AN ANALYSIS OF RETENTION OF FACTUAL MATERIAL
PRESENTED IN SONG AND STORY FORM

APPROVED:

Mark E. Bone
Major Professor

Sydney Hamilton
Minor Professor

Witt D. Blair
Director of the Department of Psychology

Robert B. Touloukian
Dean of the Graduate School
AN ANALYSIS OF RETENTION OF FACTUAL MATERIAL
PRESENTED IN SONG AND STORY FORM

THESIS

Presented to the Graduate Council of the North Texas State University in Partial Fulfillment of the Requirements

For the Degree of

MASTER OF ARTS

By

Kathryn Walker Pinson, B. A.

Denton, Texas
August, 1964
# TABLE OF CONTENTS

**LIST OF TABLES** ........................................... iv

**Chapter**

I. INTRODUCTION ........................................... 1
   - Theoretical Considerations
   - Original Study
   - Criticisms

II. EXPERIMENTAL DESIGN ................................. 18
   - Statement of Purpose and Hypotheses
   - Subjects
   - Task Variable and Procedure

III. ANALYSIS OF DATA AND DISCUSSION ............... 28
   - Analysis of Data
   - Discussion

IV. SUMMARY AND RECOMMENDATIONS .................... 38
   - Summary
   - Recommendations

APPENDIX (A, B, and C) ................................. 43

BIBLIOGRAPHY .............................................. 47
LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Summary Chart Showing Range, Median and Mean of Subjects' Ages and IQ</td>
<td>20</td>
</tr>
<tr>
<td>II. Description of Subjects by Selected Characteristics</td>
<td>22</td>
</tr>
<tr>
<td>III. Complete Analysis of Variance for Immediate Recall Scores</td>
<td>29</td>
</tr>
<tr>
<td>IV. Complete Analysis of Variance of Recent Recall Scores</td>
<td>30</td>
</tr>
<tr>
<td>V. Complete Analysis of Variance for Delayed Recall Scores</td>
<td>31</td>
</tr>
<tr>
<td>VI. Mean Scores for Immediate, Recent and Delayed Recall by Groups</td>
<td>32</td>
</tr>
<tr>
<td>VII. Subjects by Treatment, Giving Raw Scores for Immediate, Delayed and Recent Recall</td>
<td>35</td>
</tr>
</tbody>
</table>
CHAPTER I

INTRODUCTION

Retention was one of the first aspects of human learning and verbal behavior to be subjected to laboratory analysis. Ebbinghaus (6) published in 1885 Über das Gedächtnis, translated as Memory: A Contribution. This monologue remains a classic in the field. From this beginning point, many subsequent studies have been carried out.

In the area of retention, as in many other areas of psychology, a difference exists between the psychological laboratory and the everyday world in which retention is of practical concern. Bartlett (2) argued that psychologists who use meaningless stimuli and who force subjects to learn in unusual or artificial ways (e.g., by means of a memory drum) seriously hamper their chances of understanding everyday perception and memory as it is actually effected at the human level of experience. Though aware of the necessity for psychology to investigate phenomena in its simplest form, Bartlett reasoned that the processes of learning and retention can be best understood in a more natural setting with close-to-everyday material.

Before proceeding further retention must be defined. McGeoch says:
Retention is a general term referring to the persistence of those modifications of behavior which have been learned . . . . Learning and retention are continuous processes, being separated experimentally by the fact that an interval of time elapses between the attainment of an arbitrary criterion and a subsequent measurement of performance (24, p. 355).

Deese (5, p. 236) and others (26, p. 276) support this definition.

It is evident, then, that learning and retention are part of a continuous process; one cannot occur without the other. Factors which facilitate the original learning, facilitate retention. What are these factors?

Psychology has not yet isolated all of the factors that are involved in even the simplest human learning situation. It is known that some factors facilitate learning while others do not. Much critical research remains to be done in this area. Although it is beyond the scope of the present paper to discuss all of the factors that are known in the field of learning, those that are pertinent to verbal learning and retention will be reviewed.

The original learning situation is one of the major determinants of retention. The nature of the task variable is an important consideration in any learning experiment. The work of Ebbinghaus (6) dealt with nonsense syllables. He sought to eliminate prior subject knowledge by introducing a totally new and meaningless task variable. Subsequent investigators (2, p. 3; 9, p. 15; 26, p. 279) have found,
however, that subjects set up a mass of associations, some quite odd, and that these may vary from subject to subject more so than do associations aroused by common language with conventional meanings. It has been clearly demonstrated that meaningful material is learned faster than is nonsense material (17; 24, pp. 469-472; 27, p. 30).

It was Ebbinghaus (6) who first dealt with the problem of measuring retention at different time intervals following the original learning situation. The typical retention curve is based on Ebbinghaus' measurements of retention (5, p. 241). His findings that decline in retention is negatively accelerated has been corroborated by others (24, p. 358; 26, p. 359). The data from Ebbinghaus' study further indicated that the greatest amount of forgetting takes place in the first twenty-four hours (26, p. 360).

Distributed practice in learning generally results in more efficient learning than does massed practice (1, p. 187; 7, p. 324). However, there is some evidence that with some tasks massed practice may be equally or more effective (24, p. 144). Here, the nature of the task itself and the total amount to be learned are important considerations.

The number of task presentations or trials is a factor in learning and retention (1, p. 184; 20, pp. 77-78; 22, pp. 591-595). Deese states, "It is well established that amount of retention is roughly proportional to amount of original
practice" (5, p. 243). That there is not a simple one to one relationship between number of trials and subsequent retention, however, was pointed out by Estes (8, pp. 205-221). Others, too, argue either for one-trial learning or something-less than increased association with each trial (11, p. 97; 19; 28, p. 193; 29, p. 104).

