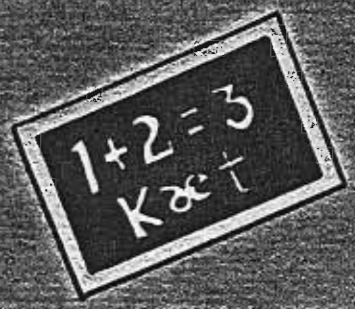
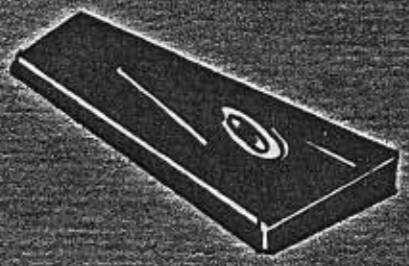


blăck board



black board

Children's Acquisition of Phonological Stress Contrasts

Kay Atkinson-King

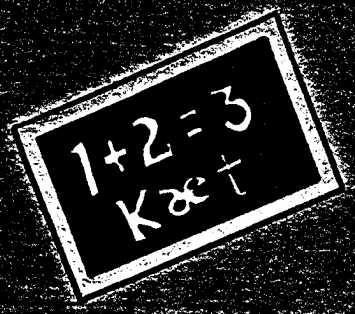
ucla

Working Papers in Phonetics 25

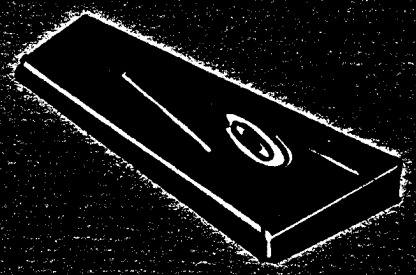
JANUARY 1973

hŏt dog

hot dŏg



black board



black board

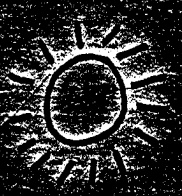
Children's Acquisition of Phonological Stress Contrasts

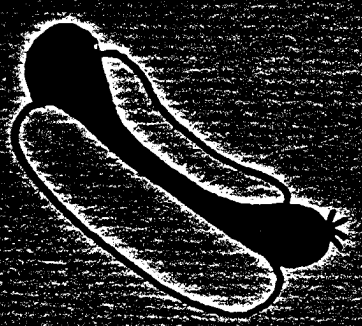
Kay Atkinson-King

ucla

Working Papers in Phonetics 25

JANUARY 1973





hot dog

hot dog

Children's Acquisition of Phonological Stress Contrasts

Kay Atkinson-King

Working Papers in Phonetics 25

January 1973

University of California, Los Angeles

Table of Contents

	Acknowledgments	iii
	Abstract	v
Chapter 1	Studies of Stress and of Noun Compounds and Noun Phrases with Adults and Children	1
Chapter 2	Identification Experiments with Compounds and Phrases	19
Chapter 3	Preference Experiment with Compounds and Phrases	61
Chapter 4	Semantic/Acoustic Experiment with Compounds and Phrases	82
Chapter 5	Production Experiments with Compounds and Phrases	102
Chapter 6	Experiments with Contrasting Stress Patterns in Other Constructions	118
Chapter 7	Summary and Discussion of Test Results	137
	Appendix	156
	List of References Cited	157

Acknowledgments

The conception, gestation, and ultimate birth of this dissertation have extended widely over time and space. Since it was conceived in Los Angeles, developed in Boston, and achieved final form in Munich, those individuals who have assisted with the delivery are numerous and scattered. The chief physician throughout the entire process was Peter Ladefoged, who sparked my interest in phonetics and linguistics and who gave me the desire, encouragement, and support to complete this project. Had he not taken the time and trouble to consult and console (everywhere from Los Angeles to London), this baby might have been thrown out with the bath water. Victoria Fromkin, who has encouraged and given helpful criticisms throughout my study of linguistics and particularly on this project, assisted generously in the delivery. I also appreciate the assistance of the other members of my committee who consulted on the case -- W.J. Dowling, Robert Wilson, and Robert Rees. In addition, I have benefited from consultations with colleagues at UCLA: Leon Jacobson, Breyne Arlene Moskowitz, Timothy Smith, and others.

The incubation period in the Boston area was particularly fruitful, and I am grateful to numerous individuals. Kenneth Stevens patiently supported me and offered useful advice which ultimately led to this subject. Merrill Garrett and Paula Menyuk contributed helpful suggestions and read and commented on an earlier draft of this work. Others who helped to make the time in Boston particularly fertile include Margaret Bullowa, Carol Chomsky, Susan Fischer, Jean Berko Gleason, William Henke, Dennis Klatt, Judy Kornfeld, Joan Mahling, Tom Shriner, and Jerry Wolf.

Richard Harshman spent many hours advising on the vital problem of statistics. A.W.F. Huggins also offered useful statistical advice. Peter Close of the MIT Day Camp and Norene Casey of the Bartlett School in Arlington, Massachusetts, cooperated generously in making subjects available. Keith North and Kathy Lach Call gave technical assistance in preparing materials which were used in gathering data. Those who edited the manuscript and assisted in its typing were Unity Evans, Julie Haaker, Barbara Hajek, Susan Hoover, Hedy Kodish, and Renee Wellin.

I owe a general debt of gratitude to those individuals who introduced me to the world of language acquisition. Among them are Susan Ervin-Tripp, Charles Ferguson, Wick Miller, and Dan Slobin. I am also grateful to the numerous teachers and colleagues who have contributed to my knowledge of linguistics, particularly at UCLA and MIT, and also at the University of California at Berkeley and Harvard University.

Last, I would like to thank my parents, James and Blanche Atkinson, for their encouragement and continuing interest in my work, and my husband, Robert King, who patiently prodded me along the painful path of delivery and held my hand during labor. This work is dedicated to them.

Kay Atkinson-King
Munich, Germany

This work was supported in part by grants from the National Institutes of Health, one administered through the Massachusetts Institute of Technology (NINDB Grant 2 ROI NB04332-07), and one administered through the University of California, Los Angeles, (NS 09780).

Abstract

Previous studies of the acquisition of linguistic stress in English have been very limited. Phonological research has generally centered around the first few years of life and thus has not dealt with complex stress rules but only, at best, with the use of lexical stress. Studies of stress with older children have focused on contrastive or emphatic stress rather than unemphatic or noncontrastive stress. Therefore the present study examined the acquisition of unemphatic stress patterns which signal differences of meaning, mark grammatical distinctions, and can be predicted by rule. Other topics considered included possible differences between the ability to comprehend and to produce stress contrasts; the relationship between syntax, semantics, and phonology in acquisition of stress patterns; the possibility of children's learning these stress patterns by lexical item rather than by rule; the relative order of acquisition of various stress rules; and the possibility of children's creating idiosyncratic rules at any time before mastering the adult rule.

The study experimentally investigated ability with these aspects of stress with approximately 300 children from the ages of five to thirteen. A number of adults were also tested to provide comparative data. The approach was one of extensive study with stress patterns in noun compounds and noun phrases (e.g., gréenhouse and green hóuse), but stress patterns with other structures were also considered (e.g., nouns and verbs such as récord, recórd). A number of experiments were conducted involving both comprehension and production skills. Three types of comprehension tasks were given: (1) identification with test items in isolation; (2) forced-choice with test items embedded in sentences; and (3) identification with test items embedded in semantically contradictory sentence frames. The production tasks included (1) free production of test items (a) in isolation and (b) in pairs in response to the experimenter's pointing to pictures; (2) imitations of the test items in isolation; and (3) production of sentences to contrast with sentences produced by the experimenter. A child's comprehension of his or her own productions was also tested.

The results of all comprehension and production tests with noun compounds and noun phrases were in basic agreement -- they indicated that the ability to identify or select these forms correctly and consistently on the basis of their distinguishing stress patterns and to produce this stress correctly is learned gradually and mainly as a function of age. Although an awareness of stress can be observed very early in life, ability with stress in these forms is not acquired until relatively late in the acquisition process. In general a child of five does not yet show ability with these stress contrasts but one of twelve does; and the closer to twelve a child is, the more likely he is to have acquired it. The main reason suggested

for this late acquisition is the interference of contrastive stress, which was found to be learned earlier than compound and phrasal stress.

The various comprehension and production tasks indicated that in general the order of mastery is: imitation, comprehension, and finally production. However, the types of task used to test skills of comprehension and production vary in difficulty, and children are able to perform some before others. The last task to be mastered was the ability to identify test items in semantically contradictory test frames; almost all but the oldest children were unable to ignore the contextual information and attend to the stress cues. The children also gave no evidence of creating idiosyncratic stress rules; they had no better success in identifying their own productions of stress contrasts than those of the experimenter.

Chapter 1

Studies of Stress and of Noun Compounds and Noun Phrases with Adults and Children

Children's acquisition of language is a process that has intrigued man for centuries. Although detailed observations of this process have been documented for two hundred years, interest in language acquisition has been particularly keen in the last decade as a result of the development of the theory of generative transformational grammar. A large portion of recent work has centered around determining how a child acquires linguistic competence with syntactic structures, but there has also been considerable investigation into the acquisition of phonology.* These and earlier investigations of phonology acquisition explore both segmental and prosodic (suprasegmental) aspects of language. However, in all studies of phonology acquisition, the concern with prosodic features has received much less attention than segmental development. Also, the work done on prosodies generally has centered on the first few years of the child's life, and thus on only the initial stages of prosodic development. The study of acquisition of linguistic stress in particular has been very limited in scope and depth. Research indicates that children both perceive and produce stress very early in life, but little has been done to trace development of ability with these features beyond this initial stage. For example, no studies have been made of when children use stress with linguistic reference and to signal meaning differences. The linguistic patternings of stress in English in particular are very complicated but systematic, a fact which has been emphasized by Chomsky's and Halle's (1968) detailed accounting of these patterns in the form of phonological stress rules. It is very probable that children do not learn to manipulate linguistic patternings of stress until well beyond the first years of life. However, up to now we have had no knowledge of what children of various ages know about linguistic stress patterns, nor have we known when and how this knowledge is acquired.

---

*A review of early studies in the acquisition of syntax and phonology can be found in McCarthy (1954). For reviews and examples of more recent work, see Smith and Miller (1966), Ervin-Tripp (1966), Hayes (1970), Reed (1970), Menyuk (1971), and Slobin (1971a).

The main purpose of this study is to provide data on these questions by investigating ability with regard to aspects of English stress in children from the ages of five to thirteen. Other topics considered include possible differences between the ability to comprehend and produce stress contrasts; the relationship between syntax, semantics, and phonology in acquisition of stress patterns; the relative order of acquisition of various stress rules; and the possibility of children's creating idiosyncratic stress rules at any time before mastering the adult rule.

The method of investigation is basically experimental in nature, although some observational data were also considered. The approach is one of extensive study with stress patterns in one set of structures, and a number of experiments were conducted involving both comprehension and production of these structures. Three types of comprehension tasks were given: (1) identification test with items in isolation; (2) preference test--i.e., forced-choice test--with test items embedded in sentences; and (3) identification test with test items embedded in semantically contradictory sentence frames. The production tasks included (1) free production of test items in isolation and in pairs in response to the experimenter pointing to pictures, i.e., naming pictures; (2) imitations of the test items in isolation; and (3) production of sentences to contrast with sentences produced by the experimenter. A unique feature of this study is that a child's comprehension of his or her own production was also tested. A large number of subjects were studied in order to give as accurate a picture as possible and also to consider individual variation. This method of study is different from the majority of studies of language acquisition since the most common approach is to concentrate only on comprehension or production skills of a few children in either an experimental or an observational setting.* It is hoped that through employing several different types of experimental tasks in the realm of both production and comprehension with a large number of children and through supplementing these data with observational information, a more complete and accurate assessment can be made concerning children's ability with stress patterns. Also, doing several tests confirms the reliability of test results. Hopefully this set of data will lead to further understanding of the general processes involved in the acquisition of stress in particular and of phonology and language in general.

The major portion of this investigation deals with children's ability with the stress patterns which distinguish noun compounds and noun phrases like gréenhouse, a hothouse for flowers, and green hóuse, a house painted

---

*McNeill (1966) commented on this lack of systematic experimental work done to explore differences between children's ability to produce and to comprehend specific aspects of language and noted that he knew of only one such study, that of Fraser, Bellugi, and Brown (1963). One more recent study which compares some elements of both comprehension and production is Shipley, Smith, and Gleitman (1969).

green. However, other constructions in which stress patterns can be used to signal a difference in meaning were also examined to a limited extent. Noun and verb pairs like récord and recórd were investigated briefly, as was contrastive or emphatic stress as it is used in sentences like "This is a réd house, not a gréen house."

There are several reasons for concentrating on stress in noun compounds and noun phrases rather than in these other constructions. First, stress is the only cue which differentiates these phrases when they are heard in isolation, i.e., they are minimal pairs with respect to stress. This is not true with constructions such as noun and verb contrasts like conduct and cónduct, which have vowel quality differences as well as stress differences even when presented in isolation. The reduced vowel which occurs in the unstressed syllable thus can act as a second cue to aid the listener in determining the part of speech. Since it is possible for a listener to ignore the stress shift and still identify the part of speech by attending only to the vowel reduction cues, these constructions are less suitable for testing stress comprehension.

Also, concentrating on stress in noun compounds and noun phrases did not appear to create independent difficulties caused by lack of knowledge of the lexical items involved. Since the test was to be one of knowledge of stress patterns and not of vocabulary items, an attempt was made to choose familiar items. This was easier in the case of compounds and noun phrases than in the case of the noun and verb pairs which are distinguished by stress and vowel reduction, since they are often more sophisticated, Latin-based vocabulary items that might not be known by a child at the age of five.* The noun compound and noun phrase pairs used did not seem to pose vocabulary difficulties for the majority of the children. In the noun phrases, although such combinations of adjective plus noun may have been new to the children (e.g., high chair, a chair that is located in a high position), all individual lexical items were easily understood and testing with individual children showed that they generally had no problems with this vocabulary or in understanding the structure and concepts. Likewise, the compounds used caused very little difficulty. A minor one seemed to be with compounds that have familiar synonyms -- for example, a few children usually used the term frankfurter or wiener for what was referred to in testing as hot dog, or they used the compound hothouse for what was called greenhouse. These dialectal or idiolectal differences did not pose any great problem, however, and testing individual children showed that such children rapidly learned the new compounds. As soon as pictures of the noun compound and noun phrase were shown by the experimenter and defined,

---

*In fact it was very difficult to find noun-verb pairs for testing which children of five years of age would know. This is one reason the testing with this construction was very limited, as will be seen in Chapter 6.

children had no difficulty with vocabulary items and the semantics of the sentence. The fact that even the youngest children tested understood both the syntactic structures and the semantics of the two types of phrase is very convincingly demonstrated by the results of the Semantic/Acoustic Test discussed in Chapter 4.

Another reason for concentrating on noun compounds and noun phrases is that they can be presented in isolation for identification. This is not true with stress in contrastive statements. Generally an entire sentence is required to test competence with contrastive stress. It could be argued that failure to give correct responses might be due not to lack of competence with the stress rules involved but rather due to confusion resulting from difficulty with the syntactic properties or vocabulary items involved in the sentence itself. That is, poor performance may reflect difficulty with the sentence frame rather than with the stress items being tested. Thus, constructions which can only be tested in a sentential context are not as sure a test of knowledge of stress rules as those which can be differentiated only by stress and can be tested in isolation.

The noun compound and noun phrase distinction was also selected for extensive testing because unlike the noun and verb pairs in isolation, these phrases were easily pictured and described. This allowed for more variation in testing procedures than would constructions that could not be pictured easily. (Pictures were also helpful in keeping children interested and entertained.)

