8111341

BRANDT, MARY ELIZABETH

EFFECTS OF TYPICALITY ON RECALL AND CLUSTERING IN A FREE RECALL TASK: A DEVELOPMENTAL STUDY

University of Hawaii PH.D. 1980

University Microfilms International 300 N. Zeeb Road, Ann Arbor, MI 48106

EFFECTS OF TYPICALITY ON RECALL AND CLUSTERING IN A FREE RECALL TASK: A DEVELOPMENTAL STUDY

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE
UNIVERSITY OF HAWAII IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN PSYCHOLOGY

DECEMBER 1980

by

Mary Elizabeth Brandt

Dissertation Committee:

Thomas Ciborowski, Chariman
Jerry Boucher
Ronald Johnson
Anthony Marsella
Paul Pedersen

ACKNOWLEDGMENTS

A research proposal becomes an actuality only with some crucial assistance and cooperation. Therefore, I would like to thank Jerry Boucher and Tom Ciborowski who offered ongoing encouragement, direction, and friendship. In addition, I am grateful to: David Fung for writing the computer program to calculate Z scores; to Lyn Walters for assisting me with numerous details and especially for her competent tallying of responses; to Geoff White for introducing me to the concept of category prototypes; and to Dorothy Frietas.

I am also indebted to those good friends and colleagues who read the initial version of the dissertation and gave helpful editorial comments and suggestions, particularly to Fran Lyons who read the whole thing.

To the principal, staff, and children of Maryknoll Elementary School for their good will and cooperation, goes a special note of thanks.

Finally, an acknowledgment means to state that one has received and I have received more than I can express from my family, George and Daniel. To them I give my deepest love and appreciation.

ABSTRACT

This study was designed to explore the effects of semantic typicality of category members upon first and third graders' recall and clustering in a multi-trial free recall task. In Experiment 1, members of 5 superordinate categories were elicited directly from 32 first grade and 33 third grade children using a category item production task. Size, diversity, and agreement among the category items were examined by age and category. Children of both age groups exhibited large, diverse, and non-idiosyncratic category conceptual schemes and no age differences occurred in number, diversity, and extent of agreement among category members first retrieved from long term memory. From among the age-appropriate category members, prototypical and nonprototypical items were selected to form all prototypical, all nonprototypical, and mixed (i.e., prototypical and non-prototypical) recall lists for the three list conditions of the free recall task used in Experiment 2. Sixty children (30 first and 30 third graders of the same ages, background, and school as the children in Experiment 1) were alternately assigned to one of the three age-approriate list conditions of Experiment 2. The items in each list condition were randomly ordered for each of the five free recall trials with the restriction that no two items from the same category were adjacent and that each trial began and ended with different items. Order of presentation of the five random-ized lists was constant across children within each list

condition. The number of items recalled and the extent of category clustering were examined. There were no significant age differences in recall and clustering with the all prototypical lists and the all nonprototypical lists. First grade children recalled as many items as third grade children and clustered their recall to the same extent as third grade children with these two types of lists. For both ages, the all prototypical list condition resulted in the highest recall and highest clustering of the three list conditions. In the all nonprototypical list condition, neither age group categorically clustered their recall. In the mixed condition, the well established age-related differences in recall and clustering occurred. Third graders recalled significantly more items and tended to have higher cluster scores, and significantly higher scores on the last two trials, than first graders. Examination of repetitions and intrusions in recall within the mixed list condition indicated that the older children used more effective retrieval cues than the younger children, while the younger children had almost twice as many poor retrieval cues than the older children.

By manipulating semantic typicality of age-appropriate categories and items, age-differences can be either produced or eliminated. These results suggest that: (1) Past free recall studies which reported significant age differences may have used mixed typicality lists and thus may not have tapped the full range of young children's competency;

(2) Production deficiency does not appear to be a child characteristic or an age-related characteristic. Rather production deficiency more aptly describes the joint functional outcome of the type of information to be recalled and the ease (or difficulty) of subsequent retrieval demands; and (3) Further work which concentrates on retrieval strategies and their interaction with typicality and age may contribute to our understanding of memory development.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

I. INTRODUCTION

Categorization and memory are basic cognitive functions of all people. Within psychology, development of these two functions and their relationship are research areas of considerable interest. Free recall research has explored the use of category structure in memory across age groups. However, any study of category use is influenced by the investigator's prior concept of what a category is. In turn, the investigator's concept of a category has been dominated by the traditional notion of category structure (Rosch, 1977). This paper proposes that a recent category model, the "family resemblance" model, challenging the traditional view has important implications for free recall design and interpretation of developmental results.

The first section of this paper contrasts the traditional and family resemblance models and presents evidence supporting three distinguishing structural features of the family resemblance model: Prototypes, degrees of membership, and indefinite boundaries. The influence of these features on dependent cognitive variables are presented. Developmental free recall studies are briefly reviewed and the implications of the family resemblance model for free recall research are discussed in the second section. The final section of the paper presents a study which explored the effects of the family resemblance model on children's free recall performance.

The Nature of Categories

The traditional Aristotelian nature of categories has dominated the perspective and treatment of categories in psychology. Categories have been treated as logical, clearly bounded entities, membership in which is determined by means of arbitrary, formal, and discrete sets of critical features (Rosch, 1975a). All items possessing a certain set of critical features have a full and equal degree of category membership (Kempton, 1978; Rosch, 1975a; Rosch & Mervis, 1975). Such a conception of category structure (definite boundaries, defining features, and equal membership) characterizes artificial categories in the standard concept identification paradigm (Bourne, 1969). In this paradigm, the experimenter determines the criteria for category membership such as color and form. A specific set of critical features, e.g., "red" and "triangle," then comprises item membership and defines the boundary of one category while another set, e.g., "blue" and "circle," performs the same function for another category. Therefore, all red triangles (regardless of size) are equal members of the former category which is distinct and clearly separated from the latter or "blue-circle" category.

The attempt to impose such criteria on various types of categories "marks virtually all definitions in the tradition of Western reason (Rosch, 1978, p. 35)." These attempts can be understood in terms of the generally recognized principle or function of categorization (Mischel, 1979). Referred to as cognitive economy (Mischel, 1979; Rosch, 1978), this principle contains:

> . . .the almost common-sense notion that, as an organism, what one wishes to gain from one's categories is a great deal of information about the environment while conserving finite resources as much as possible. To categorize a stimulus means to consider it, for purposes of that categorization, not only equivalent to other stimuli in the same category but also different from stimuli not in that category. On the one hand, it would appear to the organism's advantage to have as many properties as possible predictable from knowing any one property, a principle that would lead to formation of large numbers of categories with as fine discriminations between categories as possible. On the other hand, one purpose of categorization is to reduce the infinite differences among stimuli to behaviorally and cognitively usable proportions. It is to the organism's advantage not to differentiate one stimulus from others when that differentiation is irrelevant to the purposes at hand (Rosch, 1978, pp. 28-29).

Recent work on internal category structure (Dougherty, 1978; Kay & McDaniel, 1975; Kempton, 1978; Rosch, 1973, 1975a, 1978; Rosch & Mervis, 1975; Rosch, Simpson & Miller, 1976; Smith & Medin, 1980; Uyeda & Mandler, 1980) has provided an alternative to the traditional view of categories. This alternative treatment is also consistent with the principle of cognitive economy and at the same time, seems to map and more accurately represent category perceptions.

Alternative View of Categories

The alternative view of the nature of categories differs from the traditional in the treatment of category membership. Rosch (1973, 1975a, 1977, 1978) maintains that many categories are continuous, having overlapping areas and an internal structure "in which members are ordered according to the degree to which they are judged good examples (typical) of the category (Rosch, Simpson, & Miller, 1976; p. 491)." The "core meaning" of a category is derived from the most typical items. Thus, items within categories possess varying degrees of membership and are "considered differentially representative of the meaning of the category terms (Rosch, 1975a, p. 194)."

Such a conception of internal structure is mathematically formalized in terms of what Zadeh (1965, 1971) and Kempton (1978) call fuzzy set theory. Kay and McDaniel (1975) illustrated the notion of fuzzy sets:

> Gourmet (like many other words) seems to denote something very like a set, except that individuals appear to have different degrees of membership. Charles may be more of a gourmet than Harry and less of a gourmet than Anne. . . .Zadeh has constructed the notion of fuzzy set to formalize this sort of intuition. A fuzzy set A is defined by a characteristic function f_A which assigns to every individual x in the domain under consideration a number $f_A(x)$ between 0 and 1 inclusive, which is the degree of membership of x in A. For example, letting "f_G" symbolize the characteristic function of the fuzzy set gourmet, perhaps $f_G(\text{Harry}) = .4$, $f_G(\text{Charles}) = .7$, and $f_G(\text{Anne}) = .9$. If so, the inequalities given above in words are satisfied: $f_G(\text{Charles}) > f_G(\text{Harry})$ and $f_G(\text{Charles}) < f_G(\text{Anne})$ (p. 5).

Ordered degrees of category membership, an alternative view of category structure, was first demonstrated in the color domain. Considerable evidence suggests that color category membership is not solely determined by sets of critical features; nor are color categories non-overlapping entities. Rather, a color category is represented by a prototype or the clearest, best exemplar, surrounded by colors of decreasing similarity to the prototype and of decreasing degree of category membership (Berlin & Kay, 1969; Heider, 1971, 1972; Heider & Oliver, 1972; Kay & McDaniel, 1978; Rosch, 1975b). The difficulties people experience in making boundary judgments reflect the "fuzzy" nature of color category membership (Kay & McDaniel, 1978). Rosch extended this alternative conception of category structure to other types of categories. Specifically, her work explored whether prototype structure is applicable to semantic classifications of common objects such as birds, furniture, and clothes. She found that subjects can make meaningful judgments about the degree to which items are good or poor members of such categories (Rosch, 1973; 1975a; Rosch & Mervis, 1975). For instance, the items robin and sparrow were considered the best exemplars of the meaning of the bird category (i.e., prototypical members) with turkey and duck less representative or prototypical of bird (Rosch, 1975a; Rosch & Mervis, 1975). Such judgments had high agreement and were found reliable under changes of instructions, items, (Ribs, Shoben & Smith, 1973; Rosch, 1975a; Rosch & Mervis, 1975), and categories (Uyeda & Mandler, 1980).

Given the empirical evidence for the prototypicality structure of natural language categories, Rosch (1977; 1978) proceeded to ask two majors questions: What principles lead to formation of category prototypes; and what cognitive processing variables, if any, are affected by typicality structure?

Basis of Prototypes

It may be argued that certain linguistic variables underlie typicality or "goodness-of-example" norms.

> Typical items are claimed to be (a) the most frequent (generally the most familiar) items, or (b) the items that are experienced with greatest frequency in conjunction with the category name (associative strength). (Rosch, Simpson & Miller, 1976, p. 492).

However, several lines of research suggest that these interpretations are not adequate accounts of typicality. Measures of word frequency are not correlated with degree of typicality (i.e., goodness-of-example norms) of natural language categories (Mervis, Catlin, & Rosch, 1976; Uyeda & Mandler, 1980). The correlations were uniformly low positive to negative, none of which were significant. In a series of experiments in which frequency and nature of experience with category items were controlled, typicality structure remained an important variable in cognitive processing of categories (Rosch, 1975c; Rosch & Mervis, 1975;

Rosch, et al., 1976b). For instance, in Rosch, et al. (1976b) second experiment, the frequency of item presentation in three artificial categories (dot patterns, stick figures, and letter strings) was altered so that the less typical items were presented more frequently than the typical. In all cases, the effects of structural manipulation (i.e., analogues of typicality gradients -- prototypes and distance from the prototypes -- of semantic categories) remained. These results cannot be explained by a frequency interpretation (Rosch, et al., 1976b).

In addition, the limited evidence that exists concerning co-occurrence of items with category names does not support a frequency interpretation. Rosch, et al. (1976) summarized this evidence:

> There is no evidence for such co-occurrence frequency differences in child language protocols (Rosch, et al., 1976a). And the sentence contexts of the superordinate category names (Kučera & Francis, 1967) for which typicality norms are available (Rosch, 1975a) exhibit an <u>inverse</u> relationship between typicality ratings and co-occurrence with the category name. That is, of the small percentage (19 percent) of sentences in the Kučera and Francis corpus in which member and category names both occurred, 83 percent of the member names were items in the lowest quartile of the typicality norms. Thus, it appears that in speech of mothers and young children and in written English, names of superordinate categories tend to occur in place of, rather than co-occur with, the names of typical members of the category. Finally, persistence of the belief that frequency is the basis of typicality may itself be an effect of typicality; it has been demonstrated that subjects overestimate the frequency of typical category members relative to atypical members (Rosch, unpublished, 1976), just as they overestimate the

> probability of cognitively available and representative events . . .(p. 501).

Rosch and associates (Rosch & Mervis, 1975; Rosch et al., 1976b) explored whether Wittgenstein's (1953) structural notion of family resemblance governs the formation of category prototypes and gradients of category membership. "A family relationship consists of a set of items of the form AB, BC, CD, DE. That is, each item has at least one, and probably several, elements in common with one or more items, but no, or few, elements are common to all items (Rosch & Mervis, 1975; p. 575)." Thus, categories can be viewed as networks of overlapping attributes. Prototypical members of a category would be those items with the strongest family resemblance to (have the largest numbers of attributes which overlap those of) other members of the category.

Rosch and Mervis (1975) found that the more prototypical a category member was rated, the more attributes it had in common with other members of the category and the fewer attributes in common with members of contrasting categories. The most prototypical members also have more attributes in common with each other than do the least prototypical items. This was demonstrated

> . . .for natural language superordinate categories, (e.g., animal) for natural language basic-level categories (e.g., dog), and for artificial categories in which the definition of attributes and the amount of experience with items was completely specified and controlled (Rosch, 1978; p. 37).

If items are seen as similar to each other in proportion to the number of shared attributes, Rosch and Mervis (1975) reasoned, then the prototypical items should fall in the center of the semantic space defined by proximity scaling of category items. For all noun-type natural language categories tested, the category name and the most prototypical items appeared as the most central in the scaling solution regardless of the number of dimensions or rotations used (Rips, Shoben, & Smith, 1973; Rosch & Mervis, 1975). For example, robin (prototypical) and bird were more centrally located and closer together in semantic space than chicken (atypical) and bird.

The findings as a whole support the following conclusions: (a) Prototypical members are perceived as the most representative of the meaning of the category and as the most distinctive from members of other categories; and (b) family resemblance structure characterizes natural language categories and appears as the major basis for prototype formation (Rosch, 1978).

Variables Influenced by Typicality Structure

In terms of the effects on psychological variables, typicality structure seems to be pervasive. Using reaction time as a measure of speed of processing information, Rosch (1973) and others (Rips, Shoben & Smith, 1974) showed that adults' verification times for statements such as, "A dog (category item) is an animal (category label)", was shorter for prototypical items than for nonprototypical ones. When factors other than typicality structure, such as frequency, were controlled, the same relationship between verification time and typicality was found (Rosch, et al., 1976b). Typicality predicts verification time.

Priming or producing a certain expectation in subjects by presenting a category name in advance of a pair of category members facilitated responses of "same" to pairs of prototypical members of the category but inhibited "same" responses to atypical members in a category matching task (Rosch, 1975a, 1975b; Rosch, et al., 1976b). Thus, from hearing the category name, people tend to generate a representation of typical but not atypical category members.

Typicality influences the order of item output and the probability of producing a category item when asked to do so (Mervis, et al., 1976; Rosch, 1975; Rosch, et al., 1976b). Again, when frequency was controlled, the most prototypical members of a category were the first and most frequently given (Rosch, et al., 1976b). These findings suggest that aspects of storage, retrieval or category search may be influenced by prototypicality.