Other factors found to be of importance are length of task, serial positioning and the mode of learning as in part or whole learning (1, pp. 197-199). The greater the degree of organization in material learned, the slower the rate of forgetting and the higher the amount of retention (1, p. 215; 5, p. 242). That memory for ideas in a prose passage is better than memory for factual detail was brought out by Briggs and Reed (3, p. 516).

In general, retention favors vivid and distinct material that is rich in associative support. Isolated items placed in otherwise homogeneous material are favored in retention (26, p. 364). Material that lends itself to rhythmic patterning is more readily learned and retained (24, p. 476).

The effectiveness of rhythm probably results in part from the way in which it assists in organizing the material into units which are readily perceived together, in part from the accent it gives to certain words or serial positions, which makes them stand out as reference points, and in part from the fact that it assists an active attitude on the part of the learner (24, p. 477).

Another factor involved in the nature of the task is its affective content. Pleasant and unpleasant material is
remembered more easily than is so-called neutral material (24, p. 384). This may relate to the vividness of such material or the stronger impression such material may make.

In the preceding paragraphs, retention has been reviewed as a function of the conditions in which the original learning took place. Of paramount importance is the nature of the task itself, the mode of task presentation and the conditions of practice.

A second group of determinants of retention is the activity which intervenes between the time of the original learning and the test of retention.

Periods of waking and sleeping, which clearly differ in amounts of activity expended, affect retention in very different ways. A period of sleep immediately following the end of practice greatly reduces the amount of forgetting as compared with a period of wakefulness (26, p. 369).

In periods of wakefulness, the more similar the interpolated activity to that of the original learning the greater the forgetting and the less the retention of the original material (24, p. 130). This is usually explained in terms of retroactive inhibition (5, p. 260; 24, p. 129; 26, pp. 372-373).

A third group of determinants of retention occurs at the time memory is tested. Degree of retention is a function of the particular test used. Four measures of retention commonly used in verbal learning are recognition, reconstruction,
recall, and relearning (saving). These four were intensively studied in Luh's (23) 1922 study.

The method of recall, used in the present study, minimizes the support the subjects receive from the context of what they have learned. The result is that such recall scores are generally lower than scores received by the other three methods (5, p. 239). Hull (15, p. 164) speaks of the "threshold of response," indicating a behavioral oscillation. An item may be below the threshold of response at one moment and not be available for recall. An everyday example of this is expressed when a person verbalizes, "It's on the tip of my tongue but I can't say it." Another moment and the item may be recalled. Direct recall, providing no contextual clues that would raise the threshold level at a given moment is a less sensitive measure than the other methods mentioned.

The context in which retention is measured is also important. Associations of the learner are formed not only with the items of the task but also between the task and a definite environmental context (26, p. 380). Time of day, physical environment, personality of the experimenter—all these form a background of stimulation with which the performance of a specific task is more or less strongly associated during the practice period.

In addition to the factors associated with the nature of the material to be learned, the interpolated activity between time of learning and time of testing retention, and
the conditions present during the time of retention, a fourth important variable is the nature of the learner. Motivation, drive, fatigue, need state, prior learning or experience, knowledge of results, emotional state, mental ability and age of learner are all important in determining the amount of learning that will occur (1, pp. 115, 229-233; 18, pp. 78-81; 24, p. 172).

Underwood (31) has analyzed individual differences in speed of learning and retention. He found a zero correlation between time to learn and amount recalled but a high positive correlation between time to learn and time to relearn. He showed that very likely the critical difference between fast and slow learners is simply the associative strength of particular items at the end of learning; fast learners acquire more associative strength for particular items. In a given amount of practice, fast learners generally acquire greater associative strength for the things they are learning.

Human learning and remembering is a vast but sparsely cultivated field of study. Bartlett (2) pointed out how difficult it is to capture the everyday process of learning and remembering in a laboratory. Often, when the research is taken out of the laboratory necessary control is lost. Another source of difficulty is the inadequacy of available theory. Selected theories relevant to retention will be discussed in the following section.
Theoretical Considerations

Necessary to any theory of learning and remembering is a conception of the process which permits a person to act on the basis of his past experiences. Ever since Plato compared memory to an impression upon a wax tablet, variations of the idea of a memory "trace" have been the popular way of picturing this process (10). Vigorous criticism came from the Gestalt theorists in the early 1930's. The Gestalt modification of the trace theory reasoned that traces are "subject to two influences: communication to other traces (assimilation) and stresses inherent in the trace itself which are expressed by the Gestalt (configurational) principles, for example, sharpening and leveling in order to achieve maximum simplicity, symmetry and good form" (25, p. 3).

In 1932 Bartlett (2) reported the results of twenty years of research on perceiving, learning and retention in his book Remembering. An outstanding contributor in the field, Bartlett rejected the notion of "trace" and chose to use "schema," a term borrowed from the neurologist Head (12). Schema, Bartlett defined as

... an active organization of past reactions, or of past experiences, which must always be supposed to be operating in any well-adapted organic response. That is, whenever there is any order or regularity of behavior, a particular response is possible only because it is related to other similar responses which have been serially organized, yet which operate, not simply as individual members coming one after another, but as a unitary mass ...
Each new group of experiences . . . persists . . . as constituents of living, momentary settings belonging to the organism or to whatever parts of the organism are concerned in making a response of a given kind, and not as a number of individual events somehow strung together and stored within the organism (2, p. 201).

