TOWARD A DEVELOPMENTAL HYPOTHESIS OF CLUSTERING IN FREE RECALL

Thesis for the Degree of M. A.
MICHIGAN STATE UNIVERSITY

Robert W. Herrmann

1968

ABSTRACT

TOWARD A DEVELOPMENTAL HYPOTHESIS OF
CLUSTERING IN FREE RECALL

by Robert W. Herrmann

Sixty children-- twenty each from the 1st, 3rd and 5th grades of a local elementary school--were tested on a free recall task. The stimulus materials consisted of 20 objects--each falling into one of 4 conceptual categories. The Ss received 5 trials consisting of the presentation of each object in turn for 3 seconds, followed by a 2 minute recall period.

The data was scored for total recall and associative clustering. Analysis of the data indicated that both total recall and clustering increased as a function of trials and grade level. An attempt was made to find strategies other than clustering which might be used to organize the stimulus materials for recall. This attempt failed.

In this paper an attempt was made to reconcile the data obtained in studies of clustering in free recall with Piaget's observations on the development of classification. It was noted that Piaget has defined grouping on the basis of similarities as the basic operation necessary for classificatory behavior and that children did not consistently show this behavior until between the ages of 5½ and 8 years. Rossi, on the other hand, reports above chance clustering in children as young as 3 years old.

The model which seems to fit clustering data best is that suggested in this paper--i.e. a model which posits two types of clustering. The first type or type I clustering is an automatic, unconscious process based on the strength of verbal habits. The critical variable for the occurrence of type I clustering is the inter-item associative strength between same category

items. The greater the inter-item associative strength, the greater the amount of clustering. It was suggested that type I clustering is typical of younger children (first graders and below). Rossi's data on nursery school children suggest that the necessary verbal habit strength builds fairly early in the developmental process.

Type II clustering on the other hand requires the child to ferret out the underlying cluster strategy and use the information to organize the material for recall. Once the strategy is known to the child, this task would seem to be analogous to Piaget's free classification task. It was noted, however, that given the ability to group on the basis of similarities, the ability to perceive the categorical structure of the stimulus list is critical. It would seem likely that great individual differences in this ability exist. This would explain why a greater proportion of the 3rd and 5th graders in this study did not show type II clustering.

It was suggested that each type of clustering has a distinctive pattern of recall associated with it: the type I pattern shows a random cluster here and there among runs of unrelated words, while with type II clustering virtually all words recalled are clustered and there is but one cluster for each category.

The difference in cluster scores for 3rd and 5th graders was found to be less than that for 3rd and 1st graders. Further no 1st graders were observed to show type II clustering, but some 3rd and 5th graders did. These two findings were interpreted to indicate that between 1st and 3rd grades a sort of stage shift in cluster ability occurs--i.e. during this period some children shift from a type I performance to type II clustering.

Robert W. Herrmann

A study using cross-cultural data on clustering in free recall was discussed. It was suggested that though there may be a number of natural mental operations (clustering being one of them), it seems likely that there is a strong cultural influence upon which operations are used within a particular society.

TOWARD A DEVELOPMENTAL HYPOTHESIS OF

CLUSTERING IN FREE RECALL

By

Robert W. Herrmann

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

MASTER OF ARTS

College of Education
Department of Educational Psychology

1968

ACKNOWLEDGMENTS

The author wishes to acknowledge the help and encouragement given by Dr. Joe L. Byers and other members of the faculty of the Department of Educational Psychology in the preparation of this paper. Also thanks are due to my wife Jeanne for typing and editing the manuscript.

TABLE OF CONTENTS

TABLES AND FIGURES

CHAPTER I - INTRODUCTION

One of the first studies to appear in the literature to deal specifically with the phenomenon of conceptual clustering in the free recall of verbal materials was published by Bousfield (Bousfield, 1953). Bousfield had observed in previous studies of free recall that words which were highly associated, or which fell into the same conceptual category tended to be "clustered" or recalled in adjacent serial positions. In an attempt to isolate and quantify clustering, he presented a list of sixty nouns to one-hundred college students at the rate of one word every three seconds. Each of the words on the stimulus list fell into one of four conceptual categories--animals, vegetables, names and professions. Each category contained 15 words. Only one presentation of the list was given, followed by a ten minute recall period in which the Ss were to write as many of the words as they could remember. In spite of the fact that the words were presented in random order, the majority of the Ss tended to cluster nouns of like categories in recall. Thus, the Ss imposed an order on the stimulus material which was not inherent in the presentation.

The purpose of this chapter is threefold: I. to review the available literature on conceptual clustering and total score in free recall with emphasis on the variables which affect them; II. to discuss some of the major hypotheses advanced to account for clustering in free recall; and III. to advance the purposes and hypotheses of the present study.

I. Variables which Affect Clustering and Total Score in Free Recall

Since the appearance of Bousfield's study, many investigators have attempted to isolate the variables which affect the amount of clustering and the total score on free recall tasks. Basically, these variables are of two kinds: 1. stimulus related variables; and 2. subject related variables. In the former classification are included such things as:

1. the number of stimulus list presentations;
2. the number of categories in the list;
3. interitem associative strength;
4. word frequency;
5. blocked versus random presentation;
6. paced versus unpaced presentation;
7. visual versus aural presentation;
8. kind of stimulus material--i.e. use of verbal presentation of list of words versus pictures of objects versus objects or models of objects;
9. use of verbal versus written recall; and
10. use of "cueing" or mediators in the stimulus list.

Subject related variables include the following:

1. chronological age;
2. mental age;
3. grade level; and
4. the cultural background of the Ss.

A. Stimulus Related Variables

1. Number of stimulus presentations - Bousfield and Cohen, using Bousfield's original materials, tested 243 Ss divided into five groups on a free recall task (Bousfield and Cohen, 1953). The first group received a single presentation; the second, two and so on. However, for each group

there was but a single recall period. The words were presented at a speed of one every three seconds. In the groups receiving multiple presentations, there was a 30-second delay between the end of one presentation and the start of the next. Their findings indicated that total recall increased as a function of the number of presentations. The data when plotted showed a negatively accelerating curve as a function of the number of presentations. Total clustering was found to be positively related to the number of presentations. A number of other studies have supported the Bousfield-Cohen results for both total recall and clustering (Cofer, 1959; Cofer and Bruce, 1965; Gerjuoy and Spitz, 1966; Rossi, 1963; Rossi, 1964; Rossi and Rossi, 1965; Sharp et al., 1967). However, each of these studies used the more traditional "trial"--i.e. a recall period after each list presentation--rather than just a single period of recall.

2. Number of categories in stimulus list - Two studies have been done to show the relationship between the number of categories in the stimulus list, and clustering. The first study was done by Bousfield and Cohen, and showed a positive relationship between the number of categories and above chance clustering (Bousfield and Cohen, 1956). They used lists of 40 words with 2, 4 and 8 categories. There was a single presentation and a single recall period. One hundred sixty college students were used as Ss. They reported 50% showing above chance clustering with 2 categories, 74% with 4, and 85% with 8.

Dallet, however, reported a curvilinear relationship between both total recall and clustering, and the number of categories (Dallet, 1964). Using a list of 12 words and 2, 3, 4 and 6 categories, he found that 2 or 3 was the optimal number of categories. He reported a similar relationship using lists of 24 words with 2, 4, 6, 8 and 12 categories. Four was found to be the optimal number with the longer list. The results of the two studies seem to be in conflict. However, since there appeared to be no major difference in the methods used in the two studies (Dallet even drew his stimulus materials from Bousfield's studies), it would appear that the critical factor in determining the optimal number of categories is the number of items on the stimulus list. Thus with a short (12 items) or intermediate list (24), 3 or 4 categories are optimal. With a longer list (40 items) a greater number of categories would maximize performance. It would seem that had Bousfield and Cohen used 10 and 20 categories in addition to the numbers that they used, a curvilinear relationship similar to that reported by Dallet would have been found.

3. Interitem associative strength - Jenkins, Mink and Russell studied the effects of interitem associative strength upon clustering (Jenkins, Mink and Russell, 1958). Their lists consisted of 12 pairs of associates from the Kent-Rosanoff Association Tests. Four frequencies were used from high to low. Their results indicated that the number of clusters tends to increase as a function of the associative frequency. It was concluded that clustering is a monotonic function of the free association strength of the pairs being recalled.

Deese, also using Kent-Rosanoff associations, reported similar results for total recall. Recall varied as a function of the associates to the various Kent-Rosanoff stimuli. He reported a high correlation ($r = 0.88$) between interitem association and total recall.

Similar findings were reported by other investigators for clustering (Bousfield, Cohen and Whitmarsh, 1958; Cofer, 1959; and Wicklund, Palermo and Jenkins, 1965), and for total recall (Bousfield, Cohen and Whitmarsh, 1958).