In addition, ease of item learning (Rosch, 1973, 1975c; Rosch & Mervis, 1975; Rosch, et al., 1976b) and speed of classification of items after learning (Rosch & Mervis, 1975; Rosch, et al., 1976b) was a function of structural typicality. The more prototypical items were learned before and with fewer errors than the less typical items. "The typicality effects were obtained both when frequency of items was equated and when rates of learning were equated by inverting the relation between frequency and typicality (Rosch, et al., 1976b, p. 491)."

With young children, Anglin (1976 cited in Rosch, 1978) found that good exemplars of categories were learned before poor exemplars. Rosch (1973) also found that 10-year-old children verified prototypical members as belonging to a category faster than less typical ones and the difference in verification time between good and poor members was far more extreme for these children than for adults. Moreover, actual "errors" occurred with the less typical members, such as answering "false" to "A robin is an animal."

Convergent research with children concerning classification behavior and development of natural semantic categories supports the empirical validity of Rosch's (1977) structural typicality model. Saltz, Soller & Sigel (1972) had children sort and classify a variety of category items. They found that the youngest children (kindergartners) applied the adult-selected superordinate label to only some "core" items. A core item, in this study, was defined as that item which 75 percent of the children across all ages (kindergartners, third, and sixth graders) chose as a category instance. If an item was chosen as a category instance by most of the kindergartners, almost inevitably it was also chosen as a category instance by almost all the older children. That is, it was a core item by definition. For example, dress and shirt, selected for the clothes category by almost all the 5- to 6-year-olds, were also selected for the clothes category by almost all the older children. However, the reverse was not true. Some of the core items were clearly only "core" items for the older children, such as duck and turkey for "animal". It appears that category concepts were used by the youngest group to refer primarily to only a subset of the adult selected exemplars.

Based on items common to Saltz, et al's. (1972) study and those rated by fourth grade children for typicality (i.e., goodness-of-example) in another study (Carson & Abrahamson, 1976) Saltz, et al's "core" items were prototypical exemplars while their "non-core" items were rated as atypical. These combined findings suggest that there may be a developmental progression from first considering prototypical items as category members followed by greater willingness to include less typical items as members.

If the less typical items are viewed as nonmembers or marginal members of categories, what effect would this have on a Piagetian class-inclusion task? Carson and Abrahamson (1976) explored this question by systematically varying the typicality of majority and minority subclasses of class-inclusion questions. Four question types were produced: Typical-typical, typical-atypical, atypical-typical, and atypical-atypical. For example, when children were presented with photographs of 5 flies and 3 horses (atypical-typical animal items respectively) and were asked, "Are there more flies or more animals?", a representative response was that there were more flies because the horses are the only animals present. A real dilemma for the children occurred when presented with typical-atypical question types, e.g., 5 dogs (typical) and 3 bees (atypical) with the question, "Are there more dogs or more animals?". The most frequent response was "Same". The children seemed to consider only "dog" as a proper exemplar of animal and therefore there were the same number of animals as there were dogs. Furthermore, children's comments indicated that they were using prototypical attributes as criteria for category membership: "Well, a bee is like an animal because it's sort of furry, but you can't really call it an animal with all those legs and things (p. 1189)."

The use of atypical items significantly impaired performance. For all categories and all ages (7-,8-,9-, and 10-year-olds) performance on the atypical questions was significantly worse than on questions using only typical items. Among children who correctly answered typical-typical questions (classified as "operational"), performance on atypical questions was "worse for children unwilling to apply the adult superordinate label to atypical instances (p. 1189)." Carson and Abrahamson concluded that "semantic typicality appears to be a performance variable capable of masking underlying logical competence...which should be controlled for in tasks involving natural language categories (p. 1189)."

Rosch (1978; Rosch, et al., 1976b) also proposed that typicality structure should be acknowledged in studies of concept formation, models of the nature of categories and theories of semantic memory. To date, the sphere of developmental research has been untouched by Rosch's pioneering work. With the exception of Carson and Abrahamson's (1976) study, the potential impact of typicality structure of natural language categories remains unexplored in any developmental cognitive context. The following section addresses one developmental research area, free recall, which may benefit from such an exploration.

Implications For Free Recall

Rosch's (1977) model predicts that the ease (or difficulty) of a task involving natural language categories is a function of category item typicality. Hence, item typicality should have important implications for children's performance on a standard free recall task, a basic and commonly used tool for investigating the development of memory and mnemonic organizational processes (Jablonski, 1974; Scribner, 1974). Briefly, this task involves random presentation[1] of a list of items potentially categorizable into a priori experimenter-determined categories. After presentation of the same list to children of various ages, recall is "free". That is, the children may recall the items in any order they wish. The dependent measures are the number of items correctly recalled and the degree of clustering, where clustering refers to input-output order discrepancies due to the relationships among the items. In this context categorical clustering refers to sequential recall of items which are experimenter-determined members of the same superordinate category.

Developmental assessments of recall and clustering have been unusually consistent. Regardless of the number of recall trials (single or multiple) or the mode of list

---

1 With the constraint that no two members of the same category are adjacent.

presentation (i.e., objects, pictures or words), recall increases with age (Hagen, Jongeward & Kail, 1975; Jablonski, 1974) with significant performance differences occurring when nonadjacent age groups[2] are compared (Jablonski, 1974; Mandler & Stephens, 1967; Moely, Olson, Halwes, & Flavell, 1969). In the most current review of children's free recall, Jablonski (1974) examined the amount recalled and remarked that few studies "report results much over 50% recall whatever the manipulations. It is generally the younger children who do not typically recall more than 50%; in those cases in which 50% recall has been observed, it is usually in older children (p. 527)." For instance, the 8-year-olds in Nelson's (1969) study recalled an average of 54% of the 15 item categorizable word lists after 6 trials, while the 5-year-olds recalled 33%.

In general, category clustering has been found to follow the same age-related pattern as recall. That is, the degree of category clustering increases with age in both single and multi-trial free recall and for all stimulus modes of objects, pictures and words (Austin, Gekoski, & Zvian, 1965; Bousfield, Esterson, & Whitmarsh, 1958; Cole, Frankel, & Sharp, 1971; Hagen, et al., 1975; Jablonski, 1974; Mandler & Stevens, 1967; Moely, et al., 1969; Neimark, Slotnick, & Ulrich, 1972; Vaughan, 1968). For children above age 4, the amount

---

2 Age groups separated by at least two years, e.g., 6-year-olds vs. 8-year-olds.

recalled has been found to depend on the degree of clustering (Jablonski, 1974). This orderly relationship among age, recall, and category clustering has led to the assumption that the observed recall improvement with age is due to the age-related increases in category organization. In turn, category organization is considered a sequential function of two processes -- detection or discovery of the inherent categorizable nature of the list items followed by effective use of this structure during recall (Cohen, 1966; Cole, Gay, Glick, & Sharp, 1971; Mandler, 1966; Tulving & Osler, 1968; Tulving & Pearlstone, 1966; Scribner, 1975; Scribner & Cole, 1972). Deficiencies in either of these two processes would lead to poor clustering and consequently low recall scores. Thus younger children in comparison to older children are presumed to be "process deficient".

The current deficiency model, first proposed by Flavell and associates (Flavell, Beach, & Chinsky, 1966), is primarily a descriptive one. Referred to as the "production deficiency hypothesis", it postulates the presence of a developmental stage during which young children do not spontaneously apply appropriate skills in their cognitive repertoire to cope with a task. When applied to free recall, production deficiency refers to the non-spontaneous use by younger children of category conceptual schemes to aid recall. In other words, it is not the younger children's lack of "higher order" conceptual ability to categorize that results in both poor recall and clustering.[3] Rather this ability is present but not applied to the free recall task. When these children are experimentally induced to use their categorizing ability by drawing attention to categories in the recall list (e.g., blocking, constrained, and cued recall), their recall and clustering scores are enhanced (Cole, et al., 1971; Eysenck & Baron, 1974; Lange, 1973; Scribner & Cole, 1972). However, this enhancement occurs for all age groups and the customary age differential in performance is maintained (Arlin & Brody, 1976; Cole, et al., 1971; Moely & Shapiro, 1971; Scribner & Cole, 1972).

Although the production deficiency account adequately describes free recall findings, it does not address the puzzling problem of why young children, such as kindergartners and first graders, do not spontaneously employ category schemes in a standard free recall task. To posit a developmental stage is not explanatory. As Jablonski (1974) noted there appears to be no adequate developmental based theoretical explanation of free recall results. In fact, Moely, et al. (1969) and Flavell (1977) believe that this "superficial descriptive concept" (i.e., production

---

[3] Children as young as two years of age seem to possess some working knowledge of superordinate category concepts (Goldberg, Myers, & Perlmutter, 1974) and, at least, by age six children use categorical encoding schemes in short-term memory (Kail & Schroll, 1974; Libby & Kroes, 1971; Pender, 1969; Wagner, 1970); all of which argues against categorical concept deficiency.

deficiency) will be retired when more precise ways of characterizing and diagnosing the cognitive skills within the child are found and when the various features and focus within the task situation which interact with the child's cognitive skills become clear.

Typicality of List Items: A Free Recall Task Feature

One constant task feature concerns the typicality of items composing a free recall category list. The degree of category membership should interact with the category detection process involved in free recall and thereby influence clustering and recall performance in children. If the category list is composed of all prototypical items, which can be viewed as the most representative or salient members of their respective categories and which possess strong "relatedness" in terms of sharing more attributes in common with each other than do non-typical items, then the initial detection phase of categories within the list should be facilitated.

The hypothesis gains support, albeit indirectly, from Moely and Jeffrey's (1974) study. These authors chose items for two category lists on an intuitive basis so that the lists would vary in terms of category cohesiveness. One list was composed of category items considered by the authors to be highly related members of the same conceptual categories; the other list was judged to contain less highly related members. They found that 6½-year-olds obtained higher recall, higher clustering and better category sorting with the more highly related list.

Considering Rosch's findings (Rosch & Mervis, 1975; Rosch, et al., 1976b), it seems likely that the two lists differed in item typicality, the highly related items being prototypical and the less highly related ones being nontypical or atypical. A plausible alternative explanation of word frequency was eliminated by Moely and Jeffrey (1974). Since the two lists differed in terms of average word frequency count, the authors computed rank order correlations within each list between frequency of recall of a given item and its frequency in language according to the Kučera and Francis (1967) list. Essentially no relationship was shown for either list (r_s of .099 and .063) or for the two lists combined (r=.098).[4] In addition, word frequency did not predict which items would be sorted into groups. "In fact, higher frequency items were slightly more likely to be omitted from item groupings (p. 141)." Thus word frequency alone was not responsible for the obtained free recall and sorting differences between the two lists. Moely and Jeffrey concluded that:

---

4 Note that Mervis, et al., (1976) and Uyeda and Mandler (1980) also found essentially no relationship between word frequency and typicality.

> An important problem in investigating developmental changes or training differences in recall organization is that of quantifying item relationships so as to determine "easy" and "difficult" categories for children of various ages. Thus far, most recall studies with children have ignored this problem (p. 142).

If prototypical items enhance recall and clustering, what effect might nontypical items have? Category detection may be impeded when a category list consists of all nontypical items which are less representative members of their respective categories and are less related to one another than prototypical items. Little or no clustering and low recall are likely to occur. Moely and Jeffrey's (1974) findings suggest this may be the case, as does Lange's (1973) study.

Lange (1973) constructed category lists so that intra-list item relatedness was minimized. He used Palermo and Jenkins (1964) verbal association norms when possible.[5] But frequently he was forced to rely on his own intuitive judgment of "non-relatedness". Presentation of each of three lists to kindergartners, first, and ninth graders constituted one recall trial. He expected that the younger children would not organize their recall according to conceptual criteria with this type of list. Analyzing trial one (list one) only, Lange found that only ninth graders clustered significantly above chance levels in the standard

---

5 Palermo & Jenkins (1964) norms do not include Lange's youngest group, kindergartners.

free recall task, one of many conditions in his study. However, when the children at each grade level were experimentally induced to use a priori experimenter-determined categories, their recall and clustering scores were significantly better than their peers in the free recall condition.

These results tentatively suggest that less related items (atypical/nontypical) are still considered as category items by children (induced condition results) and that for all ages such items hinder free recall and clustering. Lange's (1973) induced condition procedures, however, appeared to actually teach the children which items belonged to what category. In addition, Lange did not statistically compare cluster or recall scores by grade within the free recall condition, nor did he compare these scores by grade within the induced condition. It is therefore unknown whether performance by grade in the standard free recall task was significantly different and whether the induced condition eliminated standard age-related performance differences.

Since Lange (1973) did not include highly related items for comparison and Moely and Jeffrey (1974) did not have a developmental aspect to their study (i.e., only one age level), the differential influence of related (prototypical) and unrelated (atypical/nontypical) category items by age remains unexplored.

Furthermore, considering the standard practice of giving the same lists to children of all ages, as Lange (1973) did, the same list may not be equivalent in typicality of its items across various ages.

While Rosch (1977) remains silent on this specific developmental issue, she does propose that the content of natural language categories (i.e., specific items) would vary with culture but that the typicality structure of the categories would be universal (Rosch, 1977). That is, categories are structured in terms of prototypes and distances from prototypes. The content and the specific prototypes of categories can be expected to vary. This variability is expected because semantic categories, in part, reflect real world phenomena. "Since the structure of the environment, both man-made objects and the flora and fauna of a region, should differ radically in different parts of the world, we would expect the categories of different cultures to differ (Rosch, 1977, p. 39)." Universal, invariant prototypical and nontypical semantic items are unlikely. In fact, regional differences within the United States seemed to produce slight variations in typicality ratings of specific items (Uyeda & Mandler, 1980). One would also expect category content and typicality of items to vary with age. Such a generalization, although logical, can be empirically validated. This is an important first step to examining typicality effects on children's free recall.

The possibility of non-equivalence of typicality across ages for the same category items has important implications for the interpretations of past age-related findings. The younger children's non-spontaneous use of category organization may, in part, be due to category content membership strength differing from adults' and older children's, i.e., typicality. Spontaneous detection of the categorical nature of a recall list may be impeded to the extent that items chosen for the categories in the list differ from the younger children's core or typical items. Indirect evidence that this may be the case comes from a study by Nelson (1974) and a replication study by Brandt (1980). It was found by both authors that core animal members[6] for the younger children (5-year-olds in Nelson's study and 6-year-olds in Brandt's study) were wild or zoo animals (giraffe, lion, tiger, elephant, and monkey). In direct contrast, the first and most frequent items for adults are dog, cat, and horse with giraffe, elephant, and tiger not even occurring in the first ten responses (Cohen et al., 1957). Dog, cat and horse were also the items considred by adults to be the best exemplars (prototypical) of the animal category (Uyeda & Mandler, 1980). Moreover, the core items for the older groups of children (8-year-olds in both studies) were

---

6 Core members were defined as those category members representative of most children's protocols of a given age group in a category item production task.

domestic animals (dog, cat, horse) and thus differed from the younger groups but were identical to the adult items.

This fact is of particular note in that category recall lists for children of all ages most often have been constructed according to the adult experimenter's formal, logical category schemes, presumably guided by the traditional notion of category membership: Either an item is or is not a member and if it is, then it is a full and equal member of the category. Degrees of membership are not recognized in the formal, traditional view of categories.