Bartlett felt however, that certain experiences could be detailed apart from the schema. These individual stimuli were called images. Summarizing, Bartlett writes:

"Remembering is not the re-excitation of innumerable fixed, lifeless and fragmentary traces. It is an imaginative reconstruction, or construction, built out of the reaction of our attitude toward a whole active mass of organized past reactions or experience and to a little outstanding detail which commonly appears in image or in language form (2, p. 213)."

According to Bartlett's theory, recall is not a reproduction of schema; it is an active construction based upon a schema. Though dominant detail—or image—does persist, the major component of the original situation that persists is the attitude which was involved. The attitude, or consciousness, is the major determinant of the way the person reproduces the original situation (2, pp. 207-214).

In 1949 Hebb's (13) book caused a renewal of interest in the neurological correlates of learning. He introduced the terms "cell assembly" and "phase sequence" to explain cortical processes.

"Any frequently repeated, particular stimulation will lead to the slow development of a "cell assembly," a diffuse structure comprising..."
cells in the cortex and diencephalon, capable of acting briefly as a closed system, delivering facilitation to other such systems and usually having a specific motor facilitation. A series of such events constitutes a "phase sequence"—the thought process. Each assembly action may be aroused by a preceding assembly, by a sensory event, or—normally—by both (13, Introduction IX).

Hebb explains the permanence of learning, or memory, by postulating that repeated stimulation of specific receptors leads slowly to the formation of cell assemblies which can act briefly as closed systems after stimulation ceases. A growth process accompanies synaptic activity making the synapse more readily traversed (13, p. 60). Thus Hebb proposes that a reverberatory trace, though transient and unstable, is able to carry the memory until growth can occur at the synaptic juncture (13, p. 62).

Another explanation of memory is given by Miller (5, p. 237). Though not a complete theory, it is given here because of its concern with retention. Miller says that what people remember in the immediate memory span is chunks of things in which information has been encoded in various ways. People recode that which they hear into smaller units. In the memorizing of new material the addition of verbal items in memory is simply the formation of chunks or the reorganization of material into more inclusive chunks. The meditational principle involved in chunk formation is not made clear.

Though no theory of memory serves to wholly explain the process, the views presented agree that there is a constant,
active process occurring and that a degree of permanence is possible.

Original Study

In 1955, Betty Isern conducted a pilot study at the University of Kansas entitled "The Influence of Music Upon the Memory of Mentally Retarded Children" (16). The study was designed to test the effect of music on learning, retention and recall.

Her subjects were 104 children in a state institution for the mentally retarded. Their chronological ages ranged from 7-19 years; IQ's ranged from 27-99. All subjects were considered to have emotional problems, varying from mild to very severe.

Five zoo animal facts were individually presented in song form to all subjects. Recall was tested immediately and, again, three days later. Following the second testing for recall, all subjects were individually told a story of five farm animals. They were then tested for immediate recall and three days later, for recent recall. Six months later, all subjects were tested for remote recall of both the song and the story.

The song and story were equated by five experts according to length, similarity of content, difficulty of text, appeal of content and the number of items to be remembered in each. The experimenter adhered to exact testing procedures for both treatments.
Using the $t$-test of significance, she found the song superior to the story at the time of immediate recall, recent recall and remote recall. Significance was beyond the .001 level of confidence.

She suggested that music increased the vividness, organization and patterning, and affective appeal of the animal facts presented in song to such a degree that the memory of the mentally retarded children was improved. Isern's conclusion that the memory of mentally retarded children in her study was improved because of the introduction of music was based upon her statistical findings and interpretation.

Criticisms

Isern's study was an original one and, as such, was difficult for no pattern had been set that she might follow. However, the following criticisms seem valid and lay open to question the results of her study.

1. All subjects received Treatment A (song), were tested twice and then given Treatment B (story), and tested twice. Four points of difficulty arise:

   a) Is it certain that the subjects were unchanged as a result of Treatment A and were, therefore, the same subjects entering Treatment B as they were when entering Treatment A? Lindquist speaks to this point, saying:

     When a number of treatments are administered in succession to the same subject, the response of the subject to any one treatment is often conditioned by the fact that other treatments have
previously been administered to him ... the criterion measure for any subject following a given treatment may be in part determined by his experience in having been measured previously, quite apart from the effect of the previous treatments themselves ... it is obvious that this design can rarely be employed in learning experiments (21, pp. 160-161).

b) Is the history (i.e., the specific events occurring between the first and any later measurement) controlled in any way? Nearness to holidays, or special occasions, weather conditions, home visits (assuming these are on a regularly scheduled basis) etc. may be different for Treatment A and Treatment B, introducing a variable quite apart from the experiment, itself.

c) Does the testing procedure introduce a confounding variable? Effects of taking one test do influence a second test (4, p. 176). What of the testing-retesting problem with six tests of the same subject? This is not simply a matter of repeated measurements as in a trend analysis. Rather, the first test (Immediate Recall of Song) is directly compared with the third test (Immediate Recall of Story) and the second test (Recent Recall of Song) is directly compared with the fourth test (Recent Recall of Story).

d) What is the likelihood that separate presentations of a song can be presented by the same experimenter in an identical manner—to all 104 subjects? Or, following these first 104 presentations, what is the probability that separate
presentations of a story can be then presented in an identical manner to the same 104 subjects?

The present study seeks to replicate in as far as possible the study outlined. However, specific changes have been made so as to effect a greater experimental control.
CHAPTER I BIBLIOGRAPHY


CHAPTER II

EXPERIMENTAL DESIGN

Statement of Purpose and Hypotheses

The purpose of the present study is to determine if music is effective in increasing the learning and retention of meaningful, verbal material with emotionally disturbed children of normal intelligence.

