This suggests that the performance in

Shepard-Metzler tasks reflects both "Spatial Relations" and "Visualization" factors.

Although the MRT was developed based on the Shepard-Metzler task, it is doubtful if the MRT can be considered as a replica of the Shepard-Metzler task, because the change of response method in the MRT (choice of alternatives) may bring about a different solving process from the original Shepard-Metzler task. The relation between the Vandenberg MRT and the original Shepard-Metzler task remains unclear.

Juhel[5] conducted a factor analysis based on a matrix of correlations between accuracy scores for 5 spatial tests and 4 laboratory memory-tasks. Among four first-order factors, two were associated with the spatial tests. The MRT was highly loaded on both, which Juhel labeled as visualization and speeded rotation. According to his interpretation, the speeded rotation factor was mainly evoked by a simple two-dimensional test rather than by the MRT. All three tests which had a high-loading on the visualization factor shared a common property: they require transformations and movements performed on internal parts of mental representations. So he stated that the MRT reflects the visualization factor.

Zimovski et al.[2] classified the MRT as an analog measure because the distractors which are mirror images of the criterion figure minimize "feature-extraction" strategies. However, Freedman and Rovegno[21] and Pezaris and Casey[22] reported strategy differences to solve the MRT between subject groups.

In the investigation by Freedman et al.[21], the subjects filled out a questionnaire on strategy after finishing the MRT. The males reported that they counted blocks less, pictured in their minds more than the females.

Pezaris et al.[22] estimated strategy preference by comparing subjects' MRT scores between the control condition and the interference condition, in which the subjects must perform either a verbal or a visual-spatial interference task concurrently with the MRT. They indicated that the males depended more on visual-spatial strategies and did less on verbal ones than the females. They also reported the existence of females who depended more on visual-spatial strategies than the other females.

In these studies by Freedman et al.[21] and Pezaris et al.[22], their main interest was sex-based strategy differences. As Wilson et al.[19] indicated, males have the advantage of the scores in the MRT. The subjects in their studies differed not only in sexes but also in scores. It is necessary to be conscious of strategy differences between high- and low-score subjects.

Above studies didn't pay attention to the relation between strategies and performance in mental rotation. The relation remains unclear.

In order to make clear the ability reflected in the scores in the MRT, it is necessary to clarify the following points;

- strategy differences between high- and low-score subjects, and
- relation between strategy preference and performance in mental rotation.

2.3 MRT and graphics education

2.3.1 Graphics education and spatial ability

In early undergraduate graphics curricula, descriptive geometry, i.e., the treatment of points, lines and objects through orthogonal projections, is taught to students in engineering (or art) courses. It is difficult for students to jump the gap between the three dimensional object and its two dimensional representation, although they have mathematical knowledge about solid geometry.

The purpose of descriptive geometry education is not only the acquisition of the skill required in technical drawing but also the enhancement of students' spatial ability which means "the ability to think in three dimensions" in this case.

Recently, some quasi-experimental studies were made to evaluate students' spatial ability from the view point of how they are related to graphic curricula[6] [23]−[25]. In these research, some psychometric spatial tests were used as a measure of spatial ability. But it is not clear if students with high or low scores in spatial tests show a high or a low performance in descriptive geometry. The relation between spatial abilities reflected in spatial tests and those required to get through graphics curricula is open to question.

2.3.2 MRT and graphics curricula

Among many psychometric spatial tests, the MRT has been widely used in recent quasi-experimental research, whose objective is to evaluate the spatial abilities of students from the viewpoint of their relationship to graphics curricula[6] [23]−[25].

Miller et al.[6] analyzed the relation between scores in the MRT and the test performance in traditional graphics curricula. Miller et al.[6] reported that many students who scored in the preceding MRT 1 standard deviation or more below the mean, had difficulty with graphics curricula and had a high failure rate.

Test performance reflects students' achievement in graphics curricula as a whole. Graphics curricula contain various descriptive geometry concepts. It remains unclear which concept taught through graphics curricula is related to scores in the MRT and which concept is not related.

Recently, researchers have been making attempts to introduce new teaching methods into graphics education. In order to compare the effect of a new teaching method with a traditional one, the effect needs to be evaluated by an objective measure. Because of its high reliability [1],the MRT has been most widely used as an objective measure of spatial ability.

Miller[23] showed that there was no interaction between teaching methods (using real models, computer-generated models, or traditional curricular approaches) and MRT scores. Sexton[24] found no significant difference in students' mean scores in the MRT regardless of the method of instruction, i.e., traditional or by 3D CAD. McCuistion[25] conducted research to find which method of presenting graphic images (static or dynamic) in a CAI (Computer Assisted Instruction) lesson would best enhance student achievement in the descriptive geometry course. In his research, the MRT was given to the students prior to and after each lesson course (static or dynamic). Although some students of the dynamic group made a larger gain on scores in the MRT, the static group mean score in the performance tests was significantly higher than in the dynamic group. He surmised that a larger gain on MRT scores does not guarantee high scores in performance tests and vice versa.

The above studies do not reveal whether the increase of score in the MRT means the direct effect of graphics education or not. It is not clear whether or not the spatial abilities evaluated by the MRT can be enhanced through graphics curricula.

2.4 Summary

The review on the related studies in this chapter is summarized as follows:

- The inconsistencies among the classification of spatial ability are due to the limitations of the factor analytic approach, in which interpretations of factors are inferred from apparent features common to spatial tests that load on the same factor. The solving process approach will play a key role in clarifying microstructure of spatial ability.

- There are two sources of individual differences in the solving process; the difference in strategies and that within one strategy. To discuss spatial ability based on the solving process, it is necessary to be conscious of the two sources of the differences. As for the MRT, it has not yet been proved that the test reflects the ability to mentally rotate the objects, though it was named "Mental Rotations".

- Relationship between "spatial ability" reflected in psychometric spatial test scores and "spatial ability" required to get through graphics curricula is not clear at present.

Chapter 3

ANALYSIS OF PROBLEM SOLVING PROCESS OF MRT

Clarifying the solving process of the MRT is a strong hint on the ability reflected by the MRT. In this chapter, the solving process of the MRT for both experts and novices is compared by using two methods as follows: error analysis of the MRT in the paper-and-pencil version and direct monitor of eye fixation during the solution of each question. The former is suitable for collecting data from a large number of subjects. Statistic analysis will suggest the probabilistic nature of the solving process, although the solving process of each subject isn't clear. The latter is suitable for analyzing the microstructure of the solving process, although data from a small number of subjects is obtained at one time. By a combination of these methods, the ability reflected by the score in the MRT will be discussed.

3.1 Scores in MRT

3.1.1 Subjects

The MRT was given to the students at 6 universities. The total number of subjects participating in the investigations was more than 2000. The age curves of the MRT score from the study by Wilson et al.[19] showed a gradual decline after the peak in the early 20's and consistent sex differences over the entire range of ages investigated (14-60 years old). The ages of the subjects in the present investigations approached the peak age in the previous study by Wilson et al.[19].

3.1.2 Difference between sexes and within each sex

The group mean scores in the MRT in each course at each university are summarized in Table 3.1. As clearly seen in Table 3.1, the mean scores at the same university were quite similar to each other, although the subjects were different in each test. In other words, the mean scores in the MRT are quite stable within the same course. The high stability in the same group through years suggests that it is appropriate to consider the MRT as an indicator of some kinds of spatial abilities.

In Fig. 3.1, the group mean scores in the MRT are plotted as a function of the estimated mean Z-scores of the National Center Test for University Entrance Examination[27] (hereafter NCTUEE). The sex differences in MRT score were observed between groups whose NCTUEE Z-scores were almost same. The sex difference in MRT scores agreed with the result reported by Wilson et al.[19].

As for male groups, the mean MRT scores in groups with low Z-scores in the NCTUEE were not so different as those in groups with a high Z-score. Suzuki et al.[28] reported that scores in a Mental Cutting Test (hereafter MCT) had considerable correlations with the NCTUEE Z-scores in male groups specializing in science/engineering courses, whereas the present investigations indicated that the Z-score in NCTUEE doesn't have a strong correlation on the MRT scores in male groups. It suggests that the MRT reflects different

Table 3.2: Group mean scores in MRT

Univ.	Spec.*	Sex**	Samples	Mean (SD)	Rem.
T1	SE	M	99	29.7 (6.8)	'90
		M	111	29.9 (6.4)	'91
		M	76	28.0 (6.5)	'91
		M	105	29.9 (6.6)	'92
		M	102	27.6 (6.9)	'93
		M	100	27.7 (7.8)	'94
		F	44	22.2 (9.1)	'90-'94
T2	SE	M	163	27.4 (7.0)	'91
		M	177	28.3 (7.5)	'92
		M	118	27.9 (8.0)	'91
		M	143	24.4 (7.8)	'92
T3	SE	M	132	23.4 (7.8)	'93
		M	79	22.3 (7.2)	'94
		F	31	20.2 (7.2)	'93 & '94
T4	SE	M	2	26.5 (2.1)	'92
		F	56	17.3 (5.6)	
T5	SE	(M)	93	25.28	from [26]
F1	H	F	50	17.5 (6.1)	'91
		F	89	15.0 (6.8)	'92
		F	43	15.5 (7.3)	'93
		F	26	15.8 (6.1)	'94
		F	110	13.6 (6.5)	'92
	(SE)	F	103	17.9 (6.9)	'93
		F	53	15.8 (5.8)	'94
F2	H	F	46	16.8 (6.4)	'92

Speciality SE : Science and/or Engineering
(SE) : Similar to SE course,
H : Home Economics.

Sex M : Male
F : Female
(M) : Almost all the students are male students.

SD : Standard deviation

aspects of spatial ability from the MCT.

As for female groups, whose NCTUEE Z-scores were nearly 50, the mean scores in the MRT were almost the same. Although some female groups with relatively high NCTUEE Z-scores showed higher mean scores in the MRT than other female groups, these groups were different not only in NCTUEE Z-scores but also in their specialty: science/engineering and home economics. It remains to be tested which factor has a greater influence on the MRT, NCTUEE Z-scores or the specialty.

3.1.3 Comparison with Mental Cutting Test

A Mental Cutting Test (hereafter MCT) has been most widely used to evaluate students' spatial ability in Japan[28]. A sample question from the MCT is shown in Fig. 3.2. Subjects are given a pictorial view of object with a cutting plane and required to choose the true cutting views from amongst alternatives. In this section, results of the MRT obtained at several universities will be compared with those of the MCT.

In three universities, both of the MRT and the MCT were given to students. The coefficients of correlations between the scores in the MRT were shown in Table 3.2. Vandenberg et al.[1] reported that the MRT showed a high correlation with other spatial ability tests and a low correlation with verbal ability tests. Other studies also reported high correlations between the MRT and other spatial tests[5][12].

Suzuki et al.[28] tentatively concluded that the MCT reflects the ability to make mental images from pictorial views. Although the coefficients of correlation between the MRT and the MCT in the present investigations indicated a considerable correlation, they were not as high as those between the MRT and other spatial tests except that at O

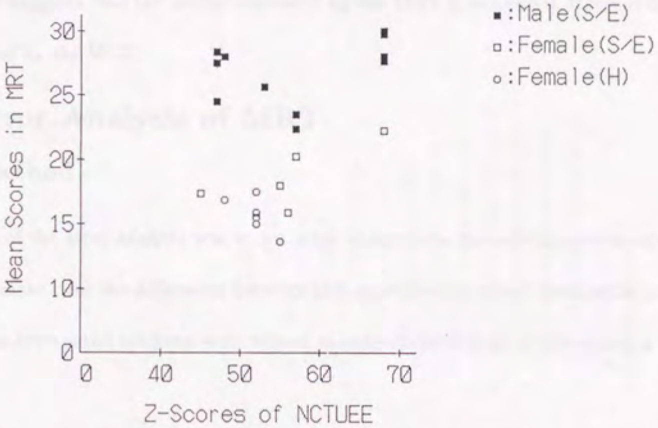


Figure 3.1: Scores in MRT and NCTUEE Z-scores

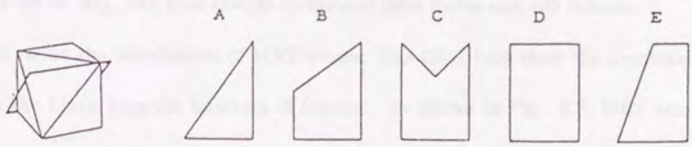


Figure 3.2: Sample question from Mental Cutting Test

Table 3.2: Coefficients of correlations between MRT and MCT

Univ.	Sample	Coefficients of Correlation
T.	99	0.29
O.	48	0.58
M.	178	0.39

university. It suggests that the ability evaluated by the MRT is somewhat different from that evaluated by the MCT.

3.2 Error Analysis of MRT

3.2.1 Method

The purpose of the error analysis was to get some insight into the solving process of the MRT and to make clear the differences between high-spatial subjects and low-spatial ones. Here, high- or low-spatial subjects were defined as subjects with high or low scores in the MRT.

The subjects were students taking descriptive geometry courses at three universities. The subjects were administered the MRT. The time limit for each part was 3 minutes, as Vandenberg recommended[29]. The investigation for naive students had been held for 4 years (from '90 to '93). The final sample consists of 1094 males and 420 females.

Fig. 3.3 shows the distribution of MRT scores. The filled bars show the numbers of males, and the blank bars the numbers of females. As shown in Fig. 3.3, MRT scores indicated normal distribution. The scores of the males were distributed over higher ranges than those of the females. The difference in the distribution between males and females was consistent to the investigation by Wilson et al.[19]. Sex differences will be discussed in a later section (Section 3.2.4).

For the sake of comparison, high-spatial and low-spatial groups were defined as follows:

- High-spatial group: The upper 10 percent of the sample.
- Low-spatial group: The lower 10 percent of the sample.

The group mean scores are shown in Table 3.3.

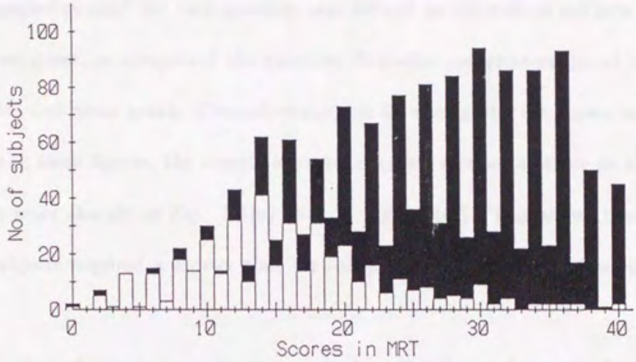


Figure 3.3: Distribution of scores in MRT

Table 3.3: Mean scores in MRT from the high-spatial and low-spatial groups

Group	Sample	Group mean score (SD)
High-spatial	206	37.5 (1.6)
Low-spatial	165	8.8 (3.0)

3.2.2 Completion-rate and mean point of complete-subjects

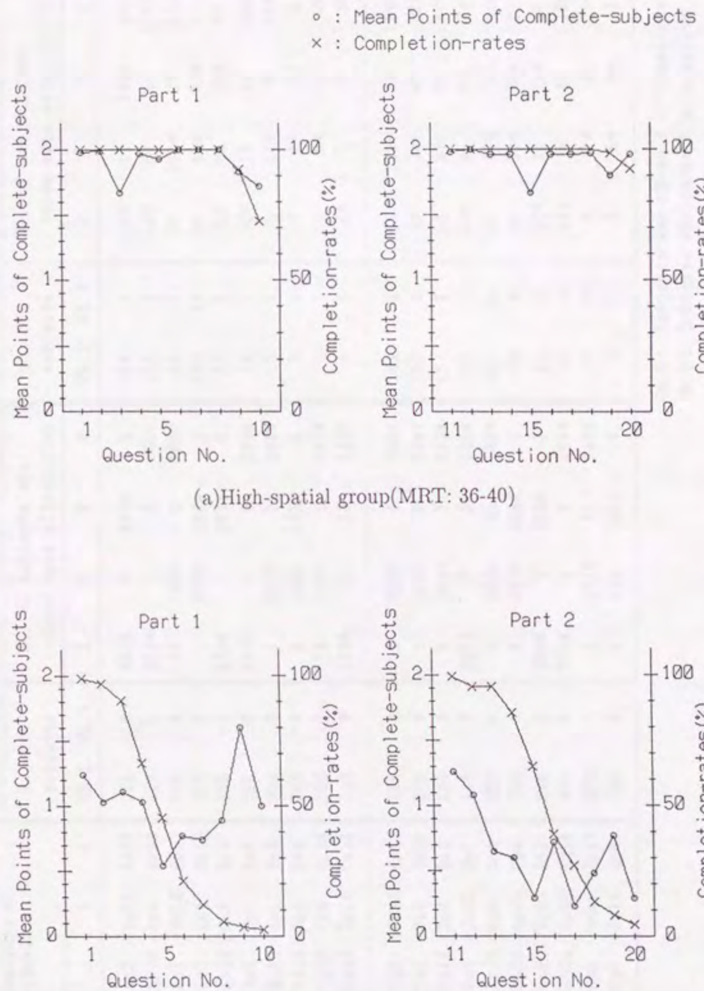
The term "completion-rate" for each question was defined as the rate of subjects who chose two alternatives, or completed the question (hereafter complete-subjects). The completion-rates and mean points of complete-subjects for each group are shown in Fig. 3.4. As shown in these figures, the completion-rates decrease as the questions go along. They decrease more sharply in Fig. 3.4(b) than in Fig. 3.4(a). This shows that the high-spatial subjects required a shorter time for solving the MRT than the low-spatial ones.

The mean points of completion-subjects are steady and high for almost all the questions as shown in Fig. 3.4(a), while they are low and fluctuate sharply as shown in Fig. 3.4(b). This indicates that the high-spatial subjects gave correct answers as long as they chose, while the low-spatial subjects tried but often failed.

Fig. 3.4 shows that there are some questions whose mean points of complete-subjects are lower than the others, and that such questions are different between the high-spatials and the low-spatials. This indicates that, although the questions in the MRT seem to be similar to each other, each question has a different relative-difficulty, and that the relative-difficulties are different for the two groups.