Consider a sample of category item selection procedures: "Common nouns" (Cole, et al., 1971); "familiar" items (Yoshimura, et al., 1971; Ornstein, et al., 1977); "familiar and easily named" (Jensen & Frederiksen, 1973); and nouns from adult category norms (Scribner & Cole, 1972; Vaughan, 1968). In all instances the determination that the items were common and/or familiar were made by the adult with little or no input from children. In all instances the adult determined whether the items were members of particular categories, not children. And in all instances the <u>same</u> list was given to all children regardless of age. Considering Nelson's (1974) and Brandt's (1980) findings, it is not suprising that the older children performed better than the younger. The standard free recall procedures (i.e., adult selection of items and categories and the same list for all children) appears biased in favor of the older children.

A striking exception to the usual item selection procedure is a free recall study by Posnansky and Pellegrino (1975) in which category items were selected from Posnansky's (1974) children's category norms. They obtained high recall, high clustering and no significant differences in amount of clustering across nonadjacent age groups (6-year-olds and 8-year-olds).[7] The authors directly attributed the high performance levels to the use of semantic categories capable of eliciting many instances from children, and to the selection of items from these categories with high normative frequencies for children. They concluded by stating:

> . . .previous developmental assessment of clustering in FR (free recall). . .may have underestimated the ability of young children to produce and use semantic retrieval schemes because the schemes available were defined by adults rather than made contingent on the children's own responses (p. 364).

Tenny (1975) also found no significant differences in clustering on category lists among kindergartners, third, and sixth graders and extremely high recall levels. Similar to Posnansky and Pellegrino (1975), her category lists were also contingent on children's own responses, but in a more direct way. Each subject received his/her own category list composed of his/her first responses to category labels.

---

7 There was also no difference in clustering between these two age groups and the 9-year-old subjects.

It is possible, given the findings on order and probability of category item output (Rosch, 1975a; Rosch & Mervis, 1975; Rosch, et al., 1976b) that Tenny's and Posnansky and Pellegrino's (1975) subjects received recall lists composed of the children's own prototypical items. In neither study were experimental techniques employed to induce spontaneous use of category organization. It appears that even the youngest children did so. Thus prototypical items may not only enhance free recall performance by facilitating category detection, age-appropriate prototypical items may also eliminate the usual age difference in recall and clustering.

Indirect evidence exists that mixed lists may adversely affect free recall performance. Given a list of prototypical and nontypical items for "Animal", such as bear, rabbit, (nontypical), and dog, horse (prototypical)[8], one would not expect children to cluster all of them together in recall. To children the nontypical items may not belong to the same conceptual groupings as the prototypical items or are not proper instances of the category (see Carson & Abrahamson, 1976). Failure to control for typicality differences within a list may be responsible for the low performance levels of children in general as noted by Jablonski (1974).

---

8 These specific exemplars were selected from Uyeda & Mandler's (1980) adult prototypical norms.

Franklin, Fulani, Henkind and Cole (1975) examined the effects of a certain type of mixed list on free recall. They reasoned that if certain items (considered members of a category in the traditional sense) have differential "saliency" within a category for two different populations (Black and Caucasian), the clustering levels would reflect these differences. Briefly, they combined category items of high and low salience for one population to form a mixed list for a standard free recall task. This list was presented to samples from two populations who differed in terms of "familiarity" with half the items at the beginning of the experiment. They found that the sample with the greater "familiarity" with all the items had significantly higher overall scores. Furthermore, this sample had higher cluster scores on both parts of the mixed list and there were no significant differences within samples in performance on the two parts of the list. To explain these interesting findings, Franklin, et al., proposed that the presence of low salient items for a particular population interferes with clustering by converting a recall list that is nominally categorizable in its entirety into one that is perceived as only partly categorizable. This explanation received support by similar results when performance on all categorizable lists was compared to another type of mixed list (part categorizable and part non-categorizable items). Clustering level for the all categorizable list was more than twice as great as that for the mixed list (Franklin, et al., 1975).

Franklin et al., (1975) indexed "item saliency" or familiarity by ease in assigning items to categories. Given the findings that ease of item classification is a function of family resemblance structure and degree of item typicality (Rosch & Mervis, 1975; Rosch, et al., 1976b), it can be argued that Franklin's et al., high and low salient items forming their mixed list may have been prototypical and nonprototypical items respectively.

Inequalities in item membership strength may disrupt clustering performance. It is possible, given the manner of item selection discussed previously, that the free recall lists presented to children in past free recall studies were actually mixed lists. If this is true, children's overall low performance is a logical result. It is also possible, given the same list procedures, that the recall list is a mixed one for the younger groups and a prototypical one for the older groups which would also produce the standard age-related findings. Alternately, children of different ages may respond differently to a mixed list even if the mixed list is composed of age-appropriate typicality of items.

Uyeda and Mandler (1980) examined the efficacy of different types of retrieval cues on recall with adults. They reported that subjects not only chose <u>the</u> most prototypical item of each of 24 categories as a retrieval cue significantly more often than any other item, but also the use of prototypical retrieval cues (either subject selected or experimenter selected; intra- or extra-list items) resulted in significantly better recall than nontypical cues in both immediate and delayed recall conditions. Furthermore, when the recall categories were composed of all nontypical items, highly prototypical retrieval cues not present in the recall list (called extra-list cues) were as effective as cues selected from within the list (nontypical intra-list cues). To account for these findings, the authors proposed the notion that cues can contain list-specific and/or category information. High prototypical cues, by their nature[9], contain category information and thus provide effective category access even when they are extra-list cues for nontypical category lists. High prototypical intra-list cues (HP-INTRA) were slightly more effective than high prototypical extra-list cues for prototypical lists because HP-INTRA cues contained list-specific information in addition to category information.

Uyeda & Mandler's (1980) study suggests an effective retrieval strategy for a mixed list. If children in a standard free recall task used the prototypical items presented in a mixed list as retrieval cues, recall would be better than if they did not use such cues. It would

---

9 Prototypical items best reflect the redundancy structure of the category by having the largest number of attributes which overlap those of other category members: They have the strongest family resemblance.

therefore be important to first determine whether older children have better recall on a mixed list than younger children when it can be insured that the recall lists are indeed mixed lists for both ages. The lists should be equivalent in terms of typicality of the items. If the age comparison within an age-appropriate or age-equivalent mixed list condition is significant, it would have great heuristic value for further developmental free recall research.

Conclusion

Family resemblance model is a competing, viable and perhaps, more psychologically valid representation of natural semantic categories than the traditional model. The tradition model treats category membership in a digital, all-or-none fashion. In contrast, the family resemblance model holds that there are gradients of membership organized around core prototypes. It permits category content variation across age and cultural groups without viewing such differences (which the traditional model must) in value-laden terms as correct and incorrect.

Degree of item prototypicality within categories has been shown to be an important independent variable influencing outcomes in a variety of cognitive processing tasks with adults. Few studies exist which examined the effects of typicality on children's performance. The one developmental study which did (i.e., Carson & Abrahamson, 1976) suggests the critical need to explore the role played by prototypicality differences of category items on the outcome of any task involving categorization, classification, and cognitive organizational skills.

One task used to track the developmental progression of categorical organization in memory is the free recall paradigm. Indirect evidence from different but related research domains (e.g., category classification, category production, and class inclusion) and from examination of free recall procedures and results indicated that degrees of typicality should have an impact on children's free recall performance. If such an impact can be demonstrated, it will substantiate, at the least, a need to control for semantic typicality in future mnemonic organizational research with children and, at the most, require a conceptual revision of developmental assessments of children's free recall performance. The latter possibility would have far reaching implications considering the current assessment of free recall research: No one currently understands what psychological variables develop to account for free recall differences and improvement with age (Jablonski, 1974).

II. THE PRESENT STUDY

The present study was designed to directly explore the effects of typicality in a free recall task upon children of two different age groups. The study had two phases: (1) In Experiment 1, category members were elicited directly from the children and the size, diversity, and agreement among items were examined by age and category. Prototypical and nonprototypical items from among the age-appropriate category pools were used in the second phase; (2) in Experiment 2, the free recall task with three conditions (prototypical, non-prototypical, and mixed list conditions) was given to children from two nonadjacent age groups. The amount recalled, degree of clustering, and improvement across trials were examined.

Experiment 1

Cateogry Item Production

To overcome the problems in item selection of past developmental free recall studies, the items for the category recall lists must be considered as members of the categories by children of the same age groups and population as the experimental free recall subjects. To minimize adult imposition, a category item production (CIP) procedure was selected. This procedure has several advantages: It allows children to provide the category items and, therefore, insures that the category items are known and familiar to each age group; it eliminates problems inherent in a classification sorting task (e.g., preselection of items and pictorial representation) and maintains the same verbal mode to be used in the free recall task; it minimizes experimenter-task demand charateristics and age-differential task demands (Nelson, 1974); and the CIP task provides a means for quantitatively determining typicality of the items for each age group.

Method

Subjects

Thirty-two first grade and thirty-three third grade children from the same private school and grade level as the experimental free recall subjects participated in the CIP task. Approximately half the children at each grade level were girls. Children who spoke English as a second language were eliminated from the sample.

Procedure

Five category labels--animal, clothes, fruit, furniture, and tools--for the CIP task were selected from among superordinate labels used in past developmental category studies (Brandt, 1980; Nelson, 1974; Rosner & Hayes, 1977; Saltz, et al., 1972; Tenny, 1975). Results of these studies established that these five labels were easily understood by early elementary school age children (i.e., no explanations or descriptions were asked for or required) and had elicited a fairly large number of category items form all the subjects.

All the children were individually interviewed during the school morning in a quiet location near their classroom. After the child was seated and appeared comfortable, he/she was given the following instructions:

> "_____________(child's name), I'm going to say the name of some things. After I say the name, I'd like you to tell me all the things you can think of that go with that name. For instance, I might say 'soda' and you might think of 'orange soda, coke, root beer, fruit punch' and lots of others. Or I might say 'food'. Tell me what things are 'food'. (Pause for the child to respond.) Those are very good. Okay, just remember each time you say all the things that go with the word I say."

The first category label was read and the child's responses were recorded verbatim in the order they were given. The child received a standard prompt to name more items if he/she remained silent for at least 20 seconds. Only one prompt per category was given. Following the second period of silence or a statement indicating that the child had completed his/her responses to the current category, the next category was read until all 5 categories were exhausted. No time limit was imposed. Approximately 10 to 20 minutes were need to complete the task. The 5 categories were presented in the following order for all children: Animals, clothes, fruit, furniture, and tools.

Results and Discussion

The CIP procedure was successful in eliciting a large number of responses (3,078). Every child provided at least three exemplars for each category and all but two children gave more than that amount. To answer the question of whether the category labels were known and appropriate ones for children of these two ages, the number of responses and the number of different items were tallied by grade and category.

These data are presented in Table 1. Both indices confirm the adequacy of the category label selection procedure. In addition, none of the children during the interview session expressed in any observable manner that the categories were unknown or unfamiliar.

For every category both the total number of responses and the number of different items were larger for the older children, replicating past findings in the literature (Brandt, 1980; Nelson, 1974; Rosner & Hayes, 1977; Saltz, et al., 1972). This size difference may reflect a performance factor in that the younger children may be unable to retrieve as many items from long term memory and/or it may reflect actual differences in the number of items stored in memory.

Rank ordering of the categories by size was identical for the two grade levels. That is, the size relationships among the categories within grade were identical. Thus possible difficulties in retrieving items for any one category relative to the others cannot be considered to differ between the older and younger children. When size is held constant by restricting examination to the first 3 items given, differences between grades in the number of different items given is essentially eliminated (see Table 2). Thus the two age groups appear to have equal diversity of items that are first retrieved from memory.

The index of commonality (IC), often used for verbal association data, was employed to examine the degree of agreement on category membership. The IC is the ratio of the frequency of the 3 most common items within each category to the total number of items given by all children restricted to the first 3 responses. It was applied to the first 3 items given by each child within each grade level for each category. ICs for each category and grade are presented in Table 3. Results showed that age difference in mean ICs (.609 vs. .535 for first and third graders respectively) was miniscule, again replicating past CIP findings (Brandt, 1980; Nelson, 1974). The younger children tended to agree among themselves on the most common items to the same extent as the older children. There are few, if any, age-related differences in either agreement or diversity for those items first retrieved from long term semantic memory.

Table 1

Size and Number of Different Items by Grade and Category

First Grade

(N=32)

	Number of Rs[a]	Number of Different Rs	Mean
Animals	400	100	12.5
Clothes	240	69	7.5
Fruit	215	52	6.7
Furniture	187	67	5.8
Tools	154	41	4.8
Total	1196	329	Mean 7.5

[a]"Rs" stand for responses.

Table 1 (continued)

Size and Number of Different Items by Grade and Category

Third Grade

(N=33)

	Number of Rs[a]	Number of Different Rs	Mean
Animals	683	185	20.7
Clothes	336	111	10.1
Fruit	324	70	9.8
Furniture	294	125	8.9
Tools	245	99	7.4
Total	1882	590	Mean 11.4

[a] "Rs" stand for responses.

Table 2

Number of Different Items

	First Grade	Third Grade
Animals	34	37
Clothes	20	19
Fruit	16	16
Furniture	25	19
Tools	16	19
Mean	22.2	22.0

Note. Number of different items given within the first three responses by grade and category.

Table 3

Index of Commonality (IC)

for the First Three Items Given

by Grade and Category

	First Grade	Third Grade
Animals	.375	.333
Clothes	.663	.576
Fruit	.708	.626
Furniture	.655	.571
Tools	.645	.567
Mean IC	.609	.535

Note.

$$IC = \frac{\text{Frequency of the 3 most common items}}{\text{Total number of items}}$$

Restricted to the first 3 given.

Experiment 2

Free Recall

The present experiment explored the relative influence of age-appropriate prototypical, nonprototypical, and mixed category lists on recall and clustering for children of two nonadjacent ages in a multi-trial free recall task. Based on preceding discussion, the following hypotheses were advanced:

(1) Categories and items that are age-appropriate and familiar result in recall exceeding 50% of the total number of possible recall items for both ages in all list conditions.

(2) The prototypical list condition (PLC) would enhance recall and clustering over both the nonprototypical list condition (NLC) and the mixed list condition (MLC) with the pattern being: PLC > MLC > NLC.

(3) Recall would increase across trials for all conditions.

(4) Clustering would increase across trials for, at least, PLC and the earliest evidence of clustering would occur in PLC.

(5) There would be no significant recall and clustering differences between ages in PLC. Age-related differences for recall and clustering were expected to occur in MLC and/or NLC.

Method

Design

The experimental design was a 2 x 3 x 5 fixed factorial with repeated measures on the last factor. The three independent variables were grade level (first and third), list condition (PLC, MLC, NLC), and trials (1 to 5). Amount recalled and degree of clustering served as the dependent measures. A priori planned comparisons within list condition by grade were conducted to examine hypothesis 5 (Hayes, 1963; Keppel, 1963). Unless otherwise indicated, all results were assessed at the .05 level of significance.

Problems with various measures of category clustering have been discussed by Frankel and Cole (1971). Among the problems mentioned were: Lack of independence between clustering scores and number of items recalled; inadequate consideration of the category constitution of the recall list (i.e., in terms of the number of categories and the number of items per category); and ambiguity of the obtained cluster score and thus difficulty for comparing the same obtained absolute score, say, between two individuals. Frankel and Cole developed a Z score which overcomes these problems. Briefly, a Z score is a "joint function of the number of runs and length of runs comprising the recalled list and the number recalled from each category of the input list (Cole, et al., 1971, p. 112)." The number of runs is one more than the number of times during recall that the category of the items being recalled changes. From Wallis and Roberts (1956, p. 571), the following formula is used for calculating the mean and variance of the number of runs of items from within each category for each recalled list for each child and each trial.

$$M_r = \frac{N\ (N + 1)\ - \sum_j N_j^2}{N}$$

and

$$V_r = \frac{\sum_j N_j^2 \left[\sum_j N_j^2 + N\ (N + 1)\right] - 2N\sum_j N_j^3 - N^3}{N^2\ (N-1)}$$

where: N = total number of items recalled on one trial

N_j = number of items of j^{th} class for that trial

M_r is normally distributed so that knowing M_r and V_r and the observed number of runs, O_r, for that trial, a Z score can be calculated.

$$Z = \frac{O_r - M_r}{V_r}$$

Clustering is defined as the presence of "too few" runs. For convention, a positive Z score indicates some degree of clustering. To calculate a group's cluster score, the formula reduces to $n.\bar{Z}$ where n is the number of subjects and $\bar{Z}$ is the mean score for the group. Z scores wre used to index categorical organization in the present experiment.

