DOCUMENT RESUME

| | | |
|---|---|---|
| ED 070 063 | 24 | CS 000 274 |

| | |
|---|---|
| AUTHOR | Alexander, Ella M. |
| TITLE | Task and Intra-Task Differences in Vocabulary Performance. |
| INSTITUTION | Wisconsin Univ., Madison. Research and Development Center for Cognitive Learning. |
| SPONS AGENCY | Office of Education (DHEW), Washington, D.C. Bureau of Research. |
| REPORT NO | WRDCCL-TR-186 |
| BUREAU NO | BR-5-0216 |
| PUB DATE | Dec 71 |
| CONTRACT | OEC-5-10-154 |
| NOTE | 83p.; Report from the Project on Reading and Related Language Arts Basic Prereading Skills: Identification and Improvement |
| EDRS PRICE | MF-$0.65 HC-$3.29 |
| DESCRIPTORS | Language Arts; *Learning; Paired Associate Learning; *Reading; Semantics; *Testing; Verbal Ability; *Vocabulary; Vocabulary Skills |

ABSTRACT

Two vocabulary tasks--one production and one recognition--were compared with the expectation that the recognition task would yield better performance than the production task. The pairs of pictures used in the recognition task were divided into eight groups defined on target and distractor frequency and same-different conceptual category membership with the exception that these groups would differ in relative error rate. Not only was the task difference confirmed, but evidence of considerable variability between test items was found, with a particularly significant effect involving category relationship. (Author)

Technical Report No. 186

# TASK AND INTRA-TASK DIFFERENCES IN VOCABULARY PERFORMANCE

Report from the Project on Reading and Related Language Arts Basic Prereading Skills: Identification and Improvement.

By Ella M. Alexander

Richard L. Venezky, Principal Investigator

Wisconsin Research and Development
Center for Cognitive Learning
The University of Wisconsin
Madison, Wisconsin

December 1971

This Technical Report is a master's thesis reporting research supported by the Wisconsin Research and Development Center for Cognitive Learning. Since it has been approved by a University Examining Committee, it has not been reviewed by the Center. It is published by the Center as a record of some of the Center's activities and as a service to the student. The bound original is in The University of Wisconsin Memorial Library.

Published by the Wisconsin Research and Development Center for Cognitive Learning, supported in part as a research and development center by funds from the United States Office of Education, Department of Health, Education, and Welfare. The opinions expressed herein do not necessarily reflect the position or policy of the Office of Education and no official endorsement by the Office of Education should be inferred.

Center No. C-03 / Contract OE 5-10-154

## STATEMENT OF FOCUS

The Wisconsin Research and Development Center for Cognitive Learning focuses on contributing to a better understanding of cognitive learning by children and youth and to the improvement of related educational practices. The strategy for research and development is comprehensive. It includes basic research to generate new knowledge about the conditions and processes of learning and about the processes of instruction, and the subsequent development of research-based instructional materials, many of which are designed for use by teachers and others for use by students. These materials are tested and refined in school settings. Throughout these operations behavioral scientists, curriculum experts, academic scholars, and school people interact, insuring that the results of Center activities are based soundly on knowledge of subject matter and cognitive learning and thay they are applied to the improvement of educational practice.

This Technical Report is from the Basic Prereading Skills: Identification and Improvement element of the Reading and Related Language Arts Project, in Program 2, Processes and Programs of Instruction. General objectives of the Program are to develop curriculum materials for elementary and preschool children, to develop related instructional procedures, and to test and refine the instructional programs incorporating the curriculum materials and instructional procedures. Contributing to these Program objectives, this element has two general objectives: (1) to develop tests for diagnosing deficits in skills which relate to reading (2) to develop a kindergarten-level program, including diagnostic tests and instructional procedures, for teaching basic prereading skills. Tests and instructional programs will be developed for visual and acoustic skills, including letter and letter-string matching with attention to order, orientation and detail, and for auditory matching, segmentation, and blending.

ACKNOWLEDGEMENTS

To Dr. Gilchrist who first believed. To Dr. Calfee whose ability to inspire is unexcelled. Particularly to Dr. Chapman whose dedication to thoroughness and details epitomizes scientific inquiry. And especially to Dr. Parke whose bravery, willingness, concern and competence in assuming another's burden will be gratefully remembered. And finally to all those past, present, and future who help to keep the faith.

CONTENTS

## LIST OF TABLES

LIST OF FIGURES

## ABSTRACT

Two vocabulary tasks--one production and one recognition--were compared with the expectation that the recognition task would yield better performance than the production task. The pairs of pictures used in the recognition task were divided into eight groups defined on target and distractor frequency and same-different conceptual category membership with the exception that these groups would differ in relative error rate. Not only was the task difference confirmed, but evidence of considerable variability between test items was found, with a particularly significant effect involving category relationship.

# I

# INTRODUCTION

Vocabulary tests are used today mainly as part of general diagnostic or evaluative testing packages, but are generally thought to have little to say about language other than how many or which words a person knows. However, the basic considerations which go into the design of vocabulary tests are questions which are basic to language. Therefore, a consideration of task differences and the underlying theoretical decisions which they represent could be informative. The recognition-production distinction has not been substantiated for small children or apart from other mechanical skills such as reading and writing. As important is the need to determine whether differences in test items can cause variability in performance. The latter is of concern not only because of implication for strategic processes in performance, but because of implications for the nature of word meaning and semantic growth.

Calfee, Chapman and Venezky (1970) found from the vocabulary section of a reading diagnostic package that a large proportion of the errors (34% for line drawing, 43% for picture naming) were intra-class confusions. These included such errors as "penny" for "half dollar", "spider" for "bee", "goat" for "cow". The authors note that within sets of stimuli these category confusions could have resulted either from the difficulty of specific categories or differential intracategory confusability or both. On the basis of the data no decision could be made between these two interpretations, but this kind of evidence indicates that proper word usage may in some sense depend on the ability to deal with the categorical relationship of item-referents. However, there was no discernible relationship in their study between sorting behavior and labeling performance by category. Further exploration of vocabulary studies for evidence of category confusions could be a starting point to a better understanding of vocabulary and general language ability.

# II

# BACKGROUND LITERATURE

## Early Vocabulary Testing

Early interest in vocabulary testing originated from concern with its use as a diagnostic tool in educational situations. The inclusion of vocabulary sections in the major intelligence tests and the high correlation between overall intelligence and verbal ability (Terman found a correlation of .90) which has been validated repeatedly, established the use of vocabulary tests as abridged intelligence tests and effective diagnostic tools for educational evaluations. The wide variety of uses include a method for evaluating the relation between vocabulary and school grades and vocabulary and major subject, an instrument for student classification and grading, a tool for assessing the proportion of the vocabulary of early readers which was familiar to the child, a qualifying examination for college entrance, a device to build vocabulary. However, statements about the depth, range, and size of word knowledge have varied greatly throughout the history of this area (Dale, 1931; Hartman, 1941; Colvin, 1951) and eventually led to a concern with word knowledge per se. Kirkpatrick (1907) was one of the first to try to determine how many words individuals of different ages and grades know, followed by others who investigated vocabulary performance as a separate and important phenomenon. But later and more in-depth research revealed larger descrepancies among vocabulary size estimates. So that while the period from 1900 to 1950 saw the greatest amount of research and discussion on the extent of vocabulary knowledge, Colvin (1951) observes that the most remarkable fact about the results was the singular lack of unanimity or general agreement among investigators, for estimates of size varied from a few thousand to 20 times as large.

The importance of accounting for such discrepancy is evident. If results from different types of studies are to be compared and tests themselves are to be used for evaluative purposes, we must assume that the measures of word knowledge are reliable and valid indicators of the same phenomenon. But as important is the fact that the unaccountable variability in results indicates the lack of clear and firm conception of the phenomenon. And this conception which is very basic to language description, is of interest to many who have other than a pragmatic interest in general language development. But particularly for linguists and psycholinguists who are concerned with describing language systems, defining linguistic units and understanding language processes the clarification of such a basic concept as word knowledge seems important.

While the reasons for the wide range of discrepancy are complex, it would seem that whether the concern is absolute size or relative range of knowledge much of the discrepancy in results stems from vague and various conceptions of the phenomenon itself--namely, the nature of word knowledge--which consequently lead to differences in testing techniques. A consideration of certain theoretical issues indicate that task and intratask variability might reasonably contribute to discrepant results. These considerations are not new but in fact seem to originate from early efforts to account for variability in vocabulary studies. The questions which were raised are, however, basic to a real understanding of the development and use of language and have not as yet been authoritatively resolved. Perhaps, then, an investigation of sources of discrepancies in vocabulary testing can lead to a better conceptualization of the nature of vocabulary ability.

Basic Conceptual Issues

Early investigators of vocabulary ability raised questions which indicate that differences in the notion of word knowledge are widespread and occur at all levels of analysis. Furthermore these questions continue to be problems even to those who have less than a strictly pragmatic interest in language.

One such consideration revolves around the word itself. At a basic level Larrick (1954) points out that researchers must come to know what a word *is*. In line with this thinking Kelley (1932) noted that theoretically the role of a word vis-à-vis a referent can be a symbol or it may represent the total fullness of meaning which is associated with it through experience. In its narrowest terms this is a contrast between a simple S-R characterization of the word stimulus and a much larger conception i.e., whether a word is simply a symbol or some objective reality covers a full range of meaning. The recent emphasis has been on the very broad role of words. Kaplan (1967) suggests that even if at some point the word functions solely as a symbol this is simply a first step to a larger conceptualization. Brown (1958) insists that even the most basic linguistic forms are categories, and describes a word as a container of meaning or a category of attributes which, by defining the important distinctions and equivalences of the culture, conveys the total expectancies of that culture. Empirically, the question still remains along with others which focus on the word unit--such as how much meaning is necessary before a word in defined and how do we define a word as a unit of measurement. Experimentally, Larrick (1954) asks, are we to count basic and derivatives, singular and plural words as different units (Thorndike, for example, was inconsistent in this regard). Seashore (1933) cites the failure to define "word" as a source of the enormous variability of size estimates, and clearly such unit specification is essential. Unfortunately, these questions have not been dealt with authoritatively so that even among scholars there exists confusion and misunderstanding about how to deal with a concept which is basic to the problem of interest (Hurlburt, 1949). And the criticism which Dale had in 1931 when he stated that a great deal of data relative to the adding of suffixes and prefixes is needed before the question of the specificity of testing can be settled, is still va[illegible].

Other considerations relevant to vacabulary performance concern the problem of defining knowledge. Dale (1931) asks directly "What do we mean when we say that a child knows a word?" or what is the nature of knowing. And closely connected with the matter of knowing is the problem of meaning, for a person is assumed to know as a item has meaning for him. Three kinds of questions arise around this issue--how much or what range of meaning must a word have before it is sufficiently defined, which behaviors reflect word knowledge, how should one deal with the fact that knowledge and meaning change over time.

All available evidence indicates that in regard to vocabulary skills, knowledge is a relative construct. For example, Kelley (1932) notes that between the extreme views of word as symbol and word as fullness of meaning lie all degrees of meaning for different individuals and concludes that the various degrees of meaning manifested show that the ways in which a word may be known can range from absolute certainty to some vague and doubtful acquaintance. In conjunction, Chambers (1904) declares that these degrees of meaning may represent different levels of accessibility which correspond to levels of knowing, those levels ranging from words which are clearly known to those which are completely inaccessible. Consequently, Cuff (1930) and Larrick (1954) note that while the ability to give one meaning of a word is often taken as knowledge this is certainly no indication of a thorough acquaintance. Hartman (1941) and Gansl (1939) cite differences in the degree of acquaintance with a word which is required in testing as one important consideration in the ambiguity of vocabulary estimates.

Experimentally, definitions of "to know" have included the ability to define a word, use it in a sentence, recognize or illustrate a situation in which the term is appropriate, recognize one meaning from several definitions, to check a word as "known" or "unknown" (Seegers and Seashore, 1949). But the question remains how can we be sure which behaviors do indeed reflect word knowledge. Furthermore, of all the ways in which knowledge can be demonstrated, which are the most efficient and direct? Which techniques best represent a subject's demonstration of word knowledge? The question has not been answered satisfactorily. Recently, however, Brown (1958) has proposed that even at the most elemental level, word knowledge, by its very nature, manifests itself in two distinct abilities--the ability to react to the word as a sign of the referent and the ability to identify new instances of a concept not labeled before. Brown's position is that understanding is a disposition rather than a distinct behavior and that word knowledge can manifest itself in a great variety of ways. Therefore, a stimulus-response model cannot possibly predict the precise behaviors which knowledge of a word will generate. However, since each name category is a recognition of the true character of the referent, evidence of knowledge (disposition to behave correctly), can be obtained directly from behavior in regard to labels, and these sorts of tasks are superior even to sorting or grouping tasks in showing sufficient evidence of an understanding of the meaning of words. And **particularly the ability to <u>name</u> referents seems** "fundamental in creating the full disposition to respond which is ultimately the only conception of meaning with which psychologists can legitimately deal."

However, the problem of measuring knowledge and meaning is complicated by the fact that both of these change over time. Dolch (1936) in answer to the question of how much meaning constitutes knowledge notes that word meaning grows continually, changing from vague familiarity to a full and exact concept. Therefore, one must recognize stages in meaning development. This is substantiated by Jersild (1940, cited in Hurlburt, 1949) who notes that developmentally there seem to be accretions to meaning throughout life and that a child's mastery of language develops not only by adding "new" words, but by an increased understanding of "old" ones.

> In much of what is presented to the child, the problem is not so much one of complete mastery as opposed to complete ignorance but rather one of varying degrees of understanding . . .(for) a certain amount of vagueness and unfamiliarity is practically inevitable during the early stages of a child's first contact with certain terms. For a time many terms are likely to be more or less meaningful or meaningless. Meanings are likely to become more comprehensible as the child makes further contacts with the term in different contexts . . .

Empirical evidence comes from Chase (1961) who found that definitions of words could be placed into at least three developmentally progressive conceptual classes. Cronbach (1942) notes that this growth and development of concepts is gradual and that the concepts which most words signify are still not complete in adulthood, therefore testing should determine the degree to which a subject's understanding is complete rather that whether he knows or does not know a word.

Most of the above considerations point to the problem of dealing with more and less meaning, and a need for an understanding of the development of knowledge and meaning. Recent efforts to deal systematically with this problem have come from the area of psycholinguistics. One basic assumption which has been characteristic of the views in this area is that perhaps people acquire the important elements of language in much the same way in which linguists describe unfamiliar languages. Therefore, an understanding of basic linguistic principles is useful to those who would understand human language usage.

There are two important ways in which linguists characterize the meaning or semantic component of a language system. One system involves representing the conceptual system of a language as a branching tree or meaning hierarchy, in such a way that each branch or marker is composed of the defining attributes represented by all the labeled modes above it. This characteristic makes the trees redundant and demonstrates the hierarchical semantic relationship between concepts (table 1). On the other hand, the distinctive feature system is more concerned with the specification of the important defining attributes or semantic features which describe a specific concept. A semantic feature table (table 2) presents a list of what is seen as the important dimensions along which all elements in the system can be defined and each item is then described as having either one or the other features of the attribute dichotomy. Such a system unlike the previous one specifies the criterial semantic features of the concepts. And since the set of semantic features is seen as constituting a large part of the meaning of a word, when words share meaning their feature sets are said to overlap. Some concepts (such as opposites) may be separated by a single distinctive feature.

Linguistic componential analysis attempts to define graphically the features which both associate and separate specific words. But how do these descriptions compare with what we know about label acquisition? There is no empirical literature directly relevant to the semantic structure of children. There are, however, some pertinent theories. Referent naming is described by Brown (1958) as the most deliberate part of first language acquisition and it might be expected that it could be taught directly. But even at this very basic level Brown stresses the categorical nature of word meaning and insists that as a word is, indeed, a category of semantic features a child must learn a word, not simply as a referent symbol, but he himself must form some conception of the categorical nature of that referent.