3.2.3 Distribution of alternatives

Constructions of the questions in the MRT (from [29]) are shown in Table 3.4. In this table, "A" to "E" denotes one of five three-dimensional objects used by Shepard et al.[10]. The subscript "p" or "n" indicates that the object is a "positive" one or "negative" one, respectively. The numerical subscripts denote the rotation angles from the home position.



(a)High-spatial group(MRT: 36-40)

(b)Low-spatial group(MRT: 0-12)

Figure 3.4: Completion-rates and mean points of complete-subjects for the high-spatial and low-spatial groups

Table 3.4: Construction of questions(from [29]) and distribution of alternatives

Q.No	Constructions of each question					High-spatial Group: No. of subjects		High-spatial Group: No. of subjects who chose each alternative				Low-spatial Group: No. of subjects		Low-spatial Group: No. of subjects who chose each alternative			
	C	1	2	3	4	Ch.2	Ch.1	1	2	3	4	Ch.2	Ch.1	1	2	3	4
1	Ap1	Ap6	An2	Ap11	An13	206	0	205*	1	204*	2	163	1	119*	34	136*	38
2	An6	An13	Ap10	Ap6	An10	206	0	204*	1	2	205*	160	1	116*	47	44	114*
3	Ap10	Bp5	Ap1	Bn12	Ap13	206	0	34	188*	1	189*	149	7	53	114*	23	115*
4	An11	Bn8	An1	An6	Bp15	205	1	2	205*	202*	2	109	11	36	89*	77*	27
5	Bn14	Bn1	Bp15	Bn6	Bp8	206	0	200*	3	203*	6	75	8	44*	34	50*	30
6	Bp1	Bp8	Bn5	Bn1	Bp5	206	0	206*	0	0	206*	34	7	22*	14	11	28*
7	Bn5	An10	Bn14	Ap17	Bn8	206	0	1	205*	0	206*	19	4	8	15*	9	10*
8	Bp6	Ap13	Bp14	Bp1	An2	206	0	1	206*	205*	0	9	1	3	6*	6*	4
9	Cp5	Cn17	Cp10	Cn6	Cp15	190	7	12	179*	5	191*	5	1	1	5*	1	4*
10	Cn17	Cn10	Cp14	Cp10	Cn14	149	9	136*	11	13	147*	4	1	4*	1	2	2*
11	Cp14	Ep3	Cp5	En15	Cp17	206	0	1	204*	2	205*	164	0	24	129*	46	129*
12	Cn15	Ep6	Cn5	Dn2	Cn10	206	0	1	205*	0	206*	157	3	40	132*	47	98*
13	Dn15	Dp11	Dn10	Dp4	Dn2	206	0	5	203*	1	203*	157	2	63	93*	63	97*
14	Dp15	Dp4	Dn1	Dn15	Dp1	206	0	206*	0	5	201*	141	5	99*	45	69	74*
15	Dn1	Cn5	Dn10	Ep6	Dn4	206	0	3	194*	32	183*	107	10	36	42*	78	68*
16	Dp2	Ep15	Dp10	Dp15	Cp6	206	0	3	203*	204*	2	64	6	25	50*	38*	21
17	Ep5	Ep10	En6	Ep15	En17	206	0	205*	2	202*	3	44	3	35*	22	16*	18
18	En10	En17	Ep14	Ep10	En14	206	0	205*	1	3	203*	21	2	14*	5	15	10*
19	Ep14	Cp3	Ep5	Dn4	Ep17	202	0	1	194*	21	188*	13	0	3	7*	6	10*
20	En15	Dn11	En5	En10	Cp6	190	7	1	190	193*	3	7	1	4	6*	3*	2

Ch.2 : Subjects who chose 2 alternatives
Ch.1 : Subjects who chose 1 alternative
* : Correct alternative

This table also shows the number of subjects who chose each alternative. As shown in Table 3.4, the distractors are rotated mirror-images of the criterion for half of the questions, while distractors in the other 10 questions are rotated images of one or two of the other structures. The author refers the former questions as "mirror-image questions" and the latter, "structural questions" as stated in the previous chapter (see Fig. 2.3). The mean scores of mirror-image questions and structural questions are shown in Table 3.5.

High-spatials

For the high-spatial group, the mean points of completion-subjects were lower in questions 3, 15, and 19 (see Fig. 3.4(a)).

In question 15 where the criterion is Dn, the number of subjects who chose the distractor "3" (Ep) was much larger than that of the subjects who chose the distractor "1" (Cn), as shown in Table 3.4. This indicates that subjects felt difficulty in distinguishing Ep from Dn. Similar partiality was observed in question 19, which contains the same pair of Ep and Dn. The figures of Dn and Ep are shown in Fig. 3.5. They are identical in topological structure, and are different only in the numbers of cubes in the central joints.

Assuming that the subjects concentrated their efforts to mentally rotate the objects and paid little attention to the numbers of cubes in the central joints, this may explain why the subjects made a lot of mistakes in questions 15 and 19.

In question 3 whose criterion is Ap, the number of subjects who chose the distractor "1" (Bp) was much greater than the number of subjects who chose the distractor "3" (Bn). This suggests that the subjects felt difficulty in distinguishing Ap from Bp in question

Table 3.5: Mean scores from mirror-image questions and structural questions

Group	Sample	Mirr.	Str.	t-value
High-spatial	206	18.7	18.8	0.466
Low-spatial	165	4.1	4.7	2.482*

Mirr. : Mean score of mirror-image questions
Str. : Mean score of structural questions
* : P<0.05

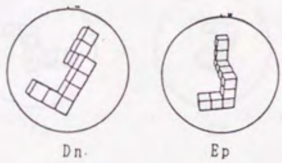


Figure 3.5: Dn and Ep

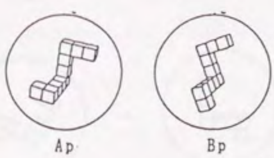


Figure 3.6: Ap and Bp

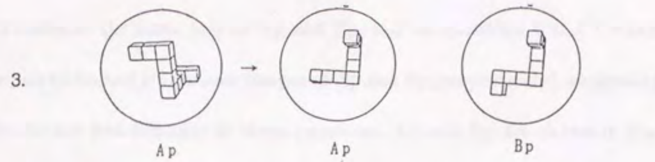


Figure 3.7: Invisible question(Question 3)

Left : Criteria, Middle : Rotated Criteria, Right : Alternatives

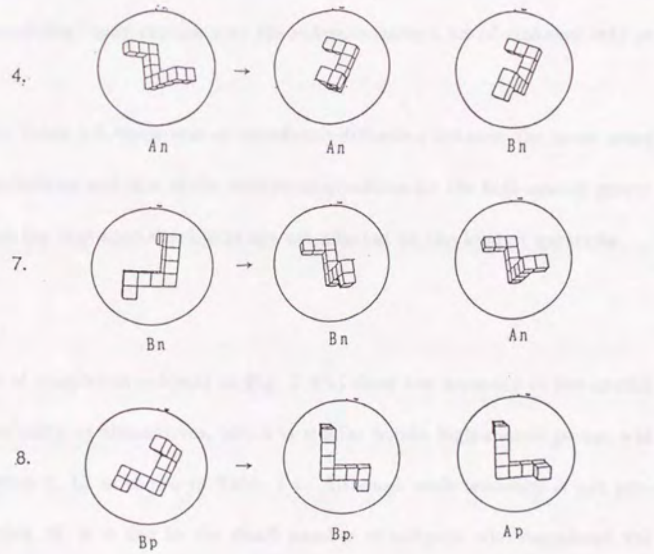


Figure 3.8: Visible questions(Questions 4, 7 and 8)

Left : Criteria, Middle : Rotated Criteria, Right : Alternatives

3. The number of subjects who chose each distractor was, however, not so different in question 8 which contains the same pair of Ap and Bp, and in questions 4 and 7 which contain a similar pair of An and Bn (mirror-images of Ap and Bp, respectively), suggesting that the subjects did not feel difficulty in these questions. Ap and Bp are shown in Fig. 3.6. They are similar in structure, and are different only in the direction of the lower arm.

When the criterion object is rotated until it locates in the same direction with the alternative, the lower arm which plays a key role in the distinction between the criterion and alternative is "invisible" in question 3 (see Fig. 3.7), while it is "visible" in questions 4,7 and 8 (see Fig. 3.8). Assuming that the subjects solve the questions by mental rotation, the "invisibility" may explain why the subjects make a lot of mistakes only in question 3.

As is shown in Table 3.5, there was no significant difference between the mean score of mirror-image questions and that of the structural questions for the high-spatial group. This suggests that the high-spatial subjects are not affected by the kind of questions.

Low-spatials

The mean points of completion-subjects in Fig. 3.4(b) show low accuracy of low-spatial subjects. The partiality of alternatives, which is similar to the high-spatial group, was observed in question 3, 15 as shown in Table 3.4. Although such tendency is not pronounced in question 19, it is due to the small number of subjects who completed the question in the low-spatial group. This suggests that the low-spatial subjects also solved the problems by mental rotation.

But as is shown in Fig. 3.4(b), in the low-spatial group, the questions whose mean scores were lower than the others were somewhat different from those in the high-spatial group. As is clearly seen in Table 3.5, the mean score of mirror-image questions was significantly lower than that of the structural questions for this group. It should be noted here that the structural questions can be solved only by detecting the differences in structure (i.e., without mental rotation) since the distractors differ from the criterion in structure. For example, in question 3, "the upper arm is parallel to the lower arm" in criterion and "the upper arm is perpendicular to the lower arm" in alternatives 1 and 3. So, subjects can easily find the distractors by detecting the relation between arms in structural questions. It is, however, difficult to solve mirror-image questions by this strategy, since the distractors have same relation between arms (parallel or not) as the criterion. Assuming that the subjects in the low-spatial group solve the questions not only by mental rotation but also by detecting the differences in structure, the lower mean scores of the mirror-image questions can be explained.

3.2.4 Comparison between males and females in the same score range

In the previous section, the comparison between subjects was made from a point of view of high- and low-spatials. As is shown in Fig. 3.3, a large proportion of the high-spatial group is shared with males and a large proportion of the low-spatial group is shared with females. The sex difference in the MRT score is the result which is common to many previous studies[1][3][9].

Which has larger effect on solving process, the differences between high- and low-spatials or gender? The error analysis was made between males and females of same range in the MRT score. Subjects whose scores are within 19 to 25 were chosen. In Fig.

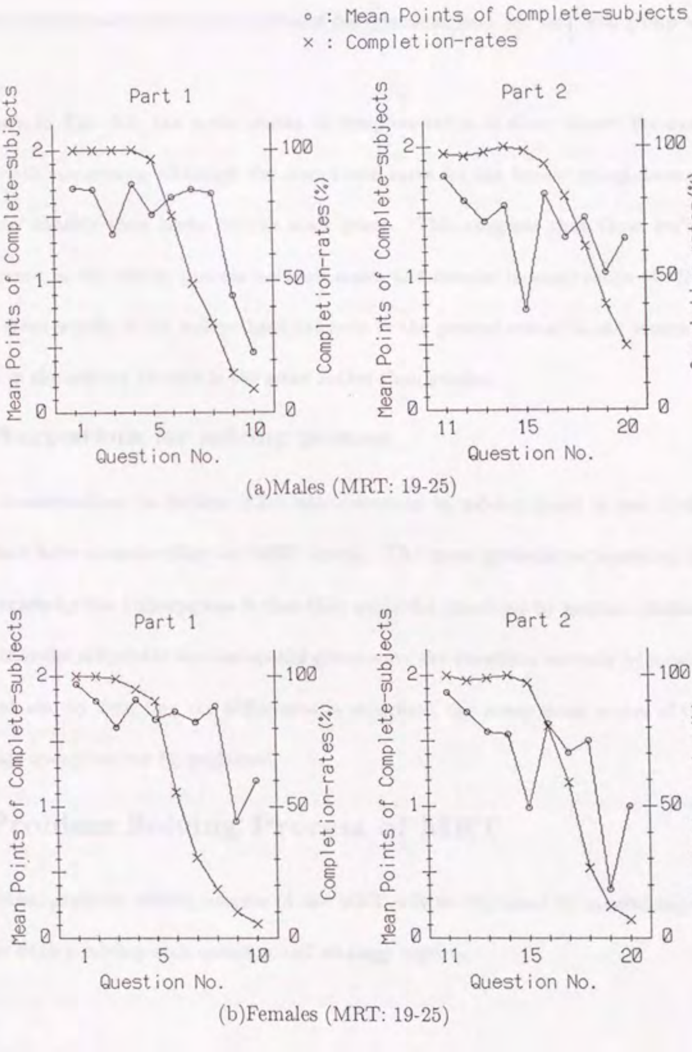


Figure 3.9: Completion-rates and mean points of complete-subjects for the male and female groups in the same score range

3.9, the completion-rates and mean points of complete-subjects for each sex group are shown.

As shown in Fig. 3.9, the mean points of complete-subjects show almost the same shape for both sex groups, although the completion-rates for the female group decrease a little more sharply than those for the male group. This suggests that there isn't a large difference in the solving process between males and females in same range of MRT scores. In other words, as for college-aged subjects in the present research, the source of differences in the solving process is the score rather than gender.

3.2.5 Suggestions for solving process

From the considerations in Section 3.2.2, the difference in solving speed is one of the factors which have a main effect on MRT scores. The most probable explanation for the errors made by the high-spatials is that they solve the questions by mental rotation. Assuming that the subjects in the low-spatial group solve the questions not only by mental rotation but also by detecting the differences in structure, the lower mean scores of the mirror-image questions can be explained.

3.3 Problem Solving Process of MRT

In this section, problem solving process of the MRT will be discussed by monitoring of eye fixation during solving each question and strategy reports.

3.3.1 Experimental method (Experiment I)

Subjects

10 subjects participated in Experiment I. 5 got full scores (40 points) in the preceding MRT. These subjects shall be called "experts". The other 5 students got low scores (less than 13 points). These subjects shall be called "novices". These experts and novices were chosen for the sake of comparison. The experts were all male and the novices were all female.

Questions

12 questions from a total of twenty on the MRT were chosen (Q.3, 4, 5, 7, 13, 14, 15, 16, 17, 18, 19, and 20). The questions were displayed on a vertical board, one question at a time. Each question was drawn on an A1 (42cm × 59cm) size sheet of paper. Fig. 3.11 shows an example of a presented question.

Instrumental system

A cornea reflectance eye tracking system was used to monitor the eye fixations while the subjects solved the problems. The optical path in the system is shown in Fig. 3.10. This system beams an near-infrared light onto the cornea and captures the reflection of the light with a CCD video camera. The reflection of the light, which is called as an "eye spot", is superimposed on the stimulus field viewed by another CCD camera. The eye fixations are monitored on-line and recorded on a videotape. When the eye moves, the eye spot is reflected at a slightly different angle and hence comes to bear at a somewhat different place on the stimulus field. Under proper calibration, the eye spot coincides essentially with the point to which the eyes are oriented. Head movements were minimized by using a chin stand and a forehead bar.

The instrumental system is shown in Fig. 3.12. The system also inputs the video signal to a digitizer, which determines the position of the eye spot relative to the vertical and horizontal synchronization pulses of the video signal. The position of the eye spot in the video frame is output as a pair of rectangular coordinates, which is then recorded in the blanking field of the videotape, namely, once every 33 msecs.

At the beginning of each trial, the subject was required to fixate 4 reference points which corresponded to the center of a criterion figure, the center of alternative 1, the middle between alternative 2 and 3, and the center of alternative 4. Simultaneously, signals (four kinds of sine wave with different frequency) corresponding to the each reference point were recorded in the left sound channel of the videotape.

In order to obtain better dissolute accuracy by maximal use of the visual field, the arrangement of the question was changed from the original MRT. The accuracy of the system analyzed by Makino et al.[30] was within 2° and enough to determine which figure (a criterion or each alternative) the subject fixated on.

When the question was presented, the system began to record the signal (1kHz sine wave) in the left sound channel. The signal continued to be recorded until the subject pushed the stop-button to choose alternatives. Then, the subject chose two alternatives by operating two toggle switches corresponding to each alternative. The toggle switches caused luminous diodes to light, so the answers were recorded in a visual field.

After the experiment, the coordinates data and signals from the videotape were transmitted to a personal computer using a data output unit, and then processed. The eye

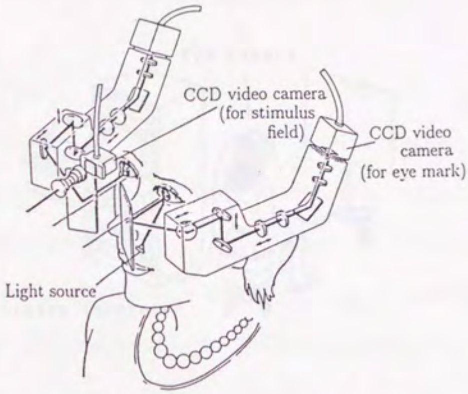


Figure 3.10: Cornea reflectance eye tracking system

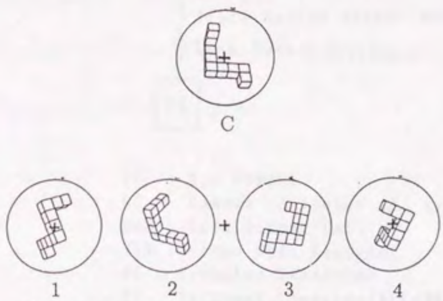


Figure 3.11: Example of MRT presentation

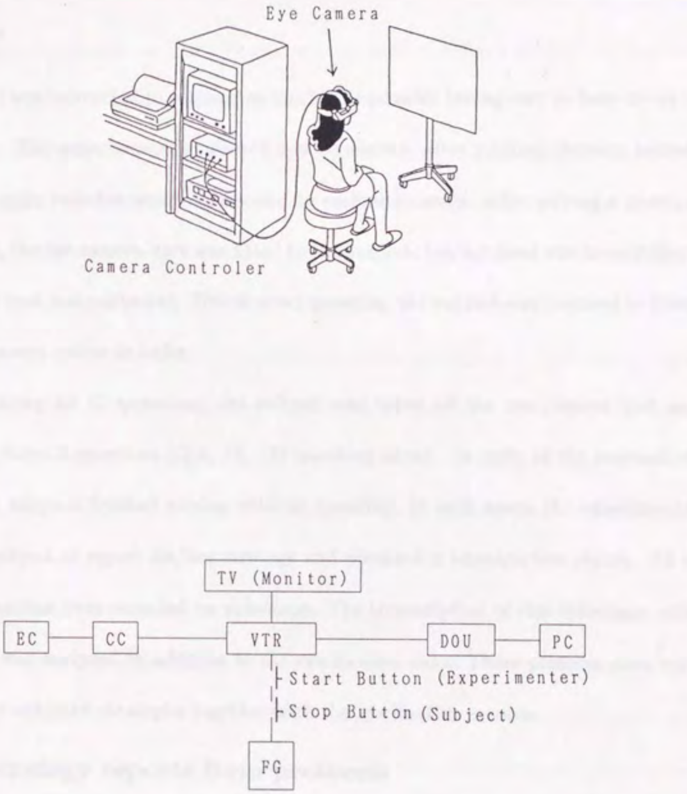


Figure 3.12: Instrumental system (Experiment I)

spot was calibrated with respect to the reference points.