Subjects

Sixty children (30 first and 30 third graders) enrolled in a private elementary school in Honolulu served as subjects.[10] Their mean ages were 6 years 7 months and 8 years 7 months for first and third graders respectively. Their ethnic backgrounds reflected the multi-ethnic composition of Hawaii. No one ethnic group predominated (see Appendix A). School records of occupational status of the parents indicated that these children were from middle-class families. The available school achievement and ability test scores indicated average to above average skills.

Recall Lists

Responses to five category labels (animals, clothes, fruit, furniture, and tools) by children from the same grade levels, background, and school as the present experimental children consituted the two age-appropriate category item pools (see Experiment 1). From these two pools, category items were selected to form 3 recall lists for each grade using criteria based on the relationships between prototypicality, item category dominance, and order of response output (Mervis, et al., 1976; Rosch, 1975b, 1975c; Rosch & Mervis, 1975; Rosch, et al., 1976b; Uyeda & Mandler, 1980).

---

[10] Four children were eliminated as potential participants since 3 spoke English as a second language and 1 was hearing impaired.

Prototypical lists. The three most common (dominant) items within the first three responses given to each of the five category labels were defined as the protoypical items for the prototypical list condition (PLC). Thus, each of the two prototypical lists, one for each grade, consisted of 15 items, three from each category.

Nonprototypical lists. The nonprototypical lists were composed of three items from each of the same five categories as the prototypical items. These items could occur no more than once within the first three responses and were equal to or less than prespecified distance scores from their respective prototypical category items. From those which met the above criteria, the items used in past developmental free recall studies were selected when possible. When none of the items was identical to past free recall items, the items with the highest number of responses were selected. These criteria for defining nonprototypical items a) were quantitative in nature, b) ensured that their selection was consistent across grade level, c) prevented selection of atypical or idiosyncratic items and potential experimenter bias, and d) established quantitative differentiation between prototypical and nonprototypical items (see Appendix B for detailed descriptions of these criteria).

Mixed lists. The mixed lists were formed from the prototypical and nonprototypical items, three items from each of the five categories. Overall, each mixed list contained approximately half prototypical and half nonprototypical items.

Thus each list condition, prototypical (PLC), nonprototypical (NLC), and mixed (MLC), consisted of 15 items, three from each category. These numbers were within the appropriate free recall range for first and third grade children. Each age group received their "own" list for each condition in order to maximize stimulus equivalence (Scribner & Cole, 1975). Table 4 contains the category items for each list condition and each grade.

Procedure

Children within each grade, in order of appearance and availability, were alternately assigned to one of the three list conditions. Both sexes were equally represented in each list condition. This resulted in 10 children (5 boys and 5 girls) in each cell.

Items for each list were randomly ordered for each of the five trials with the restriction that no two items from the same category were adjacent and that each trial began and ended with different items. The order of presentation of the five randomized lists was held constant across children within each list condition. All trials of all list conditions were taped with the approximate rate of item presentation being one every 2 seconds.

All children were individually tested during the school morning in a quiet room near their classrooms. After each child was seated and appeared comfortable, the following instructions were given:

> "________________(child's name), I'm interested in how well children remember what they hear. I have taped 15 words that you'll hear one at a time. After you hear all the words, I'll stop the tape and you tell me the words you remember hearing. I'll be writing them down. Okay? We'll do this five times. Listen carefully and try to remember all the words you hear. Ready?"

The instructions and procedures were pretested with a small sample of children from each grade who did not serve as experimental subjects.

Results and Discussion

Recall

The percentage of the total number of items recalled over trials by grade and list condition are presented in Table 5. All percentages but one (MLC, first grade) exceeded 50% with the highest percentages obtained in PLC for both grades. In addition, perfect recall scores occurred, but only with the prototypical lists. It appears that age-appropriate categories and items, especially for the younger children enhanced recall over that which would be expected given past free recall results. The use of procedures which directly insured that the items were familiar and known to the children may account for these obtained high recall scores.

Analysis of variance with repeated measures (BMDP2V) revealed that list condition ($F=3.36$, df 2, $p < .05$) and trial ($F=73.9$, df 4, $p < .0001$) factors were significant and that no interactions were significant. From Figure 1, it can be seen that the means assumed the predicted pattern of PLC > MLC > NLC, as expected. However, subsequent comparisons using the Duncan technique indicated significantly better recall for PLC than for MLC ($q=1.125$, $p < .05$) and for NLC ($q=1.183$, $p < .05$) but no significant difference between MLC and NLC. Thus, significantly better recall was obtained when all prototypical items composed the recall list. Since item familiarity was a constant across all list conditions, it appears that the strong relatedness and saliency of items within PLC provided an additional recall advantage over and above item familiarity.

In Figure 2, it is apparent that the amount recalled increased across trials for all list conditions as predicted. However, a plateau appears to have been reached by the fourth trial for MLC and NLC children. In contrast, children in PLC continued their steady improvement from trial to trial. Furthermore, for every trial recall was the highest in PLC.

The planned a priori orthonogal comparisons of grade within condition indicated no significant differences between first and third graders in PLC as predicted, and NLC. Thus all prototypical and all nonprototypical category items in a recall list tended to eliminate age-related differences in recall. It is possible that item and category familiarity alone may have accounted for these results, but this possibility is unlikely considering the mixed list result. MLC produced significantly different recall levels between the two age groups favoring the older children ($F=8.246$, $p < .01$) (see Figure 3). The younger children seemed to have a particularly difficult time with this type of list. They obtained the lowest recall scores (7.08) with the mixed list while the older children's MLC performance (9.34) was fairly equivalent to their age peer's PLC performance (9.68).

Since nonsignificant results are difficult to interpret, a more detailed examination of age similarities and differences in recall scores is needed. For this purpose, the mean number of items recalled by trial and grade for each list condition was examined. In PLC (see Figure 4), the first trial produced the greatest performance difference across age, but even that difference was slight (1.1). The most striking characteristic, however, is the overwhelming similarity of these two learning curves. Hence, it can be said with more confidence that all prototypical items in a recall list may have significantly reduced, if not eliminated, whatever developmental factors were responsible for past age-related differences in recall.

The NLC learning curves present a slightly different picture. There are essentially no differences in the mean number of items recalled between the two grades on the first two trials (trial 1: 5.2 vs. 5.1; trial 2: 6.9 vs. 7.2 for first and third graders respectively) indicating that this type of list was initially as difficult for the older children as it was for the younger group. In the later trials, especially trial 3, the older children recalled more items than the younger. By the last trial, however, only a mean difference of .7 separated the two age groups (see Figure 5).

The MLC curves presented in Figure 6 contrast sharply with the other two list conditions. On almost every trial, large differences between first and third graders occurred and by the last trial the difference between age groups is outstanding. The younger children improved only slightly from the first to the last trial (a mean increase of 2.6) while the older group displayed consistent improvement, resulting in a mean increase of over five items, twice as great as the younger group's. It appears that typicality inequalities among the items in a category list presented the younger, but not the older, children with a difficult recall task and one which was apparently more difficult than an all nonprototypical recall task.

Considering the other two list condition learning curves for the first graders, such non-cognitive factors as fatigue, motivation, and poor and/or misunderstood instructions, cannot adequately account for their performance in the mixed list condition. Nor can the cognitive factor of "limited memory span" reasonably explain the obtained difference between grades in MLC.

The significant age-related difference in recall for this mixed list, when viewed in light of the age-equivalent results in PLC and NLC, presents a problem for a "differential familiarity" hypothesis. All six experimental groups received their age-appropriate, familiar lists for recall. The items selected were contingent on the children's own responses and the semantic schemes were <u>not</u> defined by adults but by the children themselves.

It appears that typicality inequalities of the mixed list interacted with age to produce the standard age-related recall difference. One possible explanation for the obtained difference concerns the type of retrieval strategies used by these two age groups. According to Uyeda and Mandler's (1980) study, intra-list prototypical items (IP), extra-list prototypical items (EP), and intra-list nonprototypical items (IN) would all serve as effective retrieval cues for a mixed list. The IP and IN items may provide access to list specific information with IP providing additional access to category information. The EP items would provide category information only. Repetitions and intrusions during recall would indicate, although indirectly, whether these types of retrieval strategies were used.

The number and type of repetitions (prototypical and nonprototypical items) and the number and type of intrusions (category and non-category intrusions; extra-list prototypical and extra-list nonprototypical intrusions) were examined by grade within MLC to determine whether grade differences occurred on these indices.

The frequency and type of intrusions by grade for all list conditions were examined first. Intrusions were classified as category intrusions if the items occurred in the age-appropriate pool of category items obtained in Experiment 1. If there were large grade differences in PLC and NLC, then differences between the grades in MLC would be an unimportant observation.

From Table 6, it can be seen that the grades differed slightly in the total number of intrusions for each list condition. For both grades the largest number of intrusions occurred in MLC and the number of category intrusions and the total number of intrusions within MLC were essentially identical between the two grades (category intrusions: 21 vs. 19; total intrusions: 24 vs. 25 for first and third graders respectively). Thus, there appears to be no difference between grades in the number of intrusions in MLC. The grades did differ in typicality of the category intrusions: For the third and first graders' category intrusions, the percentage of intrusions that were extra-list prototypical items was 50% and 28.5%, respectively.

Repetitions were classified as either prototypical or nonprototypical repeats, and were considered as intra-list items. In a similar fashion, category intrusions were classified in terms of prototypical and nonprototypical types, and were considered as extra-list items. Table 7 presents the frequency of these retrieval types by grade. Third graders produced more prototypical intra- and extra-list items, more nonprototypical intra-list items, and less nonprototypical extra-list items than first graders. Thus, third graders had 25% more effective retrieval cues. Moreover, first graders had almost twice as many poor retrieval cues than third graders.

These data tentatively suggest that the two age groups differed in terms of the effectiveness of recall retrieval strategies with the mixed typicality lists. They also indicated the direction in which developmental free recall research may take--direct experimental manipulation of retrieval cue types in terms of typicality and their interaction with age in a free recall task.

Table 4

Prototypical, Nonprototypical, and Mixed Recall Lists for Each Grade

Prototypical Lists		Nonprototypical Lists	
Grade		Grade	
First	Third	First	Third
elephant	dog	chicken	chicken
giraffe	cat	cow	deer
tiger	tiger	mouse	fox
shirt	shirt	coat	overalls
pants	pants	hat	pajamas
dress	dress	shoes	apron
apple	apple	lime	lemon
banana	pear	coconut	coconut
orange	orange	raisins	cherries
chair	chair	dresser	dresser
couch	couch	lamp	television
table	table	stool	cabinet
hammer	hammer	axe	axe
nail	wrench	scissors	screw
screwdriver	screwdriver	ruler	shovel

Mixed Lists

Grade	
First	Third
tiger	dog
elephant	cat
cow	chicken
shirt	shirt
coat	pajamas
hat	apron
apple	apple
orange	orange
raisins	lemon
chair	chair
dresser	dresser
stool	stool
hammer	hammer
screwdriver	wrench
axe	axe

Note. See Appendix C for adult prototypical ratings of these items.

Table 5

Mean Percentage of the Total Number

of Items Recalled

by Grade and List Condition

Percent Recalled

	Grade	
List Condition	First	Third
PLC	60.6	64.5
MLC	47.2	62.2
NLC	51.9	55.8
Overall Mean %	53.2	60.8

Table 6

Number and Type of Recall Intrusions

by Grade and List Condition

First Grade

	Category	Non-category	Total
PLC	6	11	17
NLC	10	5	15
MLC	21	3	24
Total	37	19	56

Third Grade

	Category	Non-category	Total
PLC	10	2	12
NLC	12	0	12
MLC	19	6	25
Total	41	8	49

Table 7

Frequency of Intra- and Extra-List Retrieval Types for Mixed List Condition by Grade

Types of Retrieval Cues	Retrieval Effectiveness[a]		Grade	
	List-Specific	Category	First	Third
Prototypical Intra-list	+	+	37	42
Nonprototypical Intra-list	+	-	23	37
Prototypical Extra-list	-	+	6	8
Nonprototypical Extra-list	-	-	15	8
		Total:	81	95

[a] The + and - signs indicate the presence or absence of list specific and/or category information provided by the types of retrieval cues.