Vocabulary or label acquisition proceeds most directly by the naming game. The tutor in this "original word game" names things in accord with community custom, but since the meaning of a word extends beyond a single or several instances and since the criterial features of that concept are usually not explicitly stated, the tutee hypothesizes about the categorical nature of the referent to which a name is given. So that, the simple act of naming helps to establish a semantic schema of the word onto which many congitions can be fit. The semantic scheme, is, however, not completely imposed from the teacher but rather is formed and reformed by the student as he generates and tests hypotheses in response to the linguistic and nonlinguistic behaviors of others. In this way he "checks the accuracy of the fit between his own categories and those of the player (and) . . . improves the fit by correction." The point to be made is that the child plays no passive role in language acquisition and simply naming objects for him does not insure the establishment of the two behavioral dispositions (identifying instances of a label and reacting to words as a sign of the referent) which signify understanding. He must form the referent category. This implies that the language learner is continually revising his use and understanding of words in order to fully realize the defining attributes of a referent, and that long after a label is acquired the concept itself may still be incomplete.

Furthermore, in early development the word itself is seen as an attribute of the referent category. That is, the label is considered just one more of the features which define that concept. Brown cites Vygotsky (1939), for example, who observed that for children the name of an object is inseparable from and is given the same conceptual weight as functional and other defining characteristics. If indeed this is the case it is logical that items which share other characteristics might also, during the course of development, be perceived by the child as sharing a linguistic one. And this accords very well with Brown's observation that children often overgeneralize in their use of words, that they apply the same word to a great variety of referents--even those which are linguistically distinguishable. (An example is that of a child who uses the word "aunt" to refer to his aunt, his mother, and the maid.) It seems then, that in the process of coming to form the referent categories associated with a label a child overextends the meaning of a word. And while he may appear to have some of the general criterial features (like six for the example of "count"), more restrictive features might not yet be realized.

Evidence for such a position, while limited, comes from such a study as Calfee, Chapman, Venezky (1970) who found that 43% of the errors on a naming task were due to intraclass confusions of the nature of "baby" for "doll", "crib" for "bed", "spider" for "bee", and "goat" for "cow" (the errors on the labeling tasks were not related to object sorting errors for that category). However this result has been established for no other kinds of tasks and such a validation might add substantially to the position.

There are others who theoretically support this position. Anglin (1970) believes in the genetic character of a word, and in the fact that it denotes a group of referents rather than a single event. However, unlike Brown, Anglin chooses to emphasize the fact that individual word categories are systematically related to each other, and to investigate whether the process of semantic development proceeds through generaliztion to more abstract categories or differentiation from more abstract levels to more concrete ones. He, then, does not dwell on the semantic makeup of a particular word.

Anglin's use of the term generalization implies differentiation of items at a certain level of specificity or within a category as a prerequisite while Brown's use of the term is without this implication. And in fact Anglin acknowledges that Brown's observations on the overgeneralized use of words accord more with those of Lashley and Wade (1946) who found that dimensions which define a concept do not exist for an organism until it has had a chance to compare various stimuli that differ along the relevant dimension. Therefore, what Brown calls "abstraction before differentiation" is not the same as what Anglin investigates as "abstraction after differentiation".

From these two emphases we can describe meaning both as that which accrues to the word itself and the semantic relationship between words at different levels of abstraction. But whatever the interpretation, the basic theoretical questions which have been raised are involved in decisions which precede the operationalization of vocabulary tests. The possibility that knowledge of a word differs according to the demands of the situation, i.e., that knowledge exists on different levels with different degrees of accessibility has important consequences for the interpretation of vocabulary estimates. For since different methods might reflect different levels of understanding, estimates based on these responses will differ according to the facets of knowledge which they reflect.

Attention to differences in conception should lead almost immediately to concern with testing techniques. As each experimenter decides for himself how he shall deal with what it is to know or the extent of meaning, task should become an obvious and important consideration, especially in comparisons of results. However, concern with procedural differences has been limited, and most often vocabulary performance has been taken as a simplex variable which in turn fostered the assumption that all tests yield essentially the same information--how many or which words a subject knows. Thus performance definitions of knowledge have included the subject's use, recognition, discrimination, association, and definition behavior and traditional testing methods have included word counts from natural and induced speech situations, naming tasks, selection tasks, free association tasks, word association tasks, without consideration for how these procedures might activate different components of vocabulary skills or reflect vocabularies which are qualitatively different. So that, historically, the question of task has been largely ignored despite occasional evidence that other procedural variables affect vocabulary performance.

In her review, Colvin (1951) pointed to the great disparities in speaking, writing, and reading abilities as a source of the inconsistency in results and noted that different testing procedures measure a different type of vocabulary.

Bryan (1955) showed that vocabulary performance could be affected by the time of year tested, the geographical area and the response situation. By testing over a wide geographical area, during different seasons of the year, and by using multiple response situations he was able to increase substantially the estimates which had been previously accepted.

Dolch (1936) in his extensive survey of vocabulary size and range in young school children acknowledged the necessity to specify how much meaning constitutes knowledge and found that estimates of the size of vocabulary differed.

However, the most far-reaching impact on the methodology of the field was generated by Seashore (1933) who showed that for college students many size estimates were underestimates (sometimes by as much as 10%) if the size of the dictionary from which words were sample was increased. Smith (1941) substantiated these findings with school children. But as Hartman (1941) pointed out, even when the same method of dictionary sampling was used, no two procedures yielded the same values. So that while these findings generated a great deal of discussion and focused attention on procedural variables there has been little reference to the fact that techniques have ranged from checking or marking words known (Kirkpatrick, 1907; Babbitt, 1907), written and/or oral definitions (Doran, 1907), written records kept by subjects of all the words us[illegible] in conversation or writing (Brown, 1911), combinations of checking and defining (Starch, 1916), defining least familiar words (Gerlach, 1917, cited in Hurlburt, 1949), to writing sentences for words known (Brandenburg, 1918). And despite the evidence that procedural variation is reflected in differential performance the vocabulary distinctions that have been made revolve around content (scientific, historical, et al.) or mechanical skill (reading, writing, speaking). Rarely have task demands been acknowledged (Dale, 1931; Kelley, 1932; Hurlburt, 1949).

## Theoretical Considerations in the Choice of a Task

But if indeed task demands are important, on what kinds of considerations should task differences be operationalized or investigated and which tasks reflect important theoretical differences for vocabulary test results? In what ways can we answer the question of what it means to know a word? Two possibilities are to ask: What are the central abilities in vocabulary knowledge, and which of the demands which is usually placed on vocabulary knowledge is to be emphasized in an assessment of word knowledge. Tasks considerations must necessarily proceed from decisions based on such questions.

Brown has suggested two abilities as basic to vocabulary knowledge. These abilities also correspond to what are considered basic cognitive abilities. If it is true as has been suggested (Hurlburt, 1949) that vocabulary performance is composed of different skills, different tasks might presumably activate different combinations or components of such skills. And since vocabulary ability is regarded as one of the higher mental abilities (Watts, 1944) those abilities which are general to other kings of cognitive tasks might be functional in regard to vocabulary performance. The production-recognition, production-comprehension distinction is widely acknowledged as characteristic of much intellectual functioning. Such generally dichotomized intellective activity might also be apparent for vocabulary performance, and tasks which discriminate between these skills seem important considerations for vocabulary test results. Not only because they substantiate the relationship between vocabulary and general cognitive functioning but because they reflect the skills which are seen as basic, sufficient and necessary demonstrations of what it means to know a word.

Secondly, different tasks might tap different components of vocabulary knowledge. In any conception of vocabulary knowledge one must deal primarily with the word or the referent as the focus of the response though it is the word-referent relationship which is the essence of vocabulary. There is limited evidence that these can exist independently (Crosscup, 1940; Brown, 1958; McGuire, 1961) and perhaps they generate different kinds of performance. It cannot be assumed that a subject has knowledge of a word until he can correctly associate it with the proper referent (Seashore, 1933) and all definitions of knowing have emphasized the relationship between word and meaning for they are not theoretically separated in vocabulary performance, but one possible difference in the designing of tests is the relative weight given the two components. In our concern with task differences then it seems reasonable to select tasks which differ in concern with the word as opposed to object component of the word-referent relationship. Production tasks seem to reflect greater concern with word while recognition-discrimination tasks seem to emphasize the referent.

Certainly also such tasks should be as empirically valid or realistic as possible with regard to the uses of vocabulary knowledge in human situations (Seegers and Seashore, 1949). Dale (1931) reasons that our problem as researchers is to determine what reaction to words the environment can legitimately demand of the individual at every age level, for it is this reaction which determines whether a word is known to an individual. Two demands on vocabulary ability at any age (Brown, 1958) would seem to be the need to respond differentially by recognizing the meaning of words produced by others and the need to produce words for others which signify proper referents. Furthermore, these two abilities according to Brown. are the only manifestations of the "click of comprehension" with which psychologists can legitimately deal. Tasks which focus on these two different demands reflect important situational differences in word knowledge.

In sum, then, two kinds of tasks (production, recognition) seem particularly appropriate for testing task differences because in other areas they have been thought of as indicative of different performance processes and have yielded differential results, because they seem to emphasize different components of the word-referent relationship and because they seem to tap the essential demands of the environment on language ability. Thus the production-recognition difference would seem to be an important theoretical distinction as well as a performance distinction. Taking a cue from results in other areas we expect that recognition task performance will be superior to production task performance (Luh, 1922; Postman and Rau, 1957, both cited in Jung, 1968, show that measures of retention of verbal units are lowest with production procedures and highest with recognition procedures). This finding has been reported for such language components as phonology (Fraser, Bellugi and Brown, 1963; Maccoby and Bee, 1965) and morphology (Lovell and Bradbury, 1967; Lovell and Dixon, 1967) in addition to the substantial verbal learning data. Task seems to be an important variable in differential vocabulary performance. And we would expect that to the extent that these tasks reflect some basic vocabulary abilities and focus on different aspects of the vocabulary phenomenon they signal, in fact, two types of vocabulary.

## Previous Studies of Task Differences

There have been some investigations of task differences, but they have concentrated on the relationship to mechanical skills and have necessarily involved older children or adults.

Symonds early (1926) attempt to measure the size of recognition and recall vocabularies yielded a recall vocabulary one-third the size of the recognition vocabulary. However, his test was measuring not word knowledge per se, but the specialized ability to react to written sumbols of those words (reading) and is therefore inappropriate for testing at very early ages.

Seegers and Seashore (1949) cited evidence that for college students, "use" vocabularies, i.e., those words which an individual can define or illustrate in a sentence, are approximately 92% as large as recognition vocabularies (again reading recognition). And he concludes that if an adult knows a word by one criterion he is very likely to know it by other criteria so that we are not justified in specifying different types of vocabulary. But he noted that while there was great overlapping among the types of vocabularies for college students, this is not necessarily true at earlier ages.

Hurlburt (1949) reports that high school students are able to recall and write only 45% of the words they are able to recognize and associate the correct meaning with. But again these results are reflecting specialized mechanical abilities, as most of these investigators were not interested in word knowledge except as it was related to some literate skill.

Recently (1957) Templin distinguished between a vocabulary of use (based on the Seashore-Eckerson task) and a recognition vocabulary (based on the Ammons Picture Recognition task). But because each task was performed by a different age group, there is the possibility of confounding of type of task with age. And, thus, there has not been an investigation in small children of differential word knowledge as a function of task demands.

Perhaps much of the reason why task differences have not been investigated more extensively is that once obvious sorts of task differences have been demonstrated there is little inclination for repeated replication. However, the present consideration seems necessary since there has not been established the task difference in small children apart from other mechanical skills such as reading or writing abilities, and to validate the fact that issues which are basic to the vocabulary conception are reflected in tasks which yield differential performance.

Intratask Variability

In addition to task variability, performance differences might also be expected to depend in part on the characteristics of the word items composing the test, since we expect that certain characteristics of the words themselves might be associated with the degree or level of knowledge. Word frequency has been associated with performance in a variety of verbal response situations. Underwood and others in the field of verbal learning have established a relationship between word frequency and verbal learning abilities. Howes and Solomon (1951) found that the duration for which a printed English word must be presented visually to a subject in order for him to recognize it is inversely correlated with the frequency of occurence of the word in large samples of written English, i.e., the perceptual threshold is lower for words of high frequency. Solomon and Postman (1952) controlled the frequency on nonsense units and found the same inverse relationship between recognition thresholds and frequency of prior usage.

Hall (1954) found that within limits the more frequently a word appears in the language the more readily it is learned (recalled). The relationship between frequency and performance was further substantiated when Jacobs (1955) reported a correlation of .74 between Thorndike-Lorge values and correct responses on a P-A list. Furthermore, there is some evidence that frequency is associated with the semantic development. Entwhistle (1967) found that the syntagmatic-paradigmatic shift depends on the form class and word frequency. It seems, then, that the preference for associating words on a conceptual rather than a syntactic or grammatical basis is related to the frequency with which the word is used. In view of this evidence it is not inconceivable that words of higher frequency (words that have greater occurence in written and spoken language) will yield better performance in a vocabulary test.

Furthermore, a discrimination paradigm (such as will be used in our study) based on comparison of paired items, differs from some other types of tasks in that a large part of performance difficulties may be due to confusion of items (Kirkpatrick and Cureton, 1949). They write, "The difficulty of a multiple choice vocabulary item for a given group of subjects is dependent on two main factors: First, the percent of the group that could define the word correctly if asked to state its meaning and, Second the degree of discrimination required to distinguish between the correct answer and the incorrect answers, or decoys, in the item. The importance of this second point has often been overlooked with unfortunate results." So that in addition to word frequency there should be a variable which reflects differential confusability between items and thus affects vocabulary performance. Category membership is seen as such a variable. Without being able to specify exactly the nature of the similarity which same category items imply, we expect that the probability for confusion is greater between these items because of that relative similarity and increased possibility for confusion. On the other hand we would expect that words which are more dissimilar (in terms of category membership) and therefore more easily distinguished will yield higher performance on a given task. Word frequency and category membership should be variables within a task which affect performance. Aside from one investigation of the relative difficulty of lists of words within a test (Thorndike and Symonds, 1923) there has been little attention given to the matter of intratask differences in vocabulary tasks. But basically we, like Gansl, (1939) credit discrepancies in vocabulary results to the variance in item makeup as well as the operationalization of what it means to know a word.

By choosing different levels of target and distractor frequency (high and low) and varying category membership (same as target or different), we can investigate whether these characteristics are important to performance in a vocabulary test. We predicted that the factor which will be important in a discrimination task (in order of their importance) are frequency of target, frequency of distractor and category of distractor and that when these items are arranged from least important to more important (fastest to slowest moving factor) there should be a corresponding increase in performance. We are suggesting that not only the task, but the construction of the test items can have important consequences for vocabulary performance.

# III

# METHOD

## Design

A task (2) x order (2) factorial with repeated measures on the first factor was used to test the proposals. Preschool Ss were asked both to name color photographs of objects (the production task) and to select from a pair of pictures the photograph of an object named by E (the recognition task). Half of the Ss received the production task first; the other half received the recognition task first.

In the production task, a 2 x 2 design was employed with two levels of category (same as target, different from target) and frequency (higher, lower). In the recognition task, the 5 x 2 x 2 x 2 design consisted of the following variables: category of target item x frequency of target item (higher, lower) x category of distractor item (same as target or different) x frequency of distractor item (higher, lower).

## Stimuli

Twenty words designating common objects were selected on the basis of conceptual category membership, word frequency and picturability (easily and readily photographed). The items were selected from five categories which had been shown to yield a number of confusion errors in a naming task (Calfee, et al., 1970). The categories were: Insects, furniture, clothes, toys, and tableware. For this study, two high frequency and two lower frequency words were chosen to represent each category.

The criterion for high frequency category exemplars was that they be listed as one of the thousand most frequent words from the first grade sample of the Rinsland count (Rinsland, 1945), while lower frequency exemplars appeared in the second to fourth thousand of the Rinsland ranking [with the exception of one less frequent word--"pitcher" which was not listed in the 4,000 most frequent words in the Rinsland count, but appeared in the Thorndike-Lorge and Murphy counts (against which all items were compared to check for consistency) and met the other criteria]. An effort was made to keep absolute frequency rank comparable across categories. The items selected for each category, together with their frequency ranks appear in Table 3.