A final reason for studying stress patterns in compounds and phrases is that unlike other constructions which involve only single lexical items, such as noun and verb pairs, these distinctions concern stress patterns with more than one element; thus this stress is more likely to be rule-governed than memorized. Although the compound, like the noun and verb, could involve memorized stress patterns, some kind of rule is required to assign stress at least to phrases. (More will be said later about rules for assignment of stress to compounds and phrases.)

Before beginning a more detailed consideration of the experiments which were conducted, the remainder of this chapter will be devoted to a general literature review dealing with the facts and theories concerning the acquisition of stress and of noun compounds and noun phrases. Chapters 2, 3, and 4 describe the comprehension experiments with noun compounds and noun phrases. Chapter 5 deals with the production of these structures. Chapter 6 turns to comprehension of stress in noun and verb pairs and the production of contrastive stress. The concluding chapter summarizes and discusses the results of all experiments; hence the individual chapters deal primarily only with method and the results of the specific experiments.

Studies on Stress

The term "stress" has been used loosely by researchers into child language. It has acted as a general cover term to describe prosodic elements at both the phonetic and phonological levels. Even emphatic or contrastive stress is often referred to merely as "stress" with no attempt to distinguish it from normal unemphatic lexical or sentence stress.* The distinction between stress at these different levels and its different functions is rarely noted or even realized, and it is usually up to the reader to determine in what sense stress is being studied.

This lack of clarity surrounding stress in language acquisition literature is perhaps a reflection of the state of confusion and controversy concerning stress which exists in linguistics and phonetics. Many theoretical and factual disputes surround this phenomenon, and a wide variety of differences in terminology exists. (For discussion and examples, see Lieberman, 1967; Vanderslice, 1968; and Ohala, 1970).

In this study the term "stress" will refer to the phonological level of representation. This is the abstract level where stress interacts with other rules of the grammar. Chomsky and Halle (1968) have postulated rules at this level which can predict the placement of varying degrees of stress by using syntactic and phonological information and the universal principle of the transformational cycle. Like earlier linguists (e.g., Trager and Smith, 1951), they postulate a number of levels (at least four) of phonological stress. However,Chomsky and Halle claim that these levels of stress most probably will not be present in the acoustic signal but are perceived on the basis of knowledge of the language. How many abstract phonological degrees of stress there are and what their relationship is to the physical signal are controversial issues (see Ohala, 1970). For example, Vanderslice and Ladefoged (1971) object to Chomsky's and Halle's multilevels of abstract stress with no physical correlates and suggest substituting binary suprasegmental features with various phonetic correlates. However, the existence of physical correlates to the phonological levels of stress appears to be most questionable for stresses beyond two levels. Most linguists agree that there are correlates at least to primary stress, and since in this study we will deal only with the question of children's ability with primary stress, arguments concerning the existence of correlates for weaker degrees of stress need not concern us. Primary stress is said to be manifested by the interaction of variations in fundamental frequency, duration, intensity, and vowel quality. (See, for example, Fry, 1955 and 1958; Bolinger, 1958; Lieberman, 1960; and Morton and Jassem, 1965) Bolinger points out that stress involves "pitch accent," i.e., pitch contours or obtrusions, rather than pitch levels. He also distinguishes between the actual manifestation of pitch accent and the potential a syllable has to display pitch accent. Ladefoged (1967) notes that although no single acoustic parameter appears to be a correlate of stress, there is one physiological correlate to it, since "every stress is accompanied by an extra increase of subglottal pressure" (p. 46).

* Emphatic and contrastive stress should perhaps also be differentiated, but for the purposes of this study they will be considered together as they are distinguished from unemphatic stress.

When stress is being discussed from a phonetic point of view in this study, it will be so indicated. In this connection it should be kept in mind that it is possible for stress to be phonetically observable as being produced or imitated by the child on occasions when it is not motivated or connected with the adult phonological, syntactic, and semantic systems and is not being used by the child to signal a meaning difference.

In this study the emphatic or contrastive use of stress will also be distinguished from normal unemphatic or noncontrastive stress. Contrastive stress differs from unemphatic or noncontrastive stress in that it can appear on almost any syllable or word of the sentence because it is not grammatically predictable in the same sense as normal noncontrastive stress. For example:

John deceived his friend (i.e., not Bill)
John deceived his friend (i.e., not received)
John deceived his friend (i.e., not delighted)
John deceived his friend (i.e., not her friend)
John deceived his friend (i.e., not his enemy)

The relation between semantics and phonology is much more direct in the case of contrastive stress than with other kinds of stress. Contrastive stress placement is dependent more upon semantic, situational aspects than upon syntactic and phonological factors. Also, whereas the placement of normal unemphatic linguistic stress in a sentence is entirely determined by the grammar and phonology, the placement of contrastive stress in these sentences has a paralinguistic element in that the speaker selects the item which he wants to receive focus and thus to be semantically contrasted. There are obligatory grammatical rules of coreference which then accompany this decision. The exact rules for this contrastive stress shift are not formalized in Chomsky and Halle (1968) but they apparently entail both syntactic and semantic parallelism. The fact that semantic coreference plays an important role is pointed out by Vanderslice and Ladefoged (1971) with examples like Ohala's (1970). In the sentence:

Instead of jailing the thief, it's the mayor we should incarcerate .

they note that the de-accentuation takes place with incarcerate because it is a synonym of to jail. Also, when an item has been mentioned in this way, and is old information, it can not only be de-accentuated, but also dropped out entirely:

a) This is the brown house, not the gray house.
b) This is the brown house, not the gray one.
c) This is the brown house, not the gray ---.

Concerning the rules for such sentences they note,

> We assume that the surface structures generated by the syntactic component encode these obligatory accentual effects by means which take place after the word-accent rules have applied, but before the late rules which assign rhythmic and stylistic accentual variations. (Vanderslice and Ladefoged, 1971, p. 12)

An example of these rhythmic and stylistic accentual variations is Bolinger's (1965) "accent recession" (e.g., "A clárinet sólo" versus "He plays the clarinét").

Vanderslice (1968) and Vanderslice and Ladefoged (1971) also make the point that contrastive stress can make a noun compound and a noun phrase homophonous:

(a) He saw a blǽck bird but not a gréen one (NP, contrastive)
(b) He saw a blǽckbird but not a crów (N, compound).*

They differ in this claim from the rules put forward by Chomsky and Halle (1968) which would assign bird secondary stress in (a) and tertiary stress in (b). As Ohala (1970) points out, some evidence for the homophony claim is unknowingly given by Lieberman (1967). In his experiment, a speaker said the two sentences:

(a) The life of a lighthousekeeper formerly was very lonely.
(b) Our maid weighed 180 pounds, but the Joneses had a light housekeeper for more than twenty years.

Lieberman then excised the minimal pair phrases and interchanged them. He reported that no differences were heard between the stress patterns of the altered and original sentences. Ohala notes that this is due to the contrastive stress shift in sentence (b) rather than to listeners' ignoring the acoustic signal and internally generating their own stress patterns according to their knowledge of the language.

Acquisition Studies on Unemphatic Stress

As noted earlier very little attention has been paid to the study of stress in language acquisition; few (if any) researchers have concentrated exclusively on stress. The little knowledge we have concerning stress derives mainly from more general works on language development and phonology. Stress in such research is generally studied in conjunction with other suprasegmentals like intonation, and it is sometimes not clearly differentiated from the other prosodic features. Since stress is closely allied to the latter, particularly at a phonetic level, we shall consider some of the studies on suprasegmentals in general in our review of developmental research on stress in the earliest stages of language development.

---

*Example (b) which Vanderslice and Ladefoged use is somewhat confusing since it also appears to call for contrastive stress, just as does (a). A better example would be (b) without the second clause: "He saw a blǽckbird."

Suprasegmentals are commonly noted to be the first elements of speech which children produce in the first year of life. It is often noted that infants imitate adult intonation patterns (e.g., Lewis, 1936; Pike, 1949); Lenneberg (1967) notes that an infant's intonation patterns become distinct at eight months. The early use of prosodic features is also reported by many others (see Weir, 1966; Lieberman, 1967; Kaplan, 1969 and Menyuk 1971 for references). Murai (1960, 1963/64) and Nakazima (1962) show that both Japanese and American infants acquire intonation patterns very early. Weir (1966) even claims that prosodic features (and syllabic structure, to some extent) are developed in the first year of life to the extent of displaying specific language differences. Her suggestion that early samples of speech from infants learning different languages (Russian, Chinese, and English) display perceptually identifiable differences is somewhat questionable, however, since such differences were not sufficiently great to allow for language discrimination or identification by adults (Atkinson, MacWhinney, and Stoel, 1969 and 1970). Presumably, if these languages themselves are different with respect to prosodic and rhythmic features, adults should be able to make such discriminations on the basis of these cues alone (Atkinson, 1968). Just how soon specific language features like intonation, stress, rhythm, and tone are incorporated into a child's grammar to the extent of being distinguishable by adults is not known. Chao (1951) claims that tone contrasts in Chinese are acquired by 28 months. E. Pike (1949) also reported that Mixteco-speaking children learn lexical tone as they learn the segmental items in their language and that their first words show correct lexical pitch. Fitzgerald (cited in Slobin and Welsh, 1967) notes that two-year-old speakers of the tone language Gã made more errors in spontaneous imitations of segmental phonology than in imitating the tonal or rhythmic structure of sentences. Slobin and Welsh suggest that the preservation of intonational and rhythmic aspects in imitation is basic and perhaps even universal. Intonation in particular has been so generally noted as being important, that Slobin (1970 a) suggests as a universal principle the idea that intonation is the first aspect in the developmental order of acquisition of linguistic means for expressing semantic notions.

Infants have also been shown to respond differentially to the acoustic parameters which are important in prosodic features. It has been demonstrated (Eisenberg, 1967) that infants respond differentially to various intensities, durations, and frequency ranges. Other studies using conditioning techniques, cardiac rate change and E.E.G. responses are becoming popular as ways to determine infants' responses to different kinds of auditory stimuli (See Menyuk, 1971 and 1972 for a review of such studies). Kaplan (1969) used the motor behavior and cardiac rate responses to experimentally demonstrate eight-month-old infants' discriminatory responses to rising and falling intonation contours with stress on the terminal segment. Infants' responses to intonation as a meaningful signal had been noted only observationally and anecdotally in earlier literature (e.g., Lewis, 1936).

Presumably all of this evidence of children's sensitivity to prosodic features in the first year of life deals with these features at a phonetic level of representation. They may imitate and respond differentially to rising and falling intonation patterns and be sensitive to the acoustic parameters which are correlates of stress in the adult languages but there is no evidence that they are using prosodic features (particularly stress) with linguistic reference and to contrast meaning.

The literature dealing with children's acquisition of stress beyond this first year again indicates that children are sensitive to stress, but it still leaves some doubt as to when more complex phonological ability with stress patterns is acquired. It is commonly noted in data on production that children choose stressed syllables when making substitutions and that they use different substitution rules for stressed and unstressed syllables (see Ervin-Tripp, 1966 for reference). Velten (1943) in a study of his daughter from the age of 11 to 36 months reported that stress influenced the phonemic structure. She used certain vowel substitutions for stressed vowels and other vowel substitutions for unstressed vowels. Leopold (1949) noted that his bilingual child repeated the word with the strongest stress both in English and in German. Various research in imitation studies has shown that stressed words are generally retained in repetitions more often than unstressed words (e.g., Todd and Reynolds, in press). Brown and Fraser (1963) and Brown and Bellugi (1964) suggest that content words are retained in children's repetitions because they are stressed. However, Scholes (1970) has suggested that stress may not be a good explanation for the retention of content words since deletion of function words rather than content words still persists when no stress differences are present. Menyuk (1971) suggests that the Shipley, Smith, and Gleitman (1969) study shows that 15-to-30 month-old children do not attend to stress so much as to the item which is most familiar or to the structure of the utterance. Blasdell and Jensen (1970) looked at the effects of stress and position in the string with children's recall of nonsense syllables. They found that 28- to 39-month-old children most often recalled the syllable with primary stress and the last syllable in a four-syllable string. The possibly greater importance of the syllable-final position factor over the stress factor is suggested by data from Pačesová presented in Slobin (1971 b). Pačesová claims that in Czech, where all words receive initial stress, the stressed initial syllables are often omitted in child speech, while unstressed final syllables are not. On the other hand, Slobin also quotes Rūķe-Dravina as observing that in Latvian, children stress newly acquired connecting words like conjunctions, even if such words are unstressed in adult speech. Thus the importance of stress in language development is far from clear.

Moskowitz (1970) claims that a two-year-old child had a phonological rule for dropping an initial zero-stress syllable before a subsequent primary stress syllable (e.g., [mɛ́ɛ̆də̆otš̆] 'tomatoes'). This rule was apparently obligatory since it was used in 10 tokens but ignored in 8.

Weir (1962) reports that her two-and-a-half-year-old son used stress in accordance with standard English consistently and correctly.* He did not make mistakes in placing heavy stress (e.g., [rédi] 'ready'); however, he did have a tend ncy to overuse stress and sometimes placed it on an additional syllable (e.g., [šàmp^{h}ú] and [šæmp^{h}ú] both occurred for 'shampoo'). She also gives an example of a near-minimal pair with two lexical items: [ɑ́ɪdɪ̄] 'Idey' (a proper name) and [àidíə] 'idea.' She concluded from her study that her son had learned the feature of stress as opposed to no stress but not the contrastive use of various levels of stress. However, it is possible that this use of correct stress patterns within lexical items which Weir cites involves imitation of stress rather than rules for assigning stress. Miller and Ervin (1964) found that the five children in their study had all developed a two-level pitch system with primary and secondary stress around age 2.2 at the time they developed a linguistic system with marked word classes (noun and verb suffixes, etc.). They observe that although children are good mimics of prosodic features and appear to have the pitch and stress system under control, "this may be true from a phonetic, perhaps even phonemic standpoint, but does not necessarily entail the use of prosodic features in the grammatical system" (p. 29). Miller in a personal communication suggested that this statement means that children may have 4 pitch and 4 stress phonemes (in the Trager-Smith tradition), but they do not necessarily use them as the adult language user does. He and Ervin-Tripp claimed that in the early period children sometimes used prosodic elements which had no analogue in the adult language. They also state in discussing data from two of the children that "each construction in both children's linguistic system had a particular accent pattern, but the accent pattern was not always adhered to" (p. 29). Apparently some fluctuation also occurred with intonation. Although it is generally observed in the child language literature that young children may use intonation patterns to indicate sentence boundaries (Weir, 1962) and sentence types (e.g., question, declarative, imperative, etc.), Miller and Ervin note that one child was over two years of age before rising intonation consistently indicated a question. Weir (1962 and 1966) also noted that although early use of intonation is commonly reported, she was unable to find systematically contrasting patterns with a two-and-a-half-year old child.

Intonation patterns apparently remain structurally ambiguous fairly long after they first appear to be used. As Weir (1966) points out, the interrelated processes of acquisition of syntax and acquisition of phonology need to have more research to fully understand this problem. The same kind of problem also appears to arise in learning to use stress as predicted by grammatical rules. At an extremely early age, children may produce and perceive changes in the acoustic signal which correspond to what adults call stress, and they may correctly produce stress patterns

---

*She calls this "contrastive stress" but her discussion indicates that she means stress in contrast with no stress, not emphatic stress.

in lexical items; yet they may not incorporate the use of stress patterns in accordance with grammatical rules until they are much older. Using stress to contrast grammatical distinctions may therefore be a late development.