The effects of interitem similarity on nonsense materials was studied by Stimmel and Stimmel (Stimmel and Stimmel, 1967). Contrary to what one would expect, Ss, presented a low interitem-similarity-list, performed better than a group given high similarity list--with the difference increasing as a function of trials. The high similarity list was constructed from four consonants (F, S, U and X), while the low similarity list from twelve letters in the pattern consonant-vowel-consonant. One might argue that these data indicate that free-recall of nonsense material does not follow the same laws that govern the recall of meaningful material. However, if one looks at the rules for construction of the trigrams, another explanation is suggested. Even though the high similarity list is only composed of four letters (and thus has higher interitem similarity), one can raise the question as to whether this material is as meaningful as that given on the low similarity list. A run of three consonants is a construction which--though not unusual in an English word--cannot occur as a word itself. On the other hand, a CVC combination approximates more closely a real word and thus could well have more "meaning" for the subject. Thus, if the Stimmels had used a CCC or CVC construction on both high and low similarity lists, it would seem likely that their data would support the findings for meaningful material.

4. Word frequency - It is generally the practice to control for either interitem association (as in the studies cited immediately above) or for word frequency of occurrence in making stimulus lists for free recall tasks. The author could find but a single study which treated frequency of occurrence as a variable.

Bousfield and Cohen presented two groups of undergraduate psychology students with lists of 60 nouns each (Bousfield and Cohen,

1955). One list was composed of high frequency and the other of low frequency words drawn from Thorndike-Lorge. The lists were constructed so that the mean frequency of occurrence for any category was the same as the mean frequency for the entire list. They reported that total recall was higher for the group given the high frequency words. In addition, the Ss receiving the high frequency list showed significantly higher clustering than the low frequency group.

5. Order of presentation of stimulus items - Dallet, using lists of 12 and 24 items reported both higher total recall and clustering for Ss receiving a blocked presentation (all items of a given category presented serially) as opposed to Ss receiving a random presentation (no two items in the same category in adjacent serial positions) of the same items (Dallet, 1964). Cofer and Bruce reported similar findings using two 36 item lists with form class used as the dimension along which clustering could take place (Cofer and Bruce, 1965). Each list consisted of 12 nouns, 12 adjectives and 12 verbs with low interitem associative strength. Total recall was greater with blocked presentation than with random. Above chance clustering occurred only in the blocked condition.

Weingartner presented two groups of college students a list of 30 words--15 of which were associates to the word "sleep", and the other 15 to the word "sour" (Weingartner, 1964). One group received a random presentation, the other a constrained (blocked by category) presentation. His findings were consonant with those already cited for blocked versus random order.

Tulving has reported marked differences in the amount of total recall seen as a function of the order of presentation of "unrelated" words

(Tulving, 1965). With one random order, recall might be relatively good, but with another, quite poor. He concluded that a number of orders should be used in verbal learning studies to control for such effects.

6. Paced versus unpaced recall - Ekstrand and Underwood studied the effects of paced versus unpaced recall in free learning (Ekstrand and Underwood, 1963). In the paced condition, the Ss were shown 12 items on a memory drum at a rate of one item per two seconds. At the end of a presentation there was a blank space; then an asterisk would appear; and then after two seconds, another until 12 had appeared. The S was to recall a word every time an asterisk appeared. There were 60 seconds between each trial. In the unpaced condition, the S was allowed an indefinite period to recall all the items he could. Total recall was significantly higher in the unpaced conditions. They did not present materials which could be clustered; hence no data for clustering was available.

7. Visual versus aural presentation - The basic issue here is whether one obtains greater recall with visual presentation (projection of stimulus words on a screen, or presentation with a memory drum) or aural presentation (having the list read by the E). Schulz and Hopkins reported no difference between the two modes on a paced free recall task (Schulz and Hopkins, 1968). Again, clustering was not considered.

8. Kind of stimulus materials - Basically, three kinds of stimulus materials have been used in free recall tasks: 1. word lists, either read or presented visually one word at a time; 2. pictures of objects and 3. objects or models of objects (for example a model car).

Rossi and Rossi using children from two to five years reported significant differences in above chance clustering using the three types of material, with objects yielding the highest scores and words the lowest (Rossi and Rossi, 1965).

Sharp et al. reported identical results using third and sixth graders as Ss (Sharp et al., 1967). They also found significance in the same direction for total recall.

9. Verbal versus written recall - Only a single study was found which compared written and verbal recall. This study was the one cited in the previous section by Sharp et al., 1967, using 3rd and 6th graders as Ss. They reported a significant difference between the two response modes in favor of verbal recall. It would be interesting to test this variable using college students or even high school students as Ss. One might predict that the difference noted by Sharp et al. would diminish as a function of age and in college students perhaps disappear entirely.

10. Presentation of "cueing" or mediators as part of the experimental procedures - A number of studies have appeared in the literature in which attempts were made to assess the effects of presenting "cues" or mediators to the Ss on subsequent recall and clustering.

Cofer presented lists of 40 words -- 5 categories with 8 words each -- to 6 groups (Cofer, 1959). The words used in each category were synonyms. At the beginning of each recall trial the cueing groups were given one word from each category. Cofer's data did not support the notion that clustering and total recall were increased by his cueing procedure.

Rossi used another kind of cueing in a study run on children (Rossi, 1964). In the cueing condition, the category name was substituted for one item in each category--for example, animal for dog. He too reported no significant differences between the cue and no-cue group.

Weingartner presented appropriate, inappropriate or no cues both before and after presentations of a list of 30 words--either associates to the word "sleep" or the word "sour" (Weingartner, 1964). He reported that this cueing in no way affected the number of words recalled, or the number clustered; however, it did produce differences in the items recalled.

Thus, such studies as are available indicate that the effects of giving Ss cues or mediators may be minimal.

B. Subject Related Variables

Most studies of associative clustering in free recall have been primarily concerned with the stimulus related variables already discussed. The four subject related variables to be discussed here are: 1. chronological age; 2. mental age; 3. grade level; and 4. culture of the subject.

1. Chronological age - Rossi tested three groups of children with average chronological ages of 5.1, 8.0 and 11.0 years on a free recall task consisting of 20 items--each falling into one of four categories (Rossi, 1964). Each S was given five trials, with the items being presented in a different random order each time. He reported significant differences in ability to cluster between the three age groups, with the oldest children showing the best performance, the middle group the next best, and the youngest, the worst. He suggested that the relationship between clustering and chronological age is linear.

In a later study he tested children with CAs from 2 to 5 years (Rossi and Rossi, 1965). His data indicated above chance clustering at all age levels. Again, amount of clustering increased as a function of CA.

2. Mental age - Most studies dealing with mental age as a variable have used either a mixed population of mentally retarded children and normal children, or a population of just mental retardates. The latter studies have been eliminated from consideration in this paper.

Rossi tested three groups of normal and retarded children who were matched on the basis of MA (Rossi, 1963). The levels were 4-6, 7-3 and 10-0 years. All groups clustered above chance with the exception of the lowest level retardates. Otherwise, there was no difference noted between normals and retardates. Both clustering and total recall increased as a function of MA.

Gerjuoy and Spitz tested a group consisting of the following classifications:

1. middle grade retardates;
2. high grade retardates;
3. a group of normals equal in mental age to the retardates;
4. a group of normals matched with the retardates on the basis of CA; and
5. a group of college students

(Gerjuoy and Spitz, 1966). The task consisted of 20 words in 4 categories. There were five trials given. The two lower MA groups--both normal and retarded--showed little if any above chance clustering. The normals matched on the basis of CA clustered significantly above chance on trials 4 and 5. The college students clustered above chance on trials 2-5. Thus, clustering varied with MA.

On the other hand, there were no differences between the two lower and two higher MA groups on total recall. No attempt was made to explain this observation. Perhaps there was a ceiling effect across all groups. Gerjuoy and Spitz did not specify the ages of their experimental sample beyond what was presented at the beginning of this review. It seems likely that if the children used were around ten years of age, and the material used was not too difficult, little difference would be observed in total recall across age groups.

3. Grade level - Grade level is considered as a separate variable because in the two studies to be cited, the CA of the Ss is not specified. It is the usual practice when using CA as a variable to specify the mean, range and standard deviation of each age group used. The variance in age within a grade level can be quite large. Because the CA data was not specifically given in these studies, one cannot make the assumption that the groups were homogeneous in respect to age.

Three groups--3rd graders, 4th graders and college sophomores--were studied by Bousfield, Esterson and Whitmarsh (1958). The stimulus material consisted of 25 items falling into five categories. There was only a single trial. Both total recall and clustering were found to increase as a function of grade in school.

Sharp _et al_. tested groups of 3rd and 6th graders on a list of 20 items in four categories (Sharp _et al_., 1967). The Ss received five trials. On both total recall and clustering, the 6th graders were superior to the 3rd graders.

4. Cultural background of the subject - The relevance of cultural background as a variable to be considered in this review is not immediately apparent, and no attempt will be made at this point to demonstrate relevance. This will be done later in the paper.

Very little work has been done in cross-cultural studies of clustering; however, data was obtained from two individuals who have done extensive work on learning tasks with the Kpelles--a tribe of rice farmers in Liberia (Sharp et al., 1967; Coles, 1967).

Cluster studies were run on 1st and 2nd grade students (average CA = 8.0 years and 11.7 years respectively) who were attending school, and illiterate adults (average CA 28.3 years). Stimulus materials consisted of 20 objects (objects from the Kpelle society) falling into four categories. Five trials were run. For total recall there was no age effect. However, a trial effect and presentation effect (verbal presentation of words versus objects) were noted.