On the basis of the previous research (3) the following hypotheses were formulated:

1. If facts about animals are presented in song form to emotionally disturbed children designated as highly teachable then their group mean score will be significantly higher than will the group mean score of similar children who receive the same animal facts presented in story form.

2. If facts about animals are presented in song form to emotionally disturbed children designated as low teachable then their group mean score will be significantly higher than will the group mean score of similar children who receive the same animal facts in story form.

3. If the total song group, composed of the high and low teachable subjects receiving the song treatment, are combined and the ensuing group mean found, then that score will be significantly higher than will be the group mean score
found by combining the high and low teachable: subjects receiving the story treatment.

An analysis of variance will be computed for each level of recall: immediate, recent and delayed. The level of significance is to be .05 for each F test.

Subjects

All subjects for the present study were enrolled in a private residential school for emotionally disturbed children. A relatively high tuition rate would indicate that the students come from a higher-than-average socio-economic level although no special effort was made to determine the accuracy of this assumption.

Thirty-two students comprising three classroom groups were chosen to participate in the study. One subject was eliminated because of a marked hearing loss. Other subjects were lost as the experiment progressed due to leaving school, illness, and psychiatric appointments. One group, however, remained completely intact; the table of random numbers was used to reduce this group to the same number of subjects as in the other groups.

Subjects ranged in age from nine years, five months to fifteen years, eight months. The median age was twelve years 9.5 months. The mean was 12.40 years.

Subjects were of normal intelligence. All IQ's were taken from individual administrations of the Wechsler
Intelligence Scale for Children or the Revised Stanford-Binet Intelligence Test. All subjects had been tested within the preceding three-year period; several had been tested more than once. Doubtful IQ scores were checked by re-testing. IQ's ranged from 71-124. The median IQ score was 91.50 while the mean IQ score was found to be 95.58. A summary of these characteristics is found in Table I.

TABLE I
SUMMARY CHART SHOWING RANGE, MEDIAN AND MEAN OF SUBJECTS' AGES AND IQ

<table>
<thead>
<tr>
<th>Subject Characteristic</th>
<th>Range</th>
<th>Median</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>IQ</td>
<td>71-124</td>
<td>91.50</td>
<td>95.58</td>
</tr>
<tr>
<td>Age</td>
<td>9-5 to 15-8</td>
<td>12-9.5</td>
<td>12.40</td>
</tr>
</tbody>
</table>

Emotional disturbances in subjects ranged from slight to moderately severe. Diagnoses ranged from the minimally brain-damaged or mildly neurotic to paranoid schizophrenic. Brain damage was evidenced in some subjects by a motor or perceptual weakness, by marked irritability or hyperactivity in others. A number of subjects received a daily dose of drugs to control extreme behavior.

A greater number of boys were enrolled in the school than girls. In this study, four subjects were girls. Five other girls were a part of the original group but by the time of the final testing had been dropped from the study for one of the reasons cited earlier.
Because the original study had been done with the mentally retarded, it was hypothesized in the present study that music might have a different effect with the less capable student than with the more competent student. IQ's alone were judged not a sufficient way of grouping the available subjects as the presence of emotional difficulties in some subjects interfered with their learning processes. Therefore it was determined that a more global term, i.e. "teachability," might better take all factors into consideration and provide a more reliable base for grouping students.

Three educational staff meetings were held. In the first one, a general discussion brought out the factors felt to be influential in learning. Based on the comments made, a psychological rating scale was developed. Copies of this scale were distributed at the second staff meeting and discussed. A sample of the scale used is found in Appendix A. The personnel directly involved with each student rated that student.

At the third meeting, a group consensus of all teachers working with the students was taken, using the same scale. From the resulting scores for each student, a teachability rating of "High" or "Low" was assigned. IQ scores were then related to the high and low teachability groups. The mean IQ of the high teachable group was 102.7. The mean IQ score for the low teachable group was 84.1. It is interesting to note that teachers making the rating did not refer to IQ's in
any of the three staff meetings held. It is assumed that knowledge of student IQ's did not affect teacher ratings. A description of selected subject characteristics is found in Table II.

### TABLE II

**DESCRIPTION OF SUBJECTS BY SELECTED CHARACTERISTICS**

<table>
<thead>
<tr>
<th>Ss</th>
<th>CA</th>
<th>IQ</th>
<th>Teachability</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>13-10</td>
<td>100</td>
<td>L*</td>
</tr>
<tr>
<td>2</td>
<td>12-2</td>
<td>104</td>
<td>H*</td>
</tr>
<tr>
<td>3</td>
<td>14-2</td>
<td>101</td>
<td>H</td>
</tr>
<tr>
<td>4</td>
<td>14-1</td>
<td>90</td>
<td>H</td>
</tr>
<tr>
<td>5</td>
<td>10-10</td>
<td>123</td>
<td>H</td>
</tr>
<tr>
<td>6</td>
<td>15-2</td>
<td>71</td>
<td>L</td>
</tr>
<tr>
<td>7</td>
<td>10-2</td>
<td>89</td>
<td>L</td>
</tr>
<tr>
<td>8</td>
<td>14-5</td>
<td>93</td>
<td>L</td>
</tr>
<tr>
<td>9</td>
<td>12-10</td>
<td>93</td>
<td>L</td>
</tr>
<tr>
<td>10</td>
<td>9-5</td>
<td>104</td>
<td>L</td>
</tr>
<tr>
<td>11</td>
<td>13-10</td>
<td>89</td>
<td>L</td>
</tr>
<tr>
<td>12</td>
<td>11-4</td>
<td>87</td>
<td>H</td>
</tr>
<tr>
<td>13</td>
<td>12-9</td>
<td>124</td>
<td>H</td>
</tr>
<tr>
<td>14</td>
<td>9-9</td>
<td>75</td>
<td>L</td>
</tr>
<tr>
<td>15</td>
<td>12-0</td>
<td>81</td>
<td>L</td>
</tr>
<tr>
<td>16</td>
<td>11-5</td>
<td>91</td>
<td>L</td>
</tr>
<tr>
<td>17</td>
<td>12-2</td>
<td>107</td>
<td>H</td>
</tr>
<tr>
<td>18</td>
<td>15-8</td>
<td>80</td>
<td>L</td>
</tr>
<tr>
<td>19</td>
<td>15-4</td>
<td>88</td>
<td>H</td>
</tr>
<tr>
<td>20</td>
<td>14-7</td>
<td>71</td>
<td>L</td>
</tr>
<tr>
<td>21</td>
<td>11-3</td>
<td>110</td>
<td>H</td>
</tr>
<tr>
<td>22</td>
<td>13-2</td>
<td>104</td>
<td>H</td>
</tr>
<tr>
<td>23</td>
<td>11-4</td>
<td>91</td>
<td>H</td>
</tr>
<tr>
<td>24</td>
<td>15-6</td>
<td>77</td>
<td>L</td>
</tr>
</tbody>
</table>