From these twenty word items two sets of stimuli were constructed. The first consisted of 5" x 3 1/2" individual color photographs of the twenty items selected; each object was photographed against a plain background. The second set consisted of eighty pairs of photographs made up from copies of the twenty original items of the first set. One item in each pair was designated target item (that object which matched the label supplied by E) and the other was designated distractor. For each category there were 16 pairs of items comprising four subgroups, with each subgroup constructed on the basis of frequency and category membership. In the first subgroup of a given category each of the four items of that category occured paired with a distractor of the same category and same frequency. In the second subgroup of a given category each of the four items were treated as target items and were paired with distractors of the same category but of different frequency level. The third and fourth subgroups were constructed in the same manner as the first two except the distractor items were drawn from different categories rather than the same category. Within the total eighty pairs, every item appears four times as target and four times as distractor with the left-right occurence of the target randomized.

Procedure

Ss were randomly assigned to one of the two task orders. Half received the production task first, the other half received the recognition task first. In the production task, pictures were presented individually in a pre-determined randomized order and Ss were asked to name them. Exact verbal responses were recorded by E. If S failed to respond after he had been asked twice to identify the stimulus this was scored as "no response." "No response" and "I don't know" were listed as separate responses.

In the recognition task, Ss were assigned to one of four pre-determined orders of the eighty pairs of photographs and asked to point to the picture showing the object which E named. The four list orders consisted of four different Latin square permutations of blocks of twenty pairs. Each block of twenty pairs was derived by sampling one item from the first of the four sub-groups of items from a given category and one item from the second subgroup of another category and so on until twenty pairs were obtained: these pairs containing each of the twenty items occuring as targets once, and representing each of the four subgroups from each category once. Each set of twenty was then randomized and the four permutations or orders of the four randomized blocks obtained.

Each S was presented, by blocks of twenty, the entire set of item pairs with instructions to indicate ("show me") which of the pair members was that one which E named. E simultaneously circled on a scoring sheet that object to which S pointed.

The responses were scored for the number of items correctly identified (named) and the number of items correctly discriminated (recognized). In accord with general procedure, the criterion for a correct response on the production task was that S respond to the stimulus with the label as designated. Correct responding on the recognition task was simply the selection of that object-picture named by E.

Subjects

Twenty-four four and five year old pre-schoolers attending three local child development centers[1] served as subjects. Mean age was 4 yr. 11 mth., with a range of 3 yr. 11 mth. to 5 yr. 8 mth. All were children of working mothers, but were representative of diverse social backgrounds--ranging from professional to blue collar. There were 12 boys and 12 girls.

[1]Special thanks is due to Mrs. Matthews and staff of Child Development Incorporated for their patience and cooperation.

IV

# RESULTS AND DISCUSSION

## Analyses of Variance

Three analyses were performed on the data. The first, performed on the total design, was an order (2) x Ss (24) x task (2) repeated on measures analysis which revealed a significant task difference, $F(1, 22) = 136.01$, $p < .01$, with performance in favor of the recognition-discrimination task[2]. Order of presentation also proved significant, $F(1, 22) = 13.16$, $p < .01$, with an overall error rate on the first order of 10% while the error rate for second order was 5%. However, the effect of order seems to be specific to one task. The recognition task across both orders yielded 96% mean correct responding. On the production task however, those receiving the second order (discrimination first, production second) had an average of 83% correct responding, while those receiving the first order (production first, discrimination second) averaged 64% correct responding. In addition there was a task x order interaction $F(1, 22) = 22.88$, $p < .01$ reflecting the production task--second order performance.

The second analysis was run on the production data only with an order (2) x Ss (24) x category (5) x frequency (2) factorial. The effect of order on production task performance was substantiated as well as a task x order interaction since this analysis revealed differences in performance

[2] Though one or the other of these terms may be used, depending on the emphasis, they refer to the same task.

on the production tasks by order to be significant $F(1, 22) = 18.44$, $p < .01$. In addition the difference in frequency between production task items was significant $F(1, 22) = 35.49$, $p < .01^{*}$, with high frequency words yielding a mean of 77% correct and low frequency items yielding 48% mean correct responding. And finally there was a significant category x frequency interaction $F(1, 22) = 7.78$, $p < .05^{*}$ with category 2 (furniture) showing a difference of 50% in performance between high frequency and low frequency items, while category 1 (insects) showed a difference of 5%.

The third analysis was performed on the recognition task data. A target category (5) x target frequency (2) x distractor category (2) x distractor frequency (2) analysis revealed a significant main effect on this task of target word frequency $F(1, 23) = 15.027$, $p < .01$, target item category membership $F(4, 92) = 4.79$, $p < .05$ and distractor category membership (same as target or different) $F(1, 23) = 4.492$, $p < .05$ with same category items yielding 93% correct responding, different category yielding 97% correct responding. The expected distractor frequency effect was not significant. There are three significant interactions--target category x target frequency $F(4, 92) = 9.364$, $p < .01$; distractor category x distractor frequency $F(1, 23) = 10.895$, $p < .01$; and target frequency x distractor frequency $F(1, 23) = 5.250$, $p < .05$.

The evidence strongly supports our contention that a critical factor in assessing word knowledge is the task situation, while providing some partial support for intratask differences.

Task Differences

The task difference as predicted is in favor of the recognition-discrimi-

*Based on the Geiser-Greenhouse correction for nonindependence.

nation task which yields an error rate 22% lower than that for the production task, but in terms of our data the cause of this difference can only be speculated. There has been very little research into the difference between production and recognition abilities and that which is available deals with the developmental lag between the two (Maccoby and Bee, 1965; Olson and Pagliuso, 1968) as opposed to differential processes. But in addition to the conceptual distinctions already cited, three sets of performance or process factors distinguish the tests used in this vocabulary study and these differences can be interpreted in favor of the recognition task. The tasks differ in the type of response required, in the amount of information provided per stimulus set and in the type of information to which the subject must respond. In the production task the child is required to produce a different (and relatively complex) verbal response for each of the twenty items. In the recognition task S is asked to indicate his recognition of an item by the same (and relatively simple) pointing response. In the production task the single stimulus item is the sole basis for making a response. In the recognition task a compound pictorial and label stimulus set provides three important types of information: what is being requested (the target), what is not being requested (the distractor) and what is possibly being requested (the response alternatives). And finally, in the production task we ask if a subject knows a word as a response; in the recognition task the emphasis is on the referent as a response. The relative difficulty of responding with these elements may be indicated in the task difference. The difference between a verbal and pointing response, between stimuli which do and do not define the response alternatives, between requiring knowledge of a label and of a referent--all are factors which distinguish these tasks. And while we cannot say conclusively if any or all of these distinctions are important to final performance it seems clear that for four- and five-year olds, tasks with these differentiating characteristics result in a performance difference. The cause(s) of this difference remains an empirical question.

In considering further the source of task differences the significant order effect could be informative. The transfer from production task to the recognition task seems negligible--the discrimination task scores are the same regardless of order (96% correct for both orders)--but the change from the discrimination task to the production task causes considerable and significant improvement in performance (from 64% to 83% correct responding). Apparently some factor of the discrimination task helps to elevate the level of production task performance. But how does the order effect elucidate the causes of differential performance by task?

There are several ways in which factors peculiar to the discrimination task may benefit performance on the production task. If the subject comes to the production task lacking knowledge of some or all of the elements of a particular vocabulary item (a label, a referent, the label-referent association) the recognition task through a repeated and contingent presentation of items and a restricted choice situation provides an opportunity for these elements to be acquired. But most likely, [since our stimuli (referents) were chosen to be common and familiar, and because label-referent associations can be established only indirectly by S from the discrimination task presentation], the deficit in the production task is due to lack of ability to articulate, lack of knowledge of or memory failure with respect to some label. However, as this label is supplied by E during recognition testing it becomes a part of the subject's immediate response repertoire, and is available for subsequent production task performance. It is possible, then, that E provides on the recognition task as a stimulus an item which is the primary response which the child makes on production, i.e., E supplies the label. This might account for the fact that 50% of the errors on production for first-order presentation were eliminated in second order presentation. And therefore, the task difference might reflect differential ability in regard to labels.

One alternative to the preceding explanation should be dealt with here. We have suggested that the recognition task places limitations on response possibilities. This fact increases the likelihood for guessing correctly and raises the possibility that the performance difference is due mainly to this fact. Therefore, in order to determine the validity of this position it is necessary to correct the recognition score by a guessing factor. The classical formula assumes that the error rate under a two-choice situation represents 50% of the guessing rate. Doubling the error rate and subtracting from total possible correct yields a recognition task score of 92% mean correct for the production task. So it seems that though there may be an increased likelihood of guessing correctly on the recognition task, it is not an adequate explanation of the task difference.

For four- and five-year olds, then, vocabulary performance on a recognition task is superior to production task performance. And it seems that a task which requires a relatively simple indication of knowledge, provides a relatively greater amount of information, and requires a type of information which focuses on the referent will result in higher vocabulary performance. Furthermore, we are suggesting that the transfer from the recognition task to the production task is perhaps the result of presenting a label which the subject then utilizes on the production task, and that therefore the label could be a critical factor in task difference. But whatever the cause, the difference in vocabulary performance by task is confirmed, and to the extent that the tasks emphasize different vocabulary components and tap different types of knowledge we are willing to speak of separate production and recognition vocabularies in small children.

Intratask Differences in the Recognition Task

We predicted that not only task differences, but intratask differences would be associated with differential vocabulary performance. The particular predictions that were made stem from two conceptions of performance on the recognition task, namely selection or choice strategy and the processing of information bits. Strategically, it would seem that the two pivotal concerns of a subject in approaching the recognition task are how familiar he is with the individual choices and how confusable they are likely to be. It is not unreasonable to conceive of the main task in a choice situation as in some sense a matching task. The subject it would seem must match some internal conception with the available stimuli. The more certain that conception the simpler the choice-- a given image matches or it does not. So that if the S is relatively familiar with the object (that he is able to form a consistent and stable conception of the target on the basis of the label provided) there should be few errors. As the subject becomes more uncertain about the target, the conception becomes more vague and to the extent that this conception influences subsequent behavior, errors should increase.

A second concern of the subject is the distractor. After forming some conception of the target the subject is required to consider both choices (a target and a distractor) to determine which one best matches the internal conception. If the subject is uncertain of the identity of the alternative it could be a much more viable distractor, create a much more equivocal choice situation and depress performance on that item. But the more familiar the distractor, the less likely it is to be erroneously selected as the target item and the better the relative performance for that item. So that as the subject is less familiar with the distractor the uncertainty associated with his choice should lead to an increased error rate, but as the S is familiar with both the target and distractor errors should remain small.

And finally, since the subject is forced to choose between the two alternatives, and must compare the choices and decide which best fits the conception associated with the label, the extent to which the alternatives can be confused should be important to performance. Generally, the more similar items are, the more likely they are to be confused, and the more likely they are to be confused, the greater the possibility of error. It is expected, then, that test items which share common features should be more confusable than items which don't, and that, therefore, there should be a relationship between semantic similarity and performance.

As the subject focuses on the target, the distractor, and the comparison of these alternatives it would seem that target frequency, distractor frequency and distractor category relation are variables which would reflect these concerns. Frequency has been interpreted as familiarity in many verbal learning paradigms (Kausler, 1966) and by definition items which are in the same category share many more common features than items which are not. that, as the frequency of the target is high we expect that the subject is more likely to make a correct choice than if it is low. Within both lower and higher levels of target frequency, high frequency and different category distractors should be less distracting--high frequency because if S is certain of the identity of the distractor, he is equally sure that it is not the referent for the label stipulated, and different distractor category because as items are less similar they should be less confusing. If S is not certain of the identity of the alternative it becomes a viable distractor, but given that he knows the target and that target and distractor are representative of different categories, errors should remain small. Furthermore we expect that items in which both target and distractor are low in frequency and highly confusable will lead to more errors in choice than items which are low in frequency, but are not from the same category. Therefore, the performance of a S on this word discrimination by item pairs is expected to depend on the certainty of the conception of target and distractor and the confusability of the choices.

If the order of importance of word characteristics in such a strategy is correct (if frequency of target is more important than distractor category) the level of performance should be the result of an additive relationship between item characteristics and we should be able to rank pairs by combinations of these factors and predict relative error rates. The predictions were then that target frequency, distractor frequency and category of the distractor will differentially affect performance in that order of importance.

Another approach to the conception involved merits consideration. In information processing models it has been assumed that the quality of stimuli can influence the performance on a task (Sternberg, 1969, for example), and even though these studies have dealt with the rate of performance there is the underlying assumption that at a given point in time a stimulus can be more and less informative. In our task the stimulus items are seen as information units which also have this characteristic of being more or less informative. The two picture stimuli combined serve as a unit source of information which, to the extent that characteristics of the items convey information, can be quantitatively varied through the independent manipulation of such characteristics. So that by selecting and combining pictures which represent different levels of target frequency, distractor frequency and category membership the stimuli can be made to yield differing amounts of information to be utilized by the Subject in task performance. It is expected, then, that items in which the characteristics represent the upper levels of the experimental variables will yield more information than combinations based on lower levels. And though this analysis does not indicate the relative importance of the characteristics in determining performance it suggests that whatever strategy or process is involved, intratask differences should result, simply because different units yield different amounts of information. Together these conceptions yield expectation of intratask differences based on changes in the levels of each variable, independently, which allow manipulation of item difficulty. These models were seen merely as useful conceptions to guide the search for intratask differences rather than hypotheses or precise models to be proven or disproven. However, if these conceptions are correct the performance can be seen as a simple function of the factors defining the stimulus items. Table 8 contains a summary of the makeup of the items and predictions of relative performance rank.

In summary, the underlying assumptions are that in response to the label stimulus the subject forms some conception of the target item. He then compares the available choices in order to match them to the internal image or conception. Having made a choice the subject then indicates his response. Given these assumptions we can systematically order pairs on the basis of familiarity and similarity and predict relative error rates. Figure 8 indicates the expected relative distribution of errors over test items. The results are as in figure 9.

Both target frequency and distractor category are significant main effects, while distractor frequency is not. It was found that same category items yielded almost twice the number of errors as different category items (figure 3), while targets of lower frequency yielded five times the number of errors as high frequency (figure 1), these results clearly supporting the hypotheses. Furthermore, though distractor frequency was not a significant main effect, two of the three significant first order interactions involved distractor frequency. And finally while the predictions themselves break down in strict application to the data the overall pattern as shown by figure 9 is much as predicted.

In addition to the main effects, two of the significant interactions support the expectations partially--target frequency x distractor frequency and distractor frequency x distractor category --while the target category x target frequency was not anticipated.

The target frequency x distractor frequency interaction reflects the fact that high target--high distractor combinations are better distinguished than high target--low distractor combinations, while low target--high distractor pairs are not different by performance (figure 4). The interaction is based on what seems to be the differential effect of distractor frequency. Table 8 shows that the difference in high and low frequency distractors under targets of low frequency is 0%, while under targets of high frequency yielded 1% error and distractors of low frequency yield 3% error, a difference of 2%. So that, the position that the frequency of the distractor will be differentiating is supported for high frequency, but not for low frequency targets. It is not clear why high frequency distractors are not facilitative when paired with low frequency targets, but perhaps under circumstances in which there is uncertainty about the identity of the target, high frequency distractors which are more familiar become more distracting and depress the level of performance to that of low-low item pairs (figure 7).

The distractor category x distractor frequency effect seems to stem from the fact that low frequency distractor items were discriminated better if they were also of a different category from the target--as was predicted. On the other hand, high frequency items were discriminated slightly better if they were of the same category which clearly contradicts the hypothesis. An examination of the number of errors indicates that same category distractors of low frequency yield more errors than those of high frequency--almost twice as many--while different category distractors of low frequency yield fewer errors than those of high frequency--less than half as many. Table 7 shows that the difference in performance by distractor category is 1%; the difference is 4% under low frequency distractors. So it seems that the difference in distractor category is substantiated for low frequency distractors, that low frequency items are not as distracting if they are in different categories, but that different category distractors are more distracting than same category distractors if they are of high frequency (figure 5).