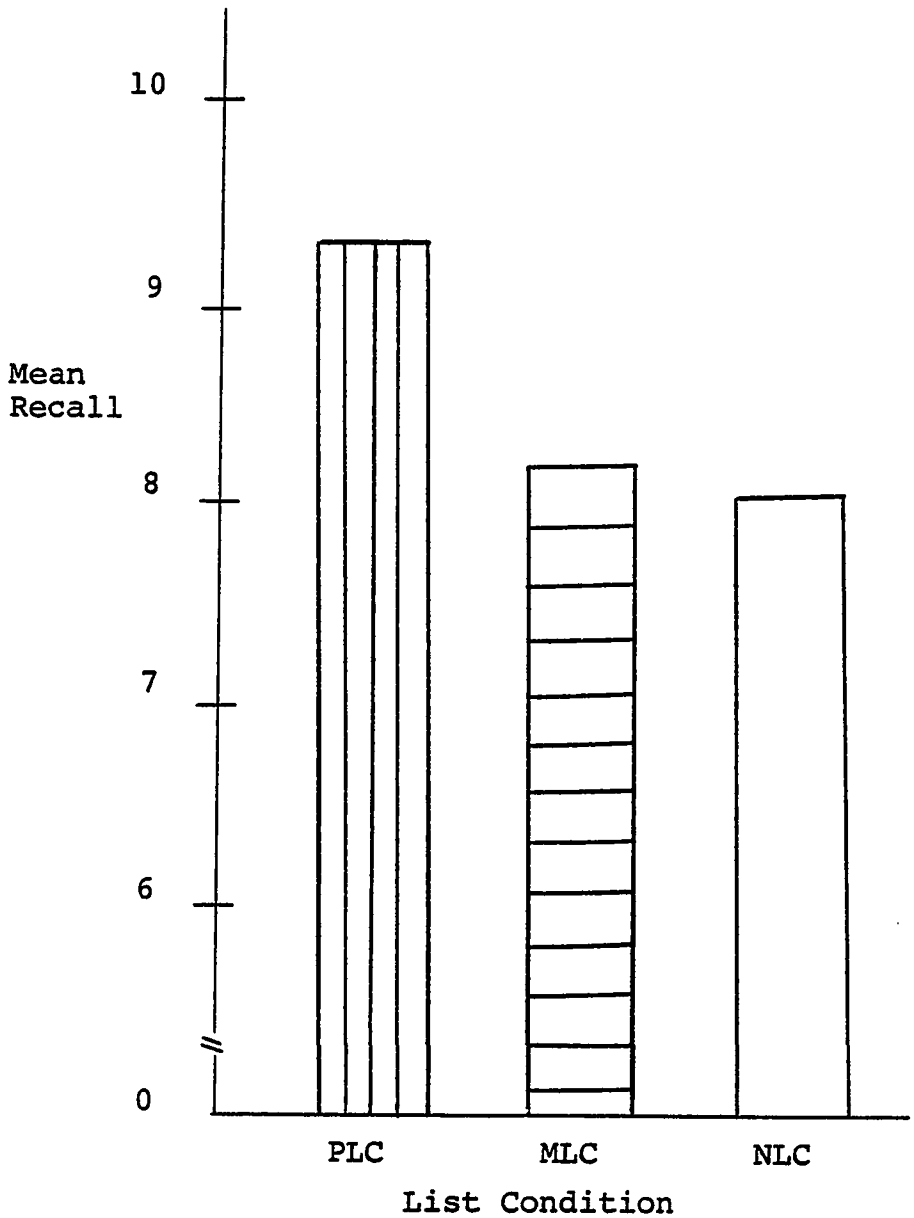

Figure 1. Histogram of mean recall scores as a function of list condition

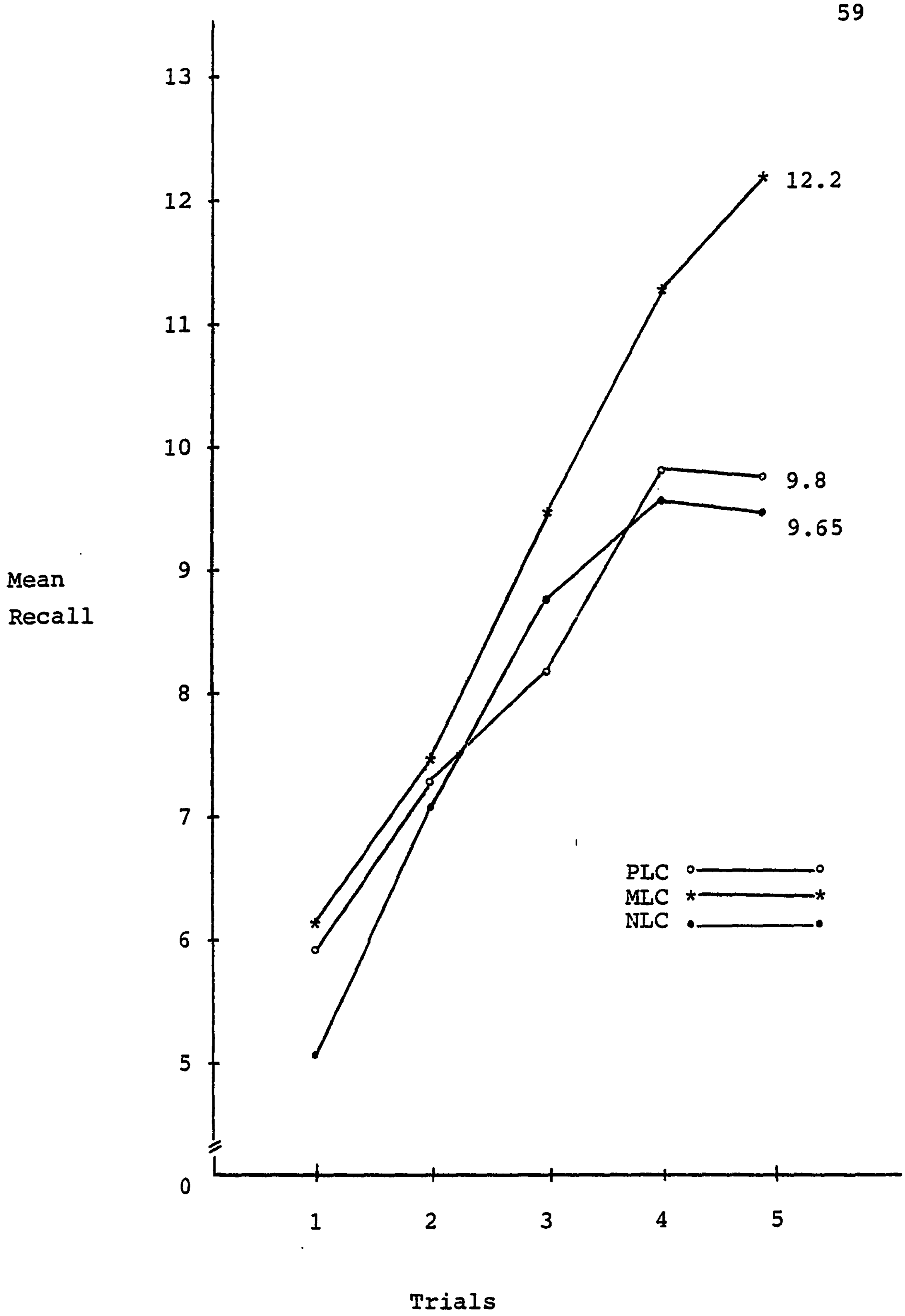

Figure 2. Mean number of recalled items for each list condition as a function of trials.

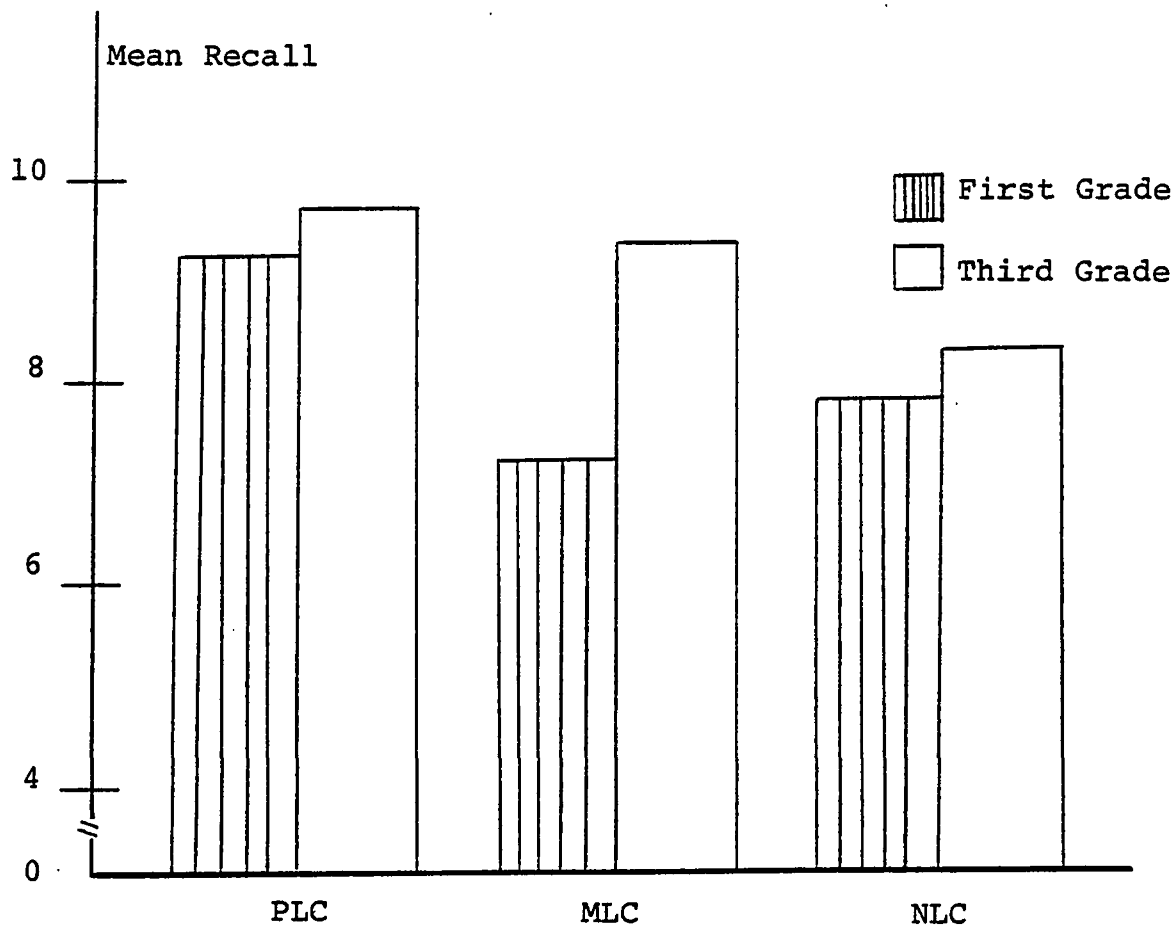

Figure 3. Mean number of recalled items for each list condition as a function of grade level.

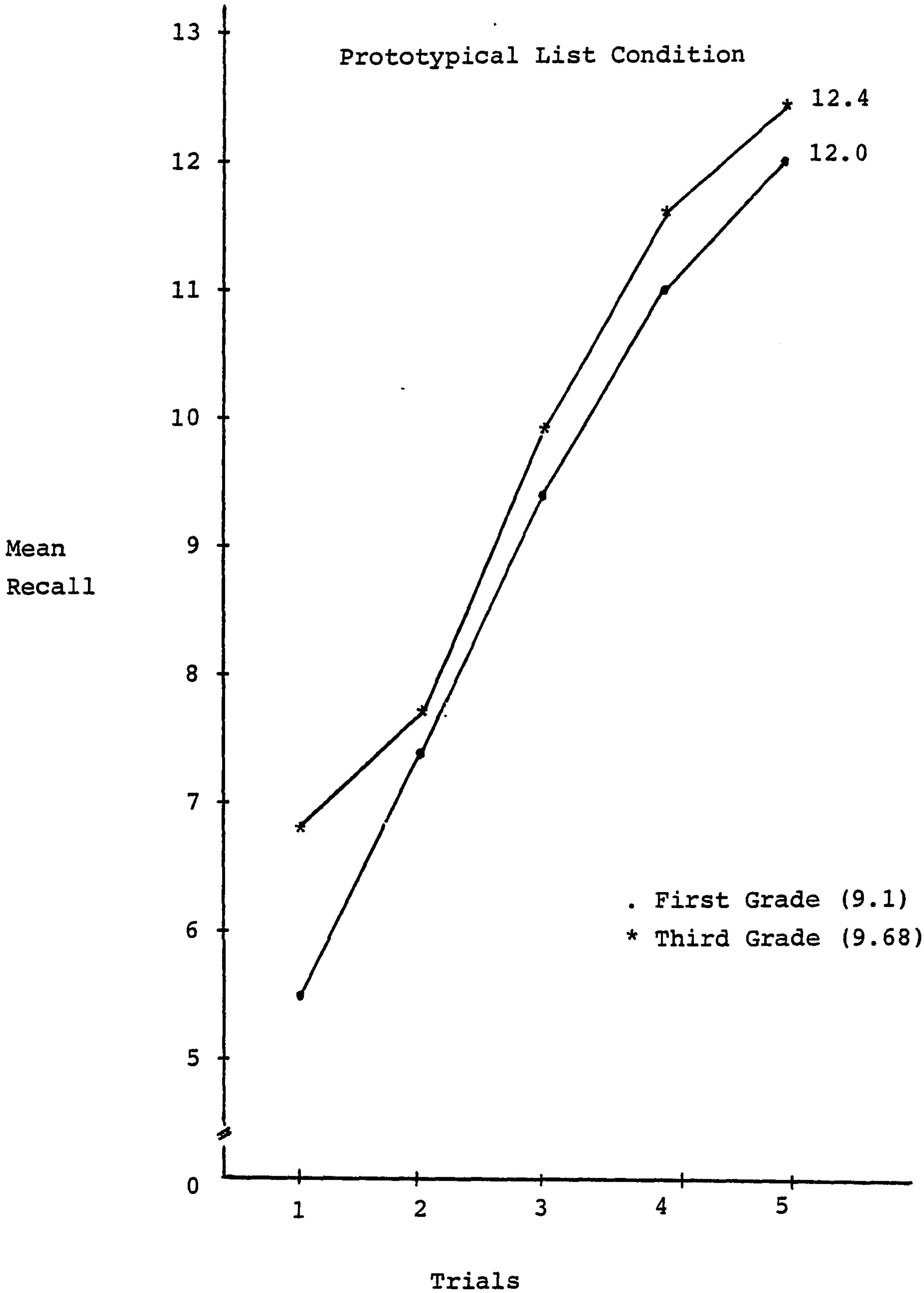

Figure 4. Mean number of items recalled in PLC by grade and trial.

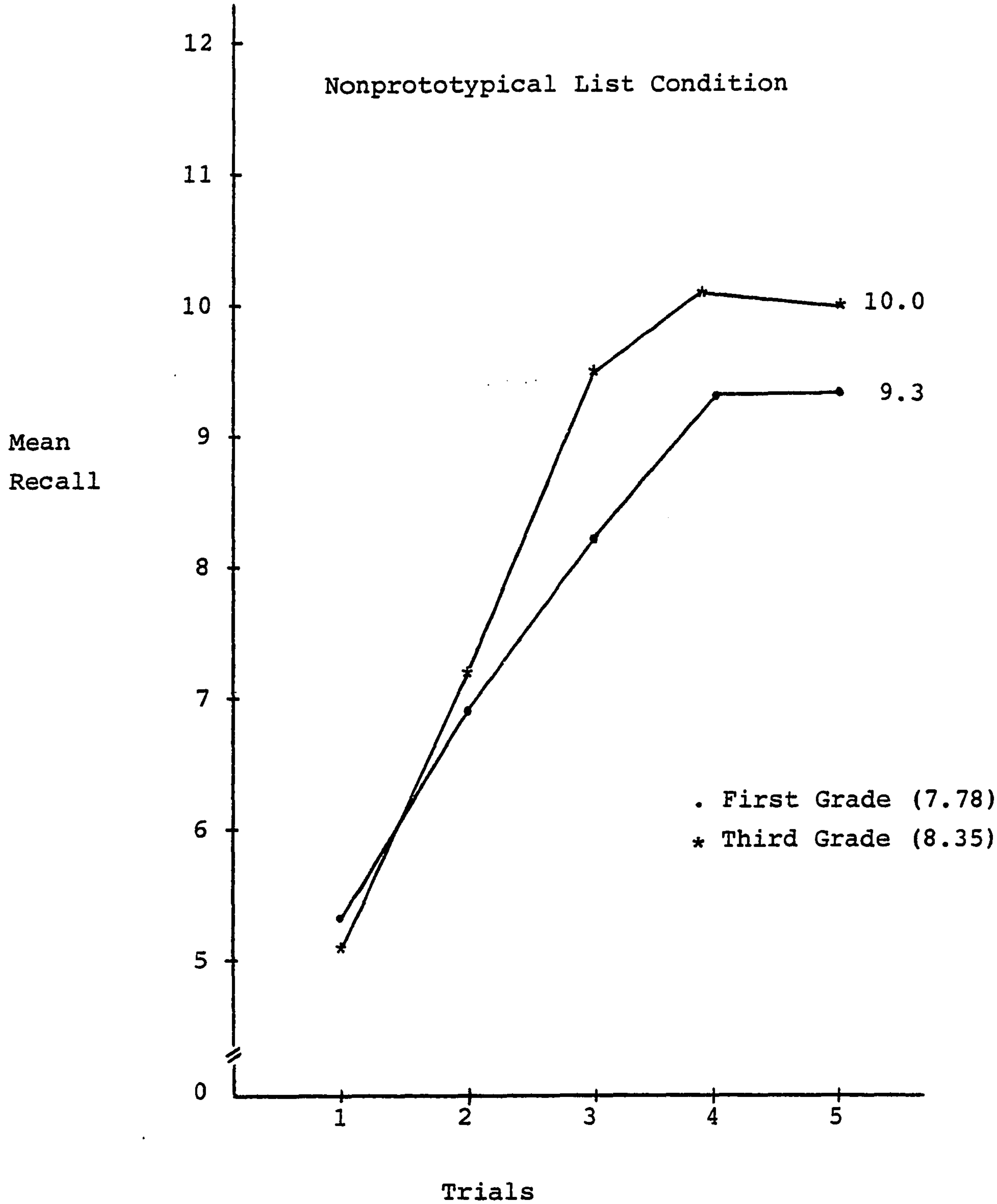

Figure 5. Mean number of items recall in NLC by grade and trials.