*L = Low

*H = High

Subjects designated as highly teachable were randomly assigned (by flipping a coin) to the song or story treatment.
Subjects designated as low teachable were assigned to one of the two treatments in similar fashion. Four groups were thus formed.

Task Variable and Procedure

The original study used two different tasks, each telling of five familiar animals and their characteristic actions. Because the subjects in the present study were of normal intelligence a more difficult task was needed. An original story was composed of unusual animal facts taken from Devoe's *This Fascinating Animal World* (2). In story form it is as follows:

**Little Known Animal Facts**

Have you ever heard of the tiglon? It's half tiger and half lion.

Another strange animal is the goby. It's the smallest fish in the world. It's found in the Phillipines. It takes four of them to make one inch.

Have you ever heard of the cattalo? It's half cow and half buffalo.

The blue whale--that's living now--is the largest animal of all time. Its baby--the biggest in the world--can be half its mother's size.

Dorothy Crocker, music therapist, wrote the music adapting the rhythm to the words. A copy of the music used is found in Appendix C. A piano accompanied the vocalist.
The song and story were performed by the same person, a professional in the field of music and speech. Both the song and story were submitted to tape so that all presentations were identical. The instructions to the subjects were also taped and were as follows:

You are going to hear a song (story) that tells some strange things about animals. All of the facts you will hear are true but they are known to very few people. See how many you can remember. The title of this story is "Little Known Animal Facts."

These instructions immediately preceded the story or song.

The room used for the study was somewhat isolated from the remainder of the school as it was on a higher level than any other classroom. The walls were unadorned, the furniture consisted simply of a window seat, table and chairs. The tape recorder was placed in position on the table; the speaker box was placed so that the experimenter's note-taking would not be distracting to the subject. The same positioning was used each time and though the tape recorder was not used in the recent and delayed testing, it was positioned as in the initial testing procedure.

The order for subject testing and task order was decided in the following manner. The names of all subjects with their task as determined by a coin flip, were written on individual slips of paper and placed in a hat. Drawing these out one at a time gave the order of subject presentation.
All subjects were tested on the same day. Subsequent tests of retention followed as closely as possible the time pattern that was established on the day of the original testing. All testing was stopped fifteen minutes prior to recess or lunch breaks. All of the testing was done during the period of academic work in the school day.

All subjects were tested individually. A uniform greeting and dismissal were followed with each subject.

After the presentation of the taped instructions and story (song), each subject was asked to tell what he remembered about the story (song) he had just heard. A check list, found in Appendix B, was used to tally responses.

Recent recall was taken three days later. The subject was asked to tell what he had heard on the tape three days earlier. Again, the check list was used to tally responses.

Delayed recall was measured two weeks after the presentation of the original tape. The subject was asked to recall what he had heard from the tape recorder two weeks previously. Responses were again recorded on the check list.

The story (song) was divided into "idea" or Babcock units (1, p. 105). Each unit furnished material which was not superfluous or repetitive in its context. Each response was scored as follows: a word or series of words in the reproduction which corresponded to an information unit of the story, even if slightly altered (in tense or by substitution of synonyms) was scored 1; a unit which was altered to a
greater extent but which still preserved the sense of the original received a score of \( \frac{1}{2} \); errors, distortions and units with no correspondence to the original received no score. The sum of these scores for each subject was taken to represent the amount of information accurately reproduced.
CHAPTER II BIBLIOGRAPHY


CHAPTER III

ANALYSIS OF DATA AND DISCUSSION

Analysis of Data

Three separate analyses of variance were computed to test if the differences found were of statistical significance. These will be discussed separately.

Immediate Recall

The first analysis of variance analyzed the data of immediate recall. The variation between the song and story treatments, between the high teachability group and the low teachability group, and the interaction between these two independent variables, were analyzed. Although all F scores were significant above the 20 per cent level, none reached the specified 5 per cent level of significance. Therefore, the data failed to support the hypotheses. In this portion of the experiment, the data indicated no statistically significant difference between the song and story treatments.