The results concerning stimulus characteristics are partially supportive of the predictions. Clearly target frequency and distractor category are important to the makeup of the test items and to test performance. And though a statistical statement of the relative importance of these variables is not possible from our data, the graphical layout indicates that target frequency separates the upper half of the distribution from the lower half with distractor category differences showing up within the upper and lower levels of the first variable. And furthermore, if we compare the difference between the binary components of each variable (the difference between high and low frequency target means should be higher than the difference between same and different category means) we find that the target frequency difference = 1.75, distractor category = .87. This suggests that target frequency supercedes distractor category in importance. And furthermore, while distractor frequency is not significant, it is a differentiating factor for low frequency distractors and for high frequency targets. However, the very presence of the interactions suggests that a model which sees performance as a simple additive function of factors defining the items is too simple to handle the complex cognitive factors involved. But the extent to which the subject is able to form some conception of the stimulus and to distinguish similar referents will be important to word knowledge.

The significant target category x target frequency effect, though not predicted, seems to reflect the fact that in the two most difficult categories (furniture, tableware) the low frequency items yielded many more errors that those of high frequency, while in the other three categories the difference is substantially smaller. This is probably due to the fact that these categories contain the word for which there was no occurence in the Rinsland count (pitcher) and is thus probably of disproportionately low frequency, and a word which seems a regional alternative to that used by these subjects (sofa). Both these low frequency responses yielded a great many errors thus exaggerating the frequency effect for those categories of which they were a member. It is possible that the elimination of these items will eliminate the significant effect (the difference in error rate between the extreme categories is halved, for example, when these items are eliminated). However, since this effect was not central to the hypotheses these analyses were not extended.

However, frequency does not account totally for the significant target category effect. We had predicted no category effects largely because of the very common items, but clearly some categories as a group yield better performance than others. The rank of categories by total performance is (best to worse) .

| | |
|---|---|
| Insects | 98% correct |
| Toys | 98% |
| Clothes | 97% |
| Tableware | 95% |
| Furniture | 93% |

This cannot be due solely to the frequency of individual items included in these categories for the average frequency of the items for these categories ranked by mean frequency (highest to lowest) are:

| | |
|---|---|
| Toys | 1425 frequency class |
| Furniture | 1500 |
| Clothes | 1750 |
| Insects | 1850 |
| Tableware | 1987 (pitcher counted as 4,500) |

The Spearman rank correlation coefficient, $r_s = .25$ is not significant. So that while a couple of low frequency words exaggerated the difference in performance by items within category, it cannot be assumed that these items account for the category effect.

In the sum we can say that target frequency and categorical relationship of the items are important within task variables while distractor frequency seems to be ancillary. In addition, the category from which items are drawn can be important to performance under the discrimination task.

Intratask Differences in the Production Task

Furthermore, while we had predicted no intratask differences within the production task and though it was not designed for that purpose, it is obvious that such differences also appear in the production task. There is a significant frequency effect within the production task with high frequency words yielding a mean percentage correct which is 30% higher than that for low frequency items. There is also a significant category by frequency interaction, with the data indicating that for some categories the difference between high and low frequency words is greater than for others. The three categories showing the greatest difference by frequency are: furniture, clothes, and toys.

The only firm statistical support for intratask differences in the production task is for the frequency effect. And it can be shown that half the errors on the production task across both orders were caused by 4 low frequency words: sofa, pitcher, rattle and skirt. Three-fourths of the errors on production were caused by 6 low frequency and two high frequency items: sofa, pitcher, rattle, skirt, blouse, bee, dress, and spider. Thus the frequency effect is relatively well substantiated for both tasks, and the influence of frequency can be shown to extend across tasks as well. If we investigate on the production task the difference in performance by items and order we find that the items showing 50% of the difference in number correct on production--order I compared with production--order I are:

pitcher
blouse
rattle
skirt

all low frequency words. So that both the more difficult items and those which seem to benefit most from second order presentation are of low frequency. It seems, then, that the main difficulty on the production task is caused by low frequency words and that whatever aspect of discrimination performance which is facilitative to the production task affects these items.

To substantiate further the effect of frequency across tasks it should be noted that the probability of getting an item correct on the discrimination task given that it is correct on the production task is .98 (and most of these items were high frequency). But given that items were incorrect on production (most of these were low frequency) the probability of getting them correct as targets and getting them correct as distractors on the discrimination task is still a high .81. Clearly low frequency words are handled better on the discrimination task than on the production task.

There is the possibility that the couple of "odd" items account for the difference in task performance by frequency. By eliminating "pitcher" and "sofa" and reevaluating errors, we find that half the errors on production are still caused primarily by low frequency words--rattle, skirt, blouse, bee while those words which showed no errors were mostly high frequency--bed, butterfly, shoe and spoon. Furthermore, half the change in performance across orders are on three low frequency words--blouse, rattle and skirt--whereas those showing no change were those mentioned above. So that clearly the frequency effect is not an artifact of the "odd" items chosen, low frequency words are differentially handled by task and high frequency words generally yield better performance.

Though the frequency and category x frequency effects are the only quantitatively distinct factors within the production task, a qualitative analysis of the responses is illuminating both of the order effect and the difficulty of producing. The errors on the production task were classified into several categories--same class, nominally descriptive, functionally descriptive, no response, "I don't know" stimulus specific, miscellaneous superordinate, subordinate.

For three of the four items showing the greatest difference across tasks the large number of incorrect responses on the production task were same class errors. For the other item (rattle) the largest number of errors were evenly divided between descriptive and "wild" or miscellaneous responses. 43% of the errors for these four items were same class, 15% were no response or "I forgot", 13% were stimulus specific, 9% were miscellaneous, 7% were functional descriptions. The proportion of same class errors for these items is even higher than for the total distribution (figure 10). And even when "pitcher" and "sofa" are eliminated as odd items we find that among the three items which caused the most (50%) errors and benefited most from the order of presentation , the percentage of same class responses is about the same as before. Even with pitcher and sofa eliminated same class errors on the "difficult" items constitutes 44% of total errors.

Implications for Theoretical Accounts of Word Meanings

These findings open up the possibility of a new interpretation of results. If we assume that rather than a deficit in knowledge of labels, performance on the both tasks reflects confusion in the use of labels and in the appropriateness with which labels which are known are applied to referents, we could account for some of the results and give support for the semantic feature or categorical interpretation of word meaning. This assumption does not seem unreasonable. If indeed the main problem in the production task was not lack of labels but improper categorical definition of the concept, and since different labels can represent overlapping attributes, we would expect that though the subject might not respond with an exactly appropriate label it should be within a certain range of similarity--perhaps within the same class of items. This was clearly the case. Neither is it surprising that this type of error should occur in a greater proportion for low frequency words since these words which are less familiar might be expected also to be those whose attributes are less clearly defined. Furthermore the fact that children are willing to assign same class words to such items suggests some readiness on their part to generalize in use of labels. But of what explanatory value is the categorical position in respect to our data and how does it fit the overall data profile?

Same class errors on the production task would seem to substantiate Brown's observation that children overgeneralize in their use of words. That as a matter of fact even though adults realize that there are many referents which, though similar, are linguistically distinguishable, the child very early grasps the fact that a word is a category and in fact exaggerates this principle. So that early in vocabulary acquisition children treat in their usage of words things which are linguistically distinguishable as equivalent.

The words which caused the greatest number of errors were low frequency words. But again the errors for these items were mostly same category which means that the subject even on the production task had some notion of the defining features of that item and saw the similarity between it and other referents included in that category.

However, the fact that same category confusions occured for the discrimination task is impressive because it indicates the general influence of this variable in word knowledge. For example, if this effect had occurred for production only we could say that the subject simply could not organize or recall the specific label for responding and simply chose a similar but more familiar label. The fact that discrimination between same category items is also difficult suggests some confusion in understanding the difference between semantically similar items so that in addition to difficulties with processes involving organizing or remembering the label, there is the possibility of misunderstanding the semantic range of the label.

We had suggested that the discrimination task is easier because by narrowing the response alternatives the probability of correct responding is increased. The fact that a subject continues to make same class errors under these conditions indicates that he still finds it difficult to match referent and labels and to distinguish item attributes by label and this suggests an incomplete understanding of the concepts represented by these labels.

However, given that class confusions are the main difficulty on the production task how is it that a discrimination task deals with these type of errors better, or of what explanatory value is the category effect in regard to the task effect? There are two possibilities. Since the discrimination task is a restricted choice situation the limitations on alternatives can lead to better performance, and the fact that a subject knows or is quite familiar with a distractor will lead him to make a correct choice. But the guessing rate is relatively low and the target frequency x distractor frequency interaction suggests that distractor frequency is not differentiating for low frequency targets for which such a consideration would be most appropriate.

The other possibility is that if a child is overgeneralizing in his use of a word and if the label which the experimenter supplies is indicative of similar semantic criterial features to the label which he would have used (same class errors on production would indicate this) then it is possible that the subject does or can logically include the picture referent with other similar referents under some broad categorical usage of the term. Thus discrimination performance would be superior because the subject can overgeneralize in his use of a label.

We are saying that, as a matter of fact, it might be easier to deal with semantic features on a discrimination task than on a production task. If the primary problem in word usage for the child is overgeneralization, the response given on production might be erroneous for that particular referent category. However, if on the discrimination task the label supplied by the experimenter can for the child cover a wide range of referents, the subject need only determine to which referent his categorical use of the stimulus label would apply. The overlapping features of attributes of the referent category and the label category supplied by the experimenter will lead him to make a correct choice. However, as the choices themselves come to have overlapping attributes (such as with same-category items) correct responding becomes more and more difficult. For as the subject is asked to discriminate between items which share many of the same criterial features in his repertoire, linguistic differentiation will be impossible. So that while the discrimination task is the easier one, the process of discriminating concepts linguistically appears to be a difficult one at this age. Perhaps erroneous usage of terms is more likely to show up on a production task but this does not preclude the same kinds of misunderstandings on a discrimination task. And if it is possible to have the same names but not the same categories, as McNeill says and Brown implied, the same category errors are an indication that while the child has many of the same names adults do, these names do not define the same categories. Therefore, it would seem that the child of 4 and 5 is still in the process of testing hypotheses concerning conceptual categories. And to the extent that the same class errors are those which are affected by order, the difference between tasks could be attributed to differential handling of categorical cues. However, this interpretation is not posited to the exclusion of contextual and process variables, but is suggested as another means of dealing with a complex phenomenon.

As has been suggested before, there is the possibility that the deficit in discriminating on the production task stems from confusion or problems in mental storage rather than in understanding or knowledge. That is, because these words are associated in experience and sorted together in memory, the confusion is in output or process rather than in knowledge. Anglin's studies which deal with the relationship between words in terms of grouping and sorting behavior suggest that not only are words which are similar placed together in a sorting task, but also in free recall, and his data suggest that shared features may play a role in the organization of responses in recall for adults, <u>but much less so for children</u>. So that most of the errors here are not performance, storage or recall problems (response organization) but are actual confusions of use. The fact that children's words are more often associated in terms of occurence (such as in grammatical patterns) rather than on a more semantic basis is substantiated by Entwhistle (1967). She and others who have investigated the developmental aspects of the syntagmatic-paradigmatic shift show that words which are associated together for children are based more on syntactic relationships than semantic ones. Thus memory and other such factors might be a secondary contribution to the outcome. The conclusions from this study are seen as semantic.

The category effect substantiates the categorical nature of the word as representing a collection of features. For if the concept is not fully formed it is most likely to be differentiation of similar items, same category items, those items which have common features which will be most difficult.

The category effect has other explanatory values. We have suggested before that the order effect stems from supplying the label on the discrimination task which is subsequently useful to the subject on the production task. However, the categorical interpretation modifies that position to suggest that we are not simply supplying a label (the subject is able to generate many closely appropriate ones) but a more proper, more restrictive, less generalized usage of a label category.

The categorical position can also be associated with a possible interpretation of the frequency effect. There is as yet no satisfactory explanation of the tendency of higher frequency words to yield better performance. The facilitative effect of frequency on memory processes and encoding processes is probably the primary explanation. However, the frequency effect in the discrimination task suggests another consideration. Jersild (1940) states that meaning is enhanced through contact with a term in different <u>contexts</u>. Werner and Kaplan (1950) found that <u>S</u>s 9 - 13 yrs. progressively assigned a meaning to artificial words which were embedded in various sentences. Such evidence indicates that the learning of reference involves learning the semantic markers of a word--the senses that it has or the contexts into which it fits--and, thus, the constraints on the conceptual range. It is possible that the more frequently a word occurs the more likely it is to occur in different contexts and the more quickly the semantic features for that concept will be established. Thus the role of frequency might very well be to accelerate the rate at which features are added to the concept (it should be noted that the syntagmatic-paradigmatic shift which signals the association of words on a semantic basis is dependent in part on word frequency).

A final thrust of support for the categorical position is that even the discrepancies from the model for intratask differences can be construed as support for the categorical view of word meaning. The deviations are difficult to explain since they involve cognitive processes within the subject which we cannot deal with here except on a highly inferential basis. However, an examination of the distribution of discrimination item pairs based on this data show that all those items which were at a different relative position from those predicted, in that they yielded better performance than expected (LL-D, HL-D), involve different category distractors. Apparently, category cues are picked up and used very effectively by subjects in a discrimination task. And perhaps that is a partial explanation of the failure of distractor frequency to reach significance, that is, that the influence of this particular variable was overshadowed by distractor category (LL-D pairs for example yield better performance than items with high frequency distractors).

But as these assumptions are valid the data also contains some evidence concerning the general growth or development of the semantic structure. Brown suggests that the problem of naming is the problem of defining the specific features which a label implies. Then it might be true that though he has labels, the attributes by which a child defines the referents in his use of labels is incomplete. Two problems which are suggested by Brown's observations (that all words can be applied to the same referent and that semantic attributes overlap) are the necessity to determine: 1) How objects which share many of the same attributes are linguistically distinguishable and 2) for which particular group of attributes a particular label is the more appropriate response (that is dealing with superordinate and subordinate relations). A word used to denote an object means cognitively attending to certain criterial properties and ignoring the irrelevant ones. Naming behavior, Brown feels, helps to establish which ones are which. But at the age of 4 and 5 it could be that the criteria which are attended to are incompletely defined, and this may contribute to what Brown calls overgeneralization, i.e., broader conception of the criterial features and broader use of labels than that which is typical of adult usage. Others (Anglin, 1970) point out that since the properties to which one attends are criterial to that concept, the word is the embodiment of a concept, but they seem to stress the fact that concepts which are hierarchically related also share overlapping attributes, and that it becomes necessary to understand the relevant criterial attributes for words which represent different levels of abstraction. If a concept is incorrectly or incompletely defined we should expect usage of the word to be somewhat inappropriate. On the other hand, if indeed a word or name is an attribute of an object, and since any one referent has several verbal attributes representing different levels of exclusiveness, each label can be seen as representing a particular subset of criterial features. The important question, then, becomes what are the criterial attributes for any specific label, and at a particular time, in a particular situation, for a particular concept which label best conveys the nature of the referent. The two developmental problems which are suggested by these emphases are the necessity to determine 1) how referents which share many of the same attributes are linguistically distinguishable and 2) for which particular group of attributes is a label the more appropriate response.

These two emphases suggest the McNeil analysis of the growth of word meaning. McNeil conceptualizes semantic feature addition or the elaboration of criterial attributes as expanding dictionary entries. The units under which meaning is filed changes from holophrases to sentences to words during the course of development, but whatever the index the addition of semantic features has important ramifications. Each new feature is a distinction which separates one class of words from another. So it is the addition of semantic features which is responsible for the separation as well as the formation of concepts through the restriction and elaboration of their meaning.