Procedure

The subject was instructed to respond as quickly as possible taking care to keep errors to a minimum. The experimenter explained how to answer: after pushing the stop button, operating toggle switches which correspond to each alternative. After solving a question for practice, the eye camera unit was fitted to the subject, his/her head was immobilized, and the eye spot was calibrated. Before every question, the subject was required to fixate on four reference points in order.

After solving all 12 questions, the subject was taken off the eye camera unit and required to solve 3 questions (Q.3, 14, 15) speaking aloud. In spite of the instruction, most of the subjects finished solving without speaking, In such cases, the experimenter asked the subject to report his/her strategy and obtained a retrospective report. All of the verbalizations were recorded on videotape. The transcription of this videotape –the protocols– was analyzed in addition to the eye fixation data. These protocol data were used to infer subjects' strategies together with the eye fixation records.

3.3.2 Strategy reports from protocols

On the basis of protocols from 29 questions (3 questions × 10 subjects, 1 protocol was lost by mistake), the reports were classified into 3 strategies.

The reports were classified as indicating "mental rotation" if the subject clearly described a rotation of figures. For example, a typical description for *mental rotation* was: "How the criterion will be seen, if I see (the criterion) located in the same angle as each alternative", or "I put two figures (a criterion and an alternative) together from their lower parts and then (I knew) they were different". If the subject without clear description of rotations reported no concrete strategy, the author inferred that the subject had unconsciously solved it by mental rotation.

A description was classified as indicating detecting structural features if the subject reported that he or she had initially detected structural features of the criterion figure and had checked whether each alternative had the same features or not. For example, a typical description for *detecting structural features* was "In the criterion, the upper (arm) and the lower (arm) were opposite (= parallel), in (alternative) 1 and 3, they (the upper and the lower arm) were perpendicular. So I chose (alternative) 2 and 3." The main feature they used was the relation between the upper arm and the lower arm in figures; for example, whether they were parallel to each other or not. This strategy was sometimes used with mental rotation.

A description was classified as indicating matching encoded descriptions if the subject clearly described taking an imaginary walk through some interior corridors. For example, a typical description for *matching encoded descriptions* was "I saw the criterion from the right side (at first). Then I imagined that I went like this (= followed the upper arm), went to the opposite, went down, then turned to the left. In (alternative) 1, I saw it from the left side. I went like this (= followed the upper arm), went to the opposite three blocks, went down, and turned to the left." She encoded the structure of each figure as if she had walked through a passage. After matching the encoded descriptions with each other, the alternatives with the same description as the criterion were chosen.

The strategies for each question of each subject are shown in Table 3.6. The numbers

Table 3.6: Strategies for each question of each subject (Protocol data)

Q. No.	Experts					Novices				
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Q. 3	?	R	(R)	R	?	R	F	F	E	R
Q. 14	R	R	(R)	R	R	(R)	R	F/R	E	(R)
Q. 15		R	(R)	(R)	R	R	R	F/R	E	(R)

R : Mental rotation with clear description
(R) : Mental rotation with no concrete strategy
F : Detecting structural features
E : Matching the encoded decription
/ : Mixture of several strategies
? : Impossoble to classify
Blank : No data

of protocols classified as each strategy are as follows:

- Mental rotation. 20/29 (69.0%)
  - with clear description. 13 (44.9%)
  - with no concrete strategy. 7 (24.1%)
- Detecting structural features (including cases used with mental rotation). 4/29 (13.8%)
- Matching encoded descriptions. 3/29 (10.3%)
- Impossible to classify. 2/29 (6.9%)

3.3.3 Typical styles and strategies

Figs. 3.13–3.21 give sample eye-movement tracks for several subjects on several MRT questions. Time flow is from the top of each figure down. Each duration is indicated by the lengths of the line. The places on the display board (a criterion figure and 4 alternatives) have been laid out horizontally. Correct alternatives are circled. The two numbers written in the rectangle indicate the alternatives chosen by the subject. The eye fixation diagrams could be classified into the following 3 patterns and 4 sub-patterns.

"R"-pattern

A typical eye fixation diagram of this pattern is shown in Fig. 3.13. The subject repeatedly looked back and forth between the criterion figure and the first alternative. Similar sequences of fixations between the criterion and the second, the third and the fourth followed in order. After these systematic sequences, the subject chose answers without repeating. This pattern is classified as "R"-pattern.

Sequences of fixations between the criterion and an alternative were similar to those of Shepard-Metzler tasks analyzed by Just and Carpenter[11]. Their subjects looked back and forth between the two figures as in "R"-pattern. They called such a sequence a "switch", and they reported that the mean number of such switches between figures increased with angular disparity as shown in Fig. 3.14. They identified the repeated fixation of corresponding segments with the transformation and comparison process. Most of the switches occurred during the transformation and comparison process, during which the number of switches increased monotonically with the angular disparity. The switching data from this stage reflects mental rotation.

From the similar sequence between the eye movements of the present investigation and those by Just and Carpenter[11], the diagram of "R"-pattern was interpreted to indicate the subject's solving process by mental rotation. Systematic sequences without repetition suggest confident judgment.

The "R"-pattern was classified into variant sub-patterns.

"RR" A typical eye fixation diagram of this pattern is shown in Fig. 3.15. The fixations similar to "R"-pattern were repeated redundantly. As is shown in Fig. 3.15, it took the subjects longer time than "R"-pattern. Redundant sequences indicate that the subject tried the next alternative before making a complete judgment.

"Rc" A typical eye fixation diagram of this pattern is shown in Fig. 3.16. Short-time fixations between the criterion and alternatives which the subject would choose were

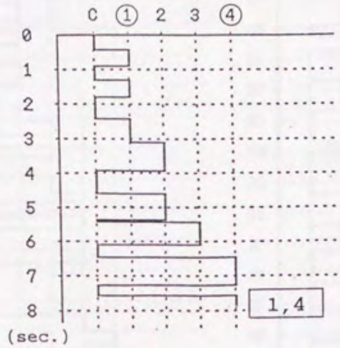


Figure 3.13: Eye fixation diagram: "R"-pattern(S1, Q18)

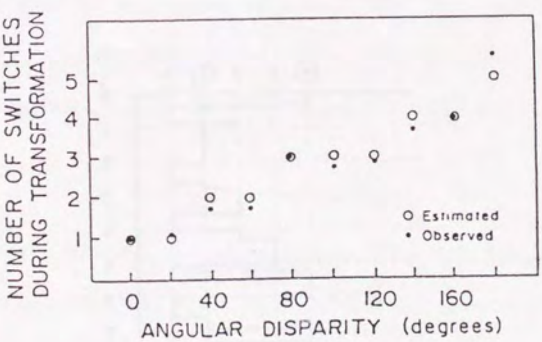


Figure 3.14: Mean number of switches for "Same" trials as a function of angular disparity(from [11])

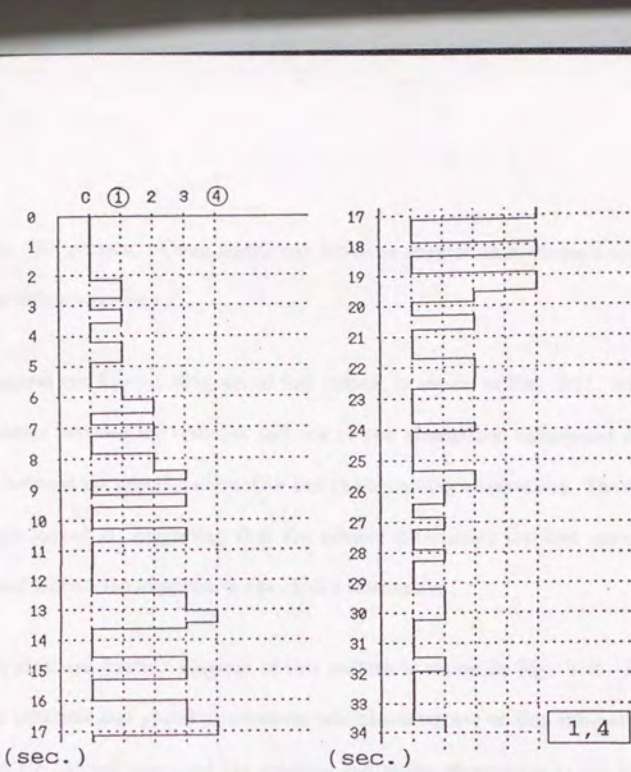


Figure 3.15: Eye fixation diagram: "RR"-pattern(S7, Q18)

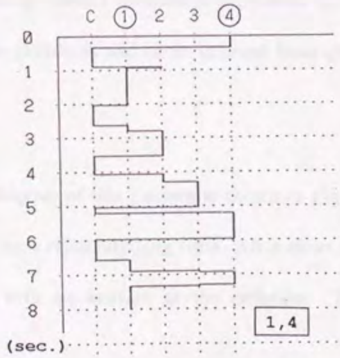


Figure 3.16: Eye fixation diagram: "Rc"-pattern(S4, Q18)

added to the "R"-pattern. These additional fixations suggest that subjects confirmed their choices before responding.

"Rs" A typical eye fixation diagram of this pattern is shown in Fig. 3.17. After systematic fixations between the criterion and one or two alternative, subsequent fixations were shown between the specific alternative and the remaining alternatives. The criterion was no longer looked at, suggesting that the subject determined the first answer with confidence and shifted the criterion to the chosen alternative.

"Rp" A typical eye fixation diagram of this pattern is shown in Fig. 3.18. Fixation between the criterion and plural alternatives was characteristic of this sub-pattern. It suggests that the subject compared the criterion and plural alternatives at one time.

In 19 of 20 protocols which were classified as mental rotation, corresponding eye fixation diagram showed either "R" or its variant subpatterns ("RR", "Rc", "Rs", or "Rp"). The subject's solving process by mental rotation is supported by the agreement between the strategies classified from protocols and those inferred from eye fixation diagram.

"F"-pattern

A typical eye fixation diagram of this pattern is shown in Fig. 3.19. The subject initially fixated on the criterion for a relatively long time. After short fixation on each alternative, the subject responded with no fixation at the criterion. This pattern is classified as "F"-pattern.

A diagram shown in Fig. 3.20 can be regarded as mixture of plural patterns. The

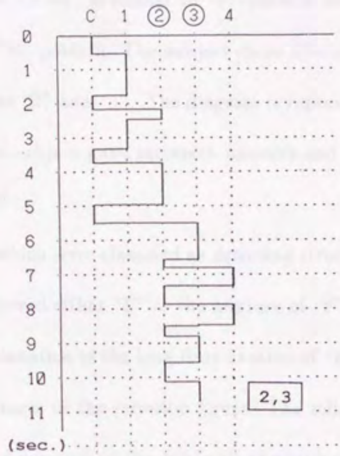


Figure 3.17: Eye fixation diagram: "Rs"-pattern(S2, Q4)

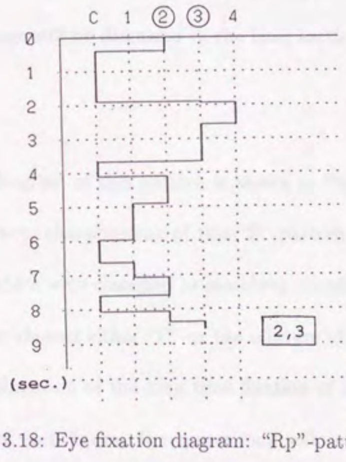


Figure 3.18: Eye fixation diagram: "Rp"-pattern(S5, Q4)

sequence from 0 sec. to 7.2 sec. is similar to "F"-pattern and the sequence from 9.3 sec. to 15.4 sec. resembles "R"-pattern. The subject chose alternatives "2" and '3", although the correct answers were "2" and "4". The diagram is represented as "−F/R". The mark "−" indicates that the subject gave incorrect answers and the mark "/" indicates the mixture of "F" and "R".

In 3 of 4 protocols which were classified as *detecting structural features*, corresponding eye fixation diagram showed either "F" or the mixture of "F" and "R" or its subpatterns.

The most likely explanation of the long time fixation of "F"-pattern is that the subject detected structural features of the criterion figure. The following short fixation on each alternative gives a good account of the subject's checking whether each alternative has the same features or not. From the above consideration, the diagram of "F"-pattern indicated the subject's solving process to have been not by mental rotation but by detecting structural features.

The mixture of the two patterns means that the subject changed strategies halfway. The reason for the change will be discussed in the later section (Section 3.4.4).

"E"-pattern

A typical eye fixation diagram of this pattern is shown in Fig. 3.21. Long solution time and long fixation time were characteristic of this "E"-pattern.

In 2 of 3 protocols which were classified as *matching encoded descriptions*, corresponding eye fixation diagram showed either "E" or the mixture of "E" and "R".

The most likely explanation of the long time fixation of diagram is that the subject encoded the structure of each figure. It requires long solution time to make each figure

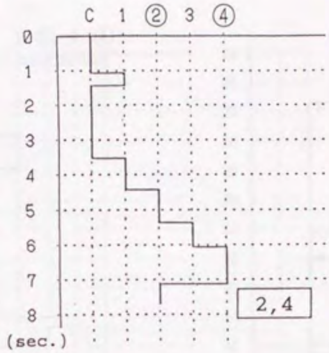


Figure 3.19: Eye fixation diagram: "F"-pattern(S7, Q3)

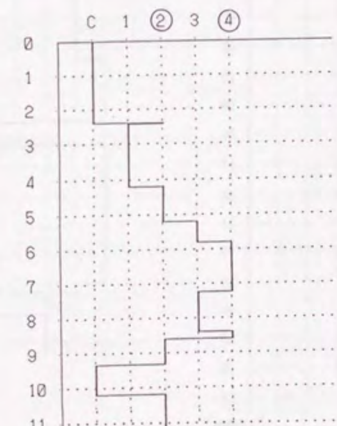


Figure 3.20: Eye fixation diagram: "–F/R"-pattern(S8, Q3)

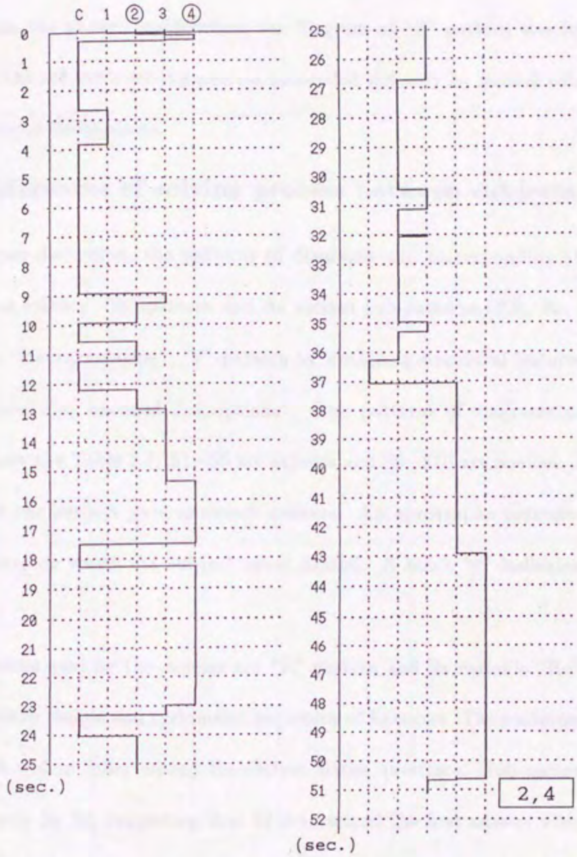


Figure 3.21: Eye fixation diagram: "E"-pattern(S9, Q7)

encoded. From the above consideration, the diagram of "E"-pattern was interpreted as showing that the subject's solving process proceeded not only by mental rotation but by matching encoded descriptions.

3.3.4 Differences of solving process between subjects

From the above discussion, the patterns of diagrams and corresponding strategies are summarized as follows. "R"-pattern and its variant sub-patterns (RR, Rc, Rs and Rp) correspond to "mental rotation", "F"-pattern to "detecting structural features", and "E"-pattern to "matching encoded descriptions". The patterns of diagrams used by each subject are shown in Table 3.7. S1−S5 are experts and S6−S10 are novices. A mark "−" indicates that the subject gave incorrect answers. An apostrophe indicates that there is an alternative on which the subject never fixated. A mark "/" indicates mixture of patterns.

Most patterns used by the experts are "R"-pattern and its variants "Rc", "Rs", and "Rp", which show simple and systematic sequences of fixations. The patterns used by the experts didn't widely differ among themselves across questions. Sub-pattern "Rs" was used almost only by S2, suggesting that S2 determined the first answer with confidence. Sub-pattern "Rp" was used only by S5, which suggests that S5 compared the criterion and plural alternatives at one time. From these results, it is considered that experts solved the MRT by simple and systematic mental rotation.