In clustering the school children's performance was similar to that of American children with age, trial and presentation effects being noted. However, their cluster score tended to be lower than their American counterparts. The adults, on the other hand, showed an age effect, but no improvement in cluster score across trials, or as a function of the kind of presentation (words or objects). However, with special training, the adults came to show clustering performance more nearly like that of Americans, complete with trial and presentation effects.

Coles attributes the lack of an age difference in total recall and the relative lack of clustering among the illiterate adults, to the high premium which Kpelle society puts on rote learning. In such a society, clustering is not a "normal" strategy for learning a series of objects. On the other hand, for the children, having received some Westernization through schooling, clustering would be a more "normal" method of learning a series of objects. Differences between Kpelle children and American children could be interpreted in part as a cultural artifact, and in part due to differences in the amount of schooling which the two groups had had.

II. Hypotheses to Explain Clustering in Free Recall

Three major hypotheses which have been advanced to account for clustering in free recall will be discussed in this section. These are the hypotheses of Bousfield, Deese and Rossi.

A. Bousfield - Superordinate and Subordinate Constructions

Bousfield based his hypothesis on the concepts of superordinate and subordinate perceptions as described by Hebb (Bousfield and Cohen, 1955). According to this hypothesis, the repeated arousal of a set of related structures will lead to the development of a large structure which will encompass all of the smaller related ones. This large structure is a superordinate structure. The smaller ones are the subordinate structures. If a subordinate structure occurs, it can lead to the perception of the superordinate, which in turn leads to the facilitation of the recall of the subordinates.

In clustering, the words are the subordinates; the categories, the superordinates. For example, if "dog" is presented, it could activate the superordinate "animal", which in turn facilitates the recall of the other subordinates--other animals.

During recall, a S gives words having high habit strength first, then ones of lower and lower strength. The strength of a superordinate structure--i.e. the ability of the structure to facilitate recall of subordinates--varies as a function of the habit strength of the subordinates. According to Bousfield's laws pertaining to the production of associations, at the start of the recall period, high frequency words will be clustered. As recall progresses, the low frequency words will likewise be clustered. As the supply of words nears exhaustion, clustering will fall to a near chance level because only the lowest frequency words are left (Bousfield and Cohen, 1955). The rate of production of associates (clustering) is proportional to the total number of remaining associates. Thus a greater rate of clustering should occur at the beginning of a recall period than at the end (Bousfield and Sedgewick, 1944; Bousfield and Cohen, 1955).

Bousfield's numerous studies on clustering have been directed towards systematically testing this hypothesis and inferences which he has made from it. He is the only one of the three theorists being discussed who has carried out such a systematic program.

B. Deese - The Free Association Hypothesis

Deese's hypothesis of clustering is based on the process of free association (Deese, 1959). He maintains that immediate memory is limited and thus can only hold a few words. Given a few words in immediate memory, the S--due to verbal habits--will respond with the words which are most highly associated with them. If the list is made up with inherent organization (the words on the list are obtained by picking the common free associations to particular words), interitem associations tend to converge on those in the list, and a few restricted items not found on the list. Thus, the higher the interitem associative strength, the greater the degree of convergence, and the greater the amount of clustering.

If, on the other hand, the list is composed of randomly selected items, no such convergence will take place. Categorical intrusions will occur more frequently, and these intrusions will be "popular" or frequent associations to the words on the list.

Recall and clustering, then, depends upon a small core of items in immediate memory, and of strong free associations to them. Recall and clustering are good or poor depending upon the tendency of free associations from items within the list to converge upon other items on the list.

C. Rossi - Miller's Chunking

Rossi, in his studies with children, maintains that clustering in free recall occurs as soon as the development of speech (Rossi and Rossi, 1965). He argues that his studies indicate that clustering is probably a basic activity in recall and can be properly called the "chunking" which is described by Miller. Miller maintains that the human mind can only handle 7 plus or minus 2 "bits of information at a time (Miller, 1956). (A bit in information theory represents a binary decision, thus a true-false question would transmit one bit of information.) In order to get around this inherent limitation on the capacity of the brain to handle input, one is able to take bits and combine them into large units which Miller calls "chunks."

For Rossi, the "bits" are the individual items on the list; and the clusters, the chunks. He maintains that the critical variable in the development of clustering is chronological age (Rossi, 1964; Rossi and Rossi, 1965).

Rossi makes no attempt to account for the process by which clustering improves with CA. Is this improvement simply a function of maturation; or is it more closely related to language development--which in turn correlates highly with CA?

III. Purposes and Hypotheses of the Present Study

In a review of the literature on verbal learning in children, Keppel poses the question as to whether it is valid to extrapolate the results of verbal learning studies on college students to children (Keppel, 1964). He points out that there has been a dearth of work in this area on children.

However, such studies as have been done suggest there is a surprising correspondence between findings with children and adults. Keppel couples his discussion with a plea for further research in the area in the hopes that this issue can be resolved once and for all.

The first purpose of this study was to provide another bit of data on verbal learning in children. To a large extent, this study was a replication of previous ones, however, using children as Ss instead of college students. In line with this purpose a single verbal learning task, namely clustering, was selected and the study designed in such fashion as to be a replication. However, an attempt was made to look at the data in new ways in addition to the standard ones.

A second general purpose of this study derives from the notion that clustering is a learning strategy. If it is indeed a learning strategy, are there alternative ones which are used on this task? Another way of putting it, if a child does not show clustering, are there other ways he can organize the material to facilitate recall? The second general purpose of this study was to attempt to establish the presence or absence of alternative strategies in free recall.

Finally, the third general purpose was to test specific hypotheses which various authors have suggested as implications of their research.

A number of specific hypotheses were to be tested in this study:

1. Total Recall should:

   A. Increase across trials, and

   B. Increase as a function of grade level.

2. Clustering should likewise show trial, grade and category effects regardless of the measure used.

3. Rossi suggests that rapid improvement in clustering ability should occur between ages 5 and 8 years, with a leveling off between 8 years and 11 years (Rossi, 1964). The following hypotheses were advanced to test Rossi's hypothesis:

A. The difference in above chance clustering between 3rd and 5th graders should be less than the difference between 1st and 3rd graders.

B. The proportion of 3rd grade children showing above chance clustering on the first trial should be significantly different from the proportions of 1st and 5th graders; however, by the last trial, there should be no difference in the proportions of 3rd and 5th graders showing above chance clustering, but both should be significantly greater than the proportion of 1st graders clustering.

4. In Rossi's work he reported that serial ordering did not occur in the recall of 2 year olds to any significant extent; and thereafter, the proportion of children showing serial ordering increased as a function of CA (Rossi and Rossi, 1965). He used a stimulus presentation method which minimized serial ordering on later trials, so he just presented data for the first trial. In the present study, the same basic method was used. His data indicate that there is an increase in serial ordering with chronological age which would imply an increase from 1st to 5th grade. however, his suggestion of a rapid increase in clustering between 5 and 8 years could mean that serial ordering will peak at about 5 years, and diminish in favor of clustering from then on. Thus, the following hypotheses regarding serial ordering were advanced:

A. A greater proportion of 1st graders would show serial ordering on the first trial, with 5th graders showing the smallest proportion; and

B. Due to the procedures used to present the stimulus material, the proportion of children using serial ordering would decrease across trials for all grades.

5. A final hypothesis was derived from the work of Bousfield. His data indicate that the "easiest" items are clustered first and then the harder ones (Bousfield and Cohen, 1955). Given a difference between the clustering of the different categories--a significantly greater proportion of initial clusters in recall should be composed of items from the category showing the greatest mean cluster score. Conversely, a greater proportion of the last clusters should be composed of items falling into the category showing the lowest mean cluster score.

CHAPTER II - METHODS

A. Subjects

Sixty children--twenty each from the first, third and fifth grades of a local public elementary school--served as the Ss for this experiment. With the exception of one child (who was of Chinese extraction) all children came from white middle class families. The occupations of the fathers of the Ss are primarily white collar or professional jobs with a fairly high proportion of teachers and college professors represented. Taking the socioeconomic backgrounds of the Ss into account, it could be expected that this group would perform somewhat above average on this task. (The question of how much above average could only be determined by obtaining normative data across all socioeconomic groups.) The children were all within the normal age range for their grade--i.e. there were no children who were repeating a grade.

Prior to beginning the experiment, class lists were obtained and the necessary number of Ss selected randomly. However, all children within each classroom used were run in an effort to eliminate possible confounding effects due to one child's having been selected for the experimental procedure while another had not.

B. Materials

The stimulus materials used consisted of twenty objects, each of which could be placed into one of four categories--five objects to a category (Table 1)

TABLE 1

Stimulus Materials

Clothing	Utensils	Transportation	Animals
shoe	pan	plane	bear
sock	spoon	boat	cat
pants	plate	truck	dog
shirt	glass	car	horse
hat	cup	train	turtle

An attempt was made to select objects which would be familiar to first grade children as well as fifth graders. One difficulty in designing a task to be given across grade levels is the avoidance of floor effects for the younger children, and ceiling effects for the older ones. To this end, the bulk of the objects used were selected from previously used lists as reported in the literature (Gerjuoy and Spitz, 1966 and Sharp *et al.*, 1967).