McNeil proposes two hypotheses regarding the addition of semantic features and the incorporation of the attributes which lead to concept or referent formation. The dictionary entry undergoes horizontal development or semantic growth if the appearance of restrictive features is sequential, and vertical development if it is directional. By definition, semantic features appear in more than one dictionary entry and in fact in a great many. If all the features necessary to define the word enter the dictionary at the same time that the word does, semantic development will consist primarily of coming to see the relationship between words which share features in some hierarchical or "vertical" fashion. But if not all the semantic features associated with a word enter the dictionary when the word itself enters, words become more restrictive as these features are sequentially added. Semantic development will then consist of horizontally completing the dictionary entries by adding new restrictions to features already acquired. This suggests that a word may be a part of one's repertoire before one has a complete understanding of its meaning. Words then can be in a child's vocabulary but have different semantic properties from the same words in the vocabulary of an older child or an adult. According to McNeil, a child who lacks knowledge of some semantic feature of a word because its entry in the dictionary is incomplete will accept word combinations that an adult with a fuller dictionary entry rejects as anomalous. And if horizontal growth is the rule, adult and child usage could be different, because of the different defining features, but not so if vertical growth is the rule. McNeil proposes that these two types of growth are not mutually exclusive in the child.

The McNeil analysis of growth of meaning into vertical and horizontal might predict that for a vertical type of development erroneous word usage might manifest itself by a preponderance of superordinate and subordinate errors because growth necessitates determining which of the particular labels under which the attributes of the referent are filed is most appropriate. Thus the handling of superordinate and subordinate designation would be the major concern. On the other hand, growth of meaning horizontally which involves adding features sequentially to a word would predict the major concern to be intra-category meaning. As these distinguishing restrictions are missing, the major problems in vocabulary building are expected to be same class confusions or errors.

The preponderance of same class as opposed to superordinate and subordinate errors in our data suggests that the more difficult problem in word usage is the designation of concepts which, though perceptually and linguistically differentiable, share certain semantic attributes. And apparently differentiating characteristics which would separate similar items are not yet a part of the semantic makeup of the words. The fact that same class errors are apparent in both tasks signals a greater need for horizontal growth of semantic features than vertical in both use and understanding of words.

But since these are not mutually exclusive we were not surprised to find that there were superordinate and subordinate confusions also. However, the overwhelming proportion of same class errors indicates that development is not so much the problem of going either up or down a meaning hierarchy but across one level. This evidence is in accordance with others (Kaplan, 1967) who have noted that abstractions do not seem to appear early in language learning and are almost nonexistent in kindergarten vocabulary. Thus they would offer little source of confusion. The additional fact that the name given to the child by an adult seems to represent maximum utility in that it anticipates the equivalent and difference that need to be observed in dealing with the object (Brown, 1970) indicates that not only are adults consistent in providing environmental contingencies, thus keeping superordinate and subordinate confusions at a minimal, but suggests that this contingency does not eliminate the active role of the child in forming the concepts which labels denote. Perhaps, the importance of superordinate and subordinate relations come as a result of seeing the need following conceptual differentiation to classify those referents which at a previous point in development are seen as having their own distinctive linguistic attributes, but share enough non-linguistic ones to have caused them to be confused.

It seems then that the sequential addition of semantic features is one of the principle problems in the development of the semantic structure.

McNeil (1966) in reference to word association tests says "We cannot tell ...what semantic markers are present or absent. What we can tell ... is that a child's dictionary entries remain incomplete well into early school years." In regard to vocabulary tests we have come to the same conclusion.

What conclusion can be made about the variables affecting vocabulary performance? In both production and recognition tasks the frequency with which a word is used will determine the level of performace, with the effect being greater for the production task. Within a task, target frequency and distractor category membership can influence performance directly and distractor frequency can interact with both the former variables. The relative importance of these factors on a discrimination task can only be inferred, but it seems that target frequency supercedes distractor category while distractor frequency is important only at certain levels of distractor category and target frequency.

The type of task is a variable which affects vocabulary performance. The evidence is that children can produce 62% of the words which they are able to recognize.

What can be said about the vocabulary of this age? A child of 4 or 5 can recognize more words than he can produce and to the extent that the tasks generating these differences reflect variation in or different emphases on certain aspects of the vocabulary phenomenon we are willing to say that these are two distinct types of vocabulary.

The original hypotheses were partially supported. Obviously the model for intratask differences is much too simple. And though its main function was simply to generate expectations of intratask variability it is obvious that complex cognitive factors will be very important in such processes.

The differential orienting response and attraction to certain stimuli might be suggested by some psychologists, for example. It is possible also that the specific discrepancies from the model for the discrimination task stem from certain methodological deficiencies such as the low error rate, and the relatively few distractors. With only two distractors (and these presented simultaneously) the choice difficulty is substantially minimized. Therefore, the distractor frequency might not have had the impact as in a more difficult task. In addition, the words were all of relatively high frequency, chosen precisely because they were common in the experience of most children. They were in effect easy words probably requiring the minimal in differential responding. Under these circumstances the frequency of the distractor might not have had the impact on performance such as in a more difficult construction.

Not to be overlooked is the lack of an adequate conception on our part. The particular predictions that were made were made on the basis of the order of importance of strategic factors and the amount of information in a stimulus set. Furthermore, we assumed that each of the three principle factors would be important on each trial and for each item of the test. It seems that at least one factor is important only at certain levels of the others.

Despite the discrepancies from the original model, the effects which are found are important. The task difference is important because it seems to point to two radically different vocabulary abilities and the necessity of dealing with task differences in comparison and diagnostic use. The frequency effect is important, resulting in differential performance on both tasks, so that, amount of usage seems to be generally facilitative across tasks. Its influence is especially significant since the difference in frequency is relatively small (Peters [1936] finding of no frequency effect on perceptual threshold for college students was thought to be due to the narrow frequency range use). After reviewing several studies Underwood and Schulz (1960) conclude that the frequency range must be rather extreme before even a small relationship (learning) emerges. Out of an estimated vocabulary of 25,000 words for first graders (Smith, 1941) the difference in frequency between the items in this study is maximal at 3,000. So, while there is no definitive theory of the frequency effect it seems that at this stage in development, differences in frequency have a profound effect on vocabulary performance.

The category effect should not be disregarded. In line with the position expressed by Brown, one of the important factors in language or word acquisition seems to be the categorization of the referent, the fact that words do not name particular things they name classes. So it seems that the present emphasis on meaning as a composite of semantic features might not be misplaced. For there is some indication that word meaning involves not just the ability to associate on a specific object to a label but to define the range of semantic features which that label implies and to delineate the boundaries between labels which might share semantic features in common.

In regard to general development it is noteworthy that if the nature of word usage requires some awareness of the generality of referents, the category effect and the seeming overgeneralized use of words suggests that not only is this principle operating at a very early age but that in fact it is exaggerated and perhaps it is such exaggeration of basic principles which promotes rapid language growth.

## Summary

In sum there are two factors affecting vocabulary performance--task or situational demands and word characteristics (of which frequency and category membership have the most unequivocal support from our data). Not only should comparisons of vocabulary performance consider these factors, but tangentially it has been indicated that a discrimination vocabulary is likely to be greater in range than a production vocabulary. And finally it suggests that a real understanding of important vocabulary differences (e.g., whether quality of the vocabulary is a developmental phenomenon) might well consider differential task situations, and that such considerations can lead to a better understanding and clarification of assumptions underlying verbal behavior.

REFERENCES

Anglin, J. M. The Growth of Word Meaning. Cambridge, Mass.:M.I.T. Press, 1970.

Babbitt, E. H. A Vocabulary Test. Popular Science Monthly, 70, (1907), 378.

Brandenburg, G. E. Psychological Aspects of Language. Journal of Educational Psychology, 9, (1918), 313-332.

Brown, R. Words and Things. New York:The Free Press, 1958.

Brown, R. How Shall a Thing Be Called. Psycholinguistics:Selected Papers, Roger Brown (Ed.), New York:The Free Press, 1970.

Brown, R. W. The size of the working vocabulary. The Nation, 93, (1911), 11.

Bryan, F. E. How Large Are Children's Vocabularies? Elementary School Journal, 55, (1953), 440-446.

Calfee, R. C., Chapman, R. S. & Venezky, R. L. How A Child Needs To Think To Learn To Read. Technical Report No. 131, Wisconsin Research and Development Center for Cognitive Learning, Madison, Wisconsin, 1970.

Chambers, W. G. How Words Get Meaning. Pedagogical Seminary, 9, (1904), 30-50.

Chase, C. I. Application of Levels of Concept Formation to Measurement of Vocabulary. Journal of Educational Research, 55, (1961), 75-78.

Colvin, C. M. A Reexamination of the Vocabulary Question. Elementary English, 28, (1951), 350-356.

Cronbach, L. An Analysis of Techniques for Diagnostic Vocabulary Testing. Journal of Educational Research, 361, (1942), 206-217.

Crosscup, R. G. A Survey and Analysis of Methods and Techniques for Fostering Growth of Meaning Vocabulary. Master's Thesis, Boston: Boston School of Education, 1940.

Cuff, N. B. Vocabulary Tests. Journal of Educational Psychology, 21, (1930), 212-219.

Dale, E. Difficulties in Vocabulary Research. Educational Research Bulletin, 10, (1931), 119-122.

Deese, J. Psycholinguistics. Boston:Allyn Bacon, Inc., 1970.

Dolch, E. W. How Much Word Knowledge Do Children Bring To Grade One? Elementary English Review, 13, (1936), 177-183.

Doran, E. W. A Study of Vocabularies. Pedagogical Seminary, 14, (1907), 401-438.

Entwhistle, D. R. Developmental Sociolinguistics:Inner City Children, Report from the Center for the Study of Social Organization of Schools, John Hopkins University, 1967.

Fraser, E., Bellugi, U. & Brown, R. W. Control of Grammar in Imitation, Comprehension and Production. Journal of Verbal Learning and Verbal Behavior, 2, (1963), 121-135.

Gansl, I. Vocabulary:Its Measurement and Growth. Archives of Psychology, 1939.

Gerlach, F. M. Vocabulary Studies. Studies in Education and Psychology, Colorado College, No. 1, 1917.

Hall, J. F. Learning as a Function of Word Frequency. American Journal of Psychology, 67, (1954), 138-140.

Hartman, G.W. Critique of the Common Method of Estimating Vocabulary Size, Together with Some Data on the Absolute Word Knowledge of Educated Adults. Journal of Educational Psychology, 32, (1941), 351-358.

Howes, D. H. and Solomon, R. L. Visual Duration Threshold as a Function of Word Probability. Journal of Experimental Psychology, 41, (1951), 401-410.

Hurlburt, L. D. The Relative Value of Recall and Recognition Techniques for Measuring Precise Knowledge of Word Meaning. Doctoral Dissertation, Boston, Mass., Boston University, 1949.

Jacobs, A. Formation of New Associations to Words Selected on the Basis of Reaction-Time-GSR Combinations. Journal of Abnormal Social Psychology, 51, (1955), 371-377.

Jersild, A. T. Child Psychology. New York:Prentice-Hall, Inc., 1940.

Jung, J. Verbal Learning. New York:Holt, Rinehart and Winston, Inc., 1968.

Kaplan, B. An Approach to the Problem of Symbolic Representation: Nonverbal and Verbal. Readings in the Psychology of Language, L. A. Jakobovits and Murray S. Miron (Eds.), Englewood Clifts, New Jersey:Prentice-Hall, Inc., 1967.

Kausler, D. H. Acquisitions As a Function of Meaningfulness and Familiarization. Readings in Verbal Learning:Comtemporary Theory and Research, D. Kausler (Ed.), New York:John Wiley & Sons, Inc., 1966.

Kelley, V. H. Techniques for Testing Word Meaning Knowledge. Elementary English Review, 9, (1932), 102-105.

Kirkpatrick, E. A. A Vocabulary Test. Popular Science Monthly, 70, (1907), 157-164.

Kirkpatrick, J. J. & Cureton, E. E. Vocabulary Item Difficulty and Word Frequency. Journal of Applied Psychology, 33, (1949), 347-351.

Lashley, E. S., & Wade M. The Pavlovian Theory of Generalization. Psychological Review, 53, (1946), 72-87.

Larrick, N. How Many Words Does a Child Know? Education Digest, 19, (1954), 42-44.

Lovell, K., & Bradbury, B. The Learning of English Morphology in Educationally Subnormal Special School Children. American Journal of Mental Deficiency, 71, (1967), 609-615.

Lovell, K. & Dixon, E. M. The growth of grammar in imitation, comprehension, and production. Journal of Child Psychology and Psychiatry, 8, (1967), 31-39.

Luh, C. W. The Conditions of Retention. Psychological Monographs, 1922, 31, (whole no. 142).

Maccoby, E. E., & Bee, H. L. Some Speculations Concerning the Lag Between Perceiving and Performing. Child Development, 36, (1965), 367-377.

McGuire, W. J. A Multiprocess Model for Paired-Associate Learning. Journal of Experimental Psychology, 62, (1961), 334-347.

McNeil, D. A Study of Word Association. Journal of Verbal Learning and Verbal Behavior, 5, (1966), 548-557.

Olson, D. R., & Pagliuso, S. M. (Eds.). From Perceiving to Performing: An Aspect of Cognitive Growth. Ontario Journal of Educational Research, Special Issue, 10, (1968) 155-247.

Peters, A. N. The Relationship Between Familiarity of Words and Their Memory Value. American Journal of Psychology, 1936, 572-584.

Postman, L., and Rau, L. Retention as a Function of the Method of Measurement. University of California Publications in Psychology, 8, (1957), 217-270.

Rinsland, H. D. A Basic Vocabulary of Elementary School Children. New York:The MacMillan Company, 1945.

Seashore, R. H. The Measurement and Analysis of Extent of Vocabulary. Psychological Bulletin, 30, (1933), 709-720.

Seashore, R. H. and Dolch, E. W. Implications of the Seashore Vocabulary Report:With Reply. Elementary English, 26, (1949), 407-413.

Seegers, J. E. and Seashore, R. H. How Large are Children's Vocabularies? Elementary English, 26, (1949), 181-194.

Solomon, and Postman, L. Frequency of Usage as a Determinant of Recognition Thresholds for Words. Journal of Experimental Psychology, 43, (1952), 195-201.

Smith, M. K. Measurement of the Size of General English Vocabulary Through the Elementary Grades and High School. Genetic Psychology Monographs, 24, (1941), 311-345.

Starch, D. Educational Measurement. New York:MacMillan Company, 1916.

Sternberg, S. The Discovery of Processing Stages:Extensions of Donder's Method. Acta Psychologica, 30, (1969), 276-315.

Symonds, P. M. Size of Recognition and Recall Vocabularies. School and Society, 24, (1926), 559-560.

Templin, M. D. Certain Language Skills in Children. Minneapolis:University of Minnesota Press, 1957.

Thorndike, E. L. & Symonds, P. M. Difficulty, Reliability and Grade Achievements in a Test of English Vocabulary. Teacher's College Record, 24, (1923), 438-445.

Underwood, B. J. & Schulz, R. W. Frequency and Meaningfulness, in Meaningfulness and Verbal Learning. Chicago:J. B. Lippincott Company, 1960.

Vygotsky, L. S. Thought and Speech. Psychiatry, 2, (1939), 29-54.

Watts, A. F. Language and Mental Development of Children. Boston:D. C. Heath and Company, 1944, p. 245.

Werner, H. & Kaplan, E. Development of Word Meaning Through Verbal Context: An Experimental Study. Journal of Psychology, 29, (1950), 251-257.