The point at which stress becomes a surface structure marker to differentiate underlying syntactic and semantic relations within sentences is not known, and there are few data on this question. Weir (1962) claims that her two-and-a-half-year-old son used stress to distinguish between noun compounds and noun phrases by stressing both syllables of the latter construction and only the first syllable of the former (e.g., two féet versus Twófeet, the name of an Indian). However, because the child was reported frequently and optionally to place an extra stress on syllables (e.g., the shàmpóo-shámpoo examples noted above), this claim can be questioned. Also, as will be seen below, the results of the present study cast even greater doubt on Weir's claims. Either the child was exceptionally advanced (which cannot be precluded) or her interpretation is incorrect.

Braine (1963) gives the examples "bàby cháir ('little chair': pivotal construction), usually without an intonation break, and báby # cháir ('the baby is in the chair')" (pp. 10-11).* These examples might be considered as evidence for use of stress patterns to differentiate meaning; however, the utterances are not minimal pairs with respect to stress, and it is not obvious what the importance of the stress pattern is in relation to other phonetic differences. Also, use of such prosodic features is apparently not consistent, since Braine adds that "the two constructions, however, certainly cannot always be distinguished from their intonation alone" (p. 11). Miller and Ervin (1964) observed that one two-year-old child used different stress patterns to indicate locative and possessive constructions. The locative had primary stress on the second word (e.g., baby róom in answer to "Where's the baby?") and the possessive had stress on the first word (e.g., báby book in answer to "Is that the baby's book?") Ervin-Tripp (1966) also states in discussing the Miller and Ervin study that "'Christy room,' if the first vowel was stressed, was a possessive, if the second, a locative construction" (p. 75). The Miller and Ervin examples of this child using stress to make these grammatical distinctions have been noted by various researchers (e.g., Slobin 1968 and 1970b; Menyuk, 1971); some researchers have concluded on the basis of these few examples from this one child that in the two-word stage children use stress to mark grammatical distinctions within sentence types. However, there is so little evidence that this conclusion appears premature. Also, Miller and Ervin's data may be questioned because their examples may not be instances of normal, unemphatic stress functioning as a consistent linguistic marker of these grammatical relations. First, they state that "the accent pattern was not

---

*# is used as a juncture symbol.

consistent in the possessive or adjectival construction" (p. 22). Second, it is possible that rather than this being an example of stress being used to distinguish these syntactic oppositions, it could involve use of focus and contrastive stress; the child may be emphasizing the word she desires in order to answer a question. This use of stress seems rather different from unemphatic syntactic stress. And as we shall see below, contrastive stress appears to be learned very early. Elsewhere, in discussing data from another child, Miller and Ervin note that this other child used the "emphatic transformation" sporadically; thus, it is obvious that Miller and Ervin were aware of the difference between contrastive stress and normal noncontrastive stress in recording their observations. It nonetheless seems possible that the stress they observed as marking grammatical distinctions in the two-word grammar does not represent use of an unemphatic phonological stress rule.

Roger Brown (1969) offers further evidence which makes the early use of stress patterns to distinguish semantic and syntactic relations appear questionable. Brown discusses Bloom's (1970) evidence that children's two-word utterances express various underlying semantic relationships like agent-object, genitive, and locative. Brown suggests that prosodic features are not used to distinguish these different semantic intentions, but rather that word order is the key factor. To demonstrate this claim, Brown compiled 132 noun plus noun constructions from Eve, a child who has been the subject of a longitudinal study. He assigned possible relational interpretations to these utterances on the basis of the mother's expansions and, infrequently, the production of a function word. He then asked Melissa Bowerman to make a narrow phonetic transcription including suprasegmental features of these 132 utterances. He gave her no prior information concerning the semantic category to which each utterance had been assigned. Brown explained it this way:

> The question was: Are there phonetic features, independent of word order, which create response types reliably associated with particular relational meanings? The answer is negative. Eve did not reliably signal her presumed specific semantic intentions with suprasegmentals or any other phonetic feature. (p. 115).

He also claims that the records of another child he has studied support this finding. Thus, whether normal unemphatic stress or even contrastive stress is used consistently to distinguish underlying syntactic relations or semantic intentions as early as the two-word stage is a question on which there is contradictory evidence.

Acquisition Studies with Contrastive Stress

Since it is not always clear whether researchers are discussing normal stress or contrastive stress (as was noted earlier), it is sometimes difficult to differentiate these phenomena in the developmental literature. However, there are a number of studies which clearly involve contrastive stress, and some use this label. Brown (1969) gives examples of young

children using stress in spontaneous speech which is most certainly contrastive, though he only refers to it as "primary stress." At the age of approximately 1.10 years, Eve went around a circle of people pointing to each person's nose and saying: "That Pápa nose; that Máma nose; that Éve nose." Each time she stressed the possessor, the word providing new information. Evidence for early use of contrastive stress also comes from Weir (1962). She cites examples of her two-and-a-half-year-old son's using "extra heavy stress" in what she calls "the emotive function of language" (p. 113). For example, Anthony used the phrase thrée trucks not as an enumeration but as a correction meaning "not two."

Further evidence of young children's early understanding and use of contrastive stress comes from Slobin and Welsh (1967). (They used both the terms "emphasis" and "stress" in the study, but it is clear that in all cases the stress studied was emphatic or contrastive stress.) In eliciting imitations from a two-year-old child, they found that contrastive stress was generally imitated and that it can lead to retention of words sometimes omitted in imitation. However, the picture is not a simple one of every emphasized word being retained. For example, they note that repeated words were generally not imitated even when they were emphasized. To the sentence, "I néed néed the ball," Echo repeated "I need the ball." However, word repetition could be recorded by the child as emphasis. Thus contrary to her usual pattern of responding with emphasis only when emphasis was used in the model utterance, Echo imitated, "Whére Kitty" to "Where where is kitty." Likewise, her imitation of the sentence "Mozart fell off off the table," was Mozart fell óff the table." Slobin and Welsh suggest that Echo's apparent equivalence of word repetition and emphasis may be analogous to the similarity between adverbial reiteration and emphasis in adult English with phrases like "very, very good" and "véry good." These data on Echo thus suggest that even at this young age she understood the semantic notion of focus and emphasis at least to some degree.

Hornby (1970, 1971) conducted an experiment with contrastive stress and other surface structure devices as they mark the topic-comment relation in sentences. In spoken language it is possible to divide sentences into two parts: what the sentence is about (the topic) and the new information which is being given by the speaker (the comment). For example, in the sentence "The bóy is riding the bicycle," boy would be the comment and bicycle would be part of the topic. This distinction is said to be indicated in surface structure by word order, grammatical structure, and contrastive stress. (For references, see Hornby 1970 and 1971). Hornby tested children on the First-,Third-,and Fifth-Grade levels to determine the relative importance of these different types of surface-structure phenomenon in making the topic-comment distinction. Results from both production and comprehension tasks indicated that children at these ages are able to utilize contrastive stress as well as other aspects of surface structure to determine

the topic-comment relation. Contrastive stress was the primary device used by these children to mark the topic-comment relation. Hornby and Hass (1970) also experimentally demonstrated that children as young as four years of age were able to use contrastive stress to mark the comment of a sentence. Here again, contrastive stress appeared to be more primary than word order in marking the topic-comment distinction. Hornby (1971) suggests that results from these studies indicate that investigations of children's two-word utterances should take into account intonation patterns as well as word order. Indeed, as noted above, the possibility of a child's using contrastive stress at this early stage does not appear unreasonable; in fact it is perhaps more likely than use of more grammatically complicated, unemphatic linguistic stress patterns.

Carol Chomsky (1971) studied the effect of contrastive stress on pronoun reference. Using the same basic type of experimental techniques as described in Chomsky (1969), she tested thirty-six children whose ages ranged from six to ten years. She asked children to respond to simple coordinated sentences like: (a) John hit Bill, and then Peter hit him. (him=Bill); (b) John hit Bill, and then Peter hit hím. (hím=John). The object pronoun in such sentences has a different referent depending on whether the pronoun is contrastively or normally stressed. (For further discussion, see Akmajian and Jackendoff, 1970). Chomsky found that all but one child tested successfully performed this task. She therefore concludes that by age six, this use of contrastive stress is well established.*

In summary, past research on children's acquisition offers some evidence that infants are sensitive to the acoustic parameters of what is called stress in the adult language. In the early stages of language development they show further sensitivity to stress in such ways as retaining stressed words more than unstressed words in imitation and "telegraphic" speech. They also appear to use contrastive stress fairly early. However, it is not clear how early children are able to use stress to differentiate meaning within a sentence. The use of stress as it is predicted by grammatical rule has been essentially unstudied until now.

Studies on Noun Compounds and Noun Phrases

For some time the stress patterns distinguishing noun compounds from noun phrases (e.g. blácкboard and black bóard) have been discussed in the

*In the same study Chomsky also performed experiments with effects of contrastive stress on focus of sentence negation, but her results were not clear. Of the 36 children tested, 34 were unable to perform the task successfully. However, she blamed this upon the difficulty of the task rather than lack of linguistic ability, since children who failed in the test responded to questions she asked in a way which led her to conclude that they have this linguistic skill. Further testing will be required to substantiate this point.

literature on English phonetics and phonology. (Henceforth, the terms "compound" and "phrase" will be used to distinguish these two structures.) Chomsky's and Halle's (1968) recent formalization of these patterns into rules has generated new interest in such stress patterns basically because of theoretical claims which they make in positing them. They account for the difference in the phonetic shape of these two constructions by utilization of a transformational cycle of ordered, recursive phonological rules, which apply to the labeled bracketing of the surface structure.

Placement of main stress in these constructions is thus predictable from the syntactic structure by means of this transformational cycle. The surface structure difference in the constructions blackboard and black board is that the former belongs to the category "noun" and the latter belongs to the category "noun phrase."

Stress is assigned to these constructions by the Compound Rule and the Nuclear Stress Rule, both of which are formalized as part of a single larger rule (p. 240, rule 18; see p. 16 for example derivations). However, because our study focuses on distinguishing between the two, they will be treated here as two separate rules. These rules are disjunctively ordered, the Compound Rule applying prior to the Nuclear Stress Rule. Chomsky and Halle claim that although these rules are specific to English and are learned as part of the mastery of that language, the principle of the transformational cycle and organization and conditions on these rules are not language specific and are not learned; they are innate and universal. The claims concerning the presence of innate structures have become the center of a great deal of controversy, and the stress rules themselves have also been widely discussed.

In addition to theoretical disputes concerning rules for assigning stress to compounds and phrases, recently there has also been controversy over the description of the compounds and phrases themselves. Some descriptions of compounds in particular have been based mainly on syntax, others mainly on semantics (e.g., see Lees, 1960 and 1970; Marchand, 1966 and 1967; Zimmer, 1970; Li, 1970; Gleitman and Gleitman, 1970). Nevertheless most linguists agree that there must be some kind of recursive generative device to account for compounding since it is a productive, creative process. Evidence that compounding is a productive process in the language is plentiful.* As Gleitman and Gleitman (1970) point out, however, there are many compounds with "frozen" or set meanings (e.g., horsecart) which could be

*An interesting example which contrasts a compound and a phrase appeared in an advertisement for Volvo, (U.S. News and World Report, February 9, 1971, p. 1). After discussing the quality and durability of this brand of automobile, the question was asked, "Are you sure you're in the market for a hardtop? Or is what you really want a hard top?"

entered into a grammar as lexical items rather than being generated by a recursive mechanism. The Gleitmans' work also suggests that there may be individual differences related to noun-compounding skills and rules where compounds of more than two words are involved.

Acquisition Studies of Compounds and Phrases

Language acquisition literature commonly reports that the structures involved in compounds and phrases are learned and used fairly early. Words which are nouns and adjectives in the adult grammar (though not necessarily in the child's) are present in the two-word utterances of children about age two (Braine, 1963; Brown and Frazer, 1963; Miller and Ervin, 1964). Shortly after this, the three-word utterance stage appears and hierarchical constructions begin to emerge. Brown and Bellugi (1964) have documented the emergence of the noun phrase in general at this stage, citing examples of the adjective-plus-noun type of noun phrase with which we are concerned in this study, and they also give examples of compound nouns. They note that in an earlier stage utterances like "That flower" and "Put hat on" were found; later, with the emergence of the noun phrase, the utterances "That a blue flower" and "Put the cowboy hat on" occurred.

Young children frequently use vocabulary items which are compounds in the adult language. This use of compounds does not necessarily reflect the use of rules for compounding, however; it is possible that such forms can be learned as set lexical items with no awareness of the compound being constructed from two words with separate meanings. Berko (1958) examined the question of whether children aged four to seven were aware of the two separate elements of certain compounds. She chose 14 test stimuli which were common in the adult language and generally had one set or "frozen" meaning (e.g., Thanksgiving). She found that although some of the older children were at a stage when they noticed the separate parts of the compounds, most of the children gave evidence of learning these compounds as simple nouns. For example, when asked why a blackboard was called a blackboard, children generally gave an answer like "because you write on it." Livant (1962) found the same general results when he gave Berko's list of compounds to four- and five-year olds. This suggests that children may learn as lexical items at least some kinds of compounds which occur with set meanings in the adult language and that they may become aware of constructional features of these compounds with set meanings only at a later stage of development. However, the issue is not clear because Weir (1962) cites evidence to indicate that her two-and-a-half-year-old son did have an awareness of the separate parts of compounds. For example, during one of Anthony's monologues the sequence "cider, applecider, cider" appeared. Also, he decomposed something into its two parts, replaced some by his own negative [nʌ́m], and created [nʌ́mfɪŋ] 'nothing.'

On the other hand, evidence that children have rules for compounding and can create unique compounds is also found in the literature. Berko (1958) suggests that compounding is a productive process in children from ages four to seven. When they were asked questions that called for compounding or deriving new words, they did not use derivational suffixes but rather a compounding pattern with primary stress on the first word. For example, she asked them "What would you call a very tiny *wug?" Of the children queried, 52 per cent formed compounds like téeny*wug, báby*wug, and líttle*wug. All had a primary stress on the first syllable. Adults tended to offer diminutive suffixes to this question. Half said *wuglet; other responses were little *wug, *wuggie, *wugette, and *wugling. Similar results were found when she asked children "What would you call a house that a *wug lives in?" No children responded with suffixes, and those older children who understood the question (the younger ones apparently did not) formed compounds. Of the First-Graders 18 per cent said *wughouse; others gave compounds like birdcage. Of the adults, 58 per cent created the compound *wughouse in response to this question, and the rest formed *wugshouse, *wuggery, and *wughut. Finally, children were asked "This is a dog with *quirks /kwərks/ on him. He is all covered with *quirks. What kind of dog is he? He is a _____dog." Berko reports that 64 per cent of them responded with the compound *quírk dog. Adults presented with this question unanimously responded with *quirky dog.* Although children were probably producing compounds in Berko's study, it is possible that the primary stress observed to be on the first syllable is not compound stress, but contrastive stress. In the case of the *wughouse example, it is possible that *wug received primary stress because it was the unusual word, the new information, and the word which received primary stress in the test question. This is even more likely in the last example noted; children would produce contrastive stress on the new information, particularly in response to the frame: "He is a _____dog."

Young children have also been observed to produce creative compounds spontaneously. For example, Weir (1962) reports that her son created bathing room and phone call book. Livant (1962) notes that after his two-year-old son was told that "a map is a paper we use when we go on vacation," the child referred to a map as "a vacation paper" (p. 18). Gleason, King, and Mandelkorn (1970) also found a similar example in a comparative study of communicative competence with advantaged and disadvantaged four- and five-year old preschoolers. Some less advanced children used noun compounding to shorten long sentences that had dependent clauses. Thus in one training task, most children were able to produce sentences of the general form, "The M & M is under the box that has the mitten on it," but less advanced children responded with a sentence like, "The M & M is underneath the mítten box." In all of these cases the stress pattern used appeared to be the compound stress pattern, but again it is possible that

*Berko does not indicate what stress pattern adults used with this construction. When we asked adults "What would you call a dog that is covered with "quirks?" (without Berko's final sentence "He is a _____dog"), we found that the majority of them place primary stress on dog.

it is an example of contrastive stress since the child was singling out one of four items in his description (i.e., it's underneath the mítten box, not the dóll box, etc.).