Objects were selected because--as noted in the introduction--of the three types of stimulus materials used for free recall tasks (objects, pictures and words), objects yield the highest recall. However, one major disadvantage with the use of objects is that it is impossible to control for word (object) frequency and interitem associative strength. This derives from the fact that a correct response to the object "cat" for example, would be scored for "cat", "kitty", "kitty cat", "pussy", and "pussy cat". However, if one were to use words, the only correct response would be "cat".

To control for the possible order of presentation effects--as discussed in the introduction of this paper--ten random presentations of the objects were generated. A random presentation was defined as one in which no two objects in the same category were allowed to occupy adjacent serial positions. This qualification was used because, if objects of the same category were allowed to remain in adjacent positions, there would have been a built-in confound between serial order effects and clustering.

Five of the orders of presentation were then randomly selected for each subject. No sequence of five was used more than once.

C. Procedure

Due to the age of the youngest Ss and the virtual impossibility of using written responses with them, the verbal response mode was used. Using verbal response, in turn, dictates the use of individual testing.

The subject was seated across a table from the experimenter. In between the two a 10-inch plywood screen was placed to hide the stimulus objects from the subject. The experimenter then told the subject:

> "I am trying to learn more about how children learn and remember things. I am going to show you twenty objects. These objects are things that you have probably seen before; in fact many of them you have in your own home. After I show them all to you, I want you to name as many of them as you can. In all I am going to show you the objects five times, and I want you to name them for me five times. You do not have to name them in any special order, but any way which is easiest for you. To help me to remember which objects you have named, I will write them down as you name them. Do you have any questions?"

After responding to questions:

> "Now before we begin, let's practice."

The subject was then shown five practice objects (book, key, belt, fork, and duck), and his responses recorded. Again he was asked if he had any questions. Then he was told that the experiment would begin. After the completion of a trial the experimenter said, "You did very

well. Now let's do it again." After all of the trials were completed, each child was given a Tootsie Pop.

A trial consisted of the presentation of each object for three seconds--for a total of one minute, plus a two minute recall period. For stimulus presentation, the experimenter took each object in turn and held it in front of the plywood screen for the required length of time. A stopwatch was used to time the presentation of each item. As previously stated, each subject received five trials.

CHAPTER III - DATA ANALYSIS AND RESULTS

A. Scoring

Data from each trial was scored in five different ways: 1. total recall; 2. total clustering; 3. above chance clustering; 4. observed-maximum cluster ratio; and 5. serial ordering.

1. Total Recall - Total recall was scored by simply counting the total number of items correctly recalled from the stimulus list, with no reference to the order in which the items were recalled. The maximum total score was 20, or 5 for each category.

2. Total Cluster Score - The total clustering score was given by the expression O(R) = N - K, where N is equal to the number of items in clusters, and K equals the number of clusters (Bousfield, 1953). A cluster is defined as two or more items in the same category which were recalled in adjacent serial positions. The logic for the subtraction of the value "K" from the number of items in clusters is simple. It is impossible to cluster a single item. There must be at least two items in the same category recalled in adjacent serial positions before clustering takes place. Categorical intrusions--i.e. words which fall into one of the categories used, but were not among the stimulus items--were not counted as correct responses. The maximum total cluster score for this experiment was (5 + 5 + 5 + 5) - 4 or 16--a maximum of 4 for each category.

3. Above Chance Cluster - Just in the act of recalling items one would expect Ss to cluster. (Clustering could not occur in this study if the items were recalled in perfect serial order because of the rules used for randomizing the stimulus lists.) Given four categories and given one item recalled--on a purely chance basis, the probability of the next item

recalled falling into the same category is one-fourth. The above chance cluster score represented the total cluster score minus this "expected" or chance cluster value. The expected cluster value is given by the expression: $E(R) = \frac{M_1^2 + M_2^2 + M_3^2 + M_4^2}{N}$ where M_1, M_2, M_3 and M_4 are the number of items recalled from each category, and N is the total number of items recalled (Gerjuoy and Spitz, 1966). The expected cluster value for a single category was found by squaring the number of items recalled in the category, and dividing by the total number of items recalled. For this study, the maximum expected cluster score was 5, or 1.25 per category. The maximum above chance cluster score was 11, or 2.75 for each category.

4. Observed-Maximum Cluster Ratio - One problem with using either total cluster or above chance cluster scores as the basis of a developmental study is that one would expect the absolute cluster score to increase as a function of age as noted in Chapter I. Thus a fourth score was used--the observed-maximum cluster ratio. Given a cluster score, what proportion of the total possible above chance clustering of the items recalled does this represent? Or to put it another way, this measure shows the extent to which Ss clustered the items which they recalled. The observed-maximum cluster ratio is given by the expression $\frac{O(R) - E(R)}{MAX(R) - E(R)}$ where the expression O(R) - E(R) represents the observed above chance clustering score and MAX(R) - E(R) represents the total possible above chance cluster score--given a total recall of N items (Gerjuoy and Spitz, 1966). Since O(R) is always equal to or less than MAX(R), the value of the observed-maximum cluster ratio must fall between 0 and 1.00. In effect, the observed-maximum cluster ratio is a proportion. Given N items recalled a certain maximum cluster score may

be obtained. This ratio is the proportion of that maximum possible score actually observed.

5. Serial Ordering - Serial ordering was scored by taking each pair of words from the recall list and comparing them with the stimulus list. If the pair appeared in contiguous serial positions in the stimulus list, the pair was scored as being serially ordered--with two further qualifications: (1) no reference was made to the numerical positions of the items on either list, and (2) the pair was scored correct if it was recalled in proper order or reversed. Thus if items A and B occurred in serial order on the stimulus list, they would be scored as serially ordered on recall if they appeared as either AB or BA (Rossi, 1967). The total number of correct pairs of items which can be derived from N things in serial fashion equals N-1 where N is the total recall. Thus, the maximum serial order score on this experiment was 19.

In order to attempt to reduce the number of variables to be discussed in this paper, the intercorrelations between total score, total cluster score, above chance cluster score and the observed-maximum cluster ratio were calculated. These data are presented in Table 2.

TABLE 2

Intercorrelations between Total Recall and Cluster Scores

	Total	Total Cluster	Above Chance Cluster	Observed-Maximum Cluster Ratio
Total	1.0000	-	-	-
Total Cluster	0.8608	1.0000	-	-
Above Chance Cluster	0.7747	0.9739	1.0000	-
Observed-Maximum Cluster Ratio	0.6272	0.9105	0.9296	1.0000

Note the high correlation between total cluster and above chance cluster. Because of the great overlap between these variables and because of the above chance score is generally used in the literature, it was decided to eliminate total cluster from further consideration in this paper. The observed-maximum ratio was found to correlate highly with above chance cluster; however, due to reasons to be developed more fully later in this chapter, it was decided to discuss this variable in more detail. Note that all correlations are high positive--with the exception of the correlation between total recall and the observed-maximum cluster ratio which could be called moderately high.

In Table 3 the intercorrelations between the several variables for each grade level are given. Note that there is a slight decrease in the relationship between total recall and both total and above chance cluster as grade level increases. The relationship between total cluster and observed-maximum cluster ratio, on the other hand, increases with grade level. To determine whether there were differences between these correlations, tests of the difference between correlations were run using Fisher's r to Z transformation (Hays, 1963). No significant differences were noted at the 0.05 level. It can be seen, then, that the intercorrelations between the variables were constant across grade levels.

In addition to the above, the following measures were obtained: 1. the proportion of children at each grade level showing significant above chance clustering on each trial; and 2. the proportion of each first cluster and last cluster on the recall lists falling into each category.

B. Data Analysis and Results

The data for total recall were analysed using a 3 way mixed model

TABLE 3

Intercorrelations between Total Recall and Cluster Scores for each Grade Level

Grade 1

	Total	Total Cluster	Above Chance Cluster	Observed-Maximum Cluster Ratio
Total	1.0000	-	-	-
Total Cluster	0.8931	1.0000	-	-
Above Chance Cluster	0.8421	0.9803	1.0000	-
Observed-Maximum Cluster Ratio	0.5386	0.8106	0.8716	1.0000

Grade 3

	Total	Total Cluster	Above Chance Cluster	Observed-Maximum Cluster Ratio
Total	1.0000	-	-	-
Total Cluster	0.8282	1.0000	-	-
Above Chance Cluster	0.7396	0.9821	1.0000	-
Observed-Maximum Cluster Ratio	0.4753	0.8533	0.8793	1.0000

Grade 5

	Total	Total Cluster	Above Chance Cluster	Observed-Maximum Cluster Ratio
Total	1.0000	-	-	-
Total Cluster	0.7508	1.0000	-	-
Above Chance Cluster	0.6948	0.9775	1.0000	-
Observed-Maximum Cluster Ratio	0.5069	0.9346	0.9484	1.0000

repeated measures analysis of variance. The degrees of freedom, computational formulas and expected mean squares were generated by use of the Millman-Glass Rules of Thumb (Millman and Glass, 1968). The main effects for the analysis were grades, categories and trials. Due to the use of two repeated measures--i.e. trials and categories--there were four error terms for the analysis. Subjects was treated as a random factor nested within grades.