Table 1. Componential Analysis by a Meaning Tree or Semantic Hierarchy (Deese, 1970)

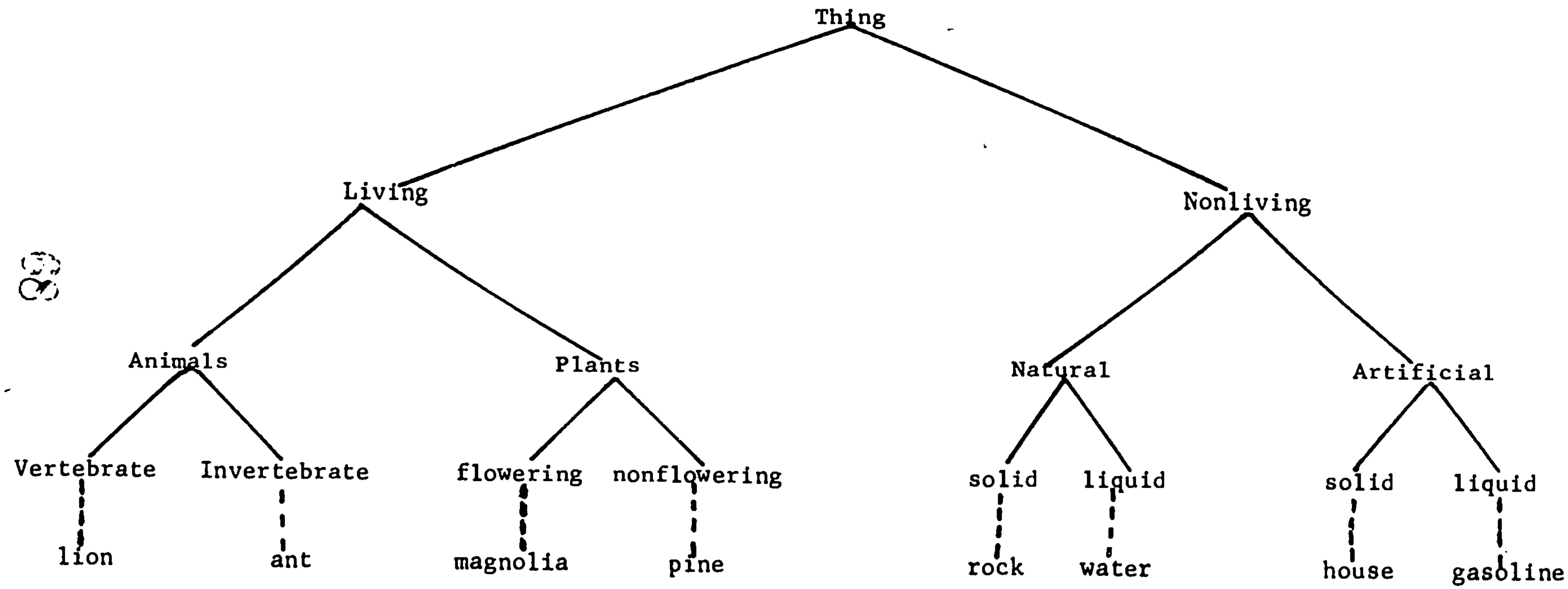


Table 2. Componential Analysis by a Semantic Feature Table ( Deese,1970)

| | | lion | ant | magnolia | pine | rock | water | house | gasoline |
|---|---|---|---|---|---|---|---|---|---|
| living | nonliving | + | + | + | + | - | - | - | - |
| animals | plants | + | + | - | - | | | | |
| vertebrate | invertebrate | + | - | | | | | | |
| flowering | nonflowering | | | + | - | | | | |
| natural | artificial | | | | | + | + | - | - |
| solid | liquid | | | | | + | - | + | - |

Table 3. Frequency Range of Stimulus Items According to the First Grade Section of the Rinsland Count

| Category | Frequency Class |
|---|---|
| Insects | |
| Butterfly | 900 - 1000 |
| Bee | 800 - 900 |
| Grasshopper | 2000 - 2500 |
| Spider | 3000 - 3500 |
| Furniture | |
| Bed | 100 - 200 |
| Table | 200 - 300 |
| Lamp | 2500 - 3000 |
| Sofa | 2500 - 3000 |
| Clothes | |
| Dress | 100 - 200 |
| Shoe | 700 - 800 |
| Skirt | 2000 - 2500 |
| Blouse | 3500 - 4000 |
| Toys | |
| Ball | 000 - 100 |
| Doll | 000 - 100 |
| Rattle | 2000 - 2500 |
| Crayon | 3000 - 3500 |
| Tableware | |
| Glass | 400 - 600 |
| Bowl | 600 - 700 |
| Spoon | 2000 - 2500 |
| Pitcher | (no occurence within first 4,000 words) |

Table 4. Description of Stimulus Item Groups for the Recognition Task and Expected Relative Performance by Groups

| Stimulus Item Groups | Description | | | Prediction of Rank by Number Correct |
|---|---|---|---|---|
| | **Target** | **Distractor** | **Category** | |
| H - H - D | High frequency | High frequency | Different | 1 (lowest no. errors) |
| H - H - S | High frequency | High frequency | Same | 2 |
| H - L - D | High frequency | Low frequency | Different | 3 |
| H - L - S | High frequency | Low frequency | Same | 4 |
| L - H - D | Low frequency | High frequency | Different | 5 |
| L - H - S | Low frequency | High frequency | Same | 6 |
| L - L - D | Low frequency | Low frequency | Different | 7 |
| L - L - S | Low frequency | Low frequency | Same | 8 (highest no. errors) |

Table 5. Mean Percent Correct Responding for Recognition Task Items Defined by Target Category

| Category | Per Cent Correct |
|---|---|
| 1 - Insects | 98 |
| 2 - Furniture | 93 |
| 3 - Clothes | 97 |
| 4 - Toys | 98 |
| 5 - Tableware | 95 |

Table 6. Mean Percent Correct Responding for Recognition Task Items Defined by Target Category and Target Frequency

| Target Category | Target Frequency | Mean % Correct |
|---|---|---|
| 1 - Insects | high | 96 |
| 1 - Insects | low | 98 |
| 2 - Furniture | high | 100 |
| 2 - Furniture | low | 85 |
| 3 - Clothes | high | 97 |
| 3 - Clothes | low | 95 |
| 4 - Toys | high | 98 |
| 4 - Toys | low | 96 |
| 5 - Tableware | high | 98 |
| 5 - Tableware | low | 88 |

Table 7. Mean Percent Correct Responding for Items defined by Distractor Frequency and Distractor Category Category

| Distractor Frequency | Distractor Category | % Correct |
|---|---|---|
| High | Same | 96 |
| High | Different | 97 |
| Low | Same | 94 |
| Low | Different | 98 |

Table 8. Mean Percent Correct Responding for Items Defined by Target Frequency and Distractor Frequency

| Target Frequency | Distractor Frequency | % Correct |
|---|---|---|
| High | High | 99 |
| High | Low | 97 |
| Low | High | 94 |
| Low | Low | 94 |

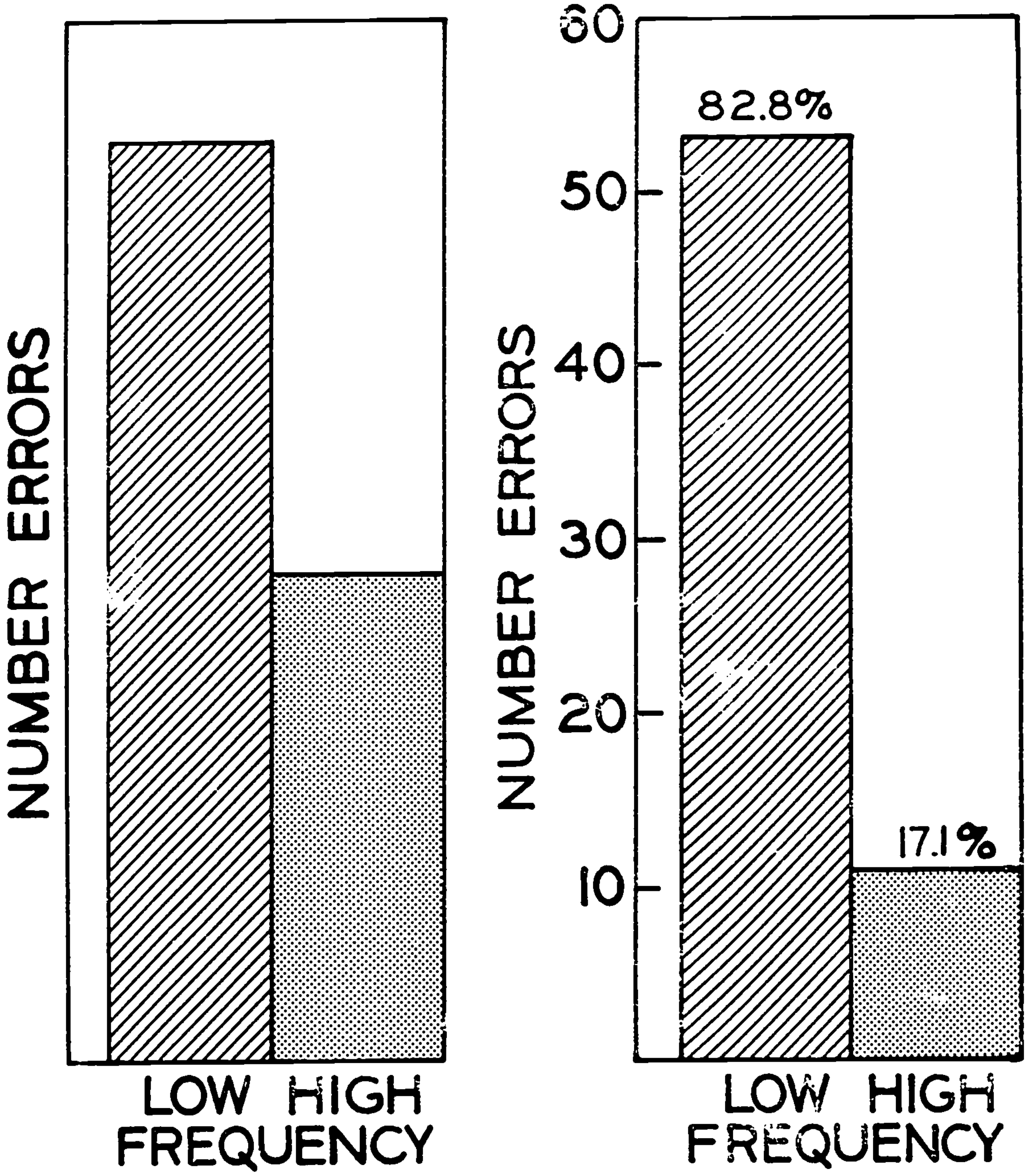



FIG. 1: EXPECTED RELATIVE DISTRIBUTION OF ERROR OF ITEMS BY TARGET FREQUENCY

OBTAINED DISTRIBUTION OF ERROR OF ITEMS BY TARGET FREQUENCY

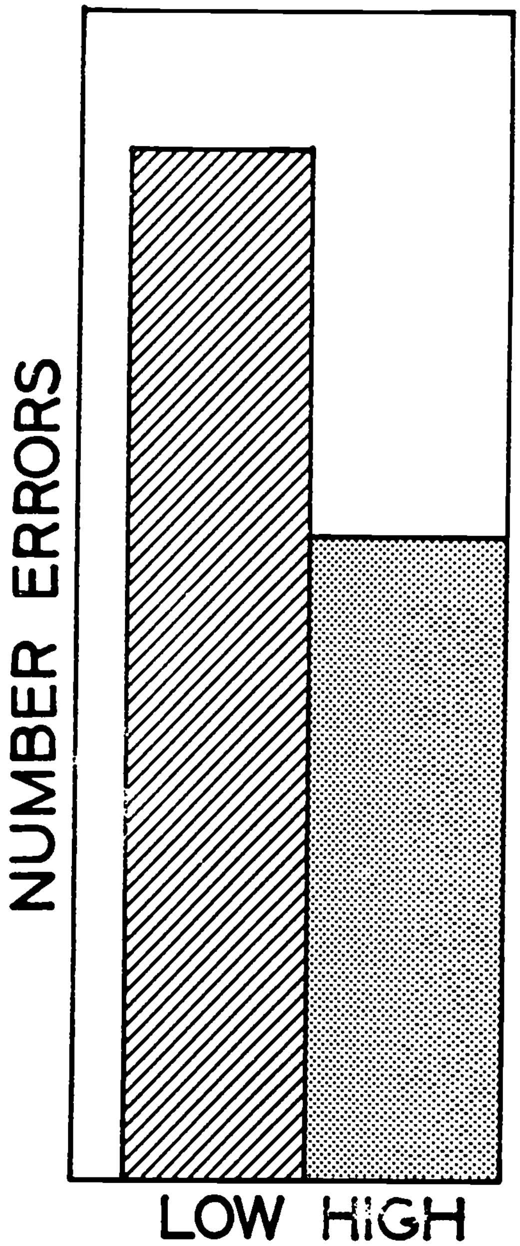
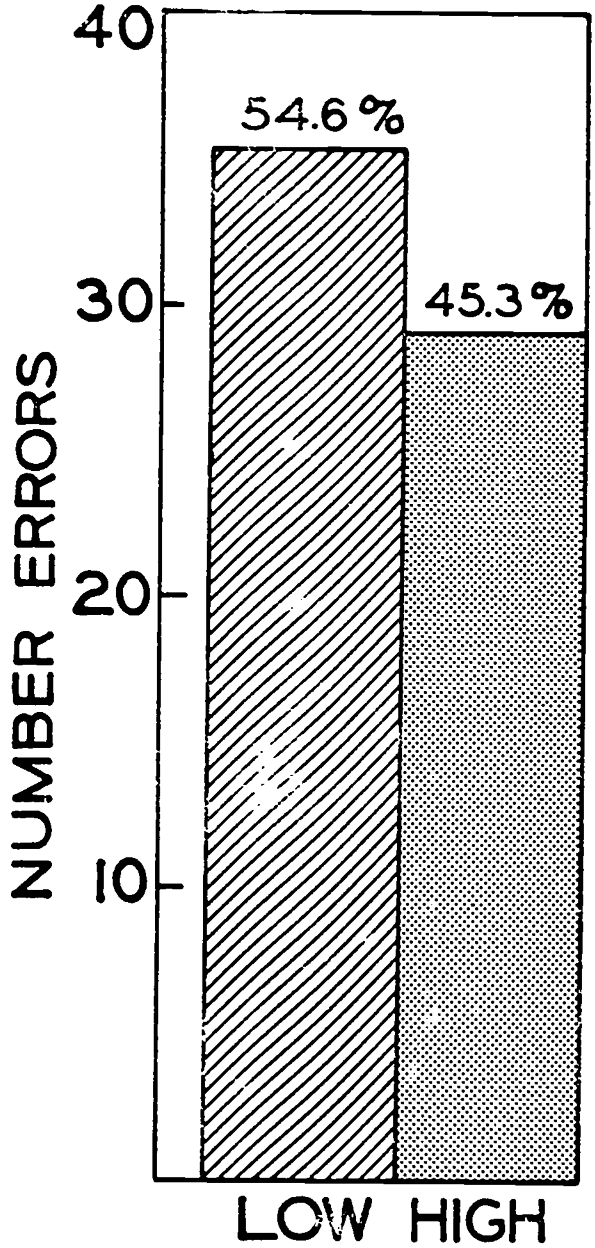