In conclusion, there is still much to be learned about the acquisition of phrases, and particularly of compounds. Exactly what kind of rules for compounding are acquired and when and how they are acquired and applied is not known. As Gleitman and Gleitman (1970) demonstrated, it is possible that not even the same compounding rules are learned or that the same degree of competence is attained by all speakers of the language. However, it has been shown that children not only use compounds which appear in the adult language (and which thus could be merely memorized as lexical items), but they also create unique compounds fairly early. Furthermore, when they use compounds, they use the correct stress pattern, though it is possible that stress is contrastive rather than compound.

The child language literature gives evidence of children's using the correct stress patterns for compounds; little has been noted concerning stress in phrases, however. Nothing is known about how children might acquire rules to assign stress in compounds and phrases. For example, if they first learn compounds as simple nouns and are not aware of the separate elements (as some studies have shown), it could be argued that children memorize the stress pattern of each lexical item rather than learn a rule such as the compound rule posited in Chomsky and Halle (1968). On the other hand, a child's use of correct stress patterns with unique compounds (which is also reported in child language studies) argues for the use of a rule for assigning stress to compounds (if it is not contrastive stress instead), since a child cannot memorize a stress pattern for a compound he has never heard before. The following study attempts to shed light on this and other questions related to children's acquisition of the stress patterns in compounds and phrases.

Chapter 2

Identification Experiments with Compounds and Phrases

The main type of testing procedure used to determine children's competence with regard to the comprehension of the stress patterns involved in distinguishing compounds and phrases was an identification test. The initial Identification Test, which will be referred to as Test 1, was given to 160 children and 18 adults and will be our main concern here. This test was later repeated, first in its entirety and then in part, with two additional test groups. These two supplemental tests were given to determine the reliability of the Test 1 results and to provide further data in areas where the initial results were inconclusive. The first of the supplemental tests, Test 2, was an exact repeat of Test 1, and it was given to 32 of the original 178 subjects several days after Test 1 had been administered. This was done in order to determine the individual and group reliability of the main test. The second, Test 3, was a partial replication of the original test consisting of the first 13 items of Test 1. It was given one year later to an additional 125 children and 19 adults.

There were several reasons for administering Test 3. First, it tested the group reliability of the results of the main test; unlike Test 2, which showed that the same children performed the same way on a second and later trial of the test, Test 3 demonstrated that a different group of children performed approximately the same way as did the original test group. Test 3 also gave additional data based on a large number of different subjects to help provide answers to questions raised by Test 1. The third and primary reason for giving Test 3 was that it permitted comparison and correlation of compound and phrase stress identification results with results from additional stress tests which were given to the Test 3 subjects. In the Test 3 session, competence with compounds and phrases was measured with additional testing procedures (the Preference Test and the Semantic/Acoustic Test), and linguistic ability with stress as it distinguished noun and verb pairs was also tested. (These tests are discussed in Chapters 3, 4, and 6.) Replicating the initial identification test with the Test 3 subjects made it possible to consider the order of acquisition and relative difficulty of these various constructions which involve stress rules both for age groups in general and for each individual.

The main concern in this chapter is Test 1, the initial and most extensive identification test. The results of Tests 2 and 3 will be considered mainly when they clarify or conflict with the results of Test 1. The slight differences in procedure for the three tests will be discussed with the procedures of Test 1.

METHOD

Subjects

The 285 children used as subjects in the three identification tests were in Kindergarten through Grade 8 (ages ranged from 5 to 13 years). All were residents of the Boston, Massachusetts, area and were native speakers of English. The majority of the children in the sample (236) were tested at the Massachusetts Institute of Technology Day Camp, a summer recreational program available to children of all employees of M.I.T. -- maintenance and secretarial staff, as well as teaching and research staff. Most children came from a middle-class environment but there was some variation in socioeconomic status and educational background of parents. The remaining 49 children were pupils at the Bartlett School, a private elementary school in Arlington, Massachusetts. These children were also from middle-class backgrounds with some socioeconomic variation, as determined by the neighborhood and profession of parents.

Thirty-seven adults, ranging from 16 to 23 years of age, were included as subjects in the study. The adult subjects, like the children, were also all native speakers of English who were residing in the Boston area. In order to provide a basis for comparison, they were tested in an identical manner to and in sessions with the children. The 37 adult subjects were counselors to the children in the M.I.T. Day Camp. They varied in socioeconomic and educational background, but in general were middle-class. Some of the younger adult subjects were still in high school. Most of the older subjects were working and a few were attending college. There was one graduate student.

For Test 1, a total of 160 children in Kindergarten through Grade 8, whose ages were from 5 to 13 years, were used as subjects. These children were tested in two main groups. From the M.I.T. Day Camp, 111 children and 18 adults were tested at the end of the summer of 1969. The remaining 49 children were in Kindergarten and Grade 1 at the Bartlett School and were tested the following winter. These additional 49 children were tested because the results from the first group of 111 children indicated that the most critical differences in performance when distinguishing the phrases and compounds by stress occurred between the ages of 5 and 7 years, but the data were somewhat sparse at these ages. Therefore additional data were obtained from children whose ages fell in this range. Since the first 111 children were tested just prior to the beginning of the school year, the data were gathered from the additional 49 children at the middle of the school year. This gave samplings of children at half-year stages in their educational development. To consider further possible differences due to formal education level, children were grouped according to grade placement in school rather than age. The children in Kindergarten and Grade 1 tested at mid-year will be referred to as the "Mid-Kindergarten" and "Mid-1" groups, whereas the First-Graders tested at the end of the summer just prior to the beginning of the school year will be called the "Pre-1" group. All

other grades will also be prefixed by "Pre-" to indicate the time of testing. The grade and sex breakdown for all subjects in Test 1 is as follows:

TABLE 2.1: Number of Test 1 Subjects by Grade and Sex

	Boys	Girls	Total
Mid-Kindergarten	10	9	19
Pre-Grade 1	5	5	10
Mid-Grade 1	15	15	30
Pre-Grade 2	13	4	17
Pre-Grade 3	10	6	16
Pre-Grade 4	11	9	20
Pre-Grade 5	17	10	27
Pre-Grade 6	3	7	10
Pre-Grade 7	1	4	5
Pre-Grade 8	4	2	6
Total	89	71	160
Adults			18

Test 2, the repeat test, was given to 32 of the original 160 test subjects of the M.I.T. Day Camp group several days after Test 1 had been administered. The breakdown by grade and number of subjects is as follows:*

TABLE 2.2: Number of Test 2 Subjects by Grade

	No. of Subjects
Grade 2	7
Grade 3	9
Grade 4	3
Grade 5	9
Grade 6	1
Total	29
Adults	3

*The repeat test is not broken down by sex as are the two larger test groups because the repeat group contained a relatively small number of subjects and its main purpose was to determine general and individual subject reliability rather than to compare specific results such as differences in performance between sexes.

The Test 3, the partial replication test which supplemented Test 1, was given in the summer of 1970 (one year after Test 1) to a total of 125 children aged 6 through 13, representing Grades 1 through 8.* These children were all participants in the M.I.T. Day Camp. The number of children by grade and sex for Test 3 are as follows:

TABLE 2.3: Number of Test 3 Subjects by Grade and Sex

	Boys	Girls	Total
Grade 1	8	0	8
Grade 2	13	14	27
Grade 3	9	7	16
Grade 4	17	8	25
Grade 5	13	5	18
Grade 6	10	10	20
Grades 7 and 8	1	10	11
Total	71	54	125
Adults			19

Again an additional 19 adults were tested in groups with the children.

Of the 125 children taking Test 3, a number had also taken Test 1 a year earlier. The number and grade of the children who took both tests are as follows:

TABLE 2.4: Number of Children Who Were Subjects in Both Test 1 and Test 3

	No. of Subjects
Pre-Grade 1	3
Pre-Grade 2	0
Pre-Grade 3	6
Pre-Grade 4	4
Pre-Grade 5	4
Pre-Grade 6	1
Pre-Grades 7 and 8	3
Total	21

*Since all children were tested at the same time, no pre- or mid- labels will be used. However, since testing was done before the school year started, the group listed here as Grade 1 is equivalent to the Pre-Grade 1 group in Test 1, etc.

The Identification Test was also given to a number of the additional children who were tested in the production of compounds and phrases; they will be considered in Chapter 5.

Stimuli and Testing Apparatus

Real words were used in testing rather than nonsense words. The use of nonsense words would have been preferable for testing children's ability with stress rules in a new situation; it would suggest that success did not indicate knowledge of the stress pattern of a particular lexical item which could have been memorized, but rather use of a rule. In attempting to create a test situation to investigate children's productive use of these stress rules, however, it was found that contrastive stress shift made this a difficult and perhaps even impossible task. With nonsense words it appeared easier to test for stress in compounds than in phrases, but contrastive stress could intervene in either case. The discussion in Chapter 1 of compounding with nonsense syllables reported in Berko's (1958) study pointed out the problems involved here.

For example, if we tried to elicit a phrase with, "This is a clean *wug. What do you call this?" (showing a picture of a dirty *wug), the most natural response would be "a dírty *wug," because of the contrastive stress rule. Since there was no obvious way to by-pass this difficulty, testing was done with real words.

Six pairs of compounds and phrases which were minimal pairs with respect to stress were selected for the identification experiments. The two main criteria for selecting these pairs were first, that they would be familiar to young children, and second, that they could be pictured easily. These pairs are found in Table 2.5.

TABLE 2.5: Compound and Phrasal Minimal Pair Test Stimuli

Noun Compounds	Noun Phrases
gréenhouse	green hóuse
rédhead	red héad
hígh chair	high cháir
blácкboard	black bóard
Réd Sox	red sócks
hót dog	hot dóg

Either one or the other member of each of the six pairs was selected for presentation on five occasions in Test 1, making a total of 30 items, of which 15 were compounds and 15 were phrases. The 30 stimuli were arranged in five sets of six in such a way that one member of each of the six contrasting pairs was presented once in each set. The order of stimuli within each set was random. The sets of six were also combined in such a way that the second set of six contained the member of each minimal pair which the first set of six had not contained. Thus, both members of all minimal pairs were presented in the first 12 items of the test, and they were presented for a second trial in the second set of 12 items in the test.

The stimuli for Test 2, the repeat test, were identical to those for Test 1. However, in Test 3, instead of having 30 items for children to identify, only the first 13 items of the original stimulus tape were presented for testing. Results from only the first 12 items will be considered in our comparison with the Test 1 results. The extra item, a repetition of the first, was included mainly to measure repetition of errors on a given lexical item.

All speech stimuli for the Identification Tests were produced by the experimenter and were recorded under low noise conditions using high-quality, wide-band width equipment. Recordings were made in the M.I.T. Research Laboratory of Electronics anechoic chamber using an Altec 684A dynamic microphone. The recordings were made at 7.5 inches per second using a Presto 800 series tape recorder with a specified frequency response of ± 2 db, 40 to 12,000 Hz, and a signal-to-noise ratio of 60 db. This recorder was located in a room adjacent to the anechoic chamber.

The main reason for using a stimulus tape rather than having the experimenter pronounce the stimuli individually during the test session was that it insured uniformity in the stimuli that all subjects heard; this made generalization between the different ages possible. It also had the advantage of leaving the experimenter free to observe the performance and reactions of the children, and to note these observations.

All subjects heard the test stimuli from a Wollensak T-1500 tape recorder which had a specified frequency response of ± 3 db, 40 to 15,000 Hz and a signal-to-noise ratio greater than 48 db. Subjects were seated in small, quiet rooms. There was no apparent problem with subjects not being able to hear the stimuli clearly, a possibility which was considered and checked throughout the testing period by the experimenter asking them if they could hear the test items.

Testing Procedure

All phrases to be tested were introduced to the children before the stimulus tape was played. The experimenter showed a picture of one member of the minimal pair, explained what it was, and then named it. The same

procedure was repeated for the picture of the other member of the pair. The pictures and pronunciations of each pair were contrasted several times to maximize the differences. The experimenter did not use terms like "stress" or "accent" at any time during the test. Children were given answer sheets which contained sets of pictures of the minimal pairs. (See Appendix for samples of pictures used). The order of these pictures matched the order of stimuli on the tape, and each set of pictures carried an item number which corresponded to the item number given verbally on the stimulus tape. Children then heard the stimulus tape and were asked to identify what they heard each time by making a mark on the line under the appropriate picture. For example, if the stimulus were hót dog, children would have to choose between a picture of a frankfurter on a bun and a dog panting in the sun. Each item was presented twice before children were asked to make a judgement. Children were always instructed to choose one of the two pictures even if they were unsure of their answer. (Because of this instruction, omissions were counted as errors in scoring.) The stimulus tape and the answer sheet included two sample items presented prior to the actual test of 30 items. These were the items híghchair and green hóuse. The examples were included to ensure understanding of the task and to give the children the opportunity to become accustomed to hearing the stimulus tape and to marking their answer sheets. After each of the sample items was presented and the subjects had made their responses, the experimenter indicated the correct answer and made certain that all children understood the task. Additional examples were given and instructions were repeated if necessary. The stimulus tape for the actual test was then presented without interruption.

The procedures used for Test 2 and Test 3 were the same as those in Test 1, with one exception. In Test 3 children heard each item produced only once before they were asked to make a judgment.*

Test sessions for Test 1 and Test 3 lasted approximately 25-30 minutes. This time period was divided between several different comprehension tests

*This change was made because the children in Test 1 appeared to make their judgments after hearing the stimulus the first time, and they did not seem to attend to the stimulus repetition. Additional pilot studies between Test 1 and Test 2 confirmed this observation (as did the results of the two tests, as will be seen below). Presenting the stimulus only once in Test 3 had the advantage of making it more valid to compare the results of the various comprehension and production tests (see Chapter 5); in the latter test, children were asked to produce each stimulus item only once and then later to identify each of these items. Earlier attempts to have children produce each item twice proved unnatural and awkward for the children.

to provide variety and avoid bordeom.* The Test 2 testing session did not include other experiments, however; it was therefore approximately 10-15 minutes long.

One important procedural question which was examined in the pilot studies was the possibility of testing children in small groups rather than individually. The children being tested were relatively old in terms of language development. All of them had started school and therefore had experience with performing activities in groups; all had had some classroom discipline and experience in using paper and pencils either for writing or drawing. The question, then, was whether children could take the identification test in small groups and indicate their responses by making a mark under one of two pictures on an answer sheet, or whether they would have to be tested individually and respond in such ways as pointing to the correct picture. The group testing situation has the obvious advantage of enabling a larger number of children to be tested; larger sample size provides a more complete and reliable picture of developmental trends versus individual differences. Therefore, in pilot studies for the identification test with compounds and phrases, a comparison was made between the performance of children tested individually and children tested in small groups. Twelve children ranging in age from five to nine were tested individually, and 14 ranging in age from 5 to 10 were tested in groups (one group of 10 and one group of 4). The explanation of items and introduction procedures were the same for all children. They were all asked to listen to the items on the stimulus tape and to identify what they heard each time by choosing between two pictures which represented the minimal pair. The main difference between testing subjects in groups and testing subjects individually was the manner of selecting the correct picture. Subjects tested individually either pointed to the correct picture on a card placed in front of them by the experimenter or pushed a button under the correct picture on a response box device (described in Menyuk and Anderson, 1969). This box had windows through which the pictures were seen, and a button under each window. When a button was pushed, it turned on a light over that particular picture, thus making the task more "game-like" and presumably motivating and interesting to the child. The experimenter then recorded each response on a score sheet. On the other hand, the responses of the children tested in groups were made on answer sheets as described above. The results showed that there were no important differences in the performance of children of the same age when tested individually and in groups. This conclusion was further supported by the testing which was done with the group of children tested individually for production data as well as comprehension data (see Chapter

*In the session in which Test 1 was administered, a second experiment on the acoustic correlates of stress using synthetic speech stimuli was given. It will not be included in this study, however. The Test 3 session included the experiments with noun and verb pairs reported in Chapter 6 and an experiment with contrastive stress and pronominal reference which will not be discussed.