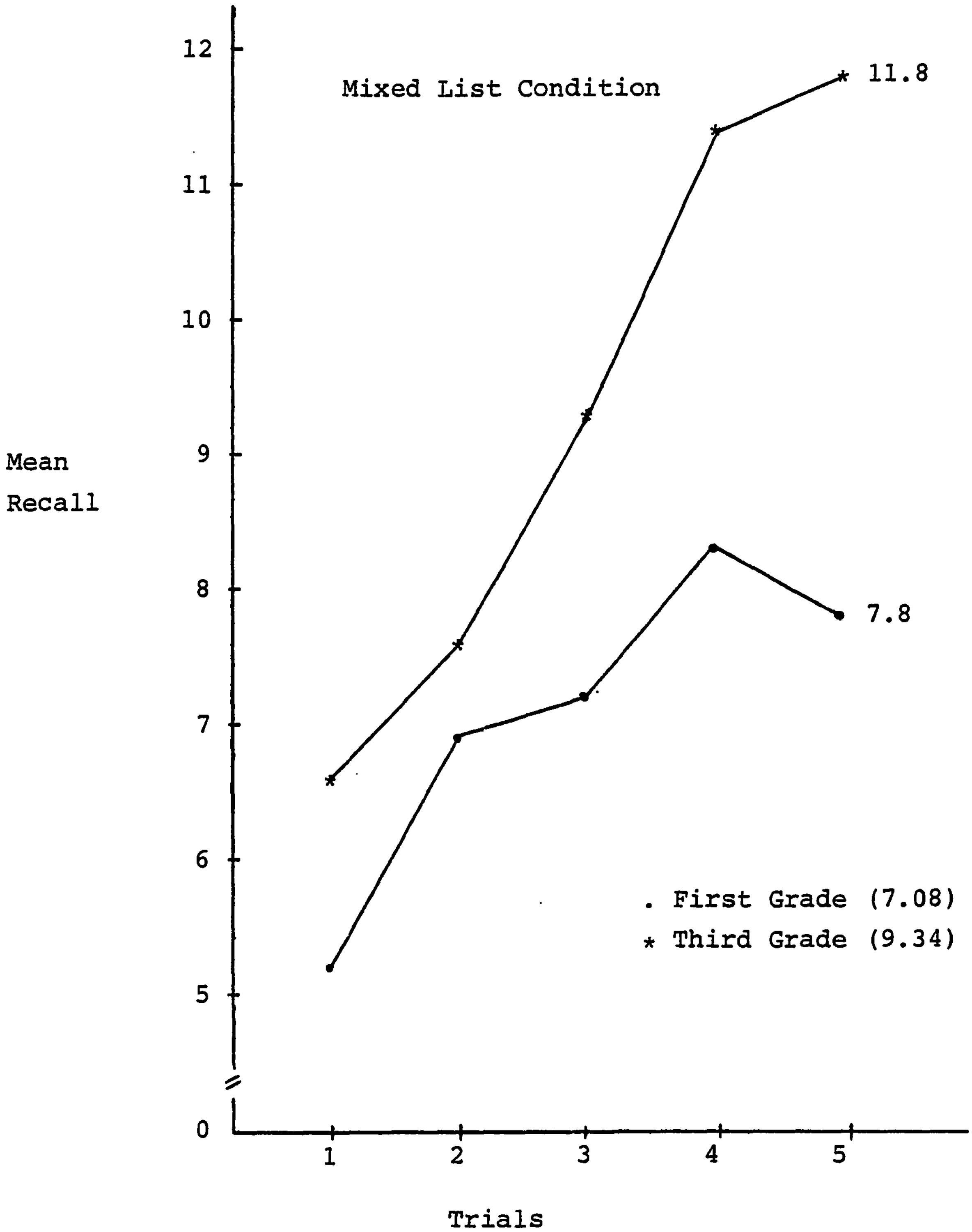

Figure 6. Mean number of items recalled in MLC grade and trials.

Clustering

List condition ($F = 15.45$, df2, p. $< .0001$), trials ($F = 3.65$, df4, p. $< .01$), and their interaction ($F = 2.41$, df4, p. $< .02$) were significant. Duncan subsequent comparisons revealed that clustering in PLC was significantly greater than in MLC ($q = .41575$, $p < .05$) which in turn was significantly greater than in NLC ($q = .43728$, $p < .05$), as predicted. An all prototypical list enhanced clustering over that obtained in the other list conditions. The presence of even some prototypical items in a list (MLC) seemed to have facilitated categorical organization relative to an all nonprototypical list (NLC).

Trial by list condition interaction is presented in Figure 7. On trial 1, clustering was equivalent and minimal for all list conditions. The earliest evidence of some degree of clustering occurred in PLC on trial 2 and clustering in PLC improved across trials, as predicted. In contrast to PLC, no clustering improvement seemed to take place in NLC. In fact, indications of any degree of clustering on any trial were absent, resulting in an overall mean Z of essentially zero (.0908). MLC cluster scores were indistinguishable from NLC scores on the first two trials. On the later trials, MLC clustering assumed an intermediate level between MLC and PLC. The presence of nonprototypical items in the mixed list appeared to mitigate the facilitative influence of prototypical items on clustering.

Figure 8 presents mean cluster scores by grade within each list condition. It is important to note that the extent of the differences between the grades within the list conditions paralleled those for recall. Since Z scores do not increase simply as a function of number of items recalled (Frankel & Cole, 1971), this parallel is striking. In contrast to the recall by grade comparisons, all a priori comparisons of clustering within list condition between grades were nonsignificant. Although this was expected for PLC and consistent with grade recall comparisons for PLC and NLC, nonsignificant grade differences remain abtruse. Therefore a more detailed picture of clustering performance is presented in Figures 9, 10, and 11.

Within PLC, first graders' clustering was slightly more variable across trials than third graders (see Figure 9). For the most part, steady improvement occurred for both grades, resulting in clustering significantly above chance levels (one tailed z, 1.65, $< .05$) for both grades.

In Figure 10, first graders' performance was again slightly more variable than third graders' in NLC. But both grades' scores hovered around zero. In fact, third graders' scores in this list condition steadily declined, albeit slightly, over trials. The general picture, however, is one of essentially no clustering for either grade.

In contrast to clustering performance in the other two list conditions, grade differences in MLC clustering increased from trial 1 to trial 5 (see Figure 11).

On trial 1, the difference between grades was nonsignificant. On the last two trials, however, third grade cluster scores jumped to above 1.000 while first grade scores dropped to zero, producing significant grade differences ($t = 11.627$, df18, $p > .01$; $t = 9.2999$, df18, $p > .01$ for trials 4 and 5 respectively). Repeated experience with a mixed typicality list tended to increase categorical organization for the older but not for the younger children.

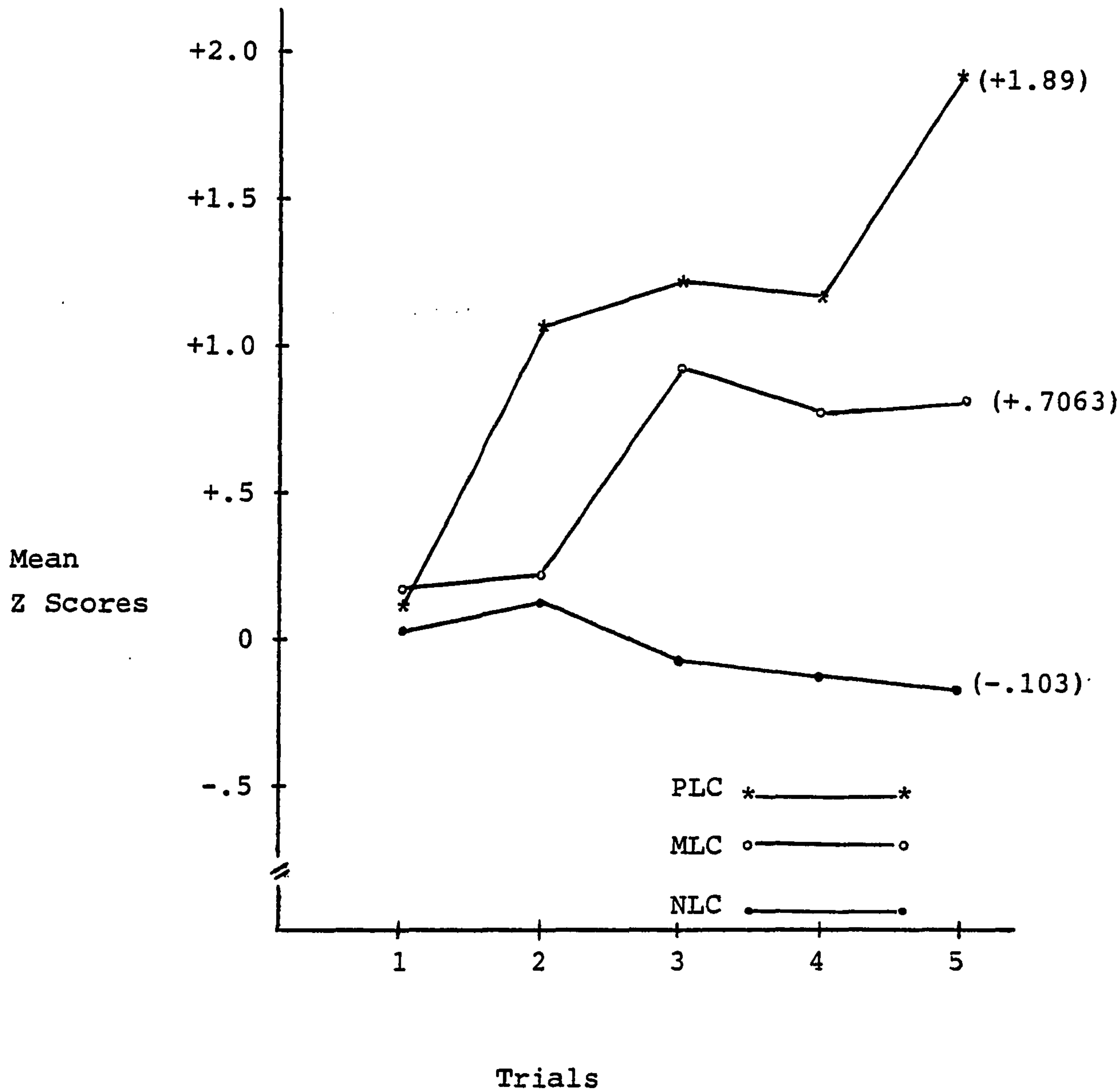

Figure 7. Mean cluster scores for each list condition as a function of trials. Positive Z scores indicate some degree of clustering; zero or negative Z scores indicate lack of clustering.

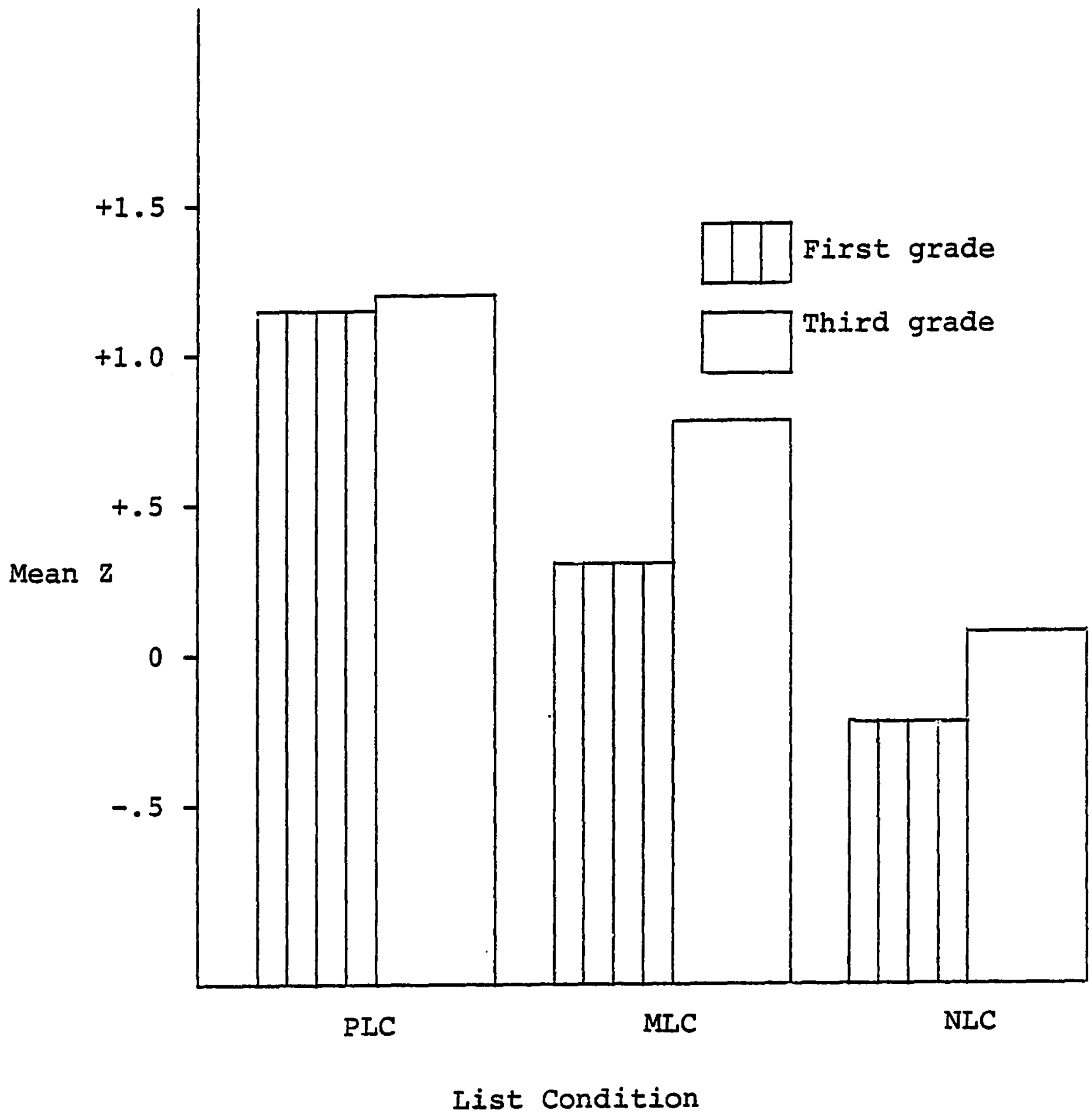

Figure 8. Mean cluster scores for each list condition as a function of grade. Positive Z score indicates some degree of clustering; zero or negative Z scores indicate lack of clustering.