No significant difference was found between the scores of the high and low teachability groups and no significant interaction was found between the treatment and teachability variables. A summary chart of this first analysis is found in Table III.
### TABLE III

COMPLETE ANALYSIS OF VARIANCE FOR IMMEDIATE RECALL SCORES

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Sq.</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Over-all Between</td>
<td>(101.21)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Between Song and Story Treatment</td>
<td>42.66</td>
<td>1</td>
<td>42.66</td>
<td>3.50</td>
</tr>
<tr>
<td>Between High and Low Teachability</td>
<td>28.13</td>
<td>1</td>
<td>28.13</td>
<td>2.31</td>
</tr>
<tr>
<td>Interaction Treatment x Teachability</td>
<td>30.42</td>
<td>1</td>
<td>30.42</td>
<td>2.50</td>
</tr>
<tr>
<td>Within Groups (error)</td>
<td>243.75</td>
<td>20</td>
<td>12.19</td>
<td>-</td>
</tr>
<tr>
<td>Total</td>
<td>344.96</td>
<td>23</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

The data analyzed in Table III do not indicate any significant difference between the recall of subjects receiving the song and story treatments or between the recall of high and low teachability groups nor in the interaction between the two.

**Recent Recall**

The second analysis of variance tested the variation found in recall three days after the original testing. Again, the variation between the song and story treatments, between the high and low teachability groupings, and the interaction between treatment and teachability were analyzed. An F score of 4.35 or higher (4, p. 197) was necessary for the differences to be significant at the 5 per cent level of confidence. In
this analysis, two conditions were found to be statistically significant. The differences found between the song and story treatments and between the high and low teachability groups were of statistical significance; that is, they occurred at greater than chance level. Recall of the story was greater than recall of the song. Recall of the high teachability group was greater than the recall of the low teachability group. No significant interaction was found. The hypothesis stating superiority of the song group over the story group was not confirmed; i.e., the story, not the song, facilitated greater recall. The data did support the hypothesis stating superior recall of the high teachability group.

A summary chart of the analysis of recent recall is found in Table IV.

TABLE IV

COMPLETE ANALYSIS OF VARIANCE OF RECENT RECALL SCORES

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Over-all Between</td>
<td>(109.20)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Between Song and Story Treatments</td>
<td>41.34</td>
<td>1</td>
<td>41.34</td>
<td>4.43*</td>
</tr>
<tr>
<td>Between High, Low Teachability</td>
<td>46.76</td>
<td>1</td>
<td>46.76</td>
<td>5.01*</td>
</tr>
<tr>
<td>Interaction: Treatments x Teachability</td>
<td>21.10</td>
<td>1</td>
<td>21.10</td>
<td>2.26</td>
</tr>
<tr>
<td>Within Groups (error)</td>
<td>186.54</td>
<td>20</td>
<td>9.33</td>
<td>-</td>
</tr>
<tr>
<td>Total</td>
<td>295.74</td>
<td>23</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

*Significant at 5 per cent level
Delayed Recall

The third analysis of variance tested the delayed recall scores, taken two weeks after the original presentation. As in the previous two analyses, the various components of the variation found between the song and story treatments, the high and low teachability groups and the interaction between treatment and teachability were analyzed. None of the F scores found were significant. A summary chart of this third analysis is found in Table V.

**TABLE V**

COMPLETE ANALYSIS OF VARIANCE FOR DELAYED RECALL SCORES

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Over-all Between</td>
<td>(56.71)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Song and Story Treatments</td>
<td>10.67</td>
<td>1</td>
<td>10.67</td>
<td>.80</td>
</tr>
<tr>
<td>Between High, Low Teachability</td>
<td>22.04</td>
<td>1</td>
<td>22.04</td>
<td>1.65</td>
</tr>
<tr>
<td>Interaction: Treatments x Teachability</td>
<td>24.00</td>
<td>1</td>
<td>24.00</td>
<td>1.05</td>
</tr>
<tr>
<td>Within Groups (error)</td>
<td>267.25</td>
<td>20</td>
<td>13.36</td>
<td>-</td>
</tr>
<tr>
<td>Total</td>
<td>323.96</td>
<td>23</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Discussion

From the statistical analyses of the preceding section, it would appear that, initially, there were no significant
differences between the recall of the song and story groups. Three days later, the story groups were significantly superior to the song groups. This advantage, however, was lost by the time of the delayed recall, two weeks later. For a possible explanation of these facts the means of the groups are presented in the following manner:

TABLE VI

MEAN SCORES FOR IMMEDIATE, RECENT AND DELAYED RECALL BY GROUPS

<table>
<thead>
<tr>
<th></th>
<th>Immediate Recall</th>
<th>Recent Recall</th>
<th>Delayed Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Song</td>
<td>Story</td>
<td>Song</td>
</tr>
<tr>
<td>HT*</td>
<td>9.67</td>
<td>10.08</td>
<td>7.67</td>
</tr>
<tr>
<td>LT*</td>
<td>5.25</td>
<td>10.17</td>
<td>3.00</td>
</tr>
</tbody>
</table>

*HT = High Teachability  
*LT = Low Teachability

Several factors now become apparent. In the case of the high-song group, low-song group and low-story group the phenomenon of reminiscence occurs. That is, more was recalled at the second or third testing than was recalled at the first testing. Kohler notes this phenomenon and explains it by saying, "Reminiscence, too, is due to the fact that disturbing effects of traces upon other traces tend to subside in the course of time" (2, p. 153).
McGeoch (3, p. 140) discusses reminiscence and points out the necessity of controlling practice or rehearsal. This was not attempted in the present study and the higher scores may only be evidence of rehearsal effects.

Three of the twelve subjects in the total story group demonstrated greater recall at the second or third testing while six of the twelve subjects in the total song group did so. Could it be that melody of the song affords a greater likelihood of rehearsal? If the tune is remembered in whole or part is it not possible that the subject would search his memory for the appropriate words?

In the test of immediate and recent recall, the group means of the story group are slightly higher. By the time of the delayed test, the high-song group has the highest group mean.