5). The same children were given the comprehension test twice -- once in a group and once individually. There were no important differences in performance. Therefore, testing was done with groups of 8 to 10 children.

Task Comprehension Test

Although there were no differences in the performance of children tested individually and children tested in groups in pilot studies with compounds and phrases, there were still some unanswered questions concerning other aspects of the testing procedures, particularly in the case of the youngest children, since these children performed randomly under both individual and group conditions. The youngest children may have failed to perform adequately because of mechanical or comprehension difficulties with the task itself. It could be argued that young children failed to perform well because they could not associate a simple line drawing with a phrase, though this seems unlikely because, as numerous language acquisition studies have shown, young children can associate names with simple drawings (e.g., Berko, 1958). Other arguments that might be raised are that children failed because of mechanical problems of manipulating paper and pencil in order to mark an answer sheet; that they were bored by such a test and put no effort into responding correctly; or that they were confused and did not understand the task completely.

In order to ensure that the experimental results would be a reflection of linguistic skill rather than of the ability to understand and perform a mechanical task, a separate Task Comprehension Test was given to 51 of the children in Kindergarten and Grade 1. In this test, the 30-item Test 1 stimulus tape was played to the children, and procedures for Test 1 were followed, with the exception that the answer sheets given them were different from those of the regular Test 1 compound and phrase identification test. In the regular test the answer sheet depicted pictures of the minimal pairs for each item, but in the Task Comprehension Test for each item it gave a choice between pictures of maximally different items. For example, in Identification Test 1 the child hearing the stimulus hót dog would have to choose between a picture of a dog in the sun and a frankfurter; in the Task Comprehension Test, if the stimulus were hót dog, the choice would be between a frankfurter on a bun and another unrelated item, such as a greenhouse.

The results of this Task Comprehension Test showed that all but two children in Kindergarten could perform the task. These two children were subsequently eliminated from the study. The average percentage of correct responses for the remaining 49 subjects was 97 per cent correct for Kindergarten children and 100 per cent for those in Grade 1. On the basis of these results it was evident that even the youngest children were able to perform the task.

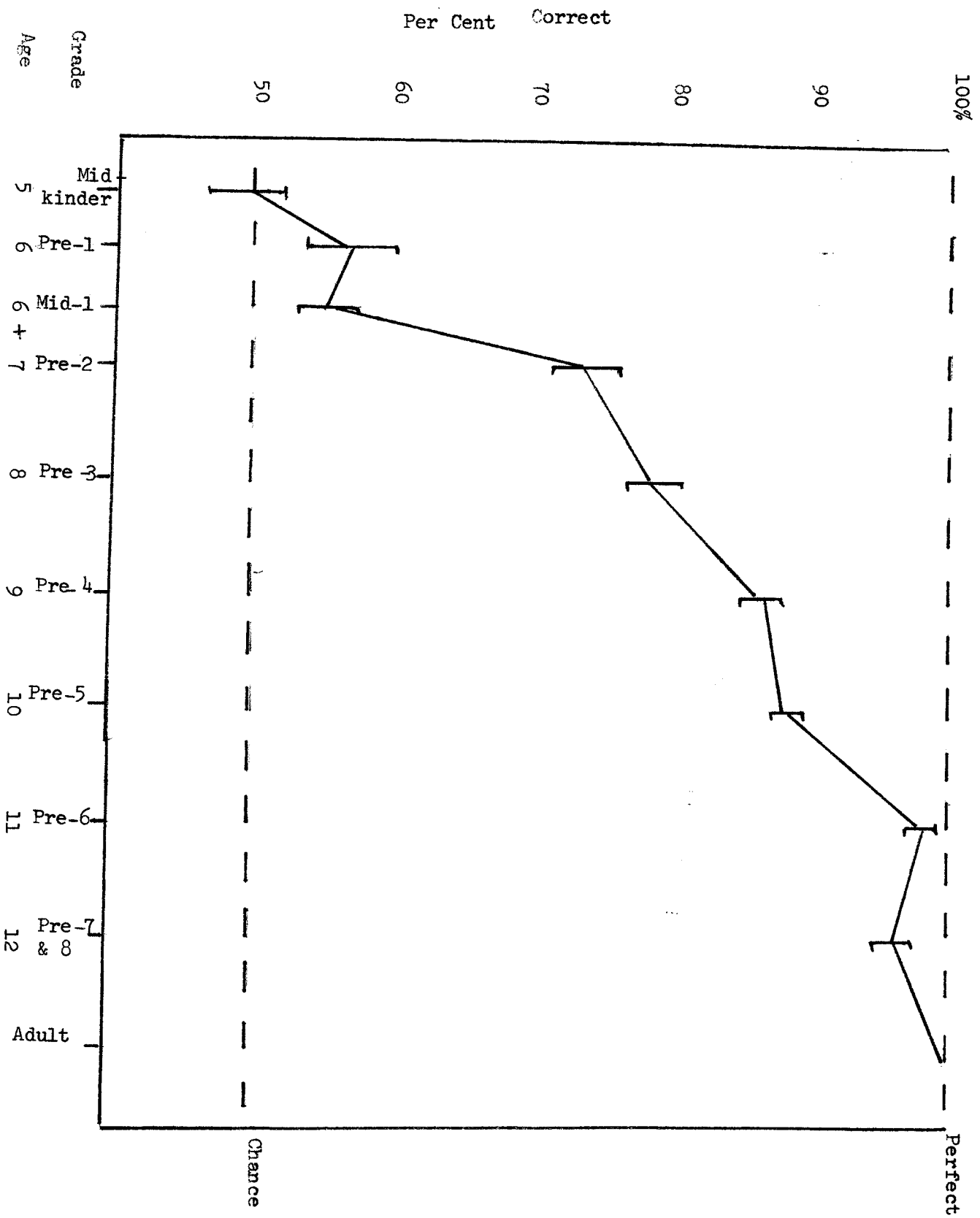


Figure 2.1: Identification Test 1 -- Percentage of Correct Responses Plotted Against Grade and Age

RESULTS

In general all of the children appeared to understand and, in most cases, to enjoy the task. Also, on the whole they appeared to be well motivated and made an effort to respond correctly to each item.

General Results

Results from Identification Test 1 indicate that the ability to correctly identify these minimal pairs increases with age. Figure 2.1 shows the percentage of correct responses for Test 1 plotted against grade and age. Measures of standard deviation are also given for each group.* This figure indicates that children in Kindergarten are only producing 49 per cent correct responses while children in Grades 6, 7 and 8, as well as adults, produce responses approaching 100 per cent correct. Between these extremes there is a consistent developmental trend with the percentage of correct responses increasing with age and grade.**

The results from Test 3, the partial replication of Test 1, support these general findings. As Figure 2.2 indicates, the percentage of correct responses generally increases with age and grade, producing the

*As noted earlier, the grades on the abscissa are prefixed by "Mid-" or "Pre-" indicating whether the children were tested in the middle of the school year or at the beginning of the school year. The ages given are average ages for the group of children in each grade. Grades Pre-7 and Pre-8 were combined since they were both smaller than the other groups and since they showed no significant differences in performance.

**There are apparently two minor exceptions to this upward trend in performance. These are the inversions which occurred between the two groups of Grade 1 and between Grade 6 and Grades 7 and 8. The difference between the two groups of Grade 1 is not significant. It is due primarily to the exceptional performance of one girl in Grade Pre-1. Her performance considerably affected the Grade Pre-1 group average because that group was smaller than the Grade Mid-1 group. The slight dip occurring at Grades 7 and 8 is also not significant and is probably due to the small number of subjects in Grades 6 through 8.

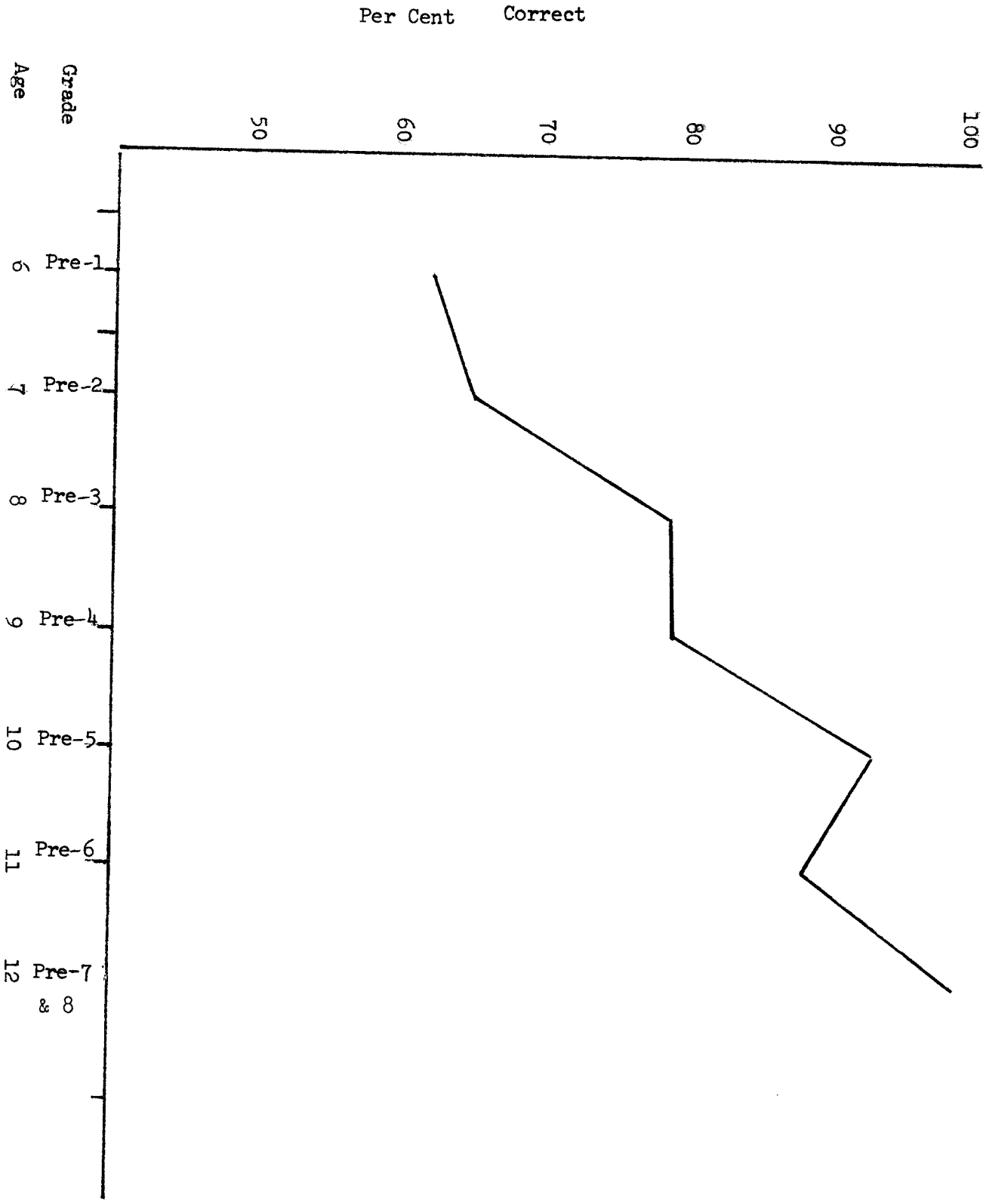


Figure 2.2: Identification Test 3 -- Percentage of Correct Responses Plotted Against Grade and Age

same type of developmental trend seen in the results of Test 1.*

In order to compare the results of these two tests for reliability, the data from only the first 12 items of Test 1 were compared with Test 3, which was a replication of these 12 items. It is probably not necessary to use only the first 12 items of Test 1 in the comparison because there is good correlation of the general performance on these first 12 items and on the entire 30-item test, as Figure 2.3 indicates.

A few minor differences are apparent between the results for each grade under these two conditions. One is that the small dip in the developmental trend which occurs in Grades 7 and 8 with all 30 items of Test 1 is not found in the first 12 items, again indicating its nonlinguistic nature. (See learning effects discussed below). Also on Test 1 the differences in performance between Grades Pre-1 and Pre-2 are not as marked with the first 12 items as they are with all 30 items, since with all 30 items the percentage of correct responses is lower for Grade Pre-1 and higher for Grade Pre-2. Nonetheless, the differences in the results of the test under these two conditions are minor. But in order to obtain the most accurate comparison possible between Tests 1 and 3, only the first 12 items of Test 1 were used.

Figure 2.4 makes possible a comparison between the percentage of correct responses by grades in the first 12 items of Test 1 and the grades in Test 3. In general, there is good correlation between the performance of the grades on both tests. The largest differences occur with Grades 4 and 6. These two grades in the Test 3 group failed to perform as well as the corresponding grades in Test 1, or as well as would have been predicted on the basis of the performance of the preceding and following grades in the Test 3 group. However, Chi-square analysis shows that none of the differences between corresponding grades in the two test groups is significantly greater than chance.

Frequency Distribution and Individual Variation

Returning once more to the results of all 30 items of Test 1 (see Figure 2.1), we might ask whether the results from Grades 1 through 5 which display an average percentage of correct responses which are better than chance but do not reach the highest level of accuracy (e.g., Grade 2 with an average of 73 per cent correct responses) are due to the majority of

*There are a few dips in the developmental trend that occur with this test group, just as there were with the Test 1 group. However, also as with the Test 1 group, these dips probably arose by chance or possibly from performance factors. Note that the dips in each of these test groups do not correspond with each other. This appears to confirm the conclusion that such dips in the general trend are not systematic errors due to some general age group factor but rather are accidental or nonlinguistic in nature.