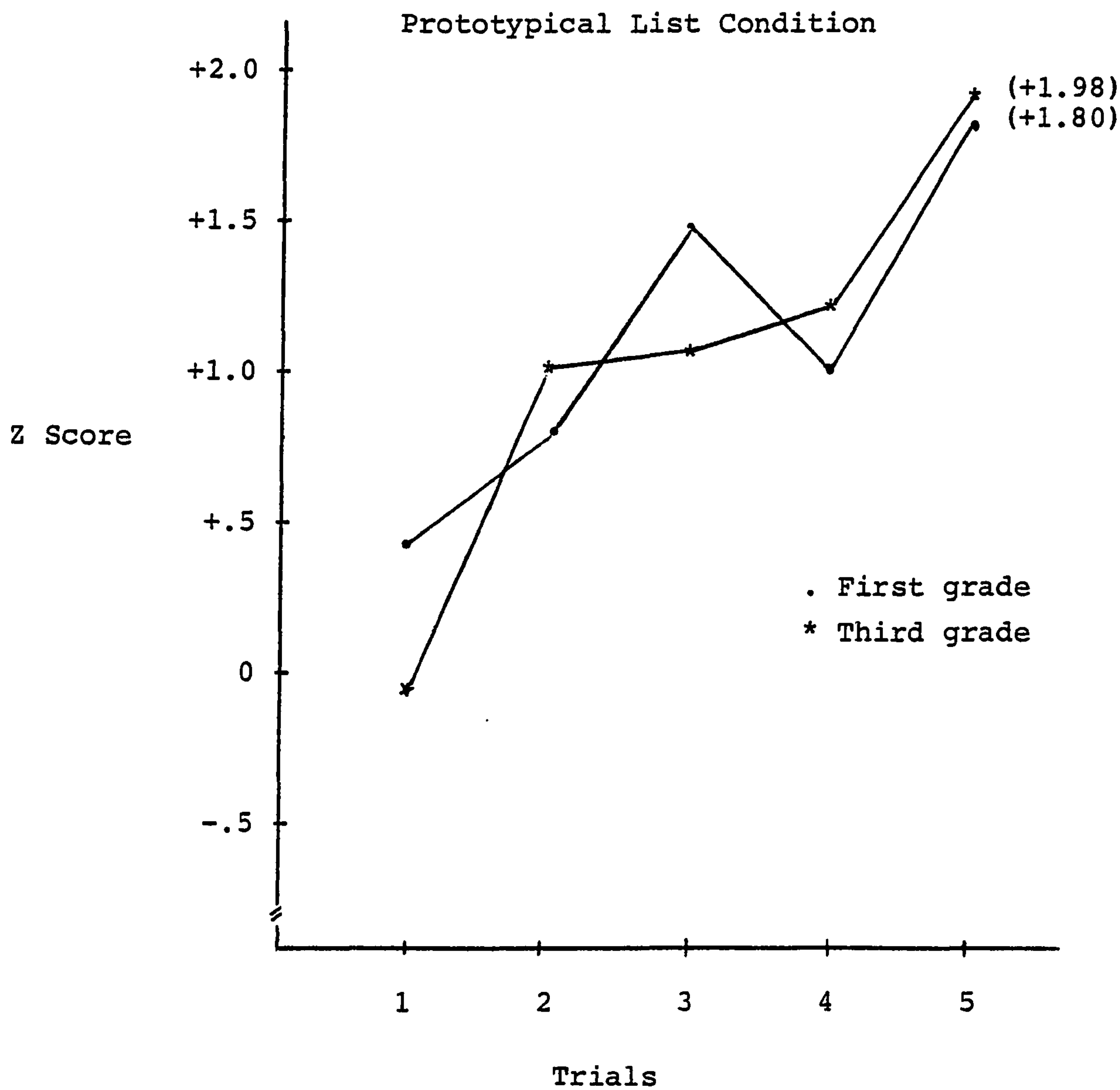

Figure 9. PLC clustering performance by grade as a function of trials. Positive Z scores indicate some degree of clustering; zero or negative Z scores indicate lack of clustering.

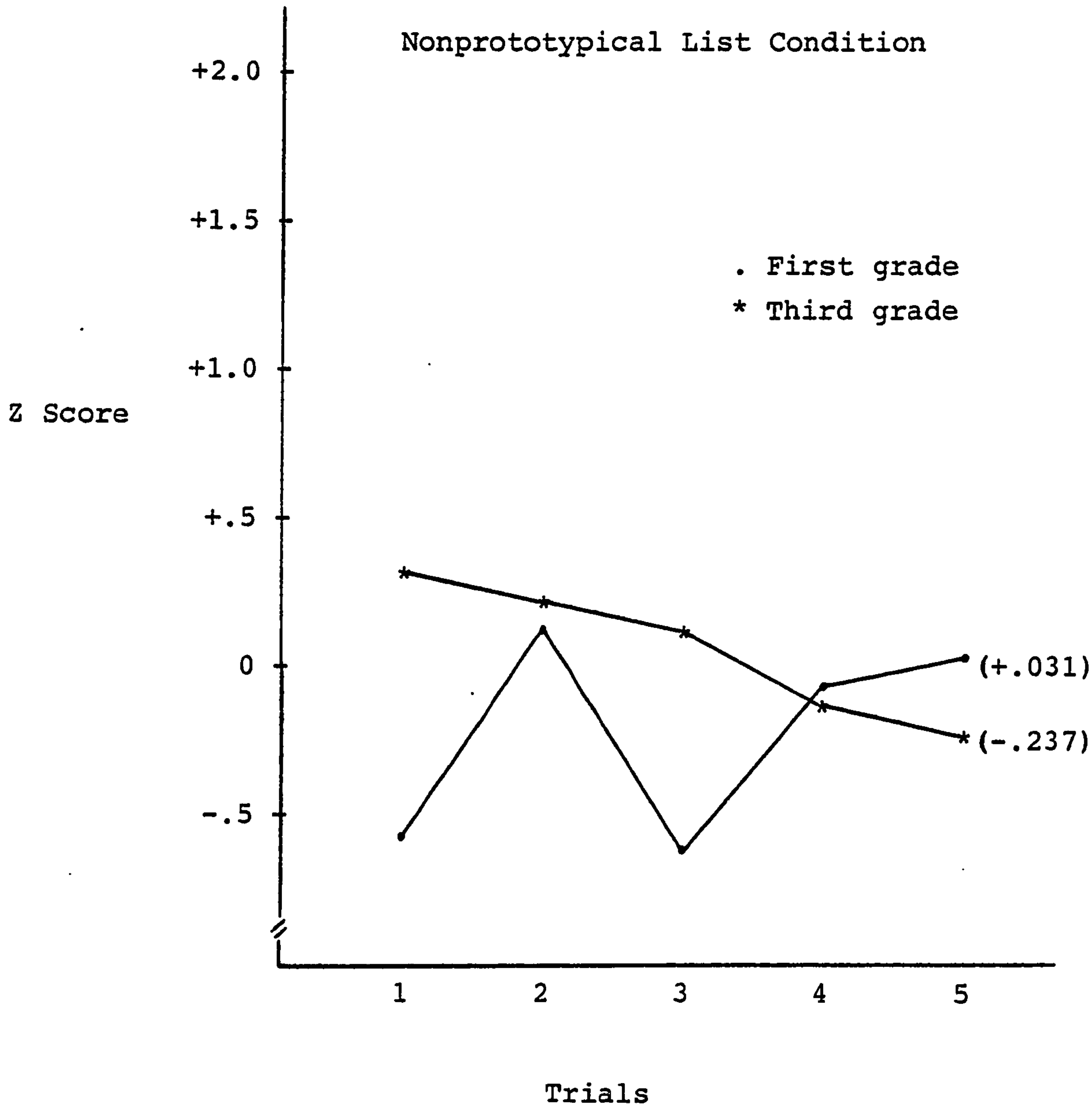

Figure 10. NLC clustering performance by grade as a function of trials. Positive Z scores indicate some degree of clustering; zero or negative Z scores indicate lack of clustering.

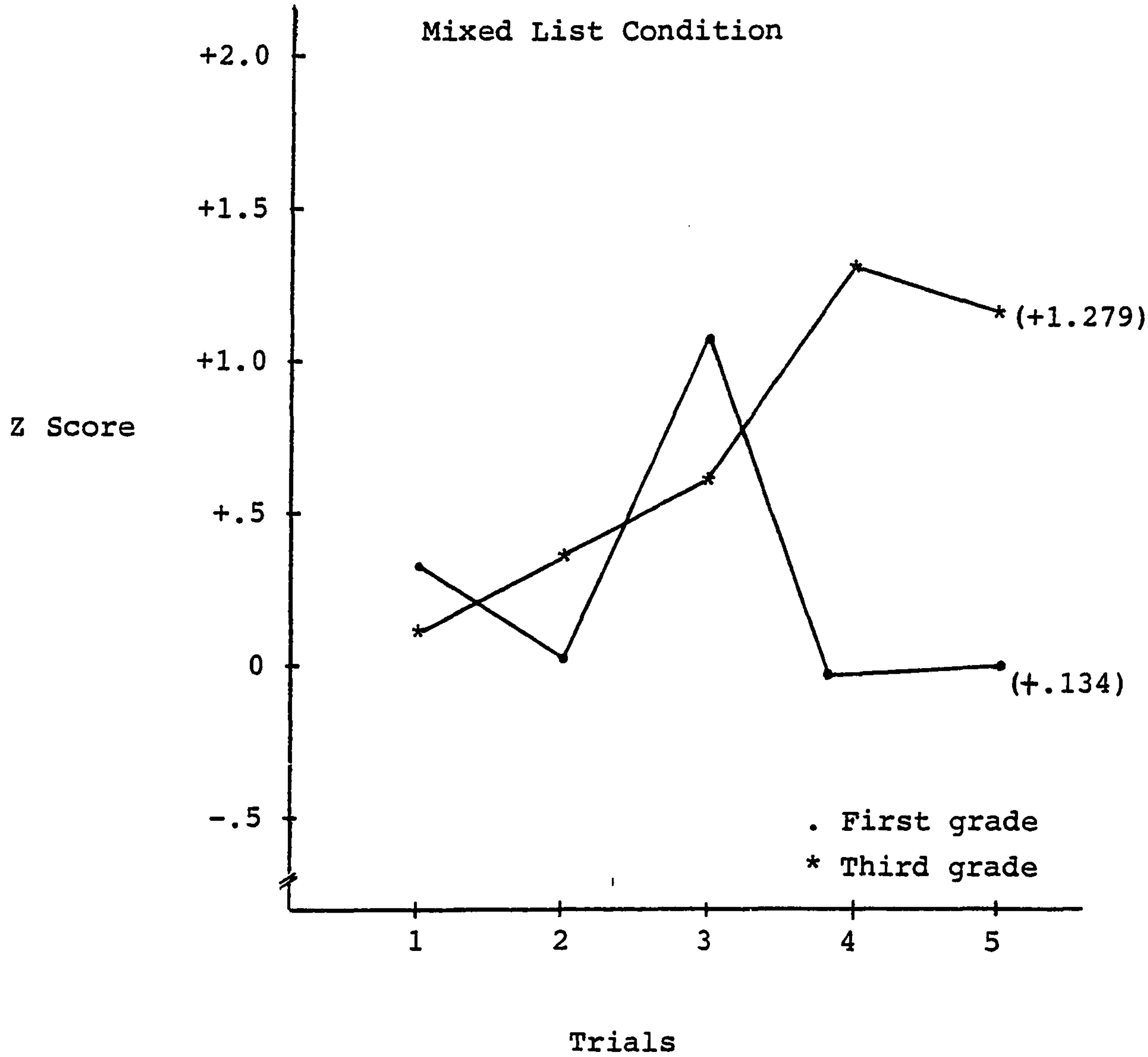

Figure 11. MLC clustering performance by grade as a function of trials. Positive Z scores indicate some degree of clustering; zero or negative Z scores indicate lack of clustering.

GENERAL DISCUSSION AND IMPLICATIONS

The consistency of past free recall developmental findings has resulted in characterizing young children as production deficient. For some, as yet unknown reason, young children do not spontaneously use their category conceptual schemes to aid recall. Their non-spontaneous use of such schemes has been conceptualized as influencing the storage, rather than the retrieval phase of memory processing (Flavell, 1977). That is, categorical organization is viewed as a storage strategy. It follows that the lack of categorical organization in recall indicates that the to-be-remembered items were not stored in a categorical way. Hence, production deficiency of young children is considered a storage strategy problem and most developmental memory investigations have concentrated on the storage phase of memory (Flavell, 1977). Results of Experiments 1 and 2 suggest that the production deficiency interpretation, while perhaps valid, is incomplete and that retrieval strategies, rather than storage strategies, are implicated.

In the category item production task of Experiment 1, both the younger and the older children exhibited (at least for the five superordinate categories used) that their category conceptual schemes were large, diverse, and non-idiosyncratic. Moreover, there were essentially no age differences in number, diversity, and extent of agreement among category members first retrieved from long term memory. Both age groups were able to effectively search and retrieve from long term storage items particular to each category label. These children did not lack the ability to retrieve items from memory, nor did they lack superordinate conceptual schemes as a basis for organizing long term storage.

In the free recall task of Experiment 2, first graders recalled as many items as third graders and organized their recall to the same extent as third graders with an all prototypical list and an all nonprototypical list. By manipulating the typicality of category items, the well established age-related performance differences can be eliminated.

For both ages, the strong relatedness and category membership saliency of age-appropriate prototypical items in the all prototypical list condition resulted in the highest recall and highest clustering of the three list conditions. The mere presence of clustering in children's recall does not prove that they consciously and intentionally used a categorical grouping strategy when storing and/or retrieving those items. (Flavell, 1977). Nevertheless, on a strictly empirical basis, the younger children cannot be labelled "production deficient". Without the usual manipulations, such as specific pre-training, category blocking, or constraining recall, the younger children in Experiment 2 appeared to spontaneously and effectively use categorical organization in recalling prototypical items.

Flavell (1977) proposed that if a strategy, like categorical organization, is effortful, challenging, and attention demanding as an act in itself, then children may have difficulty incorporating it as a subroutine within a larger cognitive program such as a memory problem. The logical and theoretical inference is that prototypical items facilitate detection of the inherent categorizable structure of the recall list and reduce the effort and attention needed to use the categorical structure in recall. Prototypical items in a recall list make the use of a categorical organizational strategy "easy".

In the all nonprototypical list condition neither age group demonstrated any evidence of categorical organization in recall. These less related and marginal category members appear to convert the same "easy" memory task as in the prototypical list condition (i.e., same setting, same instructions, same categories, same presentation rate, and the same number of items and trials) into one which was difficult. Nonprototypical items in a recall list may hinder category detection and/or require more taxing retrieval demands. Strictly speaking, both the younger and the older children in the all nonprototypical list condition can be labelled "production deficient".

These results suggest that production deficiency is not necessarily a characteristic of the child or a characteristic of a particular age. Production deficiency more aptly describes the joint functional outcome of the type of information to be recalled and the nature of subsequent retrieval demands.

The most intriguing age-related result in Experiment 2 occured in the mixed list condition. This type of recall list produced the standard age differences reported in almost all past free recall studies: The older children recalled more items and tended to cluster more than the younger children. These results raise the possibility that past free recall lists were composed of both prototypical and nonprototypical items. A separate free recall experiment employed all procedural aspects of Experiment 2 with two exceptions: The same recall list was given to both age groups and the items for the list were drawn from adult category norms,[11] a selection procedure reported in past developmental free recall studies (e.g., Scribner & Cole, 1972; Vaughan, 1968). Age-related performance differences replicated those found for the mixed list condition in Experiment 2. It was not surprising to find that the adult recall list was composed of both prototypical and nonprototypical items.

These findings strongly indicate that past free recall evidence for production deficiency, while not necessarily artifactual, is biased. The full range of young children's competency was not tapped by past procedures. Typicality, as a performance variable, should be controlled for in future free recall tasks involving natural semantic categories.