The analysis is summarized in Table 4. Note that the three main effects--grades, categories and trials--were significant at the 0.05 level. In addition, a grades by categories interaction was found.

The proportion of variance accounted for by each of the significant main effects and the one significant interaction are given in the last column. These values were computed using omega squared index of the strength of association as outlined by Hays (1963). Note that 22.5% of the variance was accounted for by the grades and trials effects, with minimal proportions accounted for by the categories effect and the grades X categories interaction.

The means and standard deviations for the three main effects and the grades X categories interaction are presented in Table 5. Note that with the 3rd and 5th grades, the means for utensils and animals are higher than those for transportation. With the first grade, transportation is higher than the other two. All grades show highest recall with the clothing category.

The grades by categories interaction is shown graphically in Figure I. Note that there is a greater similarity between the pattern of recall shown by 3rd and 5th graders, than between either one of them and 1st graders.

TABLE 4

ANOVA Summary Table for Total Recall

Source	d.f.	S.S.	M.S.	F	ω^2
Between Ss					
Grades	2	183.6467	91.8233	15.1074*	0.0961
Error (S:G)	57	346.4500	6.0780	-	-
Within Ss					
Trials	4	230.7883	57.6970	81.3550*	0.1289
Grades X Trials	8	9.4117	1.1764	1.6587	-
Error (ST:G)	228	161.7000	0.7092	-	-
Categories	3	30.1633	10.0544	8.7941*	0.0155
Grades X Categories	6	19.4267	3.2377	2.8318*	0.0070
Error (SC:G)	171	195.5100	1.1433	-	-
Categories X Trials	12	6.9451	0.5787	0.7017	-
Grades X Categories X Trials	24	29.8149	1.2422	1.5062	-
Error (CST:G)	684	564.1400	0.8247	-	-
TOTAL	1199	1777.9967			

*Significant at the 0.05 level

TABLE 5

Means and Standard Deviations Total Recall

Grades	Mean	S.D.	Trials	Mean	S.D.
1	58.00	12.30	1	10.48	2.38
3	70.20	9.45	2	13.25	3.18
5	76.90	11.14	3	14.35	3.29
			4	14.87	3.34
			5	15.45	3.17

Categories

	Mean	S.D.
Clothing	18.38	3.95
Utensils	17.10	3.92
Animals	16.43	4.52
Transportation	16.45	2.99

Grades X Categories Interaction

	Grade 1		Grade 3		Grade 5	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Clothing	15.20	3.81	19.55	3.12	20.35	3.00
Utensils	14.75	3.33	16.75	4.22	19.65	3.01
Animals	13.00	4.96	16.90	3.53	19.25	3.30
Transportation	14.90	3.42	16.50	2.77	17.70	3.82

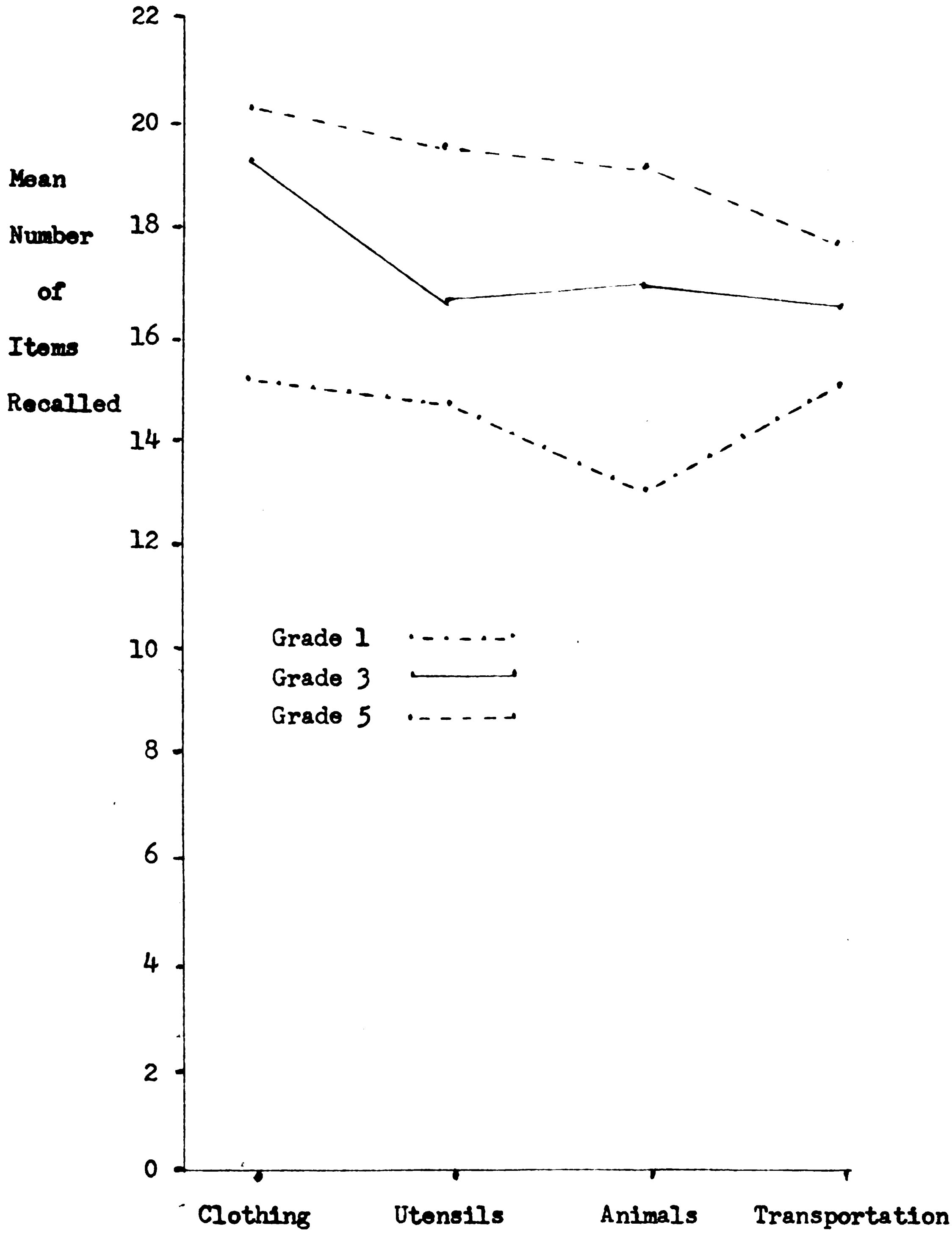

Figure 1. Grades X Categories Interaction - Total Recall

As with total recall, the data for above chance clustering were analysed using a three-way analysis of variance of exactly the same design. The same three main effects were analysed. Table 6 presents a summary of the analysis. Note that the three main effects--grades, trials and categories--were significant. No interactions were noted. The omega squared values given in the last column of the table indicate that the three main effects account for approximately 16.2% of the variance. Note that the values for both the grades effect and categories effect are similar in magnitude to those obtained in the analysis for total recall. However, the magnitude of the omega squared for trials is less than half that obtained for total recall.

In Table 7 one may find the means and standard deviations for the main effects. Note the differences in standard deviation between 1st and 5th graders. In the next chapter this observation will be discussed more fully.

The observed-maximum cluster ratio was analysed using a two-way repeated measures mixed model analysis of variance. The two factors used were trials and grades with subjects used as a random factor nested within grades. A summary of this analysis is presented in Table 8. Note that there was a significant grade effect, but the trial effect was not. The trial effect approached significance however (F observed = 2.33, F critical = 2.37). The value for omega squared (10.62) obtained for the grades effect is approximately the same as that observed for grades in the two preceding analyses.

In Table 9 the means and standard deviations for the grade main effect are presented. Note that again the standard deviation for 5th graders is greater than for the other two grades.

TABLE 6

ANOVA Summary Table for Above Chance Clustering

Source	d.f.	S.S.	M.S.	F	ω^2
Between Ss					
Grades	2	84.1598	42.0799	9.8769*	0.0967
Error (S:G)	57	242.8432	4.2604	-	-
Within Ss					
Trials	4	45.3953	11.3488	21.0631*	0.0555
Grades X Trials	8	7.6922	0.9615	1.7845	-
Error (ST:G)	228	122.8486	0.5388	-	-
Categories	3	9.2219	3.0739	6.4727*	0.0100
Grades X Categories	6	3.2937	0.5489	1.1558	-
Error (CS:G)	171	81.2210	0.4749	-	-
Categories X Trials	12	2.4520	0.2043	0.8208	-
Grades X Categories X Trials	24	7.8467	0.3269	1.3133	-
Error (CST:G)	684	170.2703	0.2489	-	-
TOTAL	1199	777.2447			

*Significant at the 0.05 level

TABLE 7

Means and Standard Deviations Above Chance Cluster

Grades	Mean	S.D.	Trials	Mean	S.D.
1	9.64	5.21	1	2.00	1.62
3	17.68	7.87	2	3.02	2.16
5	22.52	12.91	3	3.53	2.65
			4	3.80	2.94
			5	4.21	2.95

Categories	Mean	S.D.
Clothing	4.87	3.14
Utensils	4.04	3.20
Animals	3.79	2.93
Transportation	3.81	2.71

TABLE 8

ANOVA Summary Table for Observed-Maximum Cluster Ratio

Source	d.f.	S.S.	M.S.	F	ω^2
Between Ss					
Grades	2	3.6312	1.8156	3.16*	0.1062
Error (S:G)	57	13.3004	0.2333	-	-
Within Ss					
Trials	4	0.4922	0.1230	2.33	-
Trials X Grades	8	0.0971	0.0121	1.94	-
Error (ST:G)	228	12.0422	0.0528	-	-
TOTAL	299	29.5631			

*Significant at the 0.05 level

TABLE 9

Means and Standard Deviations for Observed-Maximum Cluster Ratio

Grades	Mean	S.D.
1	1.44	0.87
3	2.30	0.86
5	2.77	1.35

Serial ordering was scored for all Ss on all trials. There was a total of 259 instances of serial ordering noted--78 for 1st grade, 93 for 3rd grade and 88 for 5th grade. Considering that there were 60 Ss and 5 trials each--a total of 300 trials--that would average out to be less than one instance of serial ordering per trial. It is difficult to see how this could be above an expected level of serial ordering. Thus, no analysis was done with the serial ordering data.