Figure 2.3: Identification Test 1 -- Percentage of Correct Responses by Grade on Items 1-12 and Items 1-30

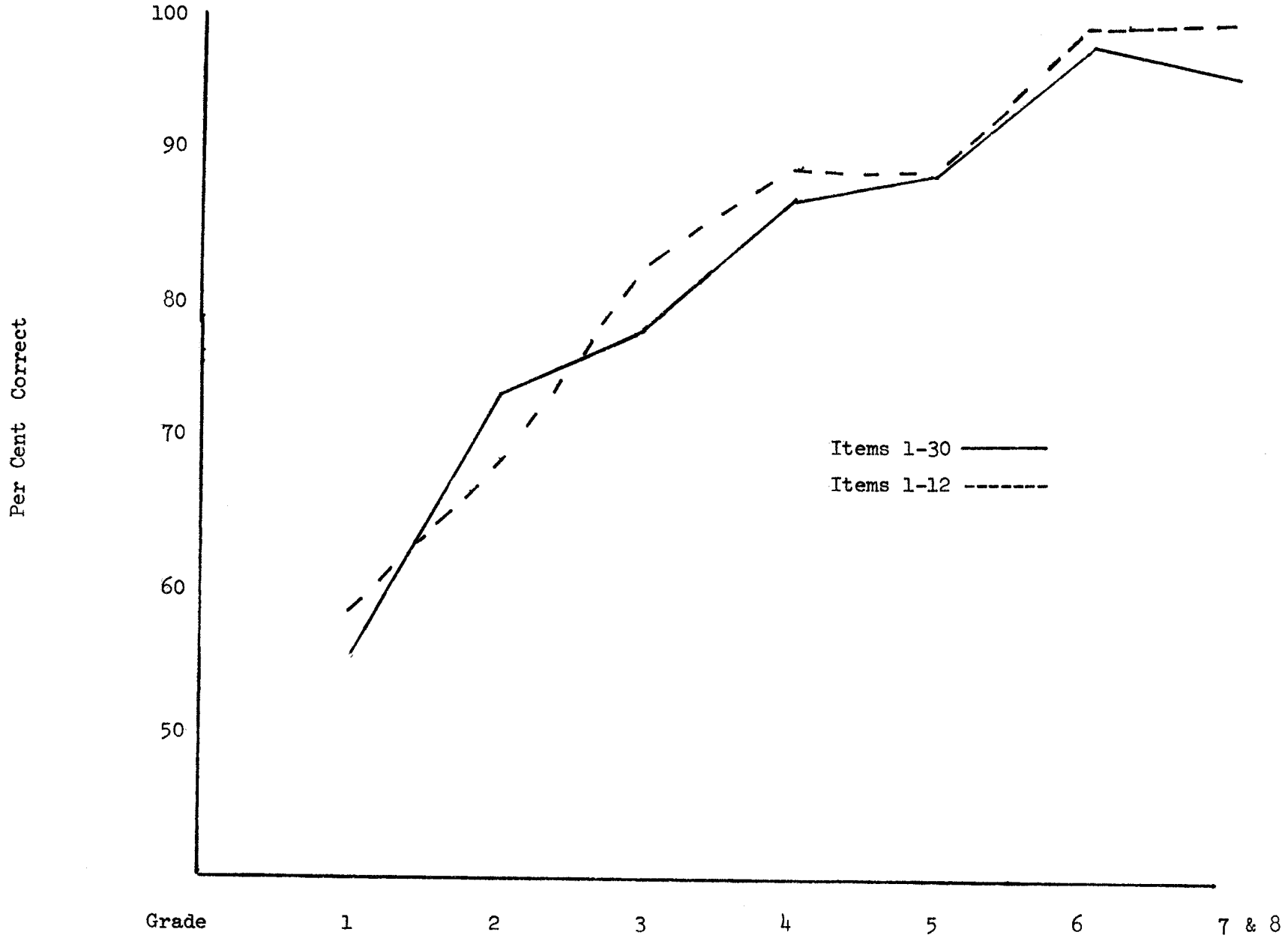
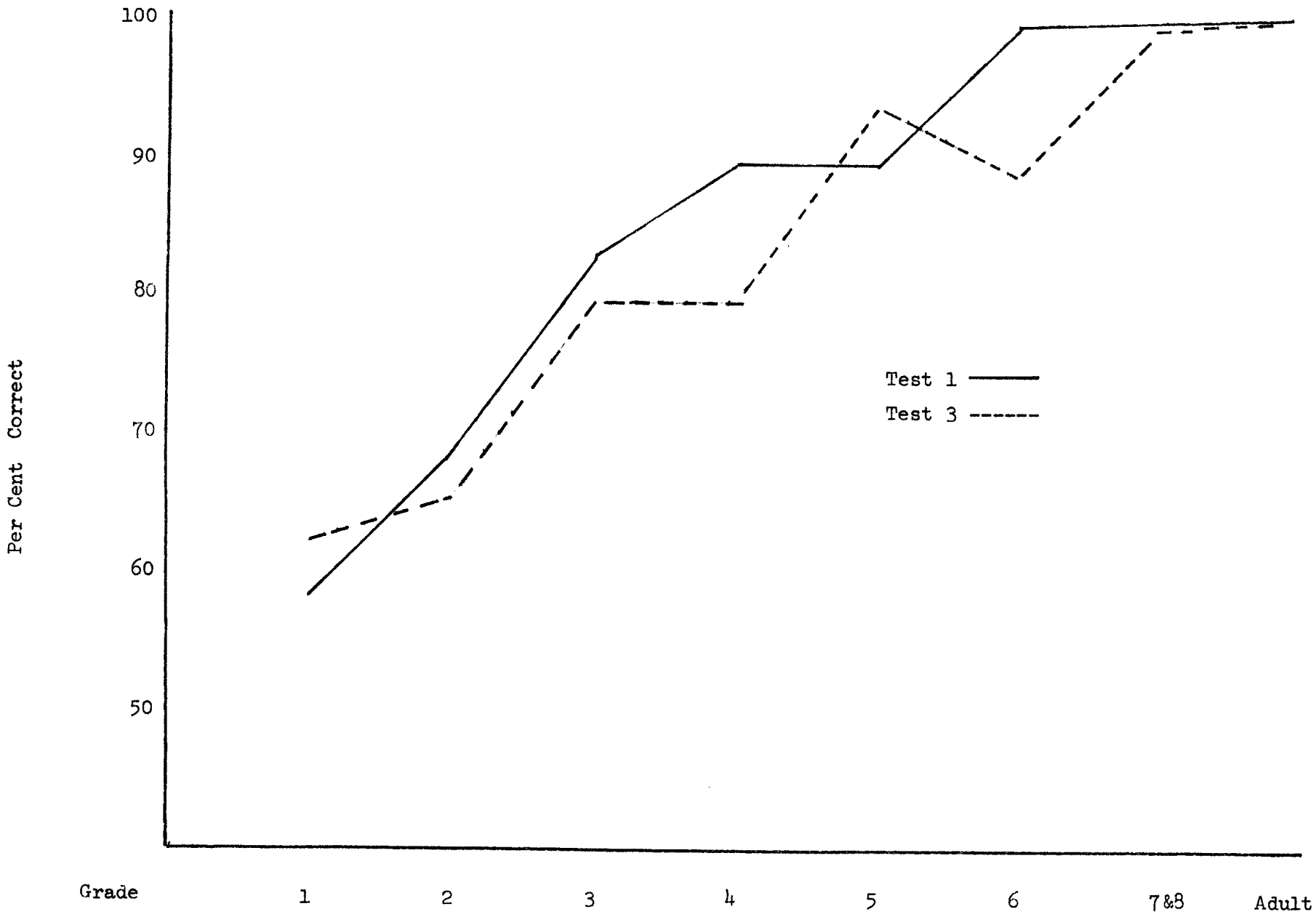


Figure 2.4: Identification Test 1 (Items 1-12) and Identification Test 3 -- Comparison of Percentage of Correct Responses by Grade



subjects in that particular grade performing at a level of accuracy which approaches the group average, or whether the group average is a result of large individual differences (e.g., a distribution pattern such as half the subjects performing poorly and half quite well). The frequency distribution of the percentage of correct responses of subjects in Grades 1 through 5 seen in Table 2.6 helps to answer this question.

Table 2.6: Identification Test -- Frequency Distribution of Percentage of Correct Responses of Subjects in Grades 1-5

% Correct Response	Number of Subjects Scoring at Each % Interval					
Grade: (Mean)	Pre-1 (57%)	Mid-1 (55%)	Pre-2 (73%)	Pre-3 (78%)	Pre-4 (87%)	Pre-5 (89%)
26 - 30%		1				
31 - 35%	1					
36 - 40%	1	1				1
41 - 45%		1		1		
46 - 50%	2	9			1	
51 - 55%	1	*6	4		1	
56 - 60%	*1	5	1	2	1	1
61 - 65%		3				
66 - 70%	3	1	2	1	1	3
71 - 75%		1	*1	2	1	
76 - 80%		1	5	*2		1
81 - 85%		1			1	3
86 - 90%	1		1	6	*3	*1
91 - 95%			2	1		3
96 - 100%			1	1	11	14

(Blank space = 0; interval closest to group mean = *)

The number of subjects whose percentage of correct scores fall at the various percentage intervals shown in this table shows that the distribution tends most often to be more normal and symmetrical than polarized at the two extremes. This indicates that most subjects in these intermediate grades have a tendency to perform at a level consistent with the group average rather than either extremely well or extremely poorly. That is, we can say that the group mode and median correspond fairly closely with the group mean.

We can see from the frequency distribution that although most subjects in Grades 1 through 5 do not deviate from the mean to a great extent, there are a few who do differ significantly from the group. Thus, although the

general picture is one of increasing ability with age, there is some individual variation in these grades. A few children in Grade 1 (6 years old) performed almost as well as adults, and a few children in Grade 5 (10 years old) did as poorly as those in Kindergarten (5 years old). However, in Kindergarten no children did significantly better than chance, and in Grade 6 and above, all subjects did well and scored above 85 per cent correct responses.

If we compare the data from Test 3, we find that they confirm the type of distribution seen in Identification Test 1. Table 2.7 shows the frequency distribution for Test 3. This table is not directly comparable to Table 2.6 because of Test 3's fewer number of test items and the resulting difference in intervals used on the frequency distribution table.

Table 2.7: Identification Test 3 -- Frequency Distribution of Percentage of Correct Responses of Subjects in All Grades

Number of Subjects Scoring at Each % Interval

% Correct Responses	Grade (Mean):	1 (61.5)	2 (64.2)	3 (79.2)	4 (78.7)	5 (92.6)	6 (88.3)	7 & 8 (98.5)
0								
8.3								
16.7								
25.0		1						
33.3		1	1					
41.7			3	2	1		2	
50.0			5	1		1		
58.3		2	3	1	3	1	1	
66.7		*1	*7	1	5			
75.0		2	2	1	4			
83.3			2	*2	*4	3	1	
91.7			3	4	2	3	7	2
100.0		1	1	4	6	*10	*9	*9

(Blank space = 0; interval closest to group mean = *)

However, in most grades the basic pattern of the two tables is similar. That is, in Test 3 the general distribution is normal and symmetrical rather than polarized, especially in the lower grades. Most subjects tend to have scores similar to that of the group mean. Also, we again find some individual variation. One difference between Test 3 and Test 1 is that in Test 3 there are three subjects in Grade 6 who do not score in the 80th percentile where-

as all subjects in Grade 6 in Test 1 scored over 85 per cent correct responses. It is not until Grades 7 and 8 that all subjects in Test 3 perform at such a high level of accuracy.

Another way of looking at this question of frequency distribution and individual differences in performance is seen in Table 2.8. This table indicates the percentage of subjects in each grade whose percentage of correct responses was higher than 70 (and therefore was significantly greater than chance at the $p<0.05$ level as was determined by chi-square

Table 2.8: Identification Test 1 -- Percentage of Subjects Scoring More or Less Than 70% Correct Responses

	Less Than 70% Correct (p =ns)	More Than 70% Correct (p = <0.05)
Kindergarten	100%	0%
Pre-Grade 1	90%	10%
Mid-Grade 1	90%	10%
Pre-Grade 2	29%	71%
Pre-Grade 3	25%	75%
Pre-Grade 4	15%	85%
Pre-Grade 5	15%	85%
Pre-Grade 6	0%	100%
Pre-Grades 7 & 8	0%	100%
Adults	0%	100%

analysis)and the percentage of subjects in each grade whose percentage of correct responses was lower than 70 (and therefore was not significantly better than chance). This table indicates that the percentage of subjects scoring significantly more correct than incorrect responses is very low in Grade 1, but there is a jump to 71 per cent in Grade 2. By Grade 6 and above, 100 per cent of the subjects have scores that are significantly better than chance.

The Test 3 data also generally confirm this pattern. Table 2.9 shows that the percentage of children who significantly score more correct than incorrect responses increases with age.*

*Because of the smaller number of items in Test 3,a minimum of 10 correct out of 12 items (approximately 80 per cent correct responses) was necessary for the correct versus incorrect responses to reach a level of less than 5 per cent probability of chance results ($p<0.02$ as was determined by the binomial test). Therefore, 80 per cent correct responses was used as a cutoff point in Test 3 instead of the 70 per cent used in Test 1.

Table 2.9: Identification Test 3 -- Percentage of Subjects Scoring More or Less Than 80% Correct Responses

Grade	Less than 80% Correct (p = ns)	More than 80% Correct (p = <0.02)
1	87.5	12.5
2	77.8	22.2
3	37.5	62.5
4	52.0	48.0
5	11.1	88.9
6	15.0	85.0
7 and 8	0	100.0
Adults	0	100.0

However, there are two exceptions to the general developmental increase in percentage of subjects scoring over 80 per cent correct responses. These occur with Grades 4 and 6. The dip in Grade 6 is slight and insignificant. However, in Grade 4 the percentage of children who score significantly more correct than incorrect responses is very low in relation to the scores of Grades 3 and 5 and to the general trend. It is also low in relation to the Grade 4 group in Test 1. There is no obvious reason for either of these dips in the developmental trend.

Not only is the poor performance of Grade 4 in Test 3 not found in the Test 1 data, but the slight dip in Grade 6 is also absent. Because of Test 3's dip in performance in Grade 6, children do not reach the 100 per cent level until Grade 7 and 8, whereas the Test 1 group reached this level in Grade 6. Another difference in Test 3 and Test 1 is that the large jump in percentage which occurred between Grades 1 and 2 in Test 1 is not found with the Test 3 group. This would thus call into question the importance of this Grade 1 to 2 age period as suggested by Test 1. It appears from the Test 3 data that the crucial age period varies between Grades 1 and 3. Although such differences exist in the actual percentages attained by corresponding grade groups in Test 3 and Test 1, they are relatively unimportant. The main consideration is the general developmental trend, and we find that both Test 3 and Test 1 display this developmental trend.