Perhaps more important than the possible bias in past developmental free recall studies are the unanticipated findings in Experiment 2 indicating a potential underlying cause of the age-related differences on a mixed typicality list. The older children's performance with this type of list was fairly

---

[11]The 15 item list was composed of the three most frequent items from each of five categories from Battig and Montague's (1969) adult category norms.

similar to their peer's performance with an all prototypical list. In contrast, the younger children's performance was more like their peer's in the nonprototypical list condition. It appeared as if the older children focused on the prototypical items in the mixed list, while the younger children attended to the nonprototypical items. However, the proposed but speculative explanation becomes more complex than this when use of retrieval cue types are considered: The third graders in the mixed list condition had more effective retrieval cues present in their recall responses than first graders; and the first graders had almost twice as many poor retrieval cues than the third graders. These data suggest that the older children's better performance may be a function of their spontaneous use of effective retrieval strategies for recalling items in a mixed list.

Currently, developmental memory studies have examined the storage side of memory strategies, and we know next to nothing about retrieval efforts (Flavell, 1977). Flavell (1977) has stated that there appears to be "no clearly interpretable developmental studies of spontaneous, uninstructed retrieval efforts as yet (p. 210)." The present study confirms that not only more work is needed on retrieval efforts, but also indicates that tasks or paradigms devised to examine retrieval, rather than storage strategies and their interaction with age and typicality may significantly contribute to knowledge of memory development.

APPENDIX A

Ethnic Backgrounds of the Free Recall Children

	School A Percentage	1976[a] Percentage	1975[b] Percentage
Caucasian	10	27.7	27.9
Chinese	21	4.3	5.6
Filipino	9	10.2	10.2
Part-Hawaiian	10	16.4	14.4
Japanese	9	26.6	24.6
Korean	2	1.3	1.5
Mixed, excl. Part-Hawaiian	36	9.2	8.2
Other	2	1.6	1.4

[a]Source: Population Characteristics of Hawaii, 1976 Hawaii Department of Health.

[b]Source: OEO 1975 Census Update Survey: Oahu

Note. Ethnic background was judged from father's surnames and mother's maiden name.

APPENDIX B

Criteria for Defining and Selecting Nonprototypical Items

Step 1. The prototypical items were eliminated from the list of potential nonprototypical items.

Step 2. Established a "distance" measure from the respective prototypical items. The distance measure was calculated in the following manner:

a) Average the frequency within the first three responses and the overall frequency for each prototypical item.

b) Subtract the lowest averaged frequency from the next to lowest averaged frequency for the prototypical items within each category.

Step 3. Eliminate from the nonprototypical pool all items that had a frequency greater than the distance measure found in Step 2b. This insures that the potential nonprototypical items were at least the equivalent distance from the lowest prototypical item as the lowest prototypical item was from the next to lowest prototypical item.

Step 4. Eliminate items which contained category label (e.g., bread<u>fruit</u>), brand names (e.g., Hang Ten), and phrases (e.g., the one to change the tire).

Step 5. Eliminate all items which occurred within the first three responses with frequencies greater than one.

APPENDIX B (continued)

Step 6. Rank order the remaining items on basis of their overall frequencies.

Step 7. Select those items with the highest rank order that were used in past free recall studies with children.

Step 8. When Step 7 was not possible, select the highest rank order items.

APPENDIX C

Adult prototypicality ratings of category items used in Experiment 2

	Category Item	Rank (1 to 30)	Average Score (Scale of 1(most) to 7(least)
	dog	1	1.50
	cat	2	1.66
	tiger	7	2.46
	elephant	19	3.34
Animal	giraffe	23	4.10
	chicken	(ranked 28th in bird category)	
	cow	4	2.30
	mouse	26	4.36
	deer	8	2.56
	fox	13	2.78
	shirt	2	1.36
	pants	1	1.20
	dress	3	1.58
	hat	28	4.42
Clothes	coat	8	2.12
	shoes	12	2.42
	overalls*		
	pajamas*		
	apron*		
	apple	2	1.18
	banana	4	1.70
	orange	1	1.14
	pear	3	1.64
Fruit	lime	22	3.24
	coconut	28	4.78
	raisins	24	4.22
	lemon	14	2.58
	cherries	6.5	2.02
	chair	1	1.24
	couch	3	1.32
	table	2	1.26
Furniture	lamp	15	2.98
	stool	18	3.32
	dresser	9	2.08
	television	27.5	4.06
	cabinet	14	2.80

*These items were not included in Uyeda & Mandler's (1980) items.

APPENDIX C (continued)

Adult prototypicality ratings of category items used in Experiment 2

	Category Item	Rank (1 to 30)	Average Score (Scale of 1(most) to 7(least)
	hammer	2	1.52
	nail	3.5	1.98
	screwdriver	3.5	1.98
	wrench	25	4.40
Tools	axe	27	4.74
	scissors*		
	ruler	10	2.76
	screw	13.5	3.14
	shovel*		

*These items were not included in Uyeda & Mandler's (1980) items.

Literature Cited

Anglin, J. Les premiers termes des reference de l'enfant. In S. Ehrlich & E. Tulving (Eds.), La memoire semantique. Paris: Bulletin de Psychologie, 1976.

Arlin, M., & Brody, R. Effects of spatial presentation and blocking on organization and verbal recall at three grade levels. Developmental Psychology, 1976, 12, 113-118.

Austin, J., Gekoski, W., & Zivian, M. Clustering in recall as a function of age and list structure. Report No. 7, October 1965, University of Michigan, Contract Nonv 5 POI HD 03168-03, National Institute for Child Health and Human Development.

Berlin, B., & Kay, P. Basic color terms: Their universality and evolution. Berkeley: University of California Press, 1969.

Bourne, L. Human conceptual behavior. Boston: Allyn and Bacon, 1968.

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

Brandt, M. A developmental study of children's category concepts: Subcultural influences. Paper presented at Western Psychological Association, Honolulu, May 1980.

Carson, M., & Abrahamson, A. Some members are more equal than others: The effect of semantic typicality on class inclusion performance. Child Development, 1976, 47, 1186-1190.

Cohen, B. Some-or-none characteristics of coding behavior. Journal of Verbal Learning and Verbal Behavior, 1966, 5, 182-187.

Cohen, B., Bousfield, W., & Whitmarsh, G. Cultural norms for verbal items in 43 categories. Technical Report No. 22, 1957, Contract Nonv-631(00), Office of Naval Research.

Cole, M., Gay, J., Glick, J., & Sharp, D. The cultural context of learning and thinking. New York: Basic, 1971.

Cole, M., Frankel, F., & Sharp, D. Development of free recall learning in children. Developmental Psychology, 1971, 4, 109-123.

Doughterty, J. Salience and relativity in classification. American Ethnologist, 1978, 5, 66-80.

Eysenck, M., & Baron, C. Effects of cueing on recall from categorized word lists. Developmental Psychology, 1974, 10, 665-666.

Flavell, J. Cognitive development. Englewood Cliffs, N.J.: Prentice Hall, 1977.

Flavell, J., Beach, D., & Chinsky, J. Spontaneous verbal rehearsal in a memory task as a function of age. Child Development, 1966, 37, 283-299.

Frankel, F., & Cole, M. Measures of category clustering in free recall. Psychological Bulletin, 1971, 76, 39-44.

Franklin, A., Fulani, L., Henkind, E., & Cole, M. Ethnic group differences in organized recall. Unpublished manuscript, Rockefeller University, 1975.

Goldberg, S., Perlmutter, M., & Myers, N. Recall of related and unrelated lists by 2-year-olds. Journal of Experimental Child Psychology, 1974, 18, 1-8.

Hagen, J., Jongeward, R., & Kail, R. Cognitive perspectives on the development of memory. In H. W. Reese (Ed.), Advances in child development and behavior (Vol. 10). New York: Academic Press, 1975.

Hayes, W. Statistics for psychologists. Holt, Reinhart, & Winston, 1963.

Heider, E. R. "Focal" color areas and the development of color names. Developmental Psychology, 1971, 4, 447-455.

Heider, E. R. Universals in color naming and memory. Journal of Experimental Psychology, 1972, 93, 10-20.

Heider, E. R., & Oliver, D. The structure of the color space in naming and memory for two languages. Cognitive Psychology, 1972, 3, 337-354.

Jablonski, E. Free recall in children. Psychological Bulletin, 1974, 9, 522-539.

Jensen, A. R., & Frederiksen, J. Free recall of categorized and uncategorized lists: A test of the Jensen hypothesis. Journal of Educational Psychology, 1973, 65, 304-312.

Kail, R., & Schroll, J. Evaluative and taxonomic encoding in children's memory. Journal of Experimental Child Psychology, 1974, 18, 426-437.

Kay, P., & McDaniel C. Color categories as fuzzy sets. Working paper No. 44. University of California, Berkeley: Language Behavior Research Laboratory, 1975.

Kay P., & McDaniel, C. The linguistic significance of the meanings of basic color terms. Language, 1978, 54, 426-437.

Kempton, W. Category grading and taxonomic relations: A mug is a sort of a cup. American Ethnologist, 1978, 5, 44-65.

Keppel, G. Design and analysis: A researcher's handbook. Englewood Cliffs, N.J.: Prentice-Hall, 1973.

Kucera, J., & Francis W. Computational analysis of present-day American English. Providence, R.I.: Boston University Press, 1967.

Libby, W., & Kroes, W. Conceptual encoding and concept recall-recovery in children. Child Development, 1971, 42, 2089-2093.

Mandler, G. Organization and memory. In K. W. Spence & J. T. Spence (Eds.), Psychology of Learning and Motivation (Vol. 1). New York: Academic Press, 1967.

Mandler, G., & Stephens, D. The development of free and constrained conceptualization and subsequent verbal memory. Journal of Experimental Child Psychology, 1967, 5, 86-93.

Mervis, C., Catlin, J., & Rosch, E. Relationships among goodness-of-example, category norms, and word frequency. Bulletin of the Psychonomic Society, 1976.

Miller, G. The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychological Review, 1956, 63, 81-97.

Mischel, W. On the interface of cognition and personality: Beyond the person-situation debate. American Psychologist, 1979, 34, 740-754.

Moely, B., & Jeffrey, W. The effect of organization training on children's free recall of category items. Child Development, 1974, 45, 135-143.

Moely, B., Olson, F., Halwes, T., & Flavell, J. Production deficiency in young children's clustered recall. Developmental Psychology, 1969, 1, 26-34.

Neimark, E., Slotnick, N., & Ulrich, T. The development of memorization strategies. Developmental Psychology, 1972, 5, 428-432.

Nelson, K. The organization of free recall by young children. Journal of Experimental Child Psychology, 1969, 8, 284-295.

Nelson, K. Variation in children's concepts by age and category. Child Development, 1974, 45, 577-584.

Ornstein, P., Hale, G., & Morgan, J. Developmental differences in recall and output organization. Bulletin of the Psychonomic Society, 1977, 9, 29-32.

Pender, N. A developmental study of conceptual, semantic differential, and acoustical dimensions as encoding categories in short-term memory. Unpublished doctoral dissertation, Northwestern University, 1969.

Posnansky, C. Category norms for verbal items for children in grades 2-6. 1974. Available from Carla Posnansky, Department of Education/Psychology, University of Rochester, Rochester, N.Y., 14627.

Posnansky, C. J., & Pellegrino, J. W. Developmental changes in free recall and serial learning of categorically structured lists. Bulletin of the Psychonomic Society, 1975, 5, 361-364.

Rips, L., Shoben, E., & Smith, E. Semantic distance and the verification of semantic relations. Journal of Verbal Learning and Verbal Behavior, 1973, 12, 1-20.

Rosch, E. On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), Cognitive development and the acquisition of language. New York: Academic Press, 1973.

Rosch, E. Cognitive representations of semantic categories. Journal of Experimental Psychology: General, 1975a, 104, 192-233.

Rosch, E. The nature of mental codes for color categories. Journal of Experimental Psychology: Human Perception and Performance, 1975b, 1, 303-322.

Rosch, E. Universals and cultural specifics in human categorization. In R. Brislin, S. Bochner, W. Lonner (Eds.), Cross-cultural perspectives in learning. New York: Halsted Press, 1975c.

Rosch, E. The good-old-days effect: Misjudgment of frequency as a function of typicality. Unpublished manuscript, 1976. (Available from Eleanor Rosch, Department of Psychology, University of California, Berkeley, California, 94720.)

Rosch, E. Human categorization. In N. Warren (Ed.), Studies in cross-cultural psychology (Vol. 1). London: Adademio Press, 1977.

Rosch, E. Principles of categorization. In E. Rosch & B. Lloyd (Eds.), Cognition and categorization. Hillsdale, N.J.: Lawrence Erlbaum, 1978.

Rosch, E., & Mervis, C. Family resemblances: Studies in the internal structure of categories. Cognitive Psychology, 1975, 7, 573-605.

Rosch, E., Mervis, C., Gray, W., Johnson, D., & Boyes-Braem, P. Basic objects in natural categories. Cognitive Psychology, 8, 1976a, 382-439.

Rosch, E., Simpson, C., & Miller, R. S. Structural bases of typicality effects. Journal of Experimental Psychology: Human Perception and Performance, 1976b, 2, 491-502.

Rosner, S., & Hayes, D. A developmental study of category item production. Child Development, 1977, 48, 1062-1965.

Saltz, E., Soller, E., & Sigel, I. The development of natural language concepts. Child Development, 1972, 43, 1191-1202.

Scribner, S. Developmental aspects of categorized recall in a West African society. Cognitive Psychology, 1974, 6, 475-494.

Scribner, S., & Cole, M. Effects of constrained recall training on children's performance in a verbal memory task. Child Development, 1972, 43, 845-857.

Scribner, S., & Cole, M. Studies of subcultural variations in semantic memory: Implications of cross-cultural research. Unpublished manuscript, Rockefeller University, 1975.

Smith, E., & Medin, D. Representation and processing of lexical concepts. Paper presented at Sloan Conference at the University of California, San Diego, 1979.

Smith, E., Rips, L., & Shoben, E. Semantic memory and psychological semantics. In G. H. Bower (Ed.), The psychology of learning and motivation (Vol. 8). New York: Academic Press, 1974.

Tenny, Y. The child's conception of organization and recall. Journal of Experimental Child Psychology, 1975, 19, 100-114.

Tulving, E. Theoretical issues in free recall. In T. R. Dixon & D. L. Horton (Eds.), Verbal behavior and general behavior theory. New York: Prentice-Hall, 1968.

Tulving, E., & Osler, S. Effectiveness of retrieval cues in memory for words. Journal of Experimental Psychology, 1968, 77, 593-601.

Tulving, E., & Pearlstone, Z. Availability versus accessibility of information in memory for words. Journal of Learning and Verbal Behavior, 1966, 5, 381-391.

Uyeda, K., & Mandler, G. Semantic categories and prototypicality: The effects of prototypical cues on recall. Paper presented at Western Psychological Association, Honolulu, 1980.

Vaughan, M. Clustering, age, and incidental learning. Journal of Experimental Child Psychology, 1968, 6, 323-334.

Wachs, T., & Gwen, G. The effects of chronological age, trials, and list characteristics upon children's category clustering. Child Development, 1971, 42, 1217-1227.

Wagner, J. A developmental study of categorical organization in short-term memory. Unpublished doctoral dissertation, University of Connecticut, 1970.

Wallace, W. Consistency of emission order in free recall. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 58-68.

Wallis, W., & Roberts, H. Statistics: A New Approach. New York: Free Press, 1957.

Wittgenstein, L. Philosophical investigations. New York: Macmillan, 1953.

Yoshimura, E., Moely, B., & Shapiro, S. The influence of age and presentation order upon children's free recall and learning to learn. Psychonomic Science, 1971, 23, 261-263.

Zadeh, L. Fuzzy sets. Information and Control, 1965, 8, 338-353.

Zadeh, L. Quantitative fuzzy semantics. Information Sciences, 1971, 3, 159-176.