FIG. 2: EXPECTED RELATIVE DISTRIBUTION OF ERROR OF ITEMS DEFINED BY DISTRACTOR FREQUENCY




OBTAINED DISTRIBUTION OF ERROR OF ITEMS DEFINED BY DISTRACTOR FREQUENCY

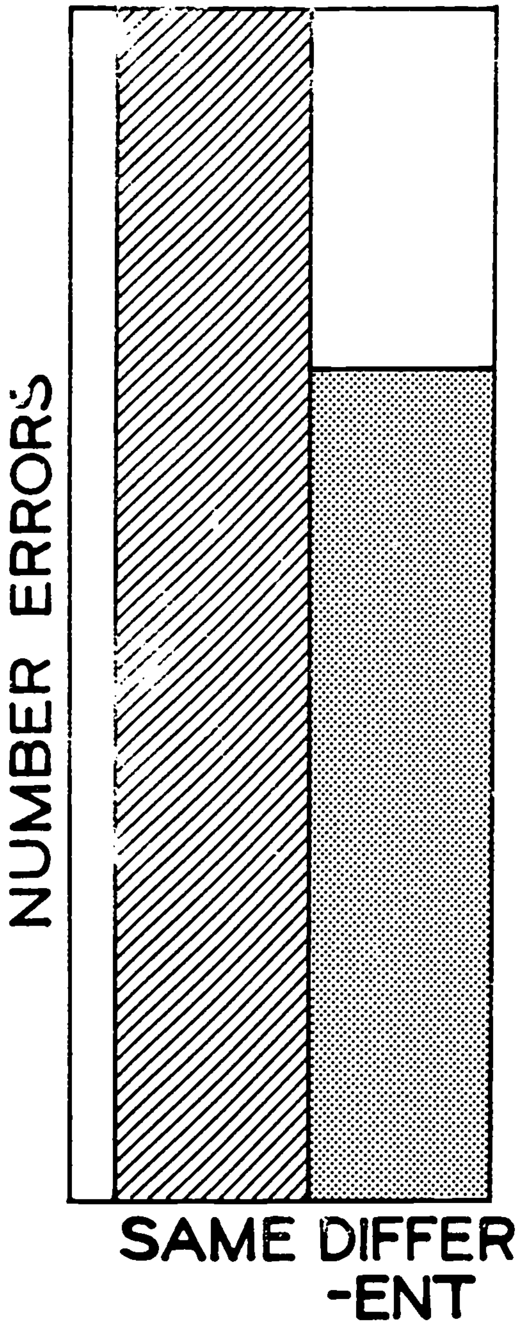
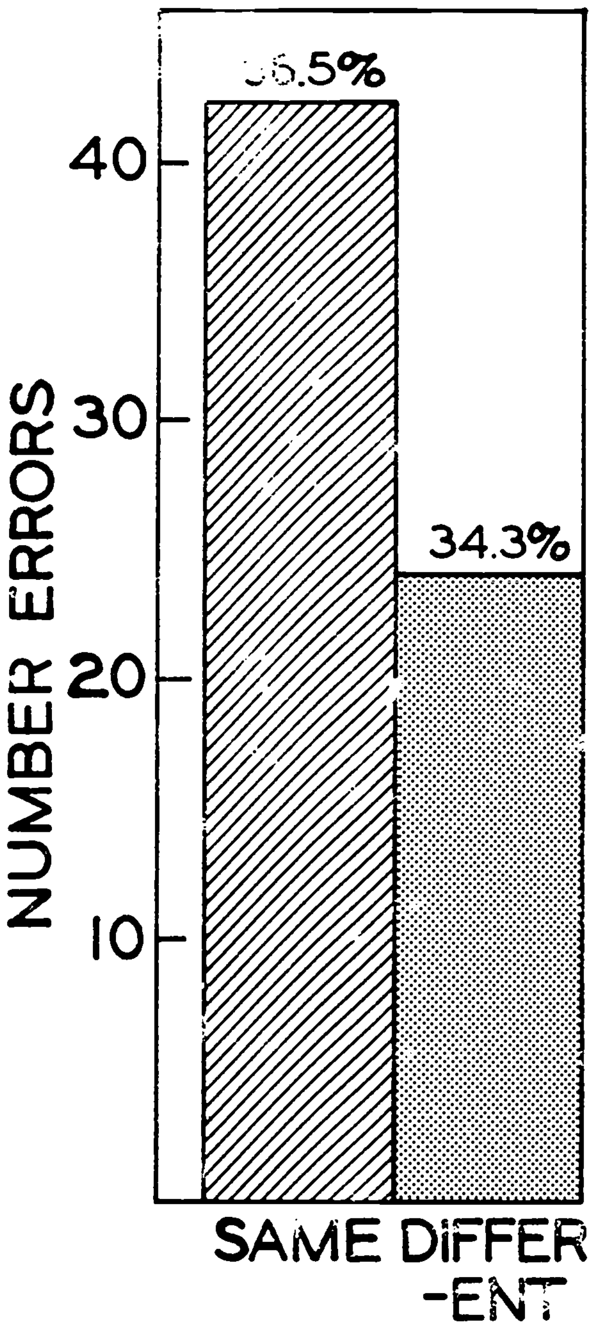