The patterns used by novices differed across the subjects and often among themselves across questions. S6 used variants of "R" and produced many incorrect answers. S10 solved most questions by "RR", showing redundant mental rotation. They solved the

Table 3.7: Patterns of experts and novices(Experiment I)

	Experts					Novices				
Sub.	S1 (M)	S2 (M)	S3 (M)	S4 (M)	S5 (M)	S6 (F)	S7 (F)	S8 (F)	S9 (F)	S10 (F)
MRT	40	40	40	40	40	12	12	11	12	6
Q. 3	R	Rs	R		Rp	R	F	-F/R	E	RR
Q. 4	R	Rs	R'		Rp	Rc	F	F/R	E	RR
Q. 5	Rc	*	RR	RR	Rp/R	-Rc	-RR	-F	R'	R
Q. 7	R	Rs	R	R	Rp/R	Rc	F	F/RR	E	RR
Q.13	RR	Rs	R	-R	Rp/Rs	-Rc	R	RR	-E/F	RR
Q.14	R	Rs	Rp/R	R	RR	-RR	RR	R	E/R	R
Q.15	Rc	Rs/RR	-R'	R	Rp/R	-R'	-R	F/RR	E	RR
Q.16	RR	Rs'	R'	RR	*	-RR	Rc	F/RR	E/F/R	R
Q.17	R	Rs'	R'	R	Rp/R	-R'	RR	R	-E	-RR
Q.18	R	Rc	R	Rc	Rp/R	-R	RR	F/RR	E	RR
Q.19	R	Rs	-R	RR	Rp/R	-F	-RR	F/RR	R	-RR
Q.20	RR	Rs'	R'	R	*	-F/R	RR	-F/RR	E	RR

Sub. (M) : Male, (F) : Female.

"/" : Mixture of several patterns, "*" : Impossible to classify, Blank : No data

MRT by mental rotation like the experts did. But mental rotation by novices was redundant and often produced wrong answers. Sequential redundancy requires longer time to solve than simple mental rotation.

S7 and S8 mainly used "F" and "RR". S8 often changed sequences of fixation halfway. When her fixation showed the "F"-pattern sequence, she was paying her attention to the relation between the arms of the figure. Structural questions could be solved easily by "F", but mirror-image questions, in which all figures had the same relation between arms, could not be solved by "F". So a change of strategy was required to get the correct answer, otherwise a wrong answer would have been produced.

S9 mainly used "E", and sometimes changed the pattern of fixation halfway. "E" required such a long time that the subjects could not solve many questions in limited time. Besides, confusion over right and left produced wrong answers in mirror-image questions.

"F"-pattern and "E"-pattern used by novices are consistent with the fact that the mean score of the mirror-image questions was lower than that of the structural questions in low score group.

Mental rotation was the main strategy used by experts and novices. Experts used simple and systematic mental rotation, while novices showed redundant sequences of fixation and sometimes produced errors. The strategies of novices differed across the subjects and often within themselves across questions. Strategies other than mental rotation, "detecting features" and "matching encoded descriptions", were indicated from eye movement and verbal protocols.

3.4 Problem Solving Process of Shepard-Metzler task

From the results of Experiment I, the solving process of the MRT was classified into the following strategies: mental rotation, detecting structural features, and matching encoded descriptions. Mental rotation was the main strategy for both experts and novices. The differences of strategies between subjects were pronounced among novices.

An MRT question consisted of four Shepard-Metzler tasks. In order to explore the source of the differences in strategies of the MRT, the performance in Shepard-Metzler tasks is compared between experts and novices in this section. In the following experiment, the solving process of Shepard-Metzler tasks will be analyzed.

3.4.1 Experimental method (Experiment II)

Subjects

12 subjects not in Experiment I participated in Experiment II(a) and (b). 4 subjects scored 37 or 38 points (near to the full score, 40) in the preceding MRT. These subjects shall be called "experts". The other 8 subjects had low scores (less than 12 points). These subjects shall be called "novices". These experts and novices were chosen for the sake of comparison.

Tasks

The main difference between the Shepard-Metzler task and the MRT is the number of alternatives. The former requires judging whether two figures are the same or not, while the latter requires choosing from alternatives.

4 Shepard-Metzler figures were chosen from those Vandenberg and Kuse used[29]. By

using the 4 kinds of objects (Ap, An, Bp, and Bn), the "Same" trials and "Different" trials were made. In the "Same" trials, the amount of angular disparity between the two figures varied from 0 to 160 degrees in 40-degree increments. The "Different" trials contained two kinds of pairs as follows : mirror-image pairs (Ap-An, Bp-Bn) and structurally different pairs (Ap-Bp, An-Bn). Fig. 3.22 shows examples of presented tasks. Fig. 3.22(a) is an example of "Same" tasks. Figs. 3.22(b) and (c) are examples of "Different" tasks, which is made of mirror-image pair, and structurally different pair, respectively.

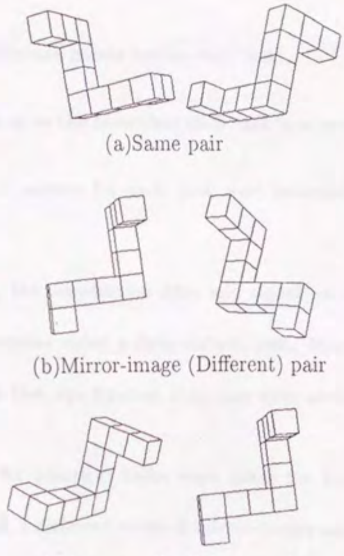
Experiment II(a)

Instrumental system The instrumental system is shown in Fig. 3.22. A cornea reflectance eye tracking system, which was almost identical to that used in Experiment I, was used to monitor the eye fixations while the subjects solved the problems. Under proper calibration, the eye spot coincided essentially with the point to which the eyes were oriented. The head movements were minimized with a chin stand.

Each task was displayed on a 29-inch display controlled by a personal computer. The each figure in the task was presented within a radius of about 10 cm. The viewing distance was approximately 90 cm, so that each figure subtended within 10° of visual angle. The system could determine which part of each figure (the upper arm, the central joint or the lower arm) the subject fixated on.

Signals which marked the following three kinds of frames were recorded in the left sound channel of the videotape.

- 2 frames corresponding to the time that the subject fixated two reference points: the center of the left figure and the center of the right figure. The subject was required



(a)Same pair

(b)Mirror-image (Different) pair

(c)Structurally different pair

Figure 3.22: Sample of Shepard-Metzler tasks presented on 29-inch display

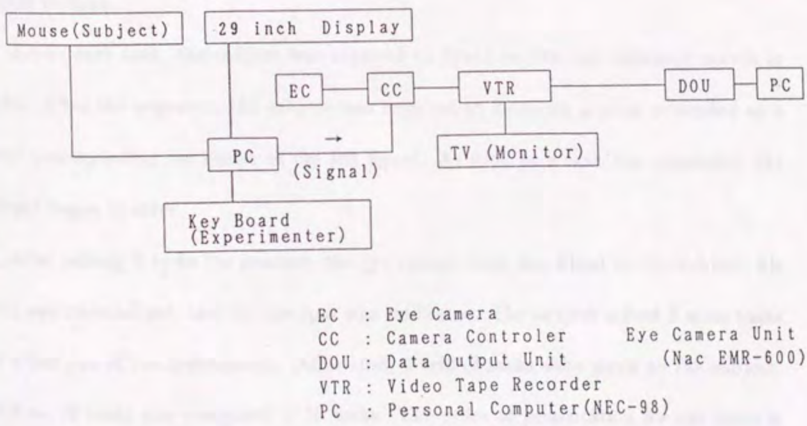


Figure 3.23: Instrumental system(Experiment II(a))

to fixate the two reference points before each task.

- 1 frame corresponding to the time that each task was presented on the display.

The solving time and the answer for each task were recorded in files by the personal computer.

After the experiment, the coordinates data and signal on the videotape were transmitted to a personal computer using a data output unit. Based on the signals and the time data recorded in the files, eye fixation diagrams were made.

Procedure 20 (and 8 for practice) tasks were made for Experiment II(a). 20 tasks consisted of 10 same trials, 7 different trials of mirror-image pair (Ap-An or Bp-Bn), and 3 different trials of structurally different pair (Ap-Bp, or An-Bn).

The subject was instructed to respond as quickly as possible taking care to keep errors to a minimum. The subject responded "Same" or "Different" by clicking one of the two mouse buttons.

Before each task, the subject was required to fixate on the two reference points in order. After the sequence, the subject was required to fixate on a cross presented at a point corresponding the center of the left figure. As soon as a task was presented, the subject began to solve.

After solving 5 tasks for practice, the eye camera unit was fitted to the subject, his head was immobilized, and the eye spot was calibrated. The subject solved 3 more tasks for a test run of the instruments. After that, 2 sets of tasks were given to the subject. Each set of tasks was composed of 10 tasks. The order of presentation for the tasks is shown in Table 3.8.

Table 3.8: Order of presentation(Experiment II(a))

	Tasks for practice (Before the eye camera)	
No.	Left	Right
1	An6	Ap12
2	Ap4	Ap16
3	Bp12	Bp10
4	Bn6	Bp4
5	Ap10	Ap2
	Tasks for practice (With the eye camera)	
No.	Left	Right
1	Bn10	Bn14
2	Bp6	Bp12
3	Ap6	Ap16
	Trial-1(Block 1)	
No.	Left	Right
1	Bn10	Bn6
2	Bp16	Bp4
3	Ap2	Bp4
4	An10	Bn16
5	Bp4	Bn12
6	Ap2	An12
7	Ap6	Ap16
8	Bn8	Bn4
9	Bp16	Bn10
10	Bp12	Bn4
	Trial-2(Block 2)	
No.	Left	Right
1	An12	An2
2	Ap2	Ap2
3	Bn16	Bp8
4	An6	Ap16
5	An12	An2
6	An6	An4
7	Bn10	Bn4
8	An4	An2
9	Bn8	Bp12
10	Ap10	Bp6

After solving all 20 tasks, the subject was taken off the eye camera unit and required to solve 4 tasks speaking aloud. In spite of the instruction, most of the subjects finished solving without speaking. In such cases, the experimenter asked the subject to report his strategy and obtained a retrospective report. All of the verbalizations were recorded on a videotape. The transcription of this videotape –the protocols– was analyzed in addition to the eye fixation data. These protocol data were used to infer subjects' strategies together with the eye fixations records.

Then, the subject participated in Experiment II(b) as follows.

Experiment II(b)

Instrumental system The eye camera and its controller were removed from the system of Experiment II(a). As same as Experiment II(a), each task was displayed on a 29-inch display controlled by a personal computer. The solving time and the answer for each task were recorded in files by the personal computer.

Procedure After Experiment II(a), the subject without the eye camera participated in Experiment II(b). The subject was instructed to solve the tasks "by mental rotation" and to respond with a mouse button. Namely, the subject was compelled to solve by mental rotation, whereas no strategy was compelled in Experiment II(a). In Experiment II(b), "Different" trials consisted of only mirror-image pairs to implicitly eliminate the strategy by detecting the relationship between arms.

The 160 Shepard-Metzler tasks were presented in a random order and both the response times and the response of the subject were recorded. The response times for all correct answers to "Same" -tasks (about half of 160 Shepard-Metzler tasks) were analyzed. The response time function was used as an index of the performance in mental rotation.

3.4.2 Typical styles and strategies

Eye fixation diagrams were made from the data of Experiment II(a) and used to infer strategies together with the protocols. Figs. 3.24–3.28 give sample eye-movement tracks for several subjects on several tasks, an eye fixation diagram and corresponding Shepard-Metzler figures. In the eye fixation diagram, the time flow is from the top of each figure down. The answer ("Same" or "Different") was indicated with the time required to solve the task. Each duration is indicated by the lengths of the line. The places on the display (a left figure and a right figure) have been laid out horizontally. The numbers in the diagram correspond to those of fixation points in the figures. When the eye spot was located within 1° in a sequence of successive video frames for at least 66.6 msecs, the frames were aggregated into a single fixation point. Circles in the figures show the eye fixation points. Each duration is indicated by the radius of the circle. The numbers in the figures indicates the sequence of the fixation points.

The eye fixation diagrams may be classified into the following patterns.

"r"-pattern and "rn"-pattern

A typical eye fixation diagram of "r"-pattern is shown in Fig. 3.24. The subject repeatedly looked back and forth between the corresponding segment of each figure. The locus of the fixations moved from the upper arms to the lower arms. Such sequences resemble those of the study by Just and Carpenter[11]. In their experiment, the mental rotation process was defined as a repeated fixation between corresponding segments.

After these systematic fixations, the subject responded without repeating. The one-way (from the upper arms to the lower arms) scan path without repeat indicates a confident judgment.

From the above consideration, the "r"-pattern is considered to reflect the solving process by mental rotation.

In Fig. 3.25, the subject also looked back and forth between the corresponding segments of each figure. The locus of the fixations moved from the upper arms to the lower arms (0 – 4.8 sec.). Then a similar sequence of fixations is repeated once again (4.8 – 8.3sec). The sequence similar to the "r"-pattern is repeated twice. The diagram with multiple cycles of "r"-pattern is classified as an "rn"-pattern. The number of repeated sequences similar to the "r"-pattern is indicated by "n" of "rn". For example, the diagram shown in Fig. 3.25 is classified as an "r2"-pattern.

An incomplete "r"-sequence, such as fixations between only upper or lower corresponding segments, is counted as an 0.5 cycle. For example, in Fig. 3.26, the fixations between the lower arms were added to the "r"-pattern. So, the diagram shown in Fig. 3.26 is classified as an "r1.5"-pattern.

These "r" and "rn"-patterns in the Shepard-Metzler task can be considered to correspond to "R" and "RR"-patterns in the MRT, respectively, namely, the strategies by mental rotation. Redundant sequences in an "rn"-pattern indicate that these subjects couldn't make a complete judgment by a one-way scan path.

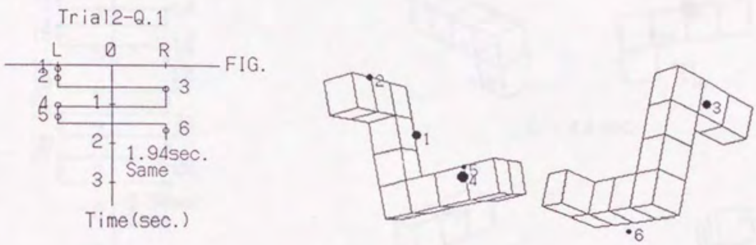


Figure 3.24: Eye fixation diagram: "r"-pattern(S13, Trial 2-1)

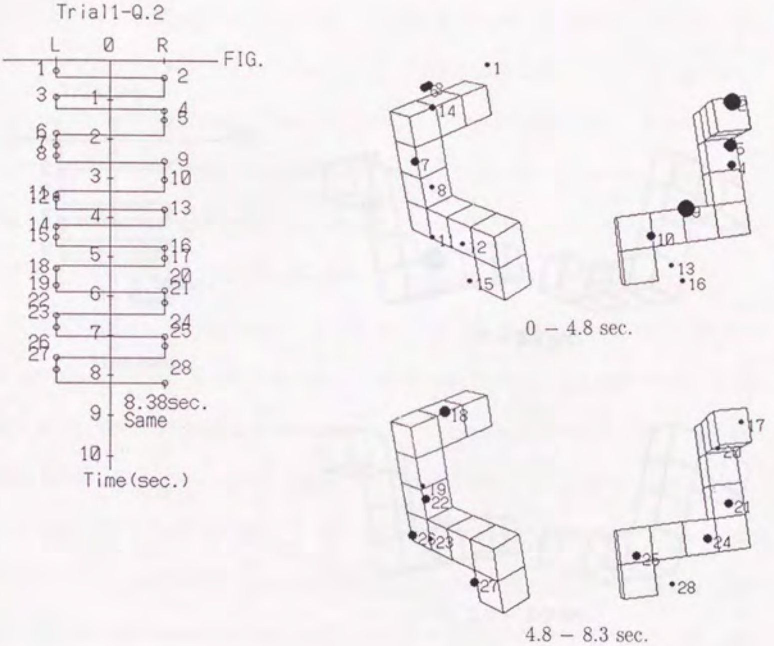


Figure 3.25: Eye fixation diagram: "r2"-pattern(S18, Trial 1-2)

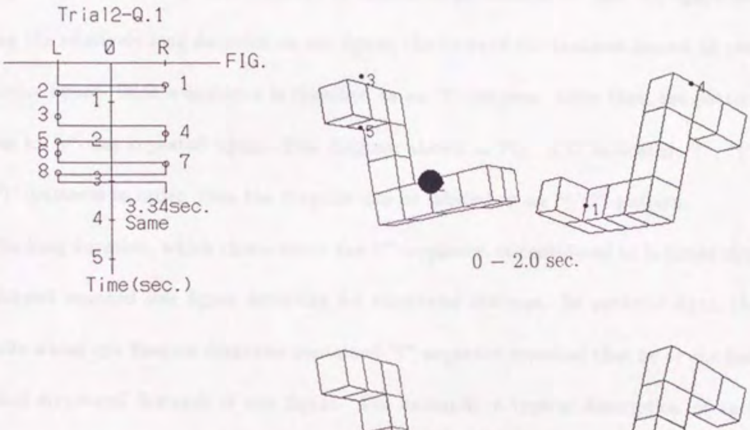


Figure 3.26: Eye fixation diagram: "r1.5"-pattern(S12, Trial 2-1)

"f/rn"-pattern

An example of eye fixation diagram of this pattern is shown in Fig. 3.27. The subject looked back and forth between the corresponding segments of each figure (0 − 3.1 sec.). Then the consecutive fixations on one figure began at 3.1sec.. It continued till 5.6 sec.. Such partial fixations were not observed in the above-mentioned "r" and "rn"-patterns. During the relatively long duration on one figure, the locus of the fixations moved all over the entire figure. Such a sequence is classified as an "f"-pattern. After that, the pattern similar to "r" was repeated again. The diagram shown in Fig. 3.27 indicated "r", "f", and "r"-patterns in order, then the diagram can be labeled as an "f/r2"-pattern.