The over-all loss of this high-song group was -1.42. The over-all loss of the high-story group was -2.50. This suggests the possibility that material, once learned by song, is more easily recalled than the same facts learned by story. A comparison of the loss factor in the low teachability groups yielded similar evidence. The low-song group over-all loss was -.92. The low-story group over-all loss was -2.50.

A confounding possibility rests with group four, the low-story group. Though the teachers' ratings of these students were low and the mean IQ of this group was the lowest of all four groups, their group mean was slightly higher than the high-story group on two out of three tests. One schizophrenic
boy, subject number twenty-four, had the third highest total recall score of the entire group of subjects. Yet this boy's illness incapacitated him to such a degree that he was unable to achieve in even a small class setting.

One other subject in the low story group exhibited surprising recall. His IQ measured 71 yet his recall scores placed him fifth highest out of the entire group of twenty-four subjects. Table VII presents the scores of each subject in immediate, recent and delayed recall.

The two subjects mentioned may have made an unusual effort to remember the material presented although neither showed a very strong interest in the subject matter involved. Their treatment was identical to that of all other subjects. However, it might be that their sensitivity to the praise "Very good," given to all subjects and the feeling of participation, "Give this note (giving the subject's name and time of leaving) to your teacher that he may know we're on schedule," might have positively affected these boys so that they tried harder than they might have otherwise done. Yet, because of the consistent high recall of this low-song group (in large part attributed to the two subjects mentioned), the differences between the other groups are not clearly demonstrated.

Another phenomena mentioned in the literature as "im-porting" was demonstrated by a number of the younger children in the study. Discussed by Bartlett (l, p. 58), importation
is the invention or adding of material in later recall that was not present in the original. The three youngest in the study, remembering that the general topic of the tape recording was animals, invented a number of animals as recalled material. Two named the common animals as "rabbit, dog, wolf, etc."; the third named unusual animals as "stegosaurus, horned lions, copperhead, rhinocerous, etc."
Another finding of the study was this: children tended to remember what they thought they heard. Those "facts" that the children recalled after just hearing the tape were by and large, the "facts" that they remembered in subsequent recall. Errors made at immediate recall were repeated at the second and third recall. One little boy, having thought he heard that the cattalo was the largest animal in the world, repeated this and added, "But I don't think that's true." At the second and third testing he reiterated the error and again voiced his doubt about its truth.

Most of the students attended to the tape very well. One, however, indicated his attention was somewhere else as he recalled at the time of the second testing, "I don't remember what the tape said but you had been using the tape recorder all morning and it was hot. This little button was red. The tape was white and brown. You wore a blue dress, etc." He was able to recall all of his own comments at the third recall test. This points up the fact that memory is best for those things attended to.

The preceding points may be summed up as follows:

1. There is no significant difference between immediate recall of animal facts taught in song form and those same facts taught in story form among the subjects of this study.

2. There is no significant difference between the immediate recall of animal facts by the low teachability groups
and the high teachability groups. This was in part the result of some factor operating in subjects of the low-story group.

3. A significant difference is found in recent recall, taken three days after the first testing, of animal facts between the song and story groups. Facts were better recalled by the story group at this level.

4. Teachability was also a significant factor at this level of recent recall. Those designated as highly teachable recalled facts to a greater degree than did those rated as low teachable.

5. Reminiscence occurred, particularly in the song group. It was suggested this might be accounted for by the greater likelihood of practice with a song than with a story.

6. Importing occurred in the youngest children of the study.

7. Children tend to remember that which they thought they heard. Material verbalized by a subject, even though in error, is the material that that subject recalls.

8. Attention of the subject is important in determining that which will be learned and later recalled.

9. Based on the over-all loss score, the song groups exhibited a definite advantage, suggesting that music may help the retention of that which is learned, though initial learning may be less than that of the story groups.
CHAPTER III BIBLIOGRAPHY


CHAPTER IV

SUMMARY AND RECOMMENDATIONS

Summary

The effect of music on learning of animal facts and subsequent retention was studied. Twenty-four subjects participated in the study. All were students in a private, residential school for emotionally disturbed children. Ages of subjects ranged from nine years, five months to fifteen years, eight months. All were presumed to have some degree of emotional difficulty, ranging from mild to moderately severe. All were of normal intelligence.

The entire group of subjects was divided into two groups on the basis of teacher ratings. These two divisions were designated as the high teachable group and the low teachable group. Each of these groups was further divided into song and story treatment groups. Assignment of subjects to treatments was on a random basis as was the order in which subjects were tested.

The task utilized for this study was an original story made up of little-known animal facts. Two groups of children (one high teachability group and one low teachability group) were presented the animal facts in story form. The same facts in song form were presented to the remaining two groups of children. As the children were assigned at random and the
task was the same for both, it was assumed that the only
difference between the groups, in terms of task, was that
in one the animal facts were sung; in the other the same
animal facts formed a story. The song, story, and instructions
were recorded by the same person. The use of a tape recorder
was used to insure uniformity of presentation.

The same room was used for testing of all subjects. The
experimenter greeted and dismissed each subject in the same
manner. Time of day was as nearly the same for each subject
at each testing period as possible.

Three tests of recall were taken. Immediate recall was
measured immediately following the presentation of the song
(story). The subject was then asked to tell what he remembered
about the song (story) he had just heard. A check list was
used to tally responses.

Recent recall was measured three days later. The subject
was asked to recall what he had heard from the tape recorder
three days previously. Delayed recall was taken two weeks
from the time of the first testing. The subject was again
asked to recall what he had heard from the tape recorder two
weeks previously.