Gradual Learning

We might ask whether a child acquires the ability to identify these stress contrasts all at once or whether he seems to develop and perfect this skill gradually. A longitudinal study with frequent samplings of individual children would be necessary to give a satisfactory answer to this question. However, there are some further frequency distribution data from this study which do suggest a possible answer. Figure 2.5 shows the percentage of children in Test 1 who individually scored a total number of correct responses significant above chance as determined by chi-square analysis, similar to Table 2.8. However, here two levels of significance are shown. The solid line indicates the proportion of children who individually scored over 70 per cent correct responses (better than chance at $p<0.05$), and the dashed line indicates the proportion who scored over 85 per cent correct responses (better than chance at $p<0.001$). In Grades 2 through 5 we observe a considerable variability between the percentage of children whose individual scores reached the two levels of significance. This variability suggests a process of gradual perfection in the acquisition of the ability to identify these contrasts. For example, in Grade 2, where the greatest difference occurs, approximately three fourths of the children had a score of at least 70 per cent correct responses ($p<0.05$), but only

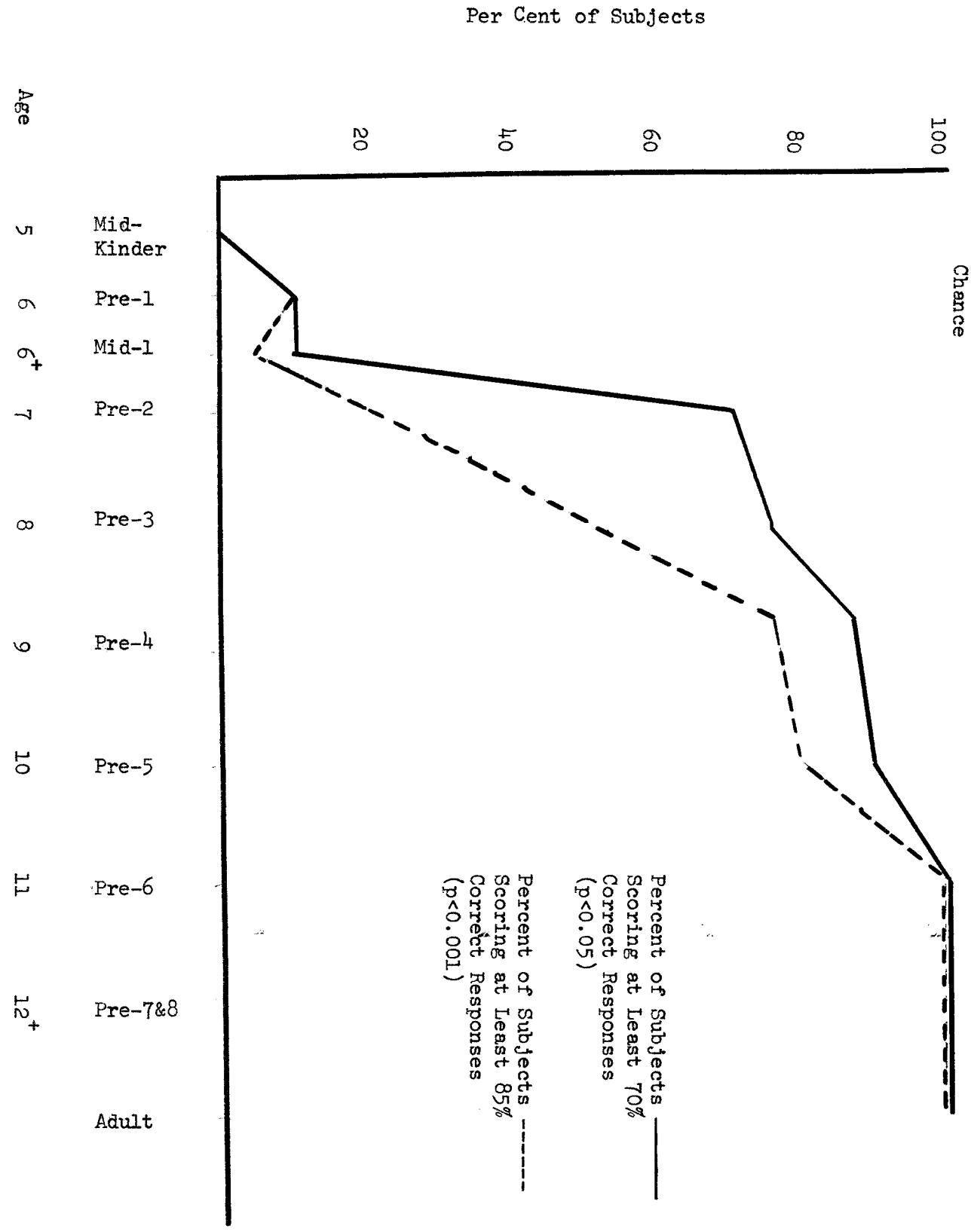


Figure 2.5: Identification Test 1 -- Percentage of Children Who Scored a Significant Number of Correct Responses Above Chance

one fourth of them scored as high as 85 per cent or more correct responses ($p<0.001$). Hence, one-half the Grade 2 children performed at a level high enough to indicate that they had some proficiency with the identification skill since they could correctly identify 70 per cent of the stimuli, but they had not mastered it completely since they could not correctly identify as high as 85 per cent of the stimuli. Thus the gap between these two levels of significance as shown in Figure 2.5 represents the percentage of children who have begun to acquire the skill of identifying these contrasts, but who are not yet able to respond correctly with a high degree of consistency.

It might be argued that the general developmental trend and, particularly, these differences in the two levels of significance seen with it do not result from linguistic factors (knowledge of the stress patterns), but rather from nonlinguistic factors (other than age) such as older children having longer attention spans and greater mechanical ability and motivation to perform this type of task well. However, there are several pieces of evidence which cast doubt on these nonlinguistic explanations. The main refutation of task-oriented explanations is the data from the Task Comprehension Test discussed above. In this test, the Kindergarten group scored 97 per cent correct responses and Grade Mid-1 scored 100 per cent correct responses. Yet the same two groups scored only 49 per cent and 55 per cent correct responses respectively on Identification Test 1. This suggests that the failure to perform well on the Identification Test cannot be attributed merely to such nonlinguistic factors and skills, but rather involves skills of a linguistic nature. If the Task Comprehension Test was easier and required less attention, etc., this was only so because knowledge of the stress patterns was not required as it was in the Identification Test.

A second type of evidence against the arguments that nonlinguistic factors like attention and motivation were the primary determinants behind the developmental trend and differences in the two levels of significance seen in the results of this stress Identification Test is the consistency in performance of both the grade groups as a whole and of individual children. If a factor like short attention span were crucial in contributing to the poor performance of the younger children, we might expect the performance of these children to get progressively worse as they became more fatigued and more inattentive. But as we saw from Figure 2.3, which compares the group's performance on the first 12 items and on the entire test, and as we will see below in the comparison of differences between a group's performance on roughly the first and second half of the test, there are no significant differences of this type in performance under these various conditions.

Factors like length of attention span and amount of motivation might make for greater variation in test results than would linguistic factors.

Yet we saw that the performances of different sets of subjects in essentially the same test, (Test 1 and Test 3) were very consistent. Moreover, as we shall discuss below, in the case of the children in the main test group who repeated the entire test several days after they took the initial one, there was a very high correlation of performance on the two. These and other such results indicate that the differences in performance seen in this stress Identification Test are primarily due to differences in linguistic ability rather than mechanical ability or performance factors.

Differences According to Sex

The results from Test 1 were analyzed to determine whether there were any significant differences between the performance of boys and girls at various ages. Figure 2.6 shows the average percentage of correct items plotted against grade and age according to sex. It indicates a consistent trend for girls to score better than boys in grades below Grade 6. Also, girls approach the 100 per cent level of correct responses earlier in the age range than boys. Another way of looking at these trends in differences between the sexes in terms of the frequency distributional type of data as seen in Figure 2.7. This figure shows the percentage of subjects scoring at least 85 per cent correct responses, (such scores being significantly greater than chance at the level of $p<0.001$ as determined by chi-square), plotted against age and grade according to sex. This way of considering the data emphasizes that girls reach a higher level of significance more quickly than boys, particularly in Grades 3 and 4. It also indicates that there is, again, individual variation in most grades. Although girls generally score better than boys, there are a certain number of girls who do not do this. Also, whereas the general group trend of girls scoring better than boys is consistent in the grades up to Grade 5, as noted, the Mann-Whitney U test indicates that these differences are significant only in Grade 4 where $\alpha < 0.02$ in favor of the girls. This therefore casts some doubt on the significance of the tendency of girls' performances to be superior to boys' as a general rule.

The data from Test 3 were analyzed to determine whether this trend existed with a different group of subjects. Figure 2.8 shows the percentage of correct responses according to grade and sex for this replication test.

Unfortunately, since no girls were in Grade 1, no comparison could be made at that age. Just as with Test 1, the other grades indicate that generally the girls tended to perform better than the boys. Some differences exist between the results of the two tests, however. As with previous comparisons of the data from Tests 1 and 3, a more reliable comparison can be obtained by again considering only the results from the first 12 items of Test 1, as shown in Figure 2.9.

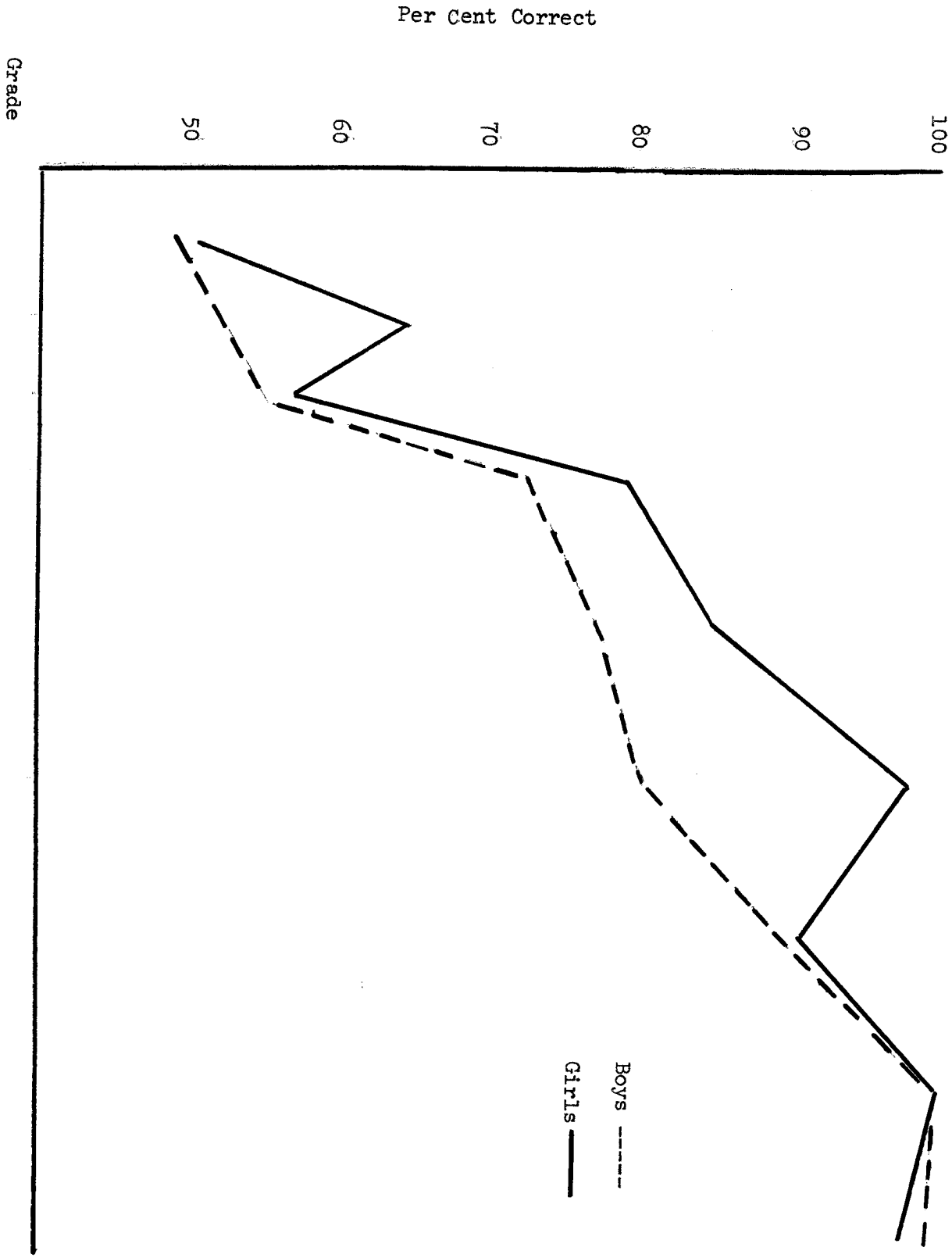


Figure 2.6: Identification Test 1 -- Average Percentage of Correct Items Plotted Against Grade and Age According to Sex

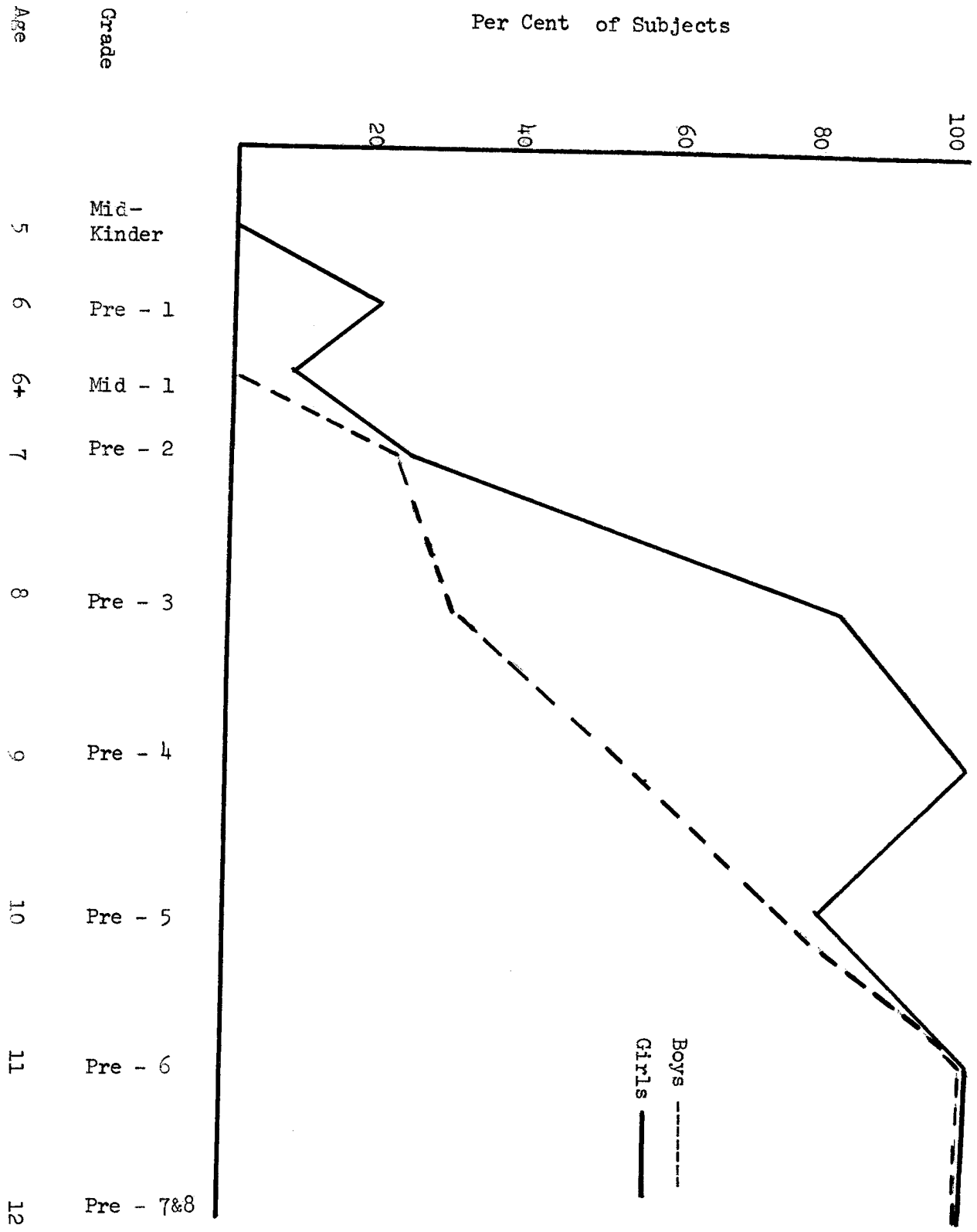


Figure 2.7: Identification Test 1 -- Percentage of Subjects Scoring at Least 85 % Correct Responses ($p<0.001$) Plotted Against Age and Grade by Sex

Figure 2.8: Identification Test 3 -- Percentage Correct Responses According to Grade and Sex (12 Items)