FIG. 3: EXPECTED RELATIVE DISTRIBUTION OF ERROR ITEMS DEFINED ON DISTRACTOR CATEGORY




OBTAINED NUMBER OF ERRORS BY ITEMS DEFINED ON DISTRACTOR CATEGORY

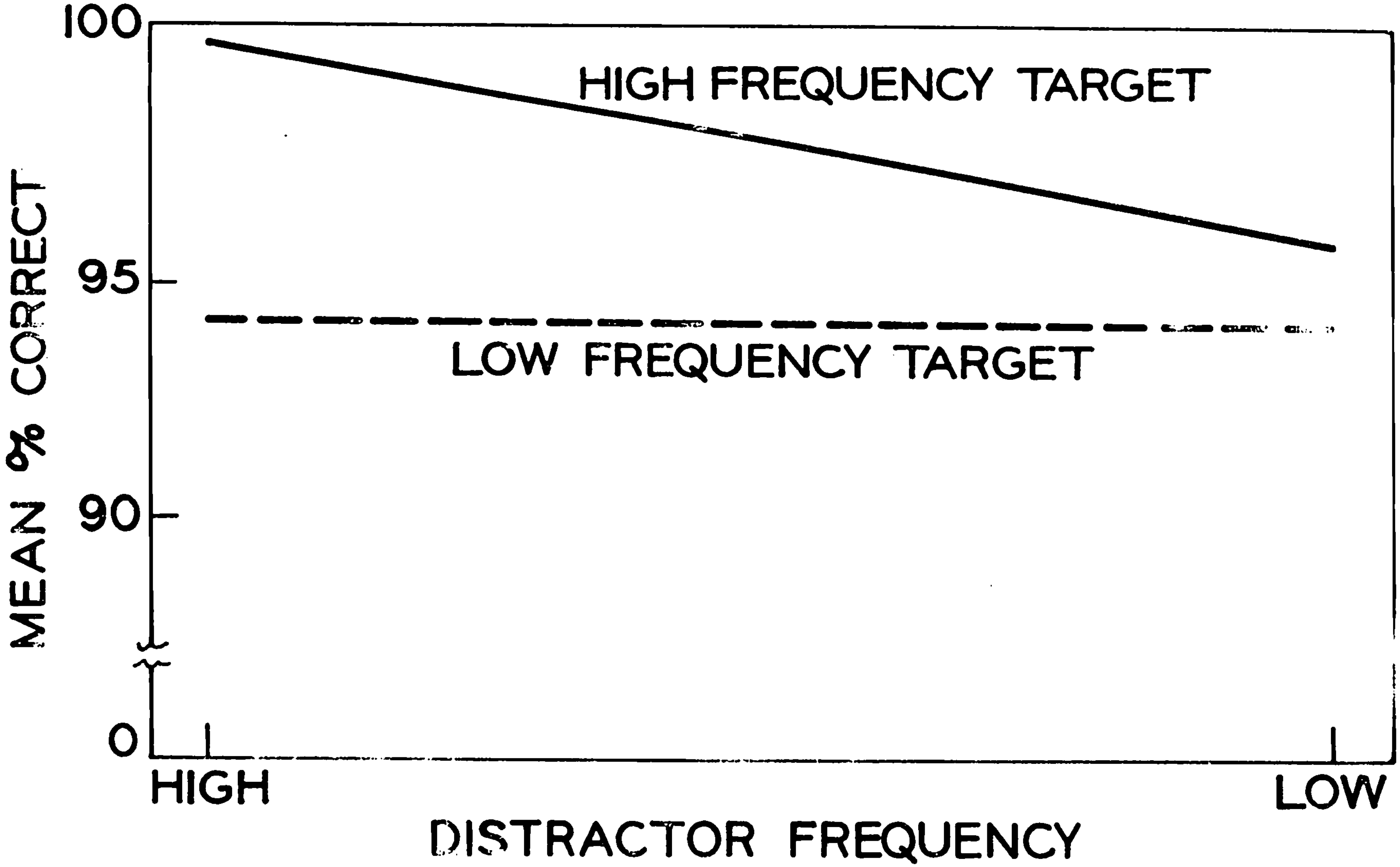



FIG. 4: MEAN PERCENT CORRECT RESPONDING OF ITEMS DEFINED BY TARGET FREQUENCY AND DISTRACTOR FREQUENCY

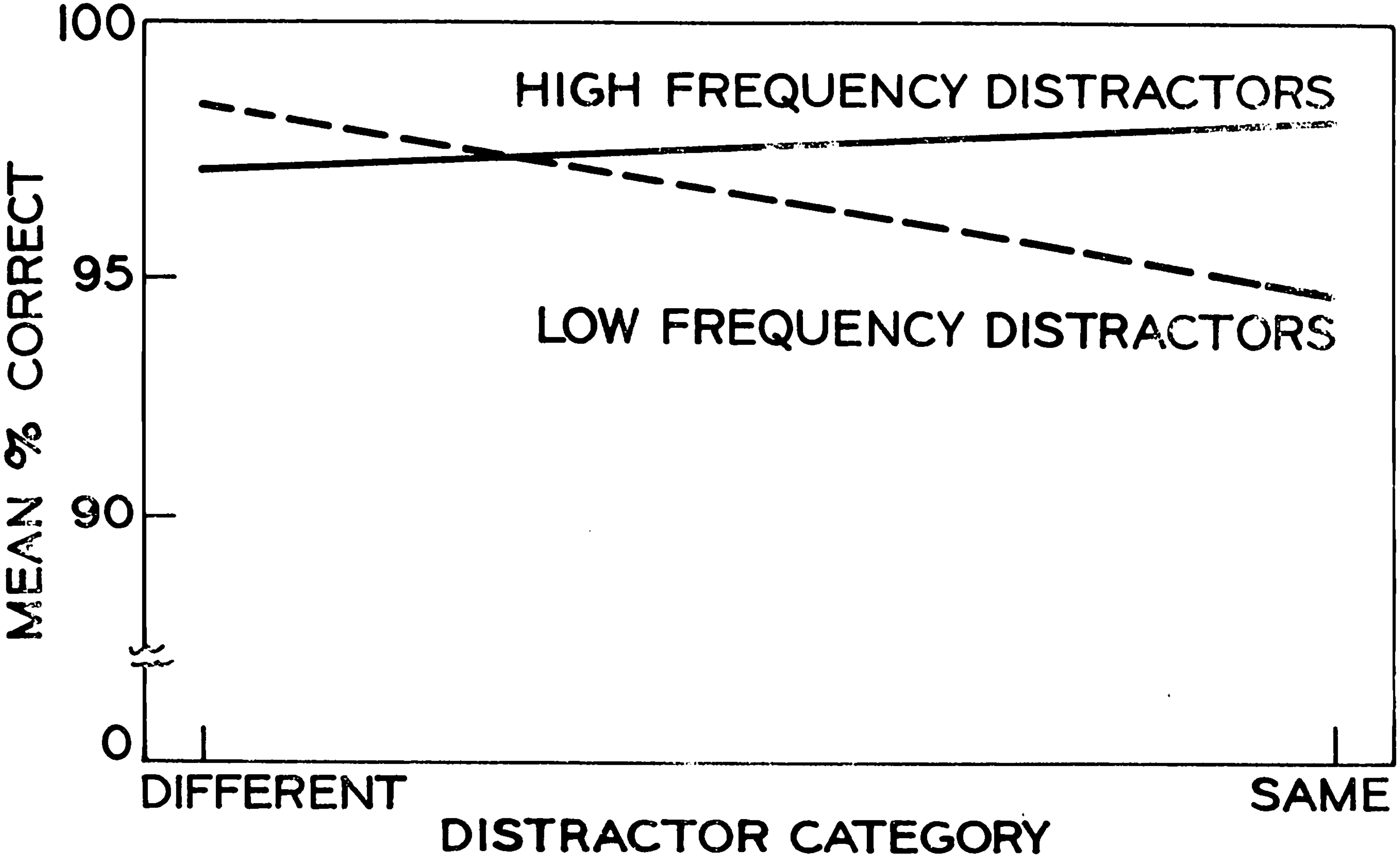



FIG. 5: MEAN PERCENT CORRECT RESPONDING BY ITEMS DEFINED ON DISTRACTOR FREQUENCY AND DISTRACTOR CATEGORY

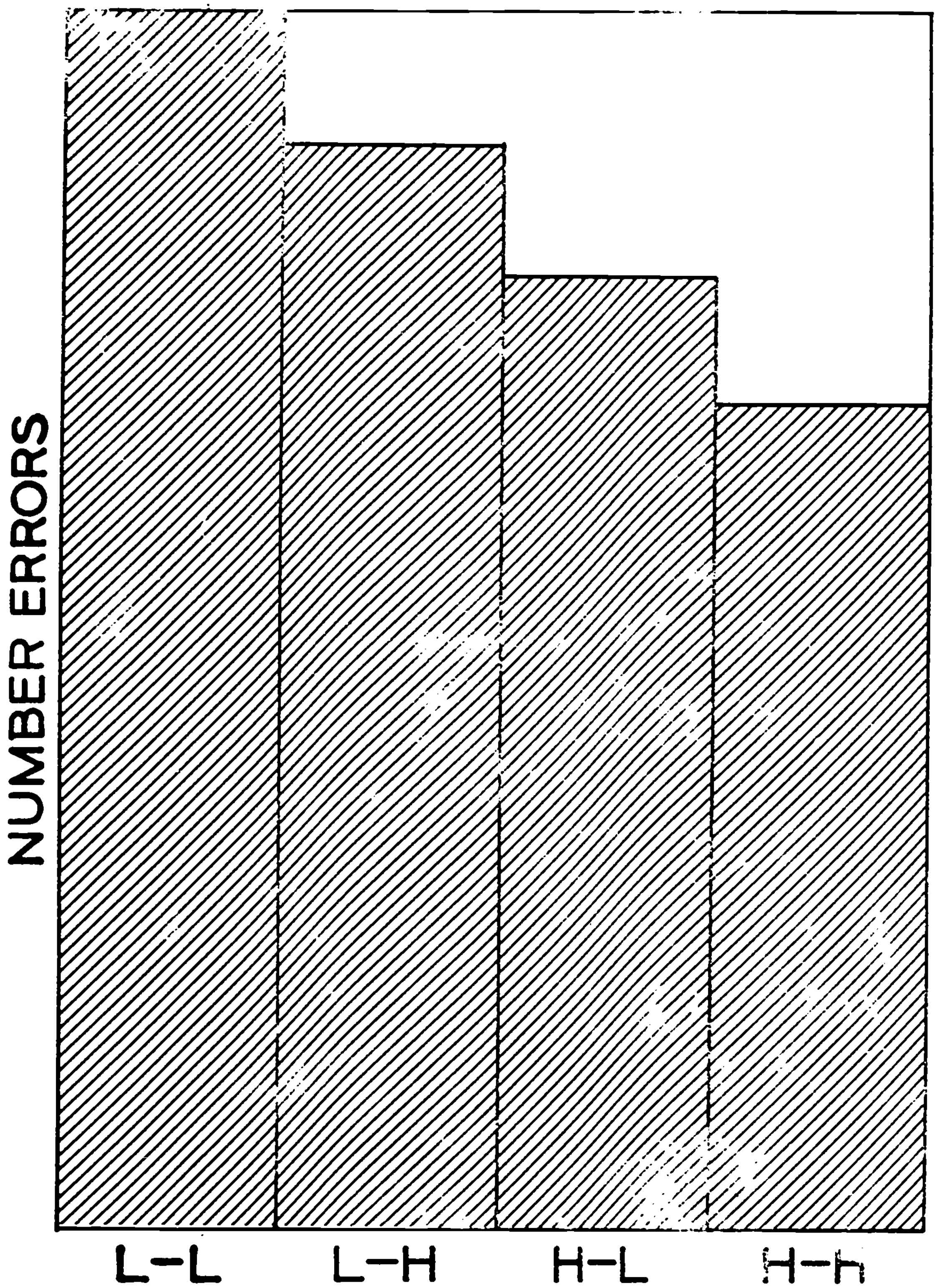



FIG. 6: EXPECTED RELATIVE DISTRIBUTION OF ERRORS BY ITEMS DEFINED ON TARGET FREQUENCY AND DISTRACTOR FREQUENCY

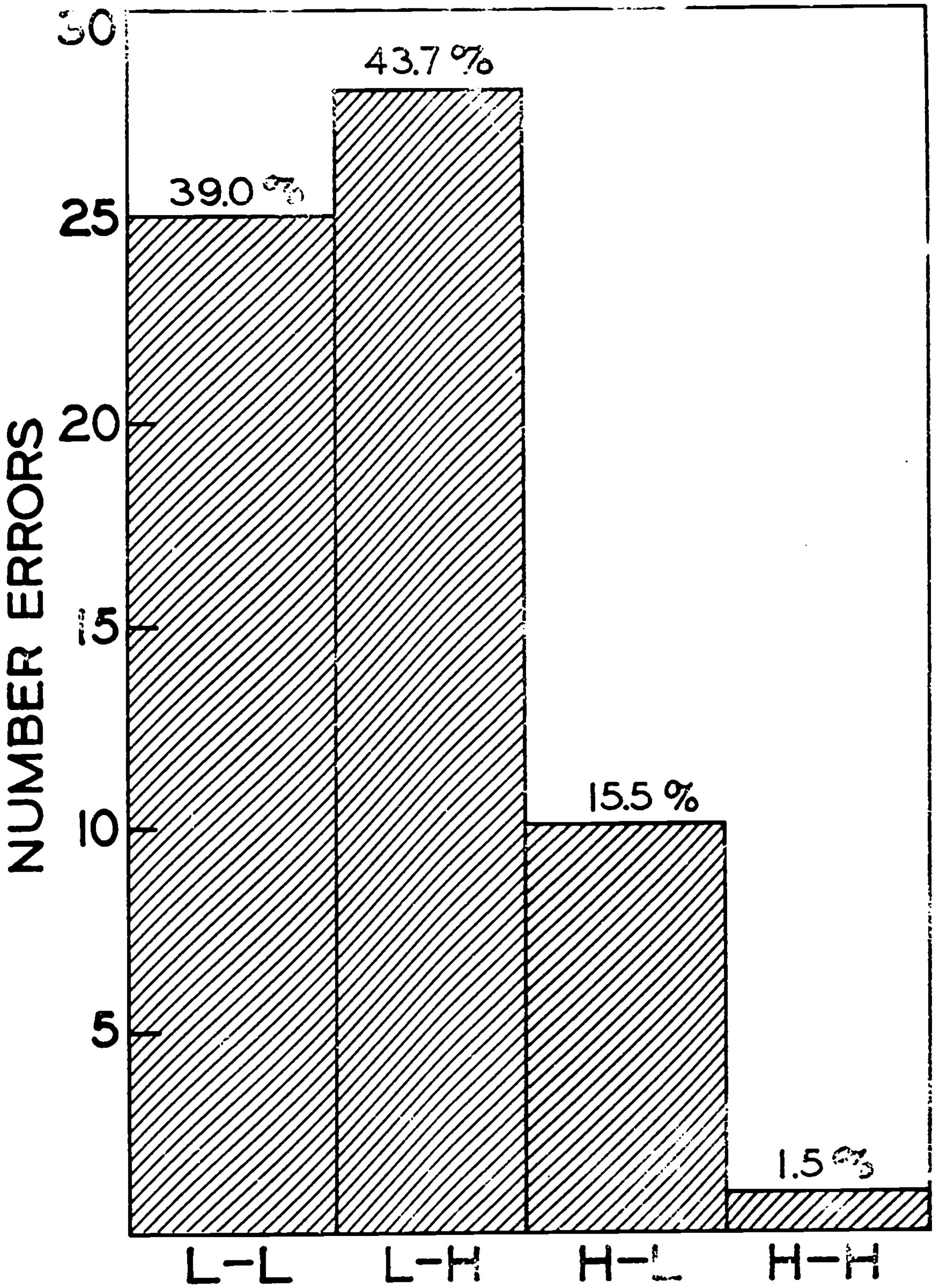



FIG. 7: OBTAINED DISTRIBUTION OF ERRORS BY ITEMS DEFINED ON TARGET FREQUENCY AND DISTRACTOR FREQUENCY

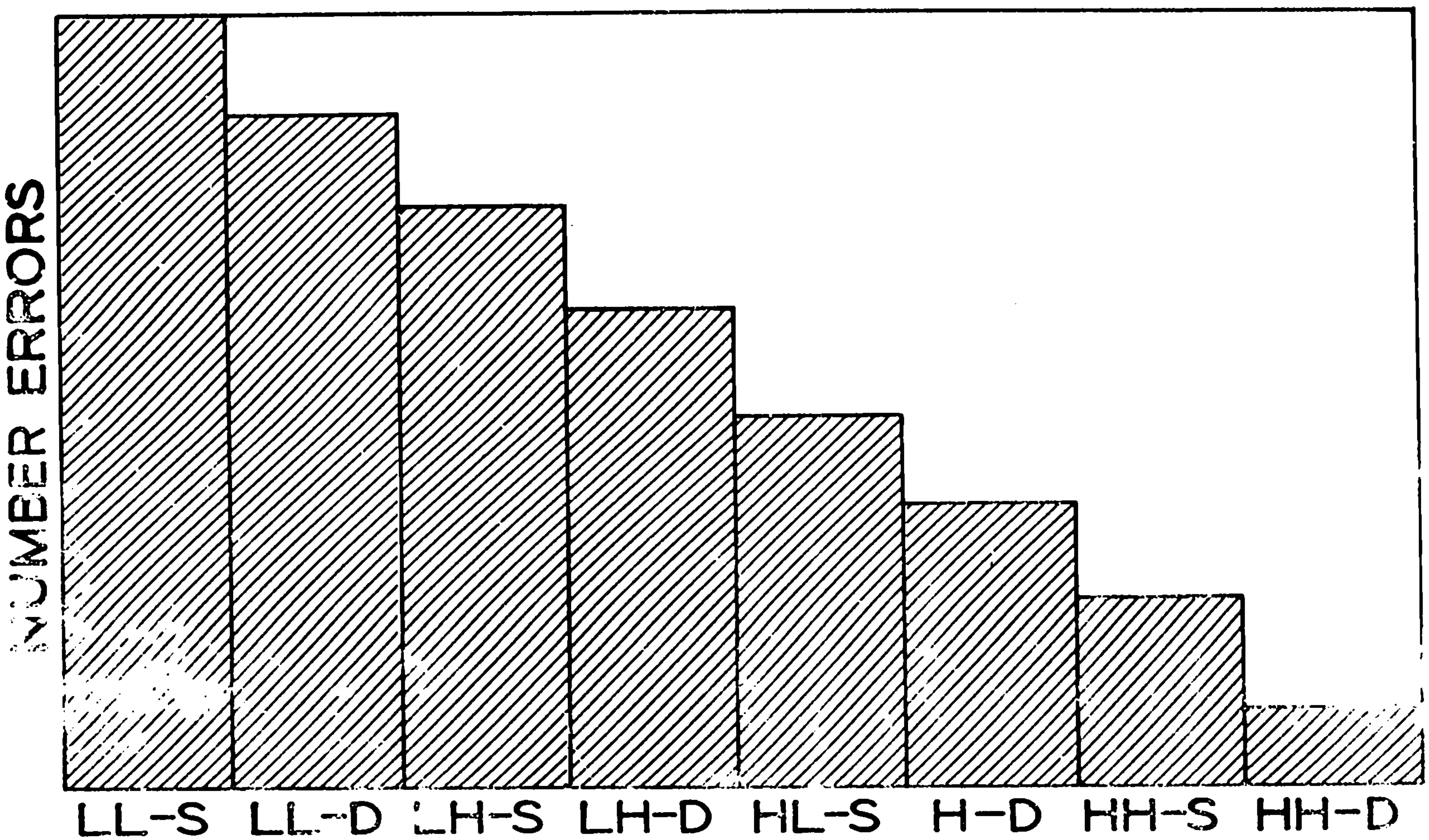



FIG. 8: EXPECTED RELATIVE DISTRIBUTION OF ERROR ON DISCRIMINATION TASK ITEMS DEFINED ON TARGET FREQUENCY DISTRACTOR FREQUENCY AND DISTRACTOR CATEGORY

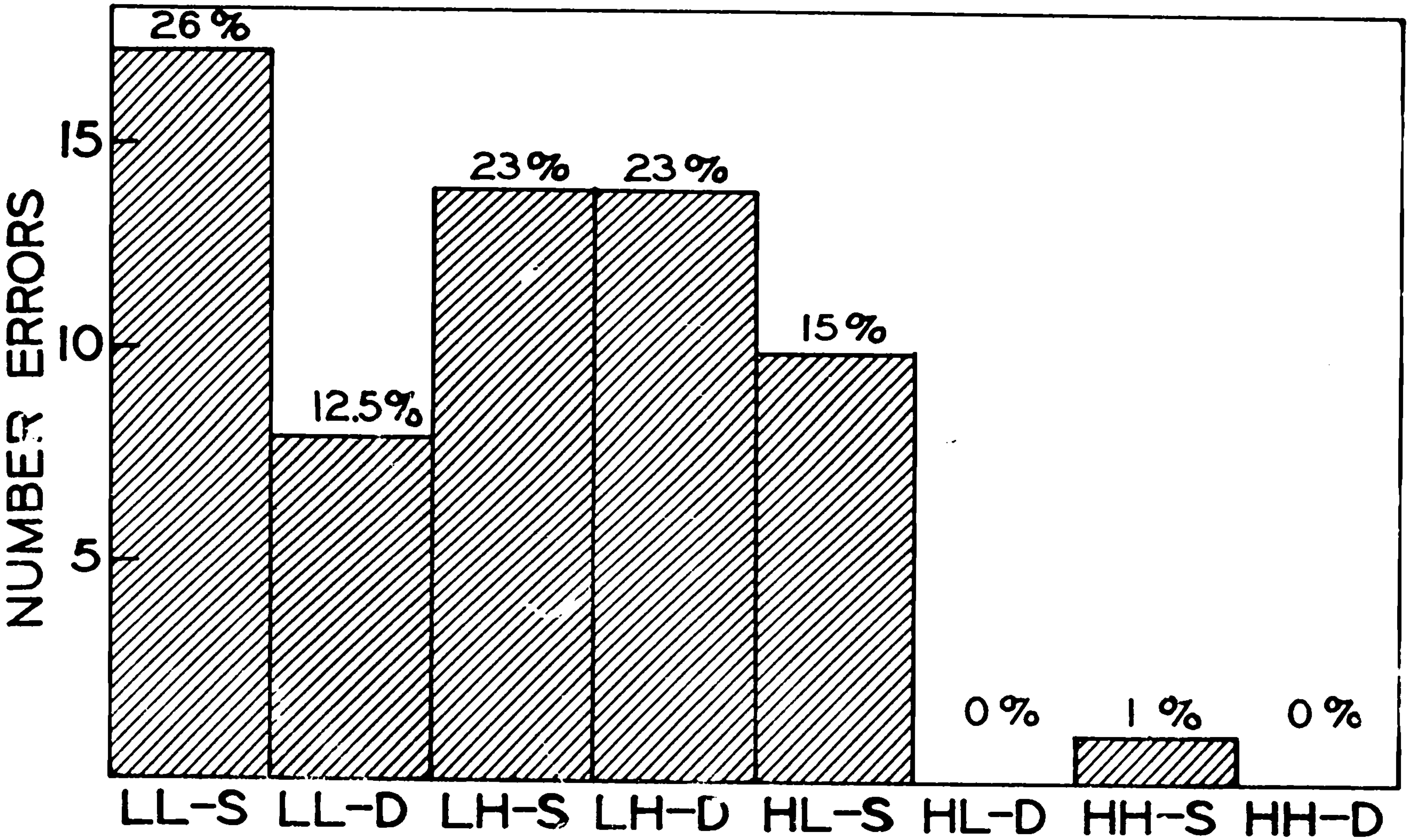



FIG. 9: OBTAINED DISTRIBUTION OF ERRORS OVER DISCRIMINATION TASK ITEMS DEFINED ON TARGET FREQUENCY DISTRACTOR FREQUENCY AND DISTRACTOR CATEGORY

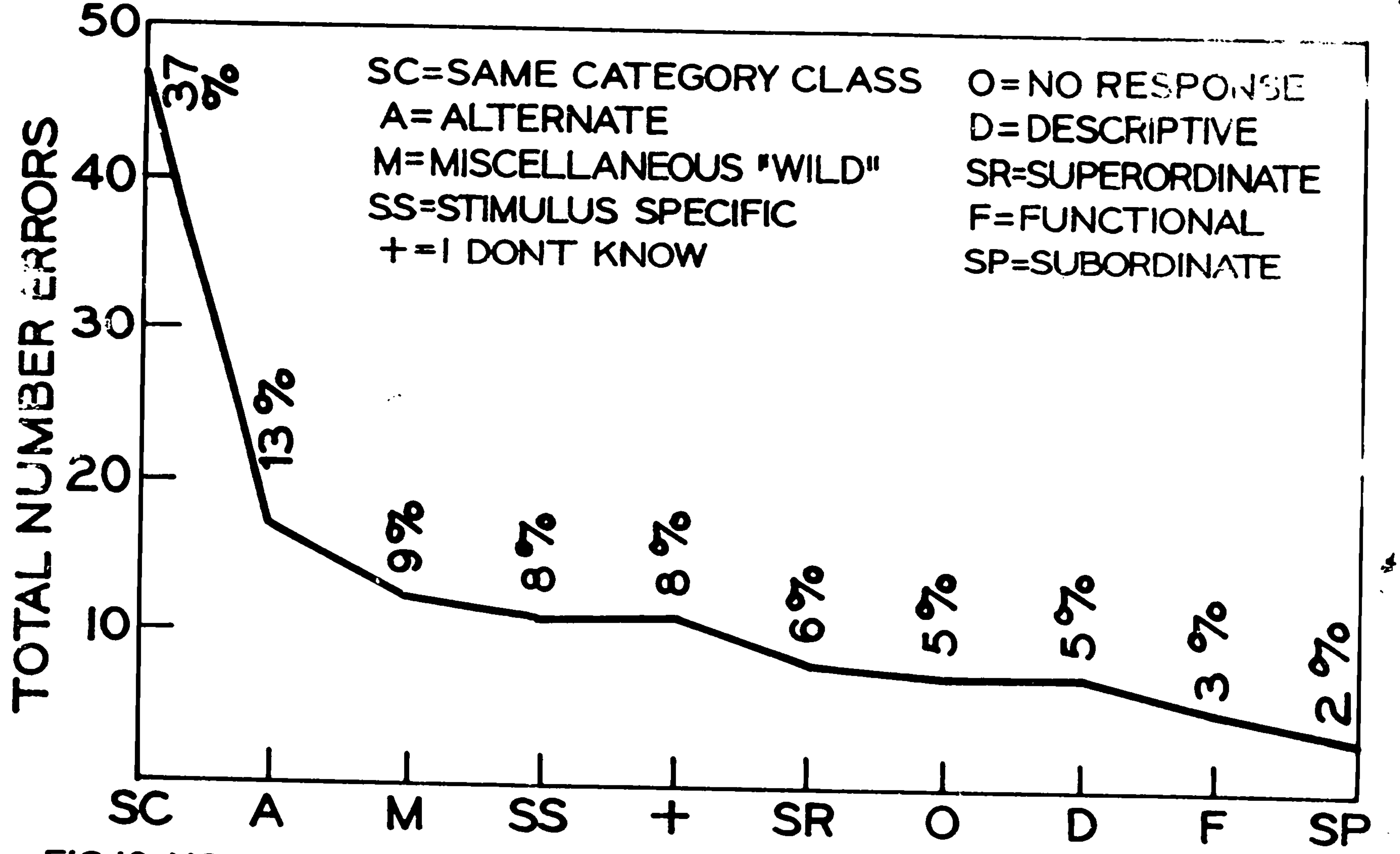



FIG.10: NO. OF ERRORS BY TYPE OF ERRORS FOR PRODUCTION TASK