Earlier in the paper it was suggested that an appropriate way to test Rossi's notion that there is a greater gain in clustering ability from 5 - 8 years than from 8 to 11 years old would be to compare the difference between clustering in 1st and 3rd graders, and 3rd and 5th graders. It was suggested that if his notion were true, the difference between 1st and 3rd graders should be greater than that between 3rd and 5th graders. To test this hypothesis, the Scheffé method of post hoc comparisons was used (Hays, 1963). The linear contrasts $(\bar{X}_3 - \bar{X}_1) + (\bar{X}_3 - \bar{X}_5)$ can be simplified to $\bar{X}_1 - 2\bar{X}_3 + \bar{X}_5$. In this form of linear contrast the relative weights for each mean are 1, -2 and 1.

It was thought that it would be interesting to see if total recall paralleled clustering on this contrast. The data for both comparisons are shown in Table 10. The first line of each comparison gives the value of the difference being tested. The second line gives the Scheffé gap--or the confidence limits associated with each test. Note that in neither case does the range of values given by the difference between the number being tested and the Scheffé value include zero. Thus, for both total recall and clustering, the difference between mean performance of 1st and 3rd graders is greater than that between 3rd and 5th graders.

TABLE 10

Post Hoc Comparisons Grades for Total Recall and Above Chance Cluster

Total Recall

$$\hat{\Psi}_g = \bar{X}_1 - 2\bar{X}_3 + \bar{X}_5 = 5.50^*$$

$$5.50 - 3.49 \leq \hat{\Psi}_g \leq 5.50 + 3.49$$

*significant at the 0.05 level

Above Chance Cluster

$$\hat{\Psi}'_g = \bar{X}_1 - 2\bar{X}_3 + \bar{X}_5 = 3.20^*$$

$$3.20 - 3.02 \leq \hat{\Psi}'_g \leq 3.20 + 3.02$$

*significant at the 0.05 level

Pairwise post hoc comparisons using the Scheffé method were carried out on the mean score for each category for both total recall and clustering (Hays, 1963). The results of these comparisons are presented in Table 11. At the left of each comparison are the first three categories, and the means associated with them. At the top are categories two through four, and below them, their respective means. The other values are the differences between the means of two categories--for example, the value below 17.10 (utensils) and to the right of 18.38 (clothing) is the difference between these two means. Since the comparisons are pairwise--and hence each mean receives the same weight in each contrast--only a single confidence interval must be calculated. This interval is given below each table. Note

TABLE 11

Post Hoc Comparisons - Categories - for Both Total Recall and Above Chance Cluster

Total Recall

		Utensils	Animals	Transportation
		17.10	16.43	16.45
Clothing	18.38	1.28*	1.95*	1.93*
Utensils	17.10	-	0.67	0.65
Animals	16.43	-	-	-0.02

$$\hat{\psi}'_g - 0.9486 \leq \psi'_g \leq \hat{\psi}'_g + 0.9486$$

*significant at the 0.05 level

Above Chance Cluster

		Utensils	Animals	Transportation
		4.04	3.79	3.81
Clothing	4.87	0.83*	1.08*	1.06*
Utensils	4.04	-	0.25	0.23
Animals	3.79	-	-	-0.02

$$\hat{\psi}'_g - 0.6079 \leq \psi'_g \leq \hat{\psi}'_g + 0.6079$$

*significant at the 0.05 level

that for both total recall and clustering, the clothing mean is significantly higher than the other three category means. No other contrasts show significance.

With the scoring methods for determining above chance clustering used in this study, one finds that often the total above chance clustering for a single subject on a single trial is less than one. Thus, one is left with the problem of defining how much clustering is significantly above chance clustering. To determine this, tables of binomial distribution were used (Burrington, 1953). The total cluster score and total recall score were determined for each subject on each trial. Then--for example--given a total cluster score of 10, total recall of 16, and the probability of an item being clustered, 0.25; what is the probability of 10 out of 16 items being clustered? Using the binomial tables the probability of 10 of 16 or more than 10 of 16 being clustered by chance--given the expected probability of 0.25--can be determined. If the table value for such a proportion was 0.05 or less, the cluster score was considered to be significantly above chance.

The proportion of children in each grade showing above chance clustering on each trial was determined in this fashion. These data appear in Table 12.

TABLE 12

Proportion of each Grade Showing Significant
Above Chance Clustering on each Trial

	Trials				
	1	2	3	4	5
Grade 1	0.25	0.25	0.50	0.45	0.55
Grade 3	0.35	0.60	0.75	0.75	0.80
Grade 5	0.55	0.75	0.80	0.70	0.80

The differences between the proportions of each grade on each trial were tested by the $\acute{U}_0$ method of multiple comparisons among a number of independent binomial populations (Marasculio, 1966). The only difference which was significant was that between 5th grade and 1st grade on the second trial. Without going into any great detail, the reason for the lack of further significant differences was that the sample size in each grade was not large enough. With a sample size half again as large as that which was used (30 instead of 20) significant comparisons would have been noted on four of the five trials. The data presented in this study--though not significant--suggest that a replication with larger sample sizes is desirable.

Table 13 shows the proportion of first clusters and the proportion of last clusters falling into each of the four categories. The differences between these proportions were tested using $\acute{U}_0$ and no significant differences were noted. This was to be expected due to the minimal differences shown in clustering for each category.

TABLE 13

Proportion of 1st and Last Clusters
Falling into each Category

1st Clusters		Last Clusters	
Clothing	0.2986	Clothing	0.2955
Utensils	0.2919	Utensils	0.2199
Animals	0.2348	Animals	0.2714
Transportation . .	0.1747	Transportation . . .	0.2132

CHAPTER IV - DISCUSSION

In assessing the results obtained in this study for total recall and above chance clustering, one might say that there were no surprises:

1. Both total recall and clustering increased as a function of grades;
2. There was an increase in both as a function of trials; and
3. There were slight differences noted in performance on the different categories.

With the observed-maximum cluster ratio a grade effect was seen, but no trial effect. This latter effect, however, did approach significance (F observed = 2.33, F critical = 2.37). The trial means increased from 0.37 to 0.49--representing an increase of 12% in the above chance clustering relative to the amount which could have taken place if all items recalled had been clustered. These data are in substantial conflict with Gerjuoy and Spitz who reported a decrease in this ratio from 0.72 on the first trial to 0.26 on the fifth trial (Gerjuoy and Spitz, 1966). They made no attempts to explain this result and there seems to be little reason to dwell on it here. Suffice it to say that their overall findings were in conflict with other studies on clustering in children. They reported no consistent clustering before age 14.

The proportions, though not showing significant differences, reveal the same two trends: 1. a greater number of the older children showing above chance clustering, and 2. an increase in the number of children in each grade showing clustering as a function of the number of trials.

In summary, then, above chance clustering increases as a function of trials, and of grade level.

The search for alternative methods of recalling the stimulus material was negative. Serial ordering was of no consequence. In addition, there was no systematic tendency to cluster certain items first and others last. This latter finding would seem to be related to the relatively small differences noted in category performance. Thus, in this study, if order were to be imposed on the stimulus materials, it was clustering.

One might argue for a mediation hypothesis to account for the lack of other types of ordering: the children were given a mediator by the presentation of the practice items (four of the five fell into the task categories). However, this does not seem likely due to the negative results for studies using mediators which were cited previously in this paper.

Clustering has been considered a classificatory task by various workers in the field of verbal learning, especially Rossi (Rossi and Rossi, 1965). According to his data, children begin to cluster significantly above chance virtually as soon as they begin to talk. If one defines classification as the ability to group objects on the basis of observed similarities, one runs into problems with Rossi's interpretation of the operations involved in clustering.

Piaget, for example, maintains that true classificatory behavior does not appear until sometime after eight years (Flavell, 1963). His results--based on experiments with free classification tasks (i.e. the child is given objects of the same and different classes and subclasses and told to place those which are similar together)--indicate that a child of $2\frac{1}{2}$ to 5 years will not place the objects into a hierarchy of superordinate and

subordinate constructions based on the similarities between objects. Rather the material is organized into figural collections which are not logical classes, but complex figures, some parts of which can be based on similarity. The sorting is typically done in bit by bit, and somewhat haphazard fashion with no overall plan. The criterion for grouping objects shifts frequently as more and more objects are added. It is not until age $5\frac{1}{2}$ to 8 years that the figural collections give way to non-figural groupings (groupings based on similarities between objects).