The long duration, which characterize the "f"-sequence, is considered to indicate that the subject scanned one figure detecting for structural features. In protocol data, the subjects whose eye fixation diagrams contained "f"-sequence reported that he or she had detected structural features of one figure. For example, a typical description of such subjects was "If the upper arm and the lower arm are parallel or not. When I can't judge only by rotation, (I see) how the arms are jointed (the central part). I don't know which I think first, the rotation or the angle of the arms" and "If I rotate the right figure, then, the lower arm faces this side. So I thought it (the right figure) was different (from the left figure)". The former description means the subject used the relation between the upper arm and the lower arm in figures to solve the task. The latter means that the subject put the central joints of each figure together and detected the difference of the direction of the lower arm of each figure. This strategy was sometimes used with mental rotation.

Two kinds of interpretation are possible for the following fixations between the corre-

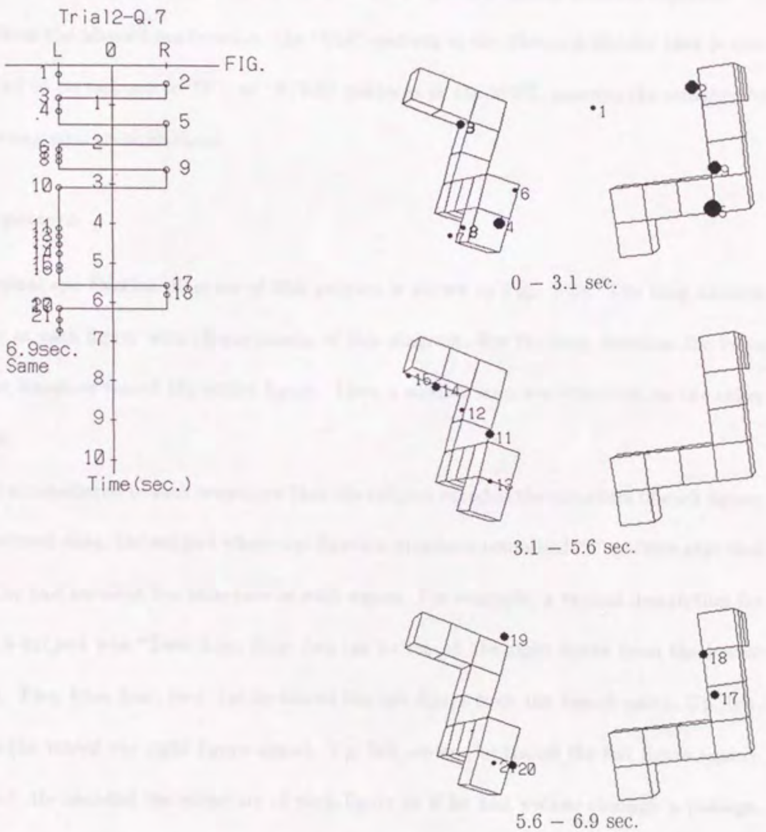


Figure 3.27: Eye fixation diagram: "f/r2"-pattern(S20, Trial 2-7)

sponding segments of each figure. First, the subject checked whether each figure had the detected features or not and answered. Second, the subject couldn't make a judgement by using the detected feature and solved the task by using mental rotation together.

From the above consideration, the "f/rn"-pattern in the Shepard-Metzler task is considered to correspond to "F"- or "F/RR"-patterns in the MRT, namely, the strategy by detecting structural features.

"e"-pattern

A typical eye fixation diagram of this pattern is shown in Fig. 3.28. The long fixation times at each figure were characteristic of this diagram. For the long duration the locus of the fixations traced the entire figure. Then a similar scan was executed on the other figure.

It is considered in such sequences that the subject encoded the structure of each figure. In protocol data, the subject whose eye fixation diagrams contained "e"-pattern reported that he had encoded the structure of each figure. For example, a typical description for such a subject was "Two, four, four, two (as he traced the right figure from the lowest part). Two, four, four, two. (as he traced the left figure from the lowest part). Up, left, down (he traced the right figure again). Up, left, down (he traced the left figure again). Same." He encoded the structure of each figure as if he had walked through a passage. After encoding both of two figures, he answered "Same" for the reason that both figures had the same description of structure.

From the above consideration, the "e"-pattern in the Shepard-Metzler task corresponds to "E"-pattern in the MRT, namely, the strategy by matching encoded descrip-

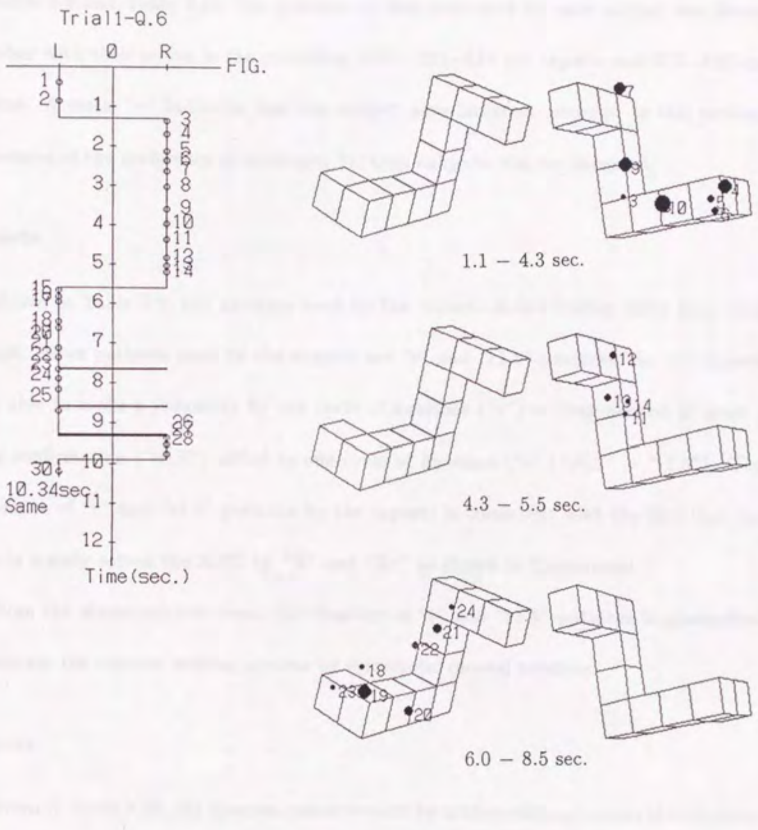


Figure 3.28: Eye fixation diagram: "e"-pattern (S19, Trial 1-6)

tions.

3.4.3 Differences of solving process between subjects

In Table 3.9 and Table 3.10, the patterns of diagrams used by each subject are shown together with their scores in the preceding MRT. S11−S14 are experts and S15−S22 are novices. A mark "−" indicates that the subject gave incorrect answers. In this section, differences of the preference of strategies between subjects will be discussed.

Experts

As shown in Table 3.9, the patterns used by the experts didn't widely differ from task to task. Most patterns used by the experts are "r" and "r1.5"-patterns. So the experts were able to make a judgment by one cycle of fixations ("r") or they needed at most a short confirmation ("r0.5") added to one cycle of fixations ("r"+"r0.5" = "r1.5"). The preference of "r" and "r1.5"-patterns by the experts is consistent with the fact that the experts mainly solved the MRT by "R" and "Rc" as shown in Experiment I.

From the above consideration, the diagram of "r" and "r1.5"-patterns is interpreted to indicate the experts' solving process by systematic mental rotation.

Novices

As shown in Table 3.10, the diagram patterns used by novices differed across the subjects. The novices are grouped as follows:

1. The subjects who mainly used "r" and "rn". (S15, S16)

2. The subjects who used "rn" and "f/rn". (S17, S18, S20, S21, S22)

Table 3.9: Patterns of experts (Experiment II(a))

	Sub.	S11	S12	S13	S14	Attributes of each task	
	Sex	M	M	M	M		
	MRT	38	37	37	37	Figures	S./D.(deg.)
Presentation No.	2- 2	r1.5	r	r	r	Ap2-Ap2	Same(0)
	2- 6	r1.5	r	r1.5	r	An6-An4	Same(40)
	2- 8	r	r	r	r	An4-An2	Same(40)
	1- 1	r1.5	r	r1.5	r	Bn10-Bn6	Same(80)
	1- 8	r	r2	r1.5	r	Bn8-Bn4	Same(80)
	1- 2	r1.5	r1.5	r	r1.5	Bp16-Bp4	Same(120)
	2- 7	r3	r1.5	r2	r1.5	Bn10-Bn4	Same(120)
	1- 7		r	r	r1.5	Ap6-Ap16	Same(160)
	2- 1	r3	r1.5	r	r1.5	An12-An2	Same(160)
	2- 5	r2	r1.5	r	r	An12-An2	Same(160)
	1- 5	r	r	r	r1.5	Bp4-Bn12	Mir.(40)
	1-10	r	r1.5	r1.5	r	Bp12-Bn4	Mir.(40)
	2- 9	r1.5	r1.5	r	r	Bn8-Bp12	Mir.(40)
	1- 6	r1.5	r1.5	r1.5	r1.5/e	Ap2-An12	Mir.(80)
	2- 4	r1.5	r1.5	r	r1.5	An6-Ap16	Mir.(80)
	2- 3	r1.5	r	-r2/e	r	Bn16-Bp8	Mir.(120)
	1- 9		r	r	r	Bp16-Bn10	Mir.(160)
	1- 3	r	r	r	r?f?	Ap2-Bp4	Str.(40)
	2-10	r	r	r	r	Ap10-Bp6	Str.(80)
	1- 4	r	-r2	r	r1.5/f?	An10-Bn16	Str.(120)

"/" : Mixture of several patterns, "?" : Vague pattern
S./D.(deg.) : Same or different(Angular disparity)
* : Mir. : Figures are mirror-images each other.(different)
Str. : Figures are different in structure.
** : Angular disparity in "Different" task is evaluated by that between upper arms, although the figures cannot be brought into congruence.

Table 3.10: Patterns of novices (Experiment II(a))

	Sub.	S15	S16	S17	S18	S19	S20	S21	S22	Attributes of each task	
	Sex	F	F	F	F	M	F	F	M		
	MRT	10	6	9	9	10	9	11	5	Figures	S./D.*(deg.)**
Presentation No.	2- 2	r	r	r1.5/f	r2.5	r	r	r	r	Ap2-Ap2	Same(0)
	2- 6	r	rn	e/r1.5	e/r2	e/r	r2	r	f/r3.5	An6-An4	Same(40)
	2- 8	r	r2.5	r	r2/f	r/f	r2/f	r1.5/f	r1.5/f	An4-An2	Same(40)
	1- 1	r2	r2	??/r1.5	r3.5	e	-r	f/r	r2	Bn10-Bn6	Same(80)
	1- 8	r2/f	??/r1.5	r2.5/f	f/r4	e	-r1.5	r3 ?	f/r4	Bn8-Bn4	Same(80)
	1- 2	r1.5	r2.5	f/r2	r2	r/f	-r2	f/r ?	r2/e	Bp16-Bp4	Same(120)
	2- 7	r	r	r2.5/f	f/r2.5	e	r2/f	r1.5	f/r1.5	Bn10-Bn4	Same(120)
	1- 7	r1.5	-f/r2	r?/f	f/r4	e	r/f	r1.5/f	r2/e	Ap6-Ap16	Same(160)
	2- 1	r	r2	r2.5/f	f/r2	e	r2.5/f	r	-r2/e	An12-An2	Same(160)
	2- 5	r	r1.5	r2.5/?	e/r4	e	r3.5/f	-r3/f	e/r2.5	An12-An2	Same(160)
	1- 5	r	r2.5	r1.5	r2/e	e/r	r3	r/e	r3.5/f	Bp4-Bn12	Mir.(40)
	1-10	r2	-r2.5	r/f	r2	e/r	r2.5/f	e? f	r4.5/f	Bp12-Bn4	Mir.(40)
	2- 9	r1.5	rr/?	r3	fe/r2.5	r/f	r2	r/f	f/r3	Bn8-Bp12	Mir.(40)
	1- 6	r1.5/f	-r3.5/f	r2/e	r6	e	r5	-r	f/r2.5	Ap2-An12	Mir.(80)
	2- 4	r	r1.5 ·	r2.5	r3.5/f	r1.5	r3	r	r4/f	An6-Ap16	Mir.(80)
	2- 3	r1.5	rr?f?	r2.5/f	??/r2	e/r/f	r1.5	-r2	r2.5/f	Bn16-Bp8	Mir.(120)
	1- 9	-r	r2.5	r2/f	r3.5	e/f	r2.5/f	r	e/r3	Bp16-Bn10	Mir.(160)
	1- 3	r	r	r2 ?	r3	e	r2/f	e	e?f?	Ap2-Bp4	Str.(40)
	2-10	r	rr	r4/f	(e)/r2	e ?	r2/f?	r1.5/f	-r2	Ap10-Bp6	Str.(80)
	1- 4	r	??/r?	-r3/f	e/r4.5	f	r2	r1.5/f	e/r2	An10-Bn16	Str.(120)

"/" : Mixture of several patterns, "?" : Vague pattern

S./D.(deg.) : Same or different(Angular disparity)

* : Mir. : Figures are mirror-images each other.(different)

Str. : Figures are different in structure.

** : Angular disparity in "Different" task is evaluated by that between upper arms, although the figures cannot be brought into congruence.

3. The subject who mainly used "e". (S19)

Group 1 (S15, S16) As shown in Table 3.10, S15 solved most of tasks by "r" and "rn"-patterns. S16 mainly used "rn"-pattern. As for S16, the degrees of redundancy indicated by "n" of "rn" are large especially in mirror-image tasks. The redundant sequences shown in "rn" are consistent with the fact that some novices solved the MRT by "RR".

Group 2 (S17, S18, S20, S21, S22) As shown in Table 3.10, S17, S18, and S22 often solved the tasks by "f/rn". S20 and S21 solved the tasks by using different patterns ("f/rn", "rn", and "e") for each task. The preference of "f/rn" by S17, S18, and S22 is consistent with the fact that some novices solved the MRT by "F" and "F/RR".

Group 3 (S19) Among 8 novices, only S19 consistently solved by "e"-pattern. The preference of "e" by S19 (novices) is consistent with the fact that some novices solved the MRT by "E".

As discussed above, the subjects were classified into 4 groups as follows : (1) Experts (simple and systematic mental rotation), (2) Novices: Group 1 (redundant mental rotation), (3) Novices: Group 2 (both feature detecting and redundant mental rotation), and (4) Novice: Group 3 (matching encoded descriptions).

The classification is very similar to that of the subjects solving the questions of the MRT. Namely, the differences in strategies of the MRT remain in Shepard-Metzler tasks. This indicates that the differences in strategies of the MRT were not evoked by the answer style (choosing from alternatives), but by individual differences in performance in solving Shepard-Metzler tasks.

3.4.4 Strategies and performance in mental rotation

In this section, the performance in mental rotation for each subject was estimated based on data of Experiment II(b) and its relation to the preference of strategies will be discussed.

Estimate of the performance in mental rotation

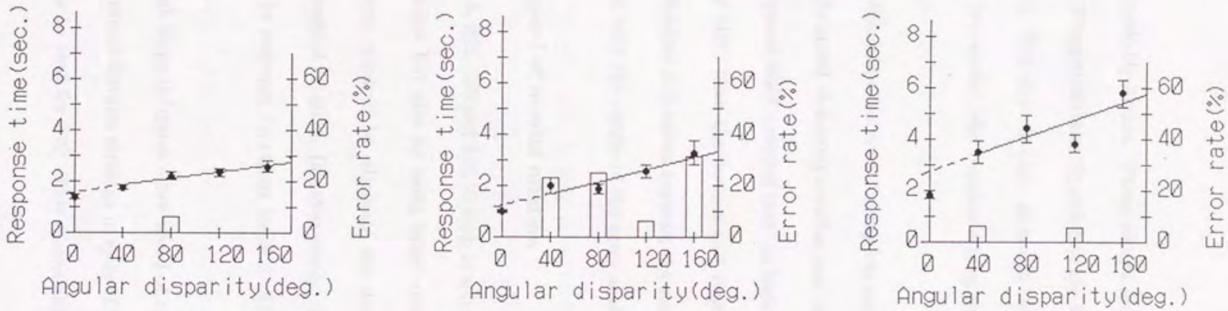
The response times for all correct answers to the "Same" trial were analyzed from data of Experiment II(b). In Figs. 3.29–3.31, the response times for each subject were plotted as a function of angular disparity. Blank bars indicate error rates, which are the rates of incorrect answers for the "Same" trials for each angular disparity. The speed of mental rotation for each subject was evaluated by the slope of the regression line between mean response times and angular disparity. The regressions were computed except for the data of 0 degrees, for the reason that the response times of 0 degrees, in which the two figures are completely same, reflect little effect of mental rotation.

Difference in the speed of mental rotation

Experiment II(a) indicated that all of the 4 experts (S11, S12, S13, S14) and 2 novices (S15, S16: Group 1) solved the tasks mainly by mental rotation. In Fig. 3.29, response functions are plotted for these subjects.

Fig. 3.29(a) is the data for a typical expert, S12. Response functions for the other experts are similar to that for S12. The slopes for the experts are small. The error rates for the experts are generally low. These results indicate that the experts are able to rotate the Shepard-Metzler figures quickly and accurately.

Fig. 3.29(b) is the data for S15(novice). Although the slope is almost equal to those



(a)S12 (b)S15 (c)S16

Figure 3.29: Response time functions (S12, S15, S16)

for the experts, the error rates are remarkably higher. These results indicate that there is a novice who is able to rotate the Shepard-Metzler figures quickly but inaccurately. Fig. 3.29(c) is the data for S16(novice). The slope is high, although the error rates are low. These results indicate that there is a novice who requires a long time to rotate the Shepard-Metzler figures with accuracy.

From the above consideration, the differences between the experts and novices, whose main strategy is mental rotation, are the speed of mental rotation and its accuracy.