It was hypothesized that the low-song group would be
significantly superior in immediate, recent and delayed re-
call than would the high-story group. It was also hypothesized
that the low-song group would be significantly superior in
immediate, recent and delayed recall than would the low-story
group. These hypotheses were not supported by the data.
This study was a factorial design. The two independent variables, each varied in two ways, were treatment (song, story) and teachability (high, low).

Three analyses of variance were computed for recall scores at the three different points of time. The analysis of immediate recall data yielded no significant results; therefore, the hypotheses regarding immediate recall were not supported.

The analysis of recent recall data yielded two significant figures. The story groups were found to have demonstrated superior recall to the song groups. Teachability entered as a significant factor with the high teachability groups superior in recall to the low teachability groups. No significant interaction was found. The superiority of the song group to the story group hypothesis was not confirmed. In fact, the evidence tended to support the opposite conclusion, i.e., that the story groups were superior to the song group in recent recall.

The analysis of delayed recall yielded no significant data. Again, the pertinent hypotheses were not supported.

An analysis of individual group means by treatment suggested that some factor(s) in the low-story group confounded the total data. This low-story group, despite a low teacher rating and the lowest mean IQ, was superior in recall to the high-story group two out of three times. Two subjects in this group, one having a severe mental illness and an IQ of 88 and
the other having a measured IQ of 71, received the third and fifth highest scores, respectively, of all twenty-four subjects. A larger number of subjects would have balanced this type of occurrence so that no one group would have been so greatly affected.

The story groups showed a higher total mean at each point in time suggesting that a single presentation of animal facts in story form was more effective in total number of facts recalled than was a song presentation of the same facts.

However, the over-all mean loss (forgetting factor) was 2.34 for the song groups and 5.00 for the story groups. This suggested the possibility that though fewer facts were learned from a single presentation of a song, those facts that were learned were better retained.

The phenomenon of reminiscence occurred in six out of twelve subjects in the song group and in three of the twelve subjects in the story group. This suggests the possibility that music invites rehearsal, thus offering one explanation for the lower loss scores.

Importation was observed in the three youngest subjects.

Recommendations

This study should be repeated with a larger number of subjects. By increasing the number of subjects the likelihood of individual variation greatly affecting group data would be reduced. It could then be determined if retention
is increased by a musical task as opposed to the same task in story form, as this study suggests.

Another way of testing the above problem would be to use the basic design as presented in this study except that the song (story) would be presented as many times as necessary to get one full recall of possible facts from each subject. This would insure equal learning. Retention might then be tested by recall as in the present study. The loss factor could be analyzed to determine if those learning the facts by song retained them longer than did those learning the facts by story.

A third possibility is suggested in the occurrence of reminiscence. This was found in 50 per cent of the song group, and in 25 per cent of the story group. Does music increase the likelihood of rehearsal, resulting in what appears to be reminiscence? Further investigation of this question would be of value.
"Teachability" may be represented by a continuum. Children are in varying degrees "teachable" and may be represented at any point along this continuum. For the purpose of this study, however, rate your students on a one to four point scale according to the following criteria: One (1) represents the highest degree of teachability and describes a student who usually attends, comprehends and retains material presented to him. Four (4), on the opposite end of the scale, represents the lowest degree of teachability and describes a student who has great difficulty in attending, comprehending and/or retaining material presented to him. Two (2) and three (3) represent varying degrees of teachability between these two extremes. A student that seems closer to the four (4) rating than the one (1) rating, would be rated as three (3). A student that seems closer to the one (1) rating than the four (4) rating, would be rated as two (2).
TEACHER RATING SCALE

1. Student usually exhibits:
   
   A. Good attention span (is an attentive listener; is alert, curious and interested).
   
   B. Good comprehension (understands directions readily; can interpret or put into own words material read, heard or experienced; has good conceptual ability).
   
   C. Good retention (retains material learned, exhibits good recall, can apply previous learning to new situations).

2. Student usually exhibits:

3. 

4. Student usually exhibits:

   A. Poor attention span (is easily distracted; easily frustrated, irritable; is inattentive, a poor listener; seems to be preoccupied or a "day dreamer").
   
   B. Poor comprehension (unable to follow directions on own; unable to interpret material read, heard or experienced; needs much repetition and individual explanation; exhibits poor conceptual ability).
   
   C. Poor retention (needs much prompting to elicit previously learned material, exhibits poor recall; has little ability to apply previous learning to new situations; demonstrates little "carry-over").
APPENDIX B

CHECK LIST

Total Facts (27)

Instructions

1. Strange facts about animals
2. True
3. Little known or known to few people

BODY of Song or Story

4. Title, "Little Known Animal Facts"
5. Tiglon
6. Half
7. Tiger
8. Lion
9. Strange animal
10. Goby
11. Smallest in the world
12. Fish
13. Found in the Phillipines
14. Takes 4 = one inch
15. Cattalo
16. Half
17. Cow
18. Buffalo
19. Blue
20. Whale
21. Living now
22. Largest animal
23. Of all time
24. Baby
25. Largest in the world
26. Half
27. Mother's size

Score 

S 

Story - Song 

(Circle One)
APPENDIX C

Have you ever heard of the tiglon? It's half tiger and half lion. Another strange animal is the goby. It's the smallest fish in the world. It's found in the Phillipines. It takes four of them to make one inch. Have you ever heard of the cattalo?

It's half cow and half buffalo. The blue whale that's living now is the largest animal of all time. Its baby the biggest in the world can be half its mother's size.
BIBLIOGRAPHY

Books


**Articles**


Reed, H. B., "Meaning as a Factor in Learning," *Journal of Educational Psychology*, XXIX (1938), 419-430.


Unpublished Materials