The findings of Piaget and Rossi seem to be in conflict but only if one considers the same operations necessary for the completion of each task. It is suggested that clustering in free recall may occur by two different operations:

1. Clustering may occur by virtue of the strength of verbal habits; or
2. It may occur as a result of the subject observing and utilizing similarities between items on the stimulus list to organize the material for recall.

Further, it is suggested that the clustering observed in Rossi's nursery school subjects and in the younger children used in this study is clustering of the first type, while some of the older children in the present study showed type two clustering.

With the development of language, associations are built up between words such that given word A and told to give an association to it, word B will often be given. This process is a relatively unconscious, automatic one. For example, if one were to ask a number of persons for an associate to the word "shoe", the author would be willing to wager that a large proportion would name "sock". However, the mere association of these two words does not in any way imply knowledge of the fact that they fall into the category "clothing".

Two of the major hypotheses of clustering summarized in the introduction of this paper emphasize the role of verbal habits as the basis for clustering (Bousfield's and Deese's). It was also noted that by manipulating the inter-item associative strengths of words on lists, one could cause the amount of observed clustering to vary considerably. Thus, one may conclude that clustering on the basis of verbal habits does occur.

For such clustering to occur, the strength of association between two items on the list must be sufficiently strong so that they automatically tend to be recalled together. With repeated trials of the same materials, this tendency would be strengthened and items of lesser associative strength would be added, yielding a trials effect. Thus one has clustering without necessarily using the knowledge that the items being clustered fall into the same category.

Type II clustering, in contrast to type I, is not an unconscious, automatic process. It requires that the S recognize that certain items within the list fall into the same category (are similar) and that he consciously organizes the material for recall on the basis of these categories. To the extent that this organizing by category can be considered a form of grouping on the basis of similarities, type II clustering can be considered to be analogous to Piaget's second stage of the development of classificatory behavior.

There is a reason why one would not expect to see all older Ss to show type II clustering. This reason pertains to the instructions given the Ss at the start of the task. On Piaget's task, the children are told to group objects on the basis of similarity. In contrast, on a free recall task the subject is left to his own designs. He must ferret out the most efficient way of dealing with the problem at hand. It seems likely that one could find marked individual differences in the ability of Ss at the same age or grade level to grasp the key to the underlying list organization. Thus, though a child might be able to group on the basis of similarity quite well, if he were to fail to recognize the cluster strategy, he would be unable to show type II clustering. Such individual differences would have tended to reduce the proportion of 3rd and 5th graders showing type II clustering in this study.

Analysis of the data indicated that there was less difference in the cluster performance of 3rd and 5th graders than 3rd and 1st graders. It seems likely that this finding relates to the fact that no 1st graders showed type II clustering, but some 3rd graders did. This "spurt" could represent a sort of stage shift in cluster performance.

Rossi considers clustering an example of Miller's chunking. The advantages of taking 20 objects, which have been presented randomly, and reorganizing them into four groups of five items each--or chunking them--are obvious. However, it also seems obvious that if one were able to consciously employ chunking, its effect would be greater. Not only would the subject show greater clustering, but total recall would tend to increase also.

The patterns of recall used by the Ss in this study offer a clue as to whether in fact the two hypothesized types of clustering exist. Two patterns of recall were noted. In the first, clusters seem to appear

at random in the recall of the material. There would be a run of unrelated words, a cluster, some more unrelated words and so on. Frequently there were two clusters of objects from the same category on the same recall trial. This pattern would seem to be associated with type I clustering. In the second pattern, however, by the second or third trial almost all words recalled were clustered. The clusters tended to be larger (i.e. more items per cluster), and there was but a single cluster per category. In addition, with this pattern of recall, the number of items recalled tended to be higher than with the first pattern.

These criteria are certainly not sufficient to demonstrate--in and of themselves--the two types of clustering hypothesized, but strongly suggest that they do occur. It seems likely from the foregoing discussion that two factors must be taken into account in any definitive study--i.e. individual differences in the ability to find the underlying category structure in the stimulus list, and the ability to group on the basis of similarities. Further research in this area should center around attempting to identify the individual difference variables critical on this type of task. Lezotte has suggested that field dependence versus field independence may be one such variable (Lezotte, 1968). He suggests that the field dependent subject (as measured by an embedded figures test) would show higher clustering on a free recall task--the assumption being that such a subject would be more likely to perceive the underlying categorical structure of the materials. The field independent subject, on the other hand, would likely overlook this structure and show type I clustering.

Two further hypotheses are suggested:

1. If one were to instruct the Ss prior to beginning a task as to the

use of clustering on the task, one would expect the older Ss to show a greater relative gain in their cluster scores.

2. The higher the inter-item associative strength the greater should be the proportion of Ss showing type II clustering.

The logic behind the first hypothesis is clear. Explaining the nature of the task eliminates one of the two operations hypothesized as necessary for type II clustering.

If indeed the other operation is grouping on the basis of similarity a la Piaget, one would not expect younger Ss (age 6 and below) to perform this operation even after having it explained to them. On the other hand, the older Ss, being able to perform the grouping operation, and having had the second operation explained to them would be able to utilize this knowledge on the task. Note that in the mediation studies cited earlier, the mediators were presented subtly in contrast to the direct approach suggested here.

A second facet of the ability to grasp the cluster strategy could well be related to inter-item associative strength. Irrespective of the field dependence--independence dimension, one would expect that the higher the relationship (association) between items, the easier such a relationship would be to see. High inter-item associative strength would tend, in part, to compensate for the field dependence-independence dimension; and some of the field independent Ss would show type II clustering. On the other hand, decreasing inter-item associative strength would decrease the probability of grasping the strategy.

At the higher associative strengths the difference in proportions of field dependent and field independent Ss showing type II clustering would likely be smaller than with lower associative strengths.

As was pointed out in the previous chapter, the standard deviation for the above chance clustering grades effect increased across grades, with the first graders showing the lowest, and the fifth graders the highest. This finding can be interpreted in light of the two level hypothesis. With first graders, few if any of the Ss would be expected to show the second type of clustering. One would expect, therefore, that their scores would tend to be more homogeneous than those of the other two grades. With third graders, a few would show type II clustering, and this would increase the range and distribution of scores. With fifth graders there would be even more showing type II and again, the distribution of scores would broaden yielding a higher standard deviation.

Rossi has suggested that clustering is a "natural" mental operation. This implies that clustering is not a cultural bound phenomenon and that it should be seen in all cultures.

One of the subject related factors affecting clustering, which was mentioned in Chapter I, was that of the cultural background of the subject. In the studies with the Kpelle tribe which were cited, it was pointed out that the illiterate adults of the tribe--while showing some clustering--did not show an increased cluster score across trials. In contrast, the children of the tribe who were attending schools did show a trial effect. Cole pointed out that the tribe stresses rote or serial learning in tasks where numbers of items or names must be learned. With a culturally ingrained strategy for approaching a task of this type, one would hardly expect them to show a high degree of clustering. Such clustering as was shown could be attributed to verbal habits rather than detected similarities.

On the other hand, the school children presumably are taught to de-emphasize this kind of rote learning. They would be less likely to approach a free recall task with the culturally ingrained set of the adults, and hence would be more likely to show performances of clustering which increased across trials. Such clustering among the adults would be seen as errors--since the ideal is perfect serial ordering.

When the adults were shown that an alternative strategy existed, they were able to cluster more, and the trial effect was noted.

While Cole's data does not disprove Rossi's suggestion that clustering is a natural mental operation, it does imply that if such operations do exist, it is possible for a given culture to pick and choose which ones it will emphasize. Thus one finds a tribe like the Kpelles who have chosen to emphasize serial ordering on list or serial type tasks. Further, the failure to find the employment of strategies other than clustering in this study could be indicative of such cultural selectivity.

Educational Implications

Drawing specific inferences from a study of this kind and applying them to the classroom setting is difficult. The prime reason for such difficulty is the artificiality of the laboratory experiment. This artificiality manifests itself not only in the nature of the environment in which the experiment is run, but equally in the task which is to be learned. The classroom is a society in miniature and consequently activities within it are performed in a social context. The first effect of the laboratory setting is to remove this context as nearly as possible. Further, while in the classroom a child is rewarded or punished according to the perceived merits of his performance--the rewards and punishments being meted out by both teacher and peer group--in most learning studies, children are rewarded for their participation as opposed to their performance. Often in this study the experimenter was forewarned about the horrendous performance to be expected from such-and-such a child, and was pleasantly surprised to find that this particular child performed better than average on the task. Thus, the laboratory experiment does tell us something about learning and the capabilities of children at various age levels, but it also removes probably one of the most potent factors at work in the classroom situation--i.e. the social context in which classroom learning takes place.