The error analysis of the paper-and-pencil MRT revealed that the high-spatial subjects required a shorter time for solving the MRT than the novices and that the low-spatial subjects often tried but failed. The individual differences between the same strategies for the Shepard-Metzler task are consistent with the results of the error analysis.

Unification of strategies and the speed of mental rotation

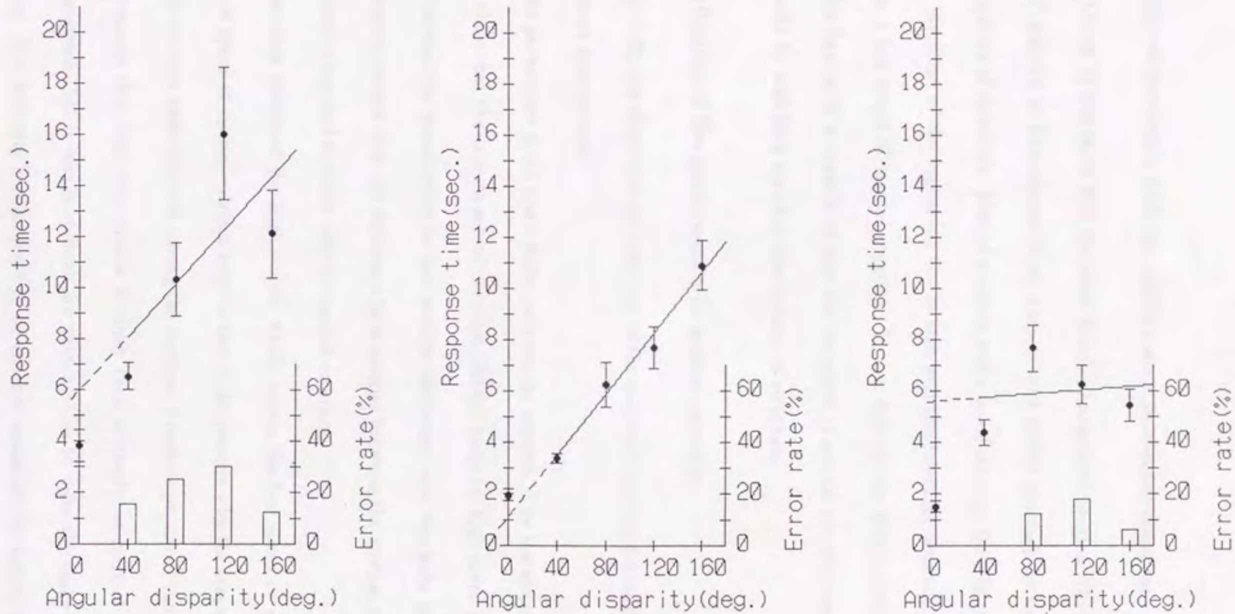
Experiment II(a) indicated that S17, S18, S22, S20 and S21 (Group 2) solved the Shepard-Metzler tasks not only by mental rotation but also by using other strategies, such as detecting structural features. They were required to eliminate the detecting features strategy and to solve the task only by mental rotation. In other words, they were forced to unify the strategy. In Fig. 3.30, the response functions for S22, S17, and S20 are shown.

As for S22 (see Fig. 3.30(a)), the high slope indicates a low speed of rotation and the high intercept reflects the detecting structural features strategy in spite of the instruction. The response function for S18 is similar to that for S22. This means that S18 and S22 was unable to unify the strategies.

The high slope and low intercept for S17 (see Fig. 3.30(b)) show that S17 was able to solve the Shepard-Metzler task without using the detecting feature, however, the speed of rotation was extremely low.

As for S20 (see Fig. 3.30(c)), the regression line is almost flat and the intercept is high. S20's error rate for "Same" tasks was 8.7 %, whereas that for "Different" tasks was 24.2 %. As for the novices other than S20 and S21, the mean error rates for "Same" and "Different" tasks were 10.1 % and 4.9 %, respectively. The much higher error rate for "Different" tasks means that S20 often misjudged "Different" (mirror-image) pair as "Same". It should be recalled here that all of "Different" tasks used in Experiment II(b) is mirror-images tasks, in which both figures have same relation between the upper arm and the lower arm. The strategy by detecting the relation between the arms misleads the subject and produces high errors in "Different" tasks. It is noted that S20 and S21 didn't solve only by feature detecting but also by mental rotation as shown in their eye fixation diagrams. It is considered that the flat slopes reflect the confusion of strategies rather than the lack of mental rotation.

From the above consideration, it may be induced that the low speed of mental rotation is one of the reasons for the novices' using other strategies than mental rotation. In addition to the low speed, it is difficult for some novices to solve the tasks without using other strategies than mental rotation. Such roundabout strategies may have much influence on novices' low performance in mental rotation.



(a)S22 (b)S17 (c)S20

Figure 3.30: Response time functions (S22, S17, S20)

Change of strategies and the performance in mental rotation

S19 (Group 3) was faced with the most drastic change in strategies. S19 had mainly used an "e"-pattern in Experiment II(a), namely, had solved the task by matching encoded descriptions of structure. Mental rotation was a novel strategy for S19. As shown in Fig. 3.31, the slope of the regression line is high and the intercept was low. The high slope means a low speed of mental rotation. It was difficult for S19 to rotate the Shepard-Metzler figures. It is considered that the low speed of mental rotation evoked S19 to solve the tasks by matching encoded descriptions of structure.

Distribution of the performance in mental rotation

In Fig. 3.32, the slope and the intercept of the response function for each subject (S11 – S22) were summarized.

The parameters don't much differ between the experts. The low slopes and low intercepts show that the experts solved Shepard-Metzler tasks by high speed mental rotation.

Whereas, the parameters for the novices distribute over the wide range. The wide distribution means that the difference in strategies between the novices remained even if they were compelled to solve only by mental rotation.

4 novices produced the high slopes, which means the low speed of mental rotation. The low speed of mental rotation may be one of the reasons why the strategies other than mental rotation were observed among the novices. 4 novices produced the high intercepts, which means that they were unable to unify their strategies in spite of the instruction. This suggests that it was so difficult for such novices to solve the tasks only by mental rotation. The difficulty is considered to be another reason of the variety of strategies.

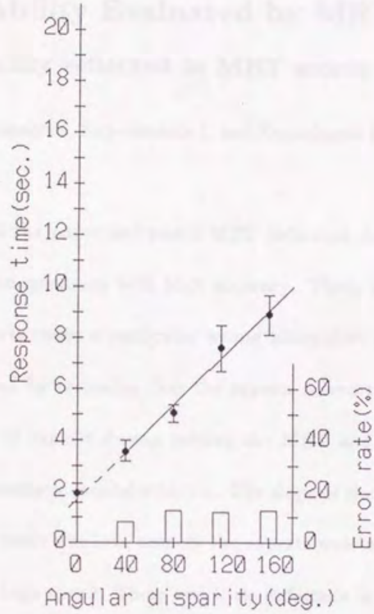


Figure 3.31: Response time functions (S19)

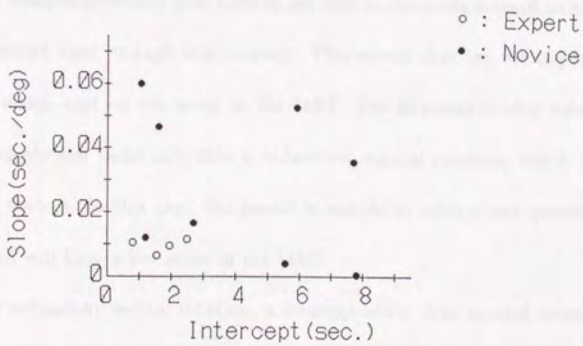


Figure 3.32: Slopes and intercepts for each subject

3.5 Spatial Ability Evaluated by MRT

3.5.1 Spatial ability reflected in MRT scores

The results of the error analysis, Experiments I, and Experiment II(a)(b) are summarized in Table 3.11.

The error analysis in the paper-and-pencil MRT indicated that the experts are able to succeed in most of the questions with high accuracy. There were a small number of tasks in which the experts chose a particular wrong alternative as a correct one. Such mistakes can be explained by assuming that the experts solved the questions by mental rotation. Eye fixations of experts during solving the MRT and Shepard-Metzler tasks indicated simple and systematic mental rotation. The slope of the response time function for the Shepard-Metzler tasks was low, namely the experts were able to rotate the three-dimensional figures at a high speed. There was little difference in strategies between the experts. The high performance in the Shepard-Metzler tasks and the stability of the strategy are bound to the high score in the MRT.

Whereas the error analysis indicated that novices are able to carry out a small number of questions in the limited time, though inaccurately. This shows that the solving time is one of the factors which load on the score in the MRT. Eye fixations during solving the MRT and Shepard-Metzler tasks indicated a redundant mental rotation, which was mainly used by some novices. In this case, the novice is unable to solve many questions in the limited time and will have a low score in the MRT.

In addition to the redundant mental rotation, a strategy other than mental rotation was suggested by the error analysis of the paper-and-pencil MRT. Assuming a strategy

Table 3.11: Summary

Subject	Error analysis of MRT	Eye fixation diagram		Strategy to solve MRT	Response time function (S. - M.*)	Performance in mental rotation
		MRT	S. - M.*			
Experts	High acccracy Errors in the invisible question	R Rc, Rs, Rp	r r1.5	Systematic mental rotation	Low slope and low intercept	High speed of mental rotation
Novices	Low speed with inaccuracy	-R RR	rn	Redundant mental rotation	High slope (or many errors)	Low speed of mental rotation
	Errors in mirror-image questions	F F/RR	f/rn	Detecting structural features Redundant mental rotation	High slope or high intercept Flat	Low speed of mental rotation Difficulty in unifying strategies
		E	e	Matching encoded descriptions	High slope and low intercept	Low speed of mental rotation

*) S. - M. : Shepard-Metzler tasks

by detecting the relation between the arms, it may be possible to explain the lower mean scores of the mirror-image questions. Because of the strategy, it is easy to make a mistake in the mirror-image questions. In these, each alternative has the same relation between its arms. Eye fixations during solving the MRT and the Shepard-Metzler tasks indicated that some novices used the detecting structural features strategy together with the mental rotation. The existence of the strategy by detecting features is consistent to the errors in mirror-image questions, suggested by the error analysis.

When the novices were compelled to solve Shepard-Metzler tasks only by mental rotation, most of the novices showed a very high slope of the response time function. This means the low speed of mental rotation. The novice is unable to finish many questions in the limited time even if the answers are correct. Then, a low score in the MRT will result in consequence of the low speed of mental rotation. In some cases, the low speed of mental rotation evokes another strategy from the novices. Some novices showed a high intercept of the response time function, which means that it was difficult for such subjects to unify the strategies.

Eye fixations also indicated that the strategies of novices sometimes differed within themselves across questions. When compelled to solve the Shepard-Metzler tasks only by mental rotation, the response time function for some of these novices showed a flat slope and high intercept. The flat slope indicates the existence of the novices for whom it is extremely difficult to unify the strategies. Such novices are unable to solve the MRT without using strategies other than mental rotation. The confusion of different strategies is also one of the sources of individual differences in mental rotation. The lack of the unification of strategies will produce a low score in the MRT.

Eye fixations indicated a strategy which hadn't been expected from the error analysis. A few novices used the strategy by matching encoded descriptions. The strategy is so complicated and takes such a long time to solve each question that the novice is unable to complete many questions in the limited time. The novice will have low score in the MRT. When such a novice was compelled to solve Shepard-Metzler tasks only by mental rotation, he had an extremely high slope of response function. Even if he solved the questions only by mental rotation, he won't be able to obtain a high score in the MRT because of the low speed of rotation.

As stated above, the source of individual differences in MRT scores is the speed of mental rotation and the variety of strategies. Most subjects solve the MRT by mental rotation. Simple and rapid mental rotation by experts produces a high score in the MRT and a redundant and slow rotation by some novices produces a low score. The speed of mental rotation is one of the factors which have an effect on the score in the MRT. There are some novices who show a variety of strategies to solve the MRT. The variety of strategies may be evoked by the low speed of mental rotation and the difficulty in unifying strategies to mental rotation.

From the above analysis, it is concluded that the score in the MRT reflects following aspects of spatial ability;

- the speed of mental rotation of three-dimensional figures and
- the difficulty in unifying strategies to mental rotation.

It is summarized that the score in the MRT evaluates the performance of mental rotation.

3.5.2 Comparison with related studies

Just et al.[12] indicated that low-ability subjects' response times for Shepard-Metzler tasks increased faster with angular disparity. Lohman[17] presented Shepard-Metzler tasks for various fixed exposures and estimated the rotation speed for a fixed level of accuracy from the speed-accuracy curves. The estimated slope (i.e. rotation speed) was generally steeper for low-ability subjects than high-ability ones. The differences in mental rotation speed between the experts and the novices in the present investigation agree with the results of the previous investigations by Just et al.[12] and Lohman[17].

The analysis of the eye fixations in the Shepard-Metzler tasks by Just et al.[12] indicated that novices did extra work at rotation and confirmation stages. Such extra work corresponds to redundant rotation of the novices observed in the present investigation.

Tapley et al.[16] reported differences in strategy preference between subjects when they solved Shepard-Metzler tasks. As for the MRT, Freedman et al.[21] and Pezaris et al.[22] explored differences in strategies between subjects. However, the methods to infer strategies used by Tapley et al.[16], Freedman et al.[21], and Pezaris et al.[22] were indirect. They used self report by their subjects or compared correct rates in dual-task conditions with that in control condition. The present experiments, in which strategy differences in the MRT were explored by more direct method, i.e., using eye fixation data, indicated the variety of strategies between novices. The eye fixations showed that the strategies of some novices differed within themselves across questions and that some novices changed strategies halfway in a question. Such instability of strategies had not been reported in the previous research[12] [16][17] [21][22].

Tapley et al.[16], Freedman et al.[21], and Pezaris et al.[22] paid attention to sex-based strategy differences in Shepard-Metzler tasks or the MRT. Freedman et al.[21] and Pezaris et al.[22] indicated sex differences in strategy preference. The present investigation directed its attention to strategy differences between high- and low-score subjects. The error analysis suggested that there isn't a large difference in strategies between males and females in same range of MRT scores. In other words, as for college-aged subjects in the present research, the source of differences in strategies is the score rather than gender. A strategy by simple and systematic mental rotation was used by the high-score subjects and strategies by redundant mental rotation, detecting structural features, and matching encoded descriptions were used by the low-score ones.

Sex differences in MRT scores were observed in the present investigation and in many previous studies[19]–[21]. A large number of males occupy high-score range and a large number of females low-score range. Then, the sex-based strategy differences indicated by Freedman et al.[21] and Pezaris et al.[22] are consistent to the strategy differences between high- and low-score subjects in the present investigation.

Pezaris et al.[22] controlled genetic and environmental factors of subjects and indicated that there were the girls whose strategy preference was similar to the boys. Casey et al.[20] reported the existence of female group whose mean MRT score was as well as male. From the results of these studies, it is considered that some females use same strategies as males and such females get similar MRT scores as males. It is consistent to that the source of differences in strategies is the score rather than gender as shown in the present investigation. Although Tapley et al.[16] reported that there was no sex differences in professed strategy, it is possible that the males and the females in their experiments might have same level of ability reflected in the MRT.

The present investigation showed that high- and low-score subjects differ not only in the slopes of response time function but also in the intercepts. The high intercept reflects strategies other than mental rotation and the difficulty in unifying strategies. The existence of strategies other than mental rotation can explain the inconsistency of the results of previous studies in which correlations between the slope of the reaction time function for Shepard-Metzler tasks and reference spatial tests (excepting the MRT) were computed[13]–[15]. Flat slopes observed in the present experiments might eliminate correlations or contribute to no or positive correlation.

Juhel[5] conducted a factor analysis for 4 laboratory memory-tasks and 5 spatial tests, in which the MRT was included. In his analysis, the MRT had high loading with both "Visualization" factor and "Speeded Rotation" factor. The present experiments revealed that the score in the MRT reflects both of the speed of mental rotation and the variety of strategies. It is considered that the speed of mental rotation and the variety of strategies correspond to "Speeded Rotation" factor and "Visualization" factor, respectively. The results of the present experiments can explain the results of Juhel's factor analysis.

Chapter 4

MRT AND UNDERGRADUATE GRAPHICS EDUCATION

The MRT has been widely used in recent quasi-experimental research [6][23]−[25], the objective being to evaluate the spatial abilities of students from the viewpoint of how they are related to undergraduate graphics curricula. Most of the researchers in the previous investigations have considered the MRT scores to be a direct indicator of the ability acquired through the curricula, without examining which aspects of spatial ability are reflected in the MRT scores.

The analysis of the solving process in the previous chapter revealed that the score in the MRT reflects (1) the speed of mental rotation and (2) the difficulty in unifying strategies to mental rotation. It is summarized that the score in the MRT evaluates the performance of mental rotation.

In this chapter, the relation between the MRT and performance in undergraduate graphics curricula will be discussed.

4.1 Relation between MRT and Term-end Test Scores

In this section, the relation between the spatial ability reflected in the scores in the MRT and the scores in the term-end tests of undergraduate graphics curricula will be analyzed.

4.1.1 Pilot study

The relationship between the scores in the MRT given before the undergraduate graphics courses and the scores in the term-end tests given after the courses was analyzed at two universities; T university and O University. Different problems were used as the term-end test at each university. The coefficients of correlation between them were 0.32 and 0.30 at T university and O university, respectively. These coefficients suggest some correlations between the MRT and the term-end tests. The following experimental research was conducted to clarify the relation between the MRT and term-end tests.

4.1.2 Experimental methods

In order to discuss a wide range of the MRT scores, two universities whose mean scores in the MRT were quite different to each other were chosen. Common problems were used to evaluate examination performance at the two universities: O university and T university.

The MRT was given to the students on the first day of the graphics courses at each university. In these courses, descriptive geometry, i.e., treatment of points, lines, planes and objects through orthogonal projections, was taught in a traditional way. Hereafter, the courses shall be called "DG courses". After DG courses, the term-end test was given. The course duration was one semester. The scores in the MRT and the term-end tests will be statistically analyzed.