As was noted already, a further source of artificiality derives from the nature of the task itself. Certainly one cannot compare the learning of a list of paired associates or the learning of the materials used in this task with such things as learning to read or doing arithmetic, etc. In our attempts to gain more experimental control we have been forced to simplify many of the tasks we use to the extent that it is improbable that they reflect the tasks of the classroom at all. Thus, most often we are left with data from which only general implications for education may be drawn rather than a specific prescription for teaching Johnny how to read. Two such implications may be drawn from this study:

1. On the surface at least, previous data in this area would indicate that children as young as 3 years old can classify objects. They are presented objects in random fashion and then on verbal recall those which are highly associated are recalled together. It is very simple; it is classification. But is it? When one compares free recall studies with Piaget's work, one has what Jung would have called a "clang" association. They just don't fit together. However, note that on free recall tasks a verbal response is required, while Piaget eliminates the overt verbal response in favor of a motor response. With elimination of the verbal response,

classificatory behavior in 3 year olds vanishes and stays vanished until 2 or 3 years later. Piaget has cautioned against using purely verbal performance as a criterion for judging the cognitive capabilities of a child. The task used in this study could well confirm his suspicions. This study, and others reviewed in the introduction of this paper, would seem to indicate that verbal habits are built up relatively early in the developmental process. Thus, what seems to be the performance of a mental operation may merely be a reflection of the strength of these verbal habits. A truer test of the ability of a child to perform a specific operation is to see whether he can perform it on demand--as in Piaget's studies. Thus, the first implication of this study is that data obtained from verbal learning tasks should be interpreted cautiously, especially when one is trying to infer the level of cognitive development in younger children.

The second implication of this study pertains to individual differences in learning. It was pointed out that the major portion of the observed variance was probably due to individual differences. These differences not only pertain to such things as variability in the strength of verbal habits, but also in differences in the approaches used to learn the task. If, for example, the field dependent-field independent variable is shown to be related to type I and type II clustering, what is being measured in part by the cluster task is a difference in the amount and kind of information being drawn from the stimulus field. Such a finding would argue strongly for differences in instruction--i.e. the field independent child would need instruction in which all elements of a task are laid out in detail and are relatively well organized. On the other hand

a field dependent child would require less explicit structure and less organization, for he is able to uncover and build the structure himself to a greater degree.

The second implication of this study, then, is that even on a relatively simple verbal learning task, individual differences would seem to be the greatest source of variability. Much more research is needed in the area. However, in spite of the little we know about individual difference variables, it seems more and more untenable to attempt to impose a universal curriculum and universal instructional techniques on all children. Individualized instruction must become a reality, not a theory.

Summary

In this paper an attempt was made to reconcile the data obtained in studies of clustering in free recall with Piaget's observations on the development of classification. It was noted that Piaget has defined grouping on the basis of similarities as the basic operation necessary for classificatory behavior and that children did not consistently show this behavior until between the ages of $5\frac{1}{2}$ and 8 years. Rossi, on the other hand, reports above chance clustering in children as young as 3 years old.

The model which seems to fit clustering data best is that suggested in this paper--i.e. a model which posits two types of clustering. The first type or type I clustering is an automatic, unconscious process based on the strength of verbal habits. The critical variable for the occurrence of type I clustering is the inter-item associative strength between same category items. The greater the inter-item associative strength, the greater the amount of clustering. It was suggested that type I clustering is typical

of younger children (first graders and below). Rossi's data on nursery school children suggest that the necessary verbal habit strength builds fairly early in the developmental process.

Type II clustering on the other hand requires the child to ferret out the underlying cluster strategy and use the information to organize the material for recall. Once the strategy is known to the child, this task would seem to be analogous to Piaget's free classification task. It was noted, however, that given the ability to group on the basis of similarities, the ability to perceive the categorical structure of the stimulus list is critical. It would seem likely that great individual differences in this ability exist. This would explain why a greater proportion of the 3rd and 5th graders in this study did not show type II clustering.

It was suggested that each type of clustering has a distinctive pattern of recall associated with it: the type I pattern shows a random cluster here and there among runs of unrelated words, while with type II clustering virtually all words recalled are clustered and there is but one cluster for each category.

The difference in cluster scores for 3rd and 5th graders was found to be less than that for 3rd and 1st graders. Further no 1st graders were observed to show type II clustering, but some 3rd and 5th graders did. These two findings were interpreted to indicate that between 1st and 3rd grades a sort of stage shift in cluster ability occurs--i.e. during this period some children shift from a type I performance to type II clustering.

A study using cross-cultural data on clustering in free recall was discussed. It was suggested that though there may be a number of natural mental operations (clustering being one of them), it seems likely that there is a strong cultural influence upon which operations are used within a particular society.

BIBLIOGRAPHY

Bousfield, W. A. The occurrence of clustering in the recall of randomly arranged associates. J. Gen. Psych. 1953, 49, 229-240.

Bousfield, W. A. and B. H. Cohen. The effects of reinforcement on the occurrence of clustering in randomly arranged associates. J. of Psych. 1953, 36, 67-81.

Bousfield, W. A. and B. H. Cohen. The occurrence of clustering in recall of randomly arranged words of different frequencies--of--usage. J. Gen. Psych. 1955, 52, 83-95.

Bousfield, W. A. and B. H. Cohen. Clustering in recall as a function of the number of word-categories in stimulus-word lists. J. Gen. Psych. 1956, 54, 95-106.

Bousfield, W. A., B. H. Cohen and G. A. Whitmarsh. Associative clustering in the recall of words of different taxonomic frequencies of occurrence. Psych. Rpts. 1958, 4, 39-44.

Bousfield, W. A., J. Esterson and G. A. Whitmarsh. A study of developmental changes in conceptual and perceptual associative clustering. J. Genetic Psych. 1958, 92, 95-102.

Bousfield, W. A. and C. H. W. Sedgewick. An analysis of sequences of restricted associative responses. J. Gen. Psych. 1944, 30, 149-165.

Burrington, R. S. and D. C. May. Handbook of Probability and Statistics with Tables. Sandusky, Ohio: Handbook Publishers, 1953.

Cofer, C. N. A study of clustering in free recall based on synonyms. J. Gen. Psych. 1959, 60, 3-10.

Cofer, C. N. and D. R. Bruce. Form-class as the basis for clustering in the free recall of non-associated words. J. Verb. Learn. Verb. Behav. 1965, 4, 386-389.

Cole, M. Personal communication with the author.

Dallett, K. M. Number of categories and category information in free recall. J. Exp. Psych. 1964, 68, 1-12.

Deese, J. Influence of inter-item associative strength upon immediate free recall. Psych. Rpts. 1959, 5, 302-312.

Ekstrand, B. and B. J. Underwood. Paced versus unpaced recall in free learning. J. Verb. Learn. Verb. Behav. 1963, 2, 288-290.

Flavell, J. H. The Developmental Psychology of Jean Piaget. New York: Van Nostrand Co. Inc., 1963.

Gerjuoy, I. R. and H. H. Spitz. Associative clustering in free recall: Intellectual and developmental variables. Amer. J. Ment. Defic. 1966, 70, 918-927.

Hays, W. L. Statistics for Psychologists. New York: Holt, Rinehart and Winston, 1963.

Jenkins, J. J., W. D. Mink and W. A. Russell. Associative clustering as a function of verbal association strength. Psych. Rpts. 1958, 4, 127-136.

Keppel, G. Verbal learning in children. Psych. Bull. 1964, 61, 63-80.

Lezotte, L. W. Personal communication with the author.

Marasculio, L. A. Large sample multiple comparisons. Psych. Bull. 1966, 70, 918-927.

Miller, G. A. The magic number seven, plus or minus two: some limits on our capacity for processing information. Psych. Rev. 1956, 63, 81-96.

Millman, J. and G. D. Glass. Rules of thumb for writing the ANOVA table. J. Educ. Meas. 1967, 4, 41-51.

Rossi, E. R. Associative clustering in normal and retarded children. Amer. J. Ment. Defic. 1963, 67, 691-699.

Rossi, E. R. Development of classificatory behavior. Child Dev. 1964, 35, 137-142.

Rossi, E. R. and S. I. Rossi. Concept utilization, serial order and recall in nursery school children. Child Dev. 1965, 36, 771-777.

Rossi, S. I. and M. C. Witrock. Clustering versus serial ordering in recall by four-year-old children. Child Dev. 1967, 38, 1139-1142.

Schultz, R. W. and R. H. Hopkins. Presentation mode and meaningfulness as variables in several verbal learning tasks. J. Verb. Learn. Verb. Behav. 1968, 7, 1-13.

Sharp, D. et al. A cross cultural study of clustering in free recall. Paper presented at the Annual Meeting of the Psychonomic Society, 1967.

Stimmel, D. T. and N. S. Stimmel. Free recall as a function of intra-list similarity and order of presentation constraints. Psych. Rpts. 1967, 21, 541-544.

Tulving, E. The effect of the order of presentation on learning of "unrelated" words. Psychonomic Science. 1965, 3, 337-338.

Weingartner, H. The free recall of sets of associatively related words. J. Verb. Learn. Verb. Behav. 1964, 3, 6-10.

Wicklund, D. A., D. S. Palermo and J. J. Jenkins. Associative clustering in the recall of children as a function of verbal association strength. J. Exp. Child Psych. 1965, 2, 58-66.

Degree of M.A.
1968

Herrmann, Robert W.