4.1.3 Design of term-end tests

The term-end tests were composed of two kinds of problems. In one, the same problems were given at both universities, while in the other, different problems requiring application abilities were given at each university. Hereafter, the former problems shall be called "common problems" and the latter, "application problems".

Fig. 4.1 and Fig. 4.2 show common problems in 1992 and 1993, respectively. Common problems consisted of the following two components.

- Component I: Making a three-view orthographic drawing from a given pictorial view of an object. (Fig. 4.1(1), Fig. 4.2(1))
- Component II: Solving basic spatial problems by constructing auxiliary views. Determination of the angle between given planes, the true length of a given line or the true shape of a given plane.(Fig. 4.1(2), Fig. 4.2(2)(3))

In descriptive geometry, three-dimensional objects are treated through three-view orthographic drawings. Making a three-view orthographic drawing from an object (Component I) is necessary to understand descriptive geometry. It is difficult for students to jump the gap between the three dimensional object and its two dimensional representation, though they have mathematical knowledge about solid geometry.

Spatial problems are solved by graphical methods; "auxiliary views are projected directly to auxiliary planes in succession until the required geometric relationships are found([31]p.3)". Basic spatial problems in Component II are solved by applying rules taught in DG courses.

In the common problems in 1993 and 1992, 10 points were given to each problem. So,

(1) Fig. I and Fig. II show sequential intersection of the planes and the parallelepiped. Object ABC... is produced by the intersection. Make a front view, a top view and a right side view of the object (drawn in bold lines).

(2) Determine the true–size view of rectangle BFG.

(3) Determine the angle between plane AEFB and plane EJKF.
<Hint> An angle between two planes appears as an intersection of the edge view of the two planes by constructing the point view of the intersection line between the two planes in the secondary auxiliary view.

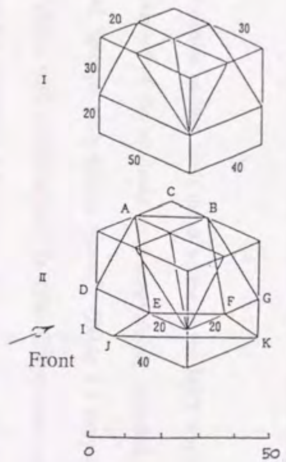
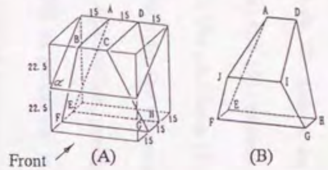


Figure 4.1: Common problems in 1992

(1) Fig.(A) shows intersection of plane α and the object drawn in bold lines. Object(B) is produced by the intersection. Make a front view, a top view and a right side view of object(B).



(2) Determine the true length of line GB and the true–size view of rectangle EDGF.

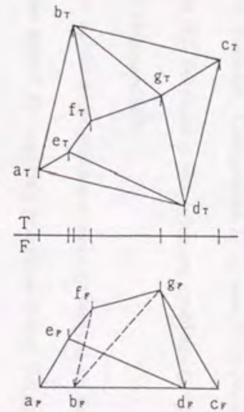


Figure 4.2: Common problems in 1993

the perfect score in the common problems in 1992 is 20, and that in 1993 is 30. As for the common problems in 1992(see Fig. 4.1), the three-view orthographic drawing which is obtained as the answer in problem (1) (namely, Component I) is used to solve the problems (2) and (3) (namely, Component II). Whereas in the common problems in 1993 (see Fig4.2), the problems (1) and (2) are independent.

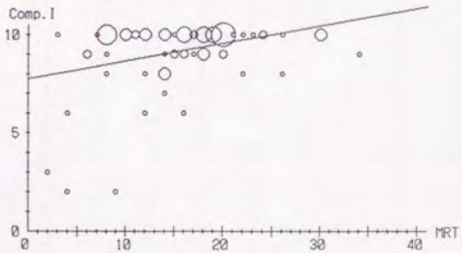
In addition to the common problems, application problems were presented at each university. Application problems were also solved by applying graphical rules. Because of complexity in problems, it is more difficult for students to find appropriate rules to apply than solving basic spatial problems. At T university, the problem was constructing a drawing of two objects intersecting. At O university, the consideration of the relation between the human body and the surface development was given.

4.1.4 Correlation between scores in MRT and term-end tests

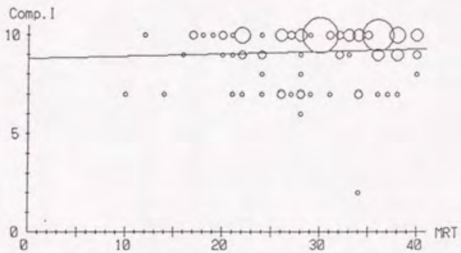
Component I

In Fig. 4.3 and Fig. 4.4, are plotted the scores in the MRT and Component I at each university. Table 4.1 shows the coefficients of correlation between the scores in the MRT and Component I. The coefficient of correlation at O university was not so low, while that at T university was very low.

Subjects from each university were chosen so that their mean scores in the MRT were almost equal and so that nearly 30 percent of the students were respectively included. Hereafter such a group shall be called "groups with similar scores". Table 4.2 shows the mean scores in Component I for "group with similar scores" in 1993. There was no significant difference in the mean score for Component I between the two universities,

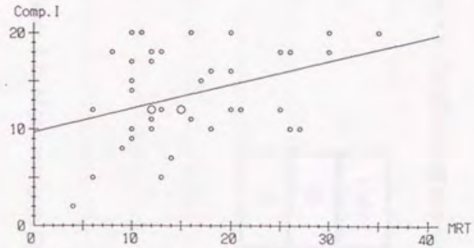


(a)O Univ.

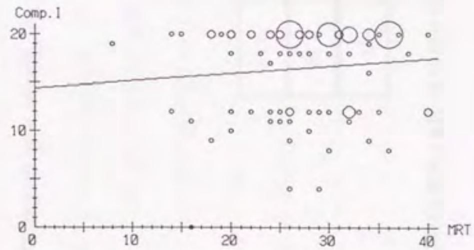


(b)T Univ.

Figure 4.3: Correlation between scores in MRT and Component I (in 1992)



(a)O Univ.



(b)T Univ.

Figure 4.4: Correlation between scores in MRT and Component I (in 1993)

Table 4.1: Coefficient of correlation between scores in MRT and Component I

Univ.	'92	'93
O.	0.35	0.38
T.	0.06	0.10

Table 4.2: Mean scores in Component I for "groups with similar scores"

Univ.	Samples	MRT	Comp.I
O.	14	24.4 (4.9)	15.1 (3.8)
T.	35	24.4 (2.2)	16.4 (4.4)

( ) : Standard deviation

although general scholastic ability of the students are quite different between the two universities. Similar results were obtained in 1992. These results show that the scores in the MRT are a factor influencing the score for Component I.

As shown in Fig. 4.4(a) and Table 4.1, the coefficients of correlation were not so low at O university. It can be concluded that for students in the low MRT score range, the scores in the MRT have some influence on the ability to make a three-view orthographic drawing from a pictorial view of an object. It can be said that the students with low scores in the MRT feel it difficult to construct an abstract two-dimensional drawing from a three-dimensional object. As shown in Fig. 4.4(b) and Table 4.1, the coefficient indicated very low correlation in the group with high scores in the MRT. The MRT may have little influence when scores in the MRT exceed a certain level.

As shown in the previous chapter, the scores in the MRT mainly reflect the ability to mentally rotate his/her image of an object; i.e. to manipulate his/her image. To make a three-view orthographic drawing, it is necessary to change the view direction or to rotate the object in his/her image. That may be the cause why the scores in the MRT have some correlations with the scores in Component I.

Component II

In Fig. 4.5 and Fig. 4.6, are plotted the scores in the MRT and Component II at each university in 1992 and 1993. The coefficient of correlation between them is summarized in Table 4.3. Table 4.4 shows the mean scores in the MRT and Component II for "groups with similar scores" in 1993.

As stated in the explanation of test design, the problems of Component I and II in

1992 are sequential, in other words, the correct answer of the problem of Component I is necessary to solve the problems of Component II. In 1992, the following two reasons are possible for students whose scores in Component II were nearly zero point.

1. Because they were unable to get the answer of Component I, they were unable to solve Component II even if they knew how to solve Component II.

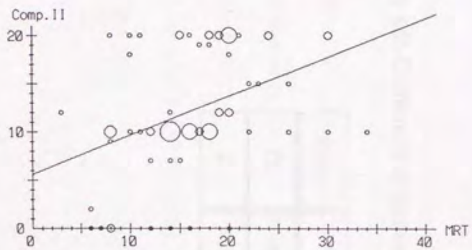
2. Because they didn't know how to solve Component II.

The former may overestimate the coefficients of correlation. To minimize the former effect, data of students whose scores in Component I were less than 7 was removed from the analysis in 1992.

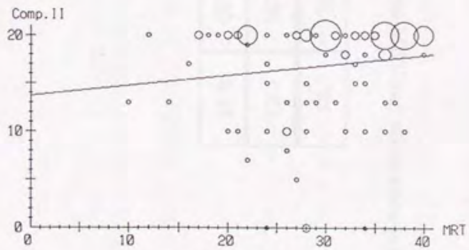
As Table 4.3 shows, the mean scores for Component II were quite different between the two universities within the groups with similar scores. So, it can be said that some factor other than the scores in the MRT had more greatly influenced the score for Component II. As shown in Fig. 4.6 and Table 4.4, the coefficients of correlation were very low except at O university in 1992.

From the above consideration, it is concluded that the scores in the MRT have little influence to solve Component II, i.e., to solve basic spatial problems by constructing auxiliary views.

The students may be able to solve basic spatial problems in Component II by only applying graphical rules without connecting drawings to three-dimensional objects in their images. This interpretation can explain low correlation between the MRT and performance in solving Component II.

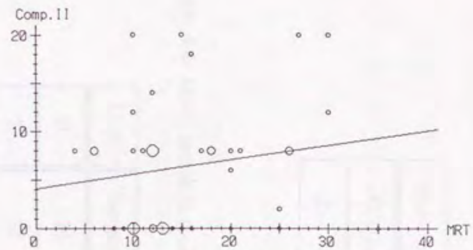


(a)O Univ.

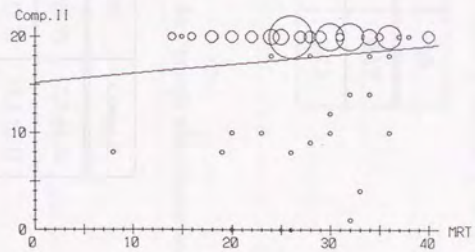


(b)T Univ.

Figure 4.5: Correlation between scores in MRT and Component II (in 1992)



(a)O Univ.



(b)T Univ.

Figure 4.6: Correlation between scores in MRT and Component II (in 1993)

Table 4.3: Coefficient of correlation between scores in MRT and Component II

Univ.	'92	'93
O.	0.39	0.17
T.	0.14	0.12

Table 4.4: Mean scores in Component II for "groups with similar scores"

Univ.	Samples	MRT	Comp.II
O.	14	24.4 (4.9)	7.7 (6.2)
T.	35	24.4 (2.2)	17.3 (6.1)

( ) : Standard deviation

Table 4.5: Coefficients of correlation between MRT and application problems

Univ.	'91	'92	'93
O.	0.30	0.36	0.08
T.	0.32	0.30	0.17

Application problems

In order to extend the discussion into more general problems on descriptive geometry, the relation between the scores in the MRT and application problems at each university shall be considered. Table 4.5 shows coefficients of correlation between the scores in the MRT and the application problems in 1992 and 1993 at each university. Data in 1991 are the results of the pilot study, in which no common problem was used. The problems of term-end tests in 1991 were similar to the application problems in 1992 and in 1993.

The coefficient of correlation in 1993 at O university was much lower than others. At O university, the main difference between 1992 and 1993 is whether application problems in each year required students to make a drawing of a surface development with their own hands or not. In 1992, students were required to make a drawing and describe a consideration of the relation between the human body and the drawing. In 1993, a drawing of a surface development of the human body was already given; the students didn't have to make a drawing by themselves. At T university, students in both years were required to make with their own hands a drawing of intersecting surfaces. Although coefficients of correlation fluctuated year by year, they were not so low.

From the above consideration, it can be concluded that there is some correlation between the scores in the MRT and abilities to solve application spatial problems in descriptive geometry.

To solve the application problems such as those at T university (intersection of two objects), students are required to find piercing points and then to connect them. They can find piercing points by applying graphical rules to given drawings. These processes are similar to those required to solve Component II. However, it is difficult to *connect* piercing points only by applying rules. It is necessary to make the image of the intersection in their minds. This may be the reason why the scores in the MRT have some correlations with the scores in application problems.

4.2 Failure rates in DG course

Experienced educators have noticed that there are students who have extreme difficulty getting over the gap between real objects and two-dimensional drawings. If educators can identify students who require more help to complete courses before they drop out, educators will be able to make special arrangements for them. The MRT, which can be administered quickly in a classroom, is one candidate for identifying such students. In this section, the relation between the scores in the MRT and acquiring credit in descriptive geometry will be analyzed in order to clarify whether the MRT can be a predictor of students at high risk of poor performance in graphics courses.

4.2.1 Research Design

The investigation was held at the same two universities in the previous section. The students took the term-end tests which consisted of common problems (Component I and Component II) and application problems stated in the previous section. Relation between the scores in the MRT and the acquiring credit in descriptive geometry will be analyzed.

4.2.2 Results

Some students dropped the course or couldn't pass the term-end test. Such students must have felt it difficult to study descriptive geometry. We will define the term "failure rate"

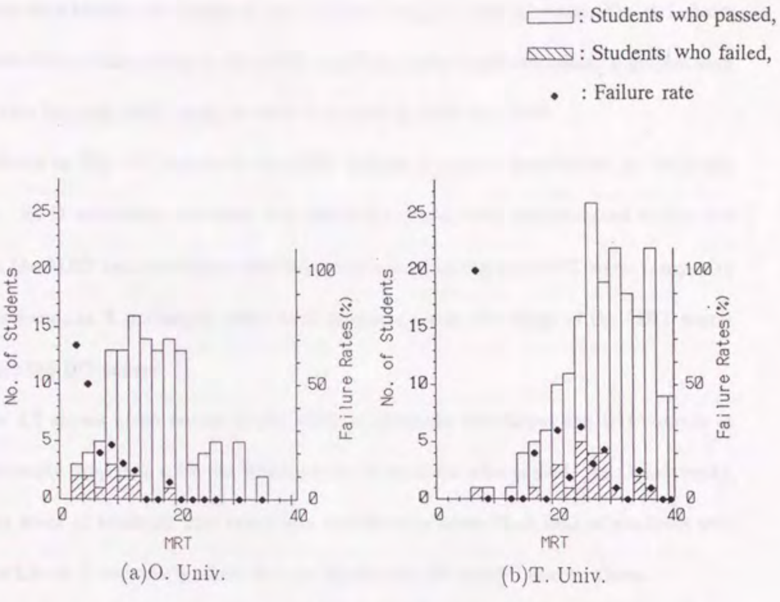


Figure 4.7: Distribution of scores in MRT and failure rates

Table 4.6: Mean scores in MRT(Students who failed and passed)

Univ.	Samples	Fail	Pass
O.	13 / 114	8.2 (4.5)	16.0 (6.6)
T.	23 / 173	25.1 (6.2)	29.1 (6.7)

Samples : No. of students who failed / passed
Fail : Mean score (sd) of students who failed
Pass : Mean score (sd) of students who passed

as the rate of students who dropped the course or couldn't pass the exam. Fig. 4.7 shows the distribution of the scores in the MRT and how many students failed, together with failure rates for each MRT score at each university in 1992 and 1993.

As shown in Fig. 4.7, scores in the MRT indicated normal distribution at both universities. At O university, students who failed the course were concentrated within low scores in the MRT and the failure rate was pronounced at the low MRT score range. On the other hand, at T university, there were students across the range of the MRT scores who failed the DG course.

Table 4.6 shows mean scores in the MRT of students who failed the DG courses at each university, together with the mean scores of students who passed. At O university, the mean score of students who failed was significantly lower than that of students who passed, while at T university, there was no significant difference between them.

From the above results, it is concluded that the students with poor scores in the MRT are likely to fail to complete the DG courses. As discussed in the previous section, there are some correlation between the scores in the MRT and performance in descriptive geometry, especially for the students with poor MRT scores. This correlation may cause the high failure rate above mentioned. The results of the present investigation suggested that the MRT can be a candidate of a predictor of students at high risk of poor performance in graphics courses.

4.3 Pretest / Posttest Comparison

As discussed in Section 4.1 and Section 4.2, there are some correlations between the ability evaluated by the MRT and performance in descriptive geometry. In this section, comparisons of MRT scores prior to and after descriptive geometry courses will be made in order to clarify whether the ability can be enhanced through DG curricula.

4.3.1 Research design

The quasi-experimental research design was chosen as the most appropriate method for this study. Subjects were students taking descriptive geometry courses at three universities. Hereafter, referred to as M university, O university, and T university. The MRT was given to subjects prior to and after the DG courses at each university. Hereafter, the tests which were given prior to and after the courses will be called "pretests" and "posttests", respectively. The interval between them was approximately 4 months, namely one semester. In DG courses, descriptive geometry, i.e., treatment of points, lines, planes, and objects through orthogonal projections, was taught in a traditional way.

Additional control experiments were held at O university and T university. In one control, the pretest and posttest were given to subjects in a course in which neither descriptive geometry nor any graphics was taught. The interval between the two tests was, again, approximately 4 months. This control experiment shall be called Control-A. Another control experiment held at another descriptive geometry course gave the posttest a week after the pretest. This control experiment shall be called Control-B.

4.3.2 Results

The mean scores in the pretests, posttests and differences between them at each university are shown in Table 4.7. As shown in Table 4.7, statistical analysis revealed that the differences in results from the DG courses at all universities were significant. The mean scores, however, increased from pretest to posttest not only in the DG courses but also

Table 4.7: Mean scores of the pretests, posttests and differences between them

Univ.	Course	Sample	Pretest(SD)	Posttest(SD)	Diff.	t-value
T.	DG course	75	29.7 (7.3)	33.8 (5.9)	+4.1	6.90*
	Control-A	102	29.6 (6.6)	34.5 (5.4)	+4.9	9.56*
	Control-B	79	27.9 (6.5)	34.9 (5.3)	+7.0	12.40*
O.	DG course	46	17.5 (6.3)	22.0 (6.5)	+4.5	6.25*
	Control-A	72	14.5 (6.6)	18.8 (7.9)	+4.3	6.01*
M.	DG course	137	27.3 (7.3)	30.3 (7.1)	+3.1	5.29*
		123	28.4 (7.1)	31.3 (6.2)	+2.9	5.20*

SD : Standard deviation
Diff. : Difference between pretest and posttest
* : $P<0.005$

in Control-A and Control-B. The analysis also revealed that the differences in the DG course were not significantly different from those in Control-A. These results suggest that the increase in the DG courses was not caused by the descriptive geometry curriculum.

The analysis showed that the differences in Control-A and Control-B were significant ($P<0.01$). As shown in Table 4.7, the increase in Control-B was much larger than that in Control-A at T university. It should be recalled that the interval between the pretest and posttest in Control-B was only a week, much shorter than that in Control-A. From these results, it can be considered that the subjects who had experienced the MRT got higher scores in the second trial, that is, the cause of the increases observed in the present experiment is the direct training effect for the MRT. Recalling that the stability of strategies has a great effect on the scores in the MRT as shown in the previous chapter, the experience of solving the MRT might promote to unify strategies. When the students solved the MRT at the second time (namely, posttest), they might be able to solve questions more efficiently than the first time.

The differences in the DG courses at M university were 3.1 and 2.9. They were nearly equal to the difference in Control-A at O university and T university. Churches et al.[32] recently reported almost same gains for students in Australia, although they didn't have control experiments. It seems likely that the direct training effect is similar in all the subject groups. If so, the difference observed at M university was not due to the descriptive geometry curricula.

Although significant gains on the MRT were made by students in the DG courses, the control experiments indicated that the gains were caused not by descriptive geometry education but by direct training effect. It is concluded that the spatial ability evaluated by the MRT is not enhanced by teaching descriptive geometry in the classroom.

As discussed in the previous section, the ability reflected by the MRT has some correlations with the performance in descriptive geometry. One may naturally suppose that teaching descriptive geometry may, then, enhance the ability reflected by the MRT. However, the ability cannot be enhanced by the teaching as shown in the present pre-post comparison. In these DG courses, lectures on descriptive geometry are given during one semester. Exercises, typically 5–7 assignments, are given as homework. The results of the present investigation indicate that basic spatial ability such as that evaluated by the MRT cannot be enhanced by such short-term education.

4.4 MRT and Graphics Education

In this section, the results of the present investigations are compared with related studies and are summarized.

4.4.1 Comparison with related studies

Some researchers tried to evaluate their undergraduate graphics courseware by using the MRT as a direct measure of spatial ability acquired through the courseware.

Miller[23] classified students in engineering graphics courses into two learning styles (visual or haptic) based on scores from preceding Successive Perception Test I. Three different instructional methods (traditional curricular approaches, computer-generated models, or real models) were randomly given to students in each group. He reported that there was no significant difference in the effectiveness of the approaches for either visual or haptic subjects nor was there any interaction between these variables in the advancement of spatial abilities evaluated by the MRT.

McCuistion[25] used the MRT as an index of spatial ability to compare two methods of presenting graphic images(static or dynamic) in a CAI(Computer Assisted Instruction). He gave the MRT prior to and after each CAI course to both a static and a dynamic group. As for students whose pre-MRT scores were in the lower 25 percent of the sample, the dynamic group made larger gains of MRT scores than the static group. However, the static group's mean score in performance tests was significantly higher than the dynamic group. He reported that the increase of scores in the MRT didn't guarantee high performance in descriptive geometry and vise versa.

Although Miller[23] nor McCuistion[25] didn't discuss clearly, their results suggested that the ability evaluated by the MRT were unable to be enhanced through ordinary graphics curricula. Their results agree with the results of the pre-post comparison of the present investigation.

Miller et al.[6] reported that the investigation at Ohio State University revealed that students whose scores of spatial ability tests were low had difficulty with engineering graphics and had a high failure rate.

Marlor and Gimmestad[33] gave spatial tests to freshmen at Michigan Technological University and those students who scored in the bottom 20th percentile were advised to take an experimental course in spatial visualization skills. In their study, although the sample was small, all of the students enrolled in the experimental group were still enrolled at the university, whereas, 28 % of the students in the control group had left the university and presumably had left engineering. Their study indicates low spatial ability students can follow up the course by predicting students who need more help and giving them additional courses.

Recently, Churches et al.[34] reported that the coefficient of correlation between the MRT and the final examinations was 0.36, which indicates a considerable correlation. The coefficient was almost same as the results of the present investigations.

As discussed in Section 4.1 and Section 4.2, the results of the present investigation indicated that the scores in the MRT have some correlations with the performance in descriptive geometry and with failure rates of graphics courses. The MRT can be used to predict students who need more help in learning graphics curricula.

4.4.2 Summary

In this chapter, the relationship between the ability evaluated by the MRT and descriptive geometry curricula was analyzed. The relation between the scores in the MRT and the term-end tests was statistically analyzed. The relation between the scores in the MRT and acquiring credit was also analyzed. Comparisons of MRT scores prior to and after descriptive geometry courses were made in order to clarify whether the ability can be enhanced through the curricula.

Following results were obtained through the analysis.

- Scores in the MRT have correlations with the performance of descriptive geometry, especially, with the ability to make three-view orthographic drawings from three-dimensional objects and to solve application problems on descriptive geometry.
- Students with extremely poor MRT scores are likely to fail to complete descriptive geometry courses.
- The spatial ability evaluated by the MRT cannot be enhanced through descriptive geometry education.

Chapter 5

CONCLUSION

In this paper, a Mental Rotations Test developed by Vandenberg et al. was discussed from the following viewpoints;

- the spatial ability reflected in MRT scores
- the relation between the MRT and descriptive geometry performance.

In Chapter 3, the solving process of the MRT was analyzed by comparison between experts and novices. The analysis was conducted by three methods as follows: the error analysis in the paper-and-pencil MRT, the eye fixations during solving questions, and the response solving time as a function of an angular disparity. The analysis in Chapter 3 revealed that the score in the MRT reflects following aspects of spatial ability;

- the speed of mental rotation of three-dimensional figures and
- the difficulty in unifying strategies to mental rotation.

It is summarized that the score in the MRT evaluates the performance of mental rotation.

In Chapter 4, the relationship between the MRT and descriptive geometry performance was discussed. The relation between the scores in the MRT and the scores in the term-end tests of descriptive geometry was statistically analyzed. The analysis of the relation between the scores in the MRT and acquiring credit in descriptive geometry was also conducted. Comparisons of MRT scores prior to and after descriptive geometry courses were made in order to clarify whether the ability reflected by the MRT can be enhanced through descriptive geometry curricula. Following results were obtained through the analysis in Chapter 4.

- Scores in the MRT have correlations with the performance of descriptive geometry, especially, with the ability to make three-view orthographic drawings from three-dimensional objects and to solve application problems on descriptive geometry.
- Students with extremely poor MRT scores are likely to fail to complete descriptive geometry courses.
- The spatial ability evaluated by the MRT cannot be enhanced through descriptive geometry education.

ACKNOWLEDGMENTS

The author wishes to express gratitude to the estate of the late Professor Steven Vandenberg for permission to use the Mental Rotations Test.

The author would like to express gratitude to Professor Kenjiro Suzuki for his support as supervisor throughout her years as a graduate student.

The author would like to thank Associate Professor Emiko Tsutsumi at Otsuma Women's University for her helpful suggestions.

The author wishes to express her gratitude to the referees Professor Saburo Nagano, Professor Ryo Hirasawa, Associate Professor Michio Kato, and Associate Professor Shinsuke Shimojo for reading the manuscript and making a number of helpful suggestions.

Many people have contributed to collecting data of the MRT. The author is indebted to Professor Kiyotaka Miyazaki at Otsuma Women's University for providing her the chance of the control experiment in Chapter 4. The author would like to acknowledge the support of Professor Shigeo Hirano and Professor Kunio Shimada at Musashi Institute of Technology. The author would like to thank Professor Takashi Jingu at Meisei University for helpful suggestions. The author wishes to thank Professor Kiyoe Fuchigami at Tokai University Junior College for her support in collecting data. The author also would like to thank the students at the above universities for their cooperation to the investigation.

The author withes to express her gratitude to Assistant Professor Kazuhiro Ueda and Mr. Takaaki Saito for frequent and helpful discussions. Thanks are extended to other graduate students of the Department of General Systems Studies, Graduate Division of

International and Interdisciplinary Studies, The University of Tokyo.

Finally, the author wishes to thank her family and friends for their continuous encouragement.

Bibliography

[1] Vandenberg, S. G., and Kuse, A. R.: "Mental rotations, a group test of three-dimensional spatial visualization". *Perceptual and Motor Skills* **47**(1978), 599-604.

[2] Zimovski, M. F., and Wothke W. "*The measurement of human variation in spatial visualizing ability: a process-oriented perspective*" (Technical Report 1986-1). Johnson O'Connor research foundation, Human engineering laboratory.

[3] McGee, M. G.: "Human spatial abilities: Psychometric and environmental, genetic, hormonal, and Neurological influences". *Psychological Bulletin* **86**(1979), 889-918.

[4] Lohman, D. F., and Kyllonen P. C. "Individual differences in solution strategy on spatial tasks". In R. F. Dillon, and R. R. Schmeck (Eds.), *Individual differences in cognition (Vol. 1).* (1983) Academic Press.

[5] Juhel, J.: "Spatial abilities and individual differences in visual information processing ". *Intelligence* **15**(1991), 117-137.

[6] Miller, C. L., Wiley, E. W., and Bertoline, G. R.: "Strategies for improving the visualization ability of engineering graphics students". Proceedings of the 4th. ICECGDG (*Miami*, 1990).

[7] Snow, R. E. "Aptitude process". In R. E. Snow, P. A. Federico, and W. E. Montague (Eds.), *Aptitude, learning, and instruction(Vol. 1)*. (1980) Hillsdale, NJ: Erlbaum.

[8] Lohman, D. F.: "Spatial abilities as traits, processes, and knowledge." In R. J. Sternberg (Eds.), *Advances in the psychology of human intelligence (Vol. 4)* (1988), Hillsdate, NJ: Erlbaum.

[9] Linn, M. C., and Petersen, A. C.: "Emergence and characterization of sex differences in spatial ability: a meta-analysis". *Child Development* **56**(1985), 1479-1498.

[10] Shepard, R. N., and Metzler, J.: "Mental rotation of three-dimensional objects". *Science* **171**(1971), 701-703.

[11] Just, M. A., and Carpenter, P. A.: "Eye fixations and cognitive processes". *Cognitive Psychology* **8**(1976), 441-480.

[12] Just, M. A., and Carpenter, P. A.: "Cognitive coordinate systems: accounts of mental rotation and individual differences in spatial ability". *Psychological Review* **92**(1985), 137-172.

[13] Lansman M.: "Ability factors and the speed of information processing". In M. P. Friedman, J. P. Das, and N. O'Connor (Eds.), *Intelligence and learning* (1981), New York: Plenum Press.

[14] Lohman, D. F.: "Estimating individual differences in information processing using speed-accuracy models". In R. Kanfer, P. L. Ackerman, and R. Cudeck (Eds.), *Abilities, motivation, and methodology: The Minnesota Symposium on Learning and Individual Differences* (1989), Hillsdale.

[15] Poltrock, S. E., and Brown, P.: "Individual differences in visual imagery and spatial ability". *Intelligence* **8**(1984), 93-138.

[16] Tapley, S. M., and Bryden, M. P.: "An investigation of sex differences in spatial ability: mental rotation of three-dimensional objects". *Canadian Journal of Psychology* **31**(1977), 122-130.

[17] Lohman, D. F.: "The effect of speed-accuracy tradeoff on sex differences in mental rotation". *Perception and Psychophysics* **39**(1986), 427-436.

[18] Sayeki, Y: " ' Body analogy' and the cognition of rotated figures". *The Quarterly Newsletter of the Laboratory of Comparative Human Cognition* **3**(1981), 36-40.

[19] Wilson, J. R., DeFries, J. C., McClearn, G. E., Vandenberg, S. G., Johnson, R. C., and Rashad, M. N. : "Cognitive abilities: use of family data as a control to assess sex and age differences in two ethnic groups". *International Journal of Aging and Human Development* **6**(1975), 261-276.

[20] Casey, M. B., and Brabeck, M. M.: "Exceptions to the male advantage on a spatial task: family handedness and college major as factors identifying women who excel". *Neuropsychologia* **27**(1989), 689-696.

[21] Freedman, R. J., and Rovegno, A. L.:"Ocular dominance, cognitive strategy, and sex differences in spatial ability". *Perceptual and Motor Skills* **52**(1981), 651-654.

[22] Pezaris, E., and Casey, M. B.: "Girls who use "masculine" problem-solving strategies on a spatial task: proposed genetic and environmental factors". *Brain and Cognition* **17**(1991), 1-22.

[23] Miller, C. L.: "The integration of real and computer generated models into a sketching based engineering graphics course". Proceedings of the 5th. ICECGDG (*Melbourne*, 1992), 96-101.

[24] Sexton, T. J.: "Effect on spatial visualization: introducing basic engineering graphic concepts using 3D CAD training". *Engineering Design Graphics Journal* **56**(1992), 36-43.

[25] McCuistion, P. J.: "Static vs. dynamic visuals in computer assisted instruction". *Engineering Design Graphics Journal* (Spring, 1991), 25-33.

[26] Chibana, K., Nishihara, K., and Yoshida, K.: "Effect of anaglyphic stereogram on the MRT (in Japanese)". Proceedings of 1993 Annual Meeting of Japan Society for Graphic Science, May 1993, 121-126.

[27] Yoyogi Seminar (Eds.) : *University entrance examinations alpha* (in Japanese) **218**(1990).

[28] Suzuki, K., Shiina, K., Makino, K., Saito, T., Jingu, T., Tsutsumi, N., Kashima, S., Shibata, M., Tsutsumi, E., and Isoda, H.: "Evaluation of students' spatial abilities by a mental cutting test". Proceedings of the 5th. ICECGDG (*Melbourne*, 1992), 277-281.

[29] Vandenberg, S. G., and Kuse, A. R.: "Answer key for Mental Rotations Test".

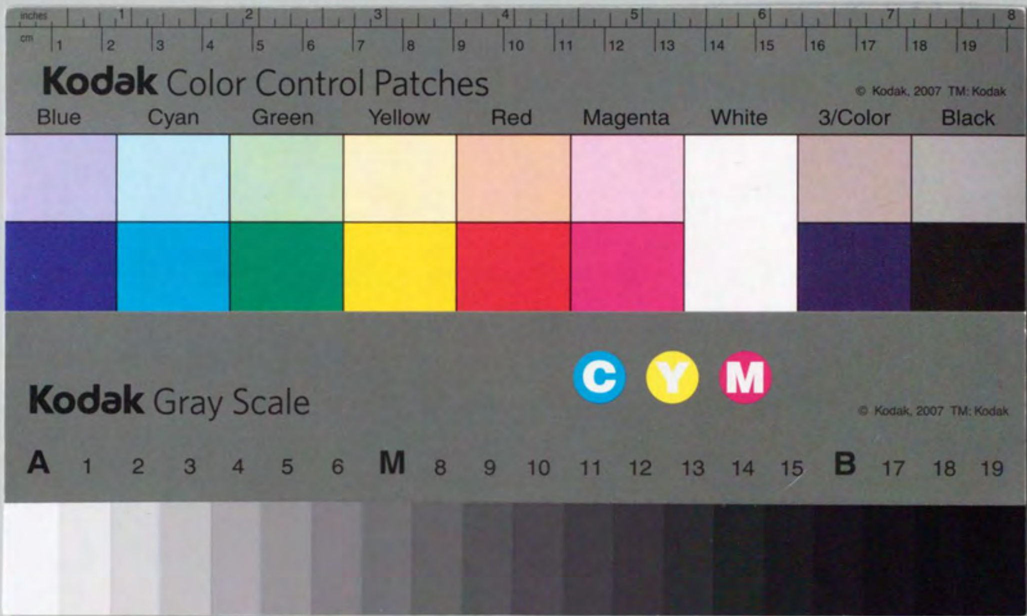
[30] Makino, K., Saito, T., Shiina, K., Suzuki, K., and Jingu, T.: "Analysis of problem solving process of a mental cutting test by the use of eye fixations data". Proceedings of the 5th. ICECGDG (*Melbourne*, 1992), 398-402.

[31] Earle, J. H.: *Engineering design graphics* (1969). Addison-Wesley Publishing Company.

[32] Churches, A. E., Barratt, A. J., Challen, J. M., Frost, R. B., Isles, J. D., Kanapathipillai S., Magin, D. J., and Platfoot, R. A. : "The impact of computer graphics in developing students' visualization and mechanical engineering design abilities". Proceedings of the 6th. ICECGDG (*Tokyo*, 1994), 771-775.

[33] Marlor, S. S., and Gimmestad, B.: "An introduction to 3-D spatial visualization – a pre-graphics course – ". Proceedings of the 6th. ICECGDG (*Tokyo*, 1994), 820-824.

[34] Churches, A. E., Magin, D. J., and Barratt, A. J. "Prediction of examination performance in drawing and descriptive geometry based on spatial ability tests". Proceedings of the 6th. ICECGDG (*Tokyo*, 1994), 796-800.


Kodak Color Control Patches
© Kodak, 2007 TM: Kodak
Blue
Cyan
Green
Yellow
Red
Magenta
White
3/Color
Black
Kodak Gray Scale
C
Y
M
© Kodak, 2007 TM: Kodak
A 1 2 3 4 5 6 M 8 9 10 11 12 13 14 15 B 17 18 19