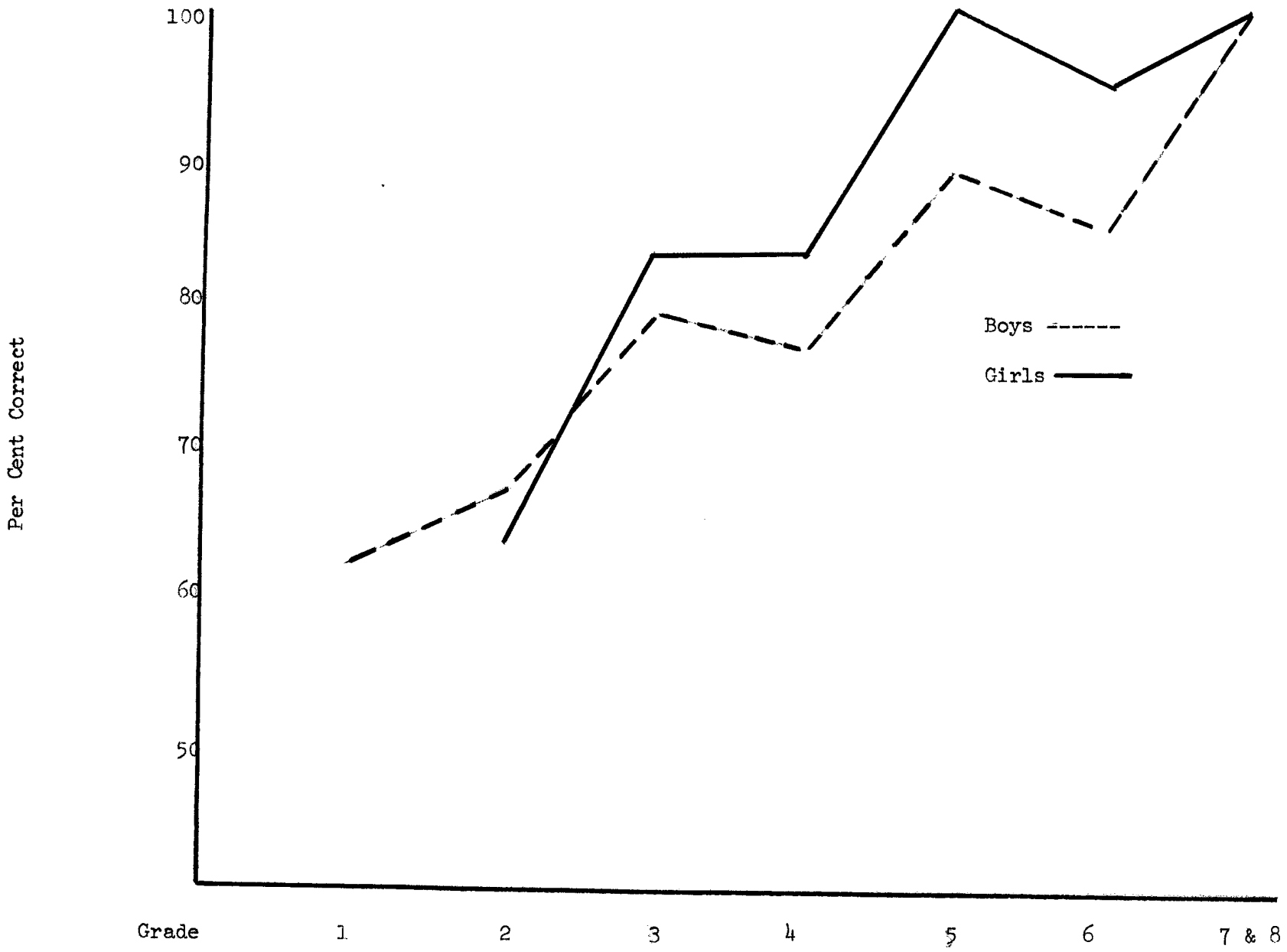
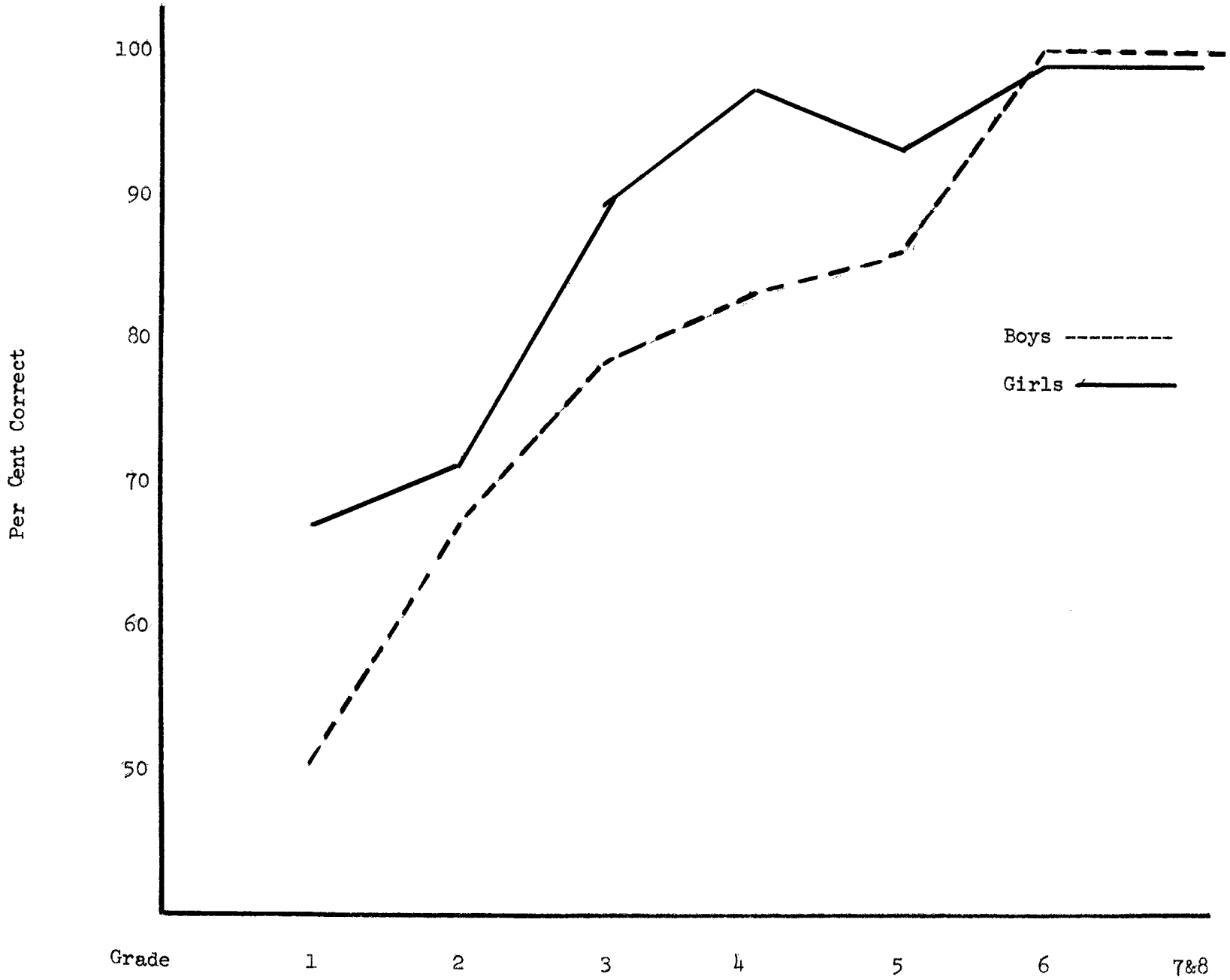


Figure 2.9: Identification Test 1 (First 12 Items -- Percentage of Correct Responses According to Grade and Sex



One difference between the performance of girls and boys on the two tests is that in Grade Pre-1, the boys in Test 3 did better than did the same grade in Test 1. Also, whereas in Test 1 the girls in Grade 2 did better than the boys in the same grade (though not significantly so), in Test 3 they did not do as well as the boys in this grade. The lower-level performance of the Grade 2 girls in the replication test is unusual in that it is the only instance in either test group where girls below Grade 6 did not do better than boys.

The Mann-Whitney U Test indicates that in Test 3 the differences between the performance of boys and girls is not significantly greater than chance in any grade. In Test 1 these differences were significant in only one grade; hence sex differences do not appear to be an important factor. However, the general tendency for girls to score better than boys does correspond in the two test groups; therefore, sex differences may play some role in the acquisition of the ability to identify stress contrasts, but it is not clear how important a role this is.

Learning Effects

The data were analyzed to determine whether there was any evidence of learning effects. Figure 2.10 displays the percentage of correct responses by grade for the first and second sets of Test 1, a set being 12 items in which each member of the six minimal pairs was heard once. This figure compares the performance of groups on the first versus the second trial of each minimal pair, or on roughly the first versus the second half of the test (items 1-12 versus items 13-24). It shows that in the majority of the grades the results were fairly consistent throughout the test. Chi-square analysis indicates that no grade displayed a difference in performance between the first and second sets of 12 items of the test which is significantly greater than chance.

Although the performance of the entire group was basically very consistent throughout the test and there were no significant differences in performance between the first and second sets for any grades, some trends did appear. With the exception of Grade Pre-1, a slight tendency existed for the younger grades, Kindergarten through Grade 2, to demonstrate over-all learning effects. Grades 3 and above showed the opposite effect -- i.e., no learning effect occurred but rather performance tended to be poorer, perhaps because of such factors as fatigue or a possible drop in motivation due to boredom. One explanation for these slight tendencies is that the younger children, who had a very limited ability to identify these contrasts, may have found the task more challenging than did the older children who had greater ability and who might have become bored with the task sooner because of its simplicity. At any rate, the fact that it was the older rather than the younger children who gave indications of fatigue or lack of attention

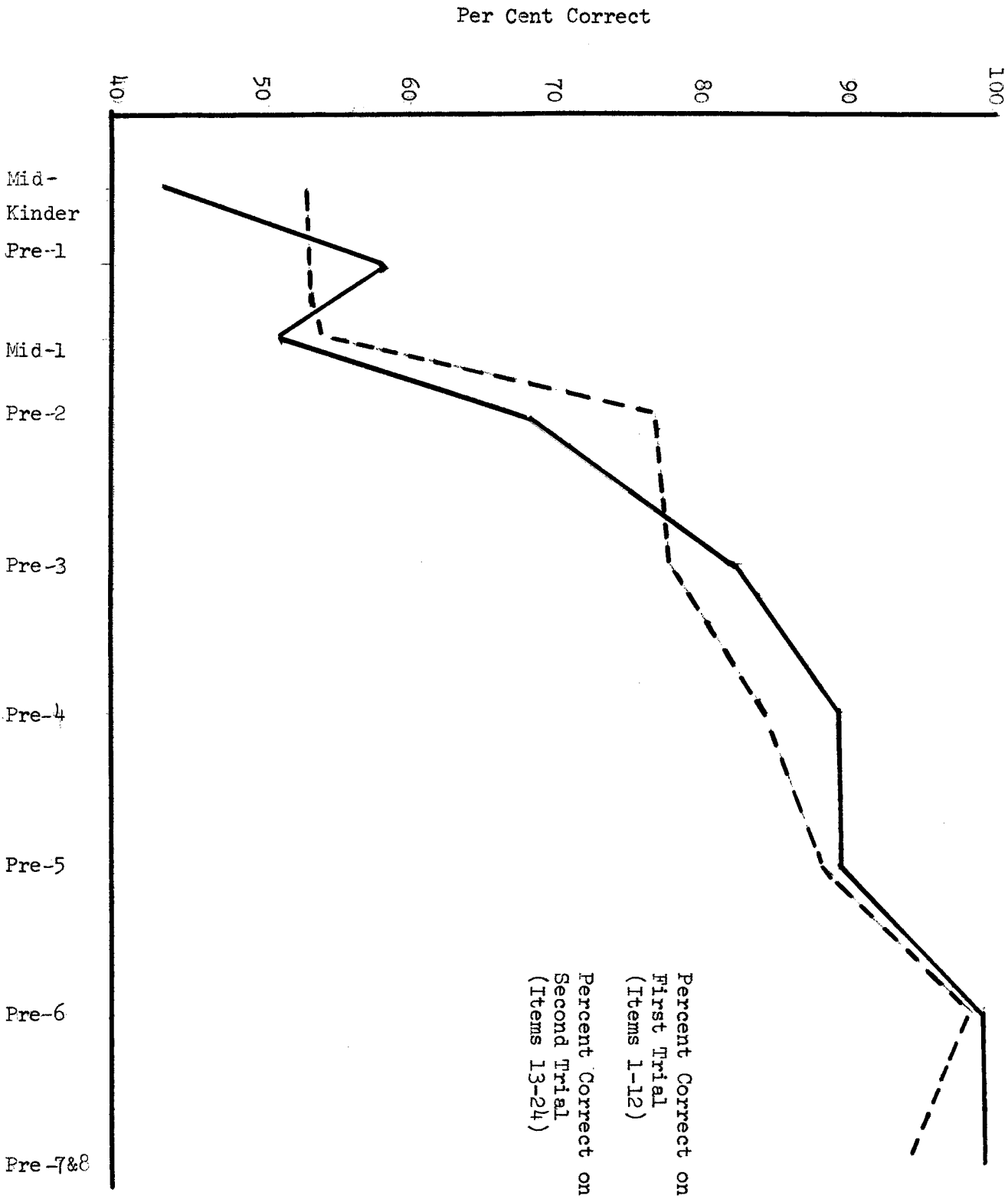


Figure 2.10: Identification Test 1 -- Percentage of Correct Responses by Grade for First Set of 12 Items Versus Second Set of 12 Items

at the end of the test reinforces the conclusions made earlier that the poor performance of the younger grades in general was due to lack of linguistic ability rather than to a nonlinguistic factor like fatigue or losing interest in the task. If such factors were primary in determining the poor performance of the younger children, we would expect the trend to be opposite and the younger children to become fatigued or bored as the test progressed. Since this was not the case, the importance of such factors appears negligible.

A comparison of the results from Tests 1 and 2 also gives no indication of learning effects. Only one of the 32 subjects, a Second Grader, performed randomly on Test 1, but significantly on Test 2. (See the discussion of the correlation of Tests 1 and 2 below for further treatment of this question.)

The first item of Test 3, hót dog, was repeated at the end of the test. An error analysis of its first and second trial indicated no significant differences in the number of errors on each trial. The first three grades and Grades 6 and 7 had slightly fewer errors on the second trial, which would indicate learning effects, and only Grade 4 had slightly more on the second trial, but all differences were minimal. Only 3 of 125 subjects missed both trials of the item hót dog.

Response Biases: Positioning of Pictures on Page and Type of Syntactic Structure

The data were analyzed to determine whether any types of response bias affected the results, since children who do not possess enough linguistic ability to correctly identify the test items may use some nonlinguistic factor in making their responses. Data related to the positioning of pictures on the children's answer sheets were analyzed for evidence of such biases. The results indicated that no grades or individual subjects either consistently marked, or showed a strong tendency to mark, the picture located closest to the outer margins or that closest to the inside of the page. There were also no grades or individual subjects who demonstrated a bias toward marking only the right member or only the left member of each minimal pair depicted on the answer sheet.

The results were also examined to determine whether there was a response bias toward choosing compounds rather than phrases or vice versa. This question was considered since it is possible that children of different ages might prefer one structure to the other because of the order of acquisition, frequency of appearance, productivity of these structures, etc. It was found that no individual subjects consistently chose all compounds or all phrases. Furthermore, as seen in Table 2.10, there was no general tendency for any grade to choose one type of response over the other.

Table 2.10: Responses Made by Subjects in Test 1 -- Total Number and Percentage of Compound and Phrasal Responses

Grade and No. of Ss	Total Number of Responses	Compound No.	Compound %	Phrase No.	Phrase %
Kind. (19Ss)	570	269	47.2	301	52.8
Pre-1 (10Ss)	300	179	59.6	121	40.4
Mid-1 (30Ss)	900	408	45.4	492	54.6
Pre-2 (17Ss)	510	251	49.2	259	50.8
Pre-3 (16Ss)	480	239	49.8	241	50.2

(Grades above Pre-3 showed an approximately equal number of responses of each type.)

Grades Pre-1 and Mid-1 displayed the greatest differences, but the tendencies were in opposite directions. Whereas Grade Pre-1 tended to give more compound than phrasal responses, Grade Mid-1 gave a few more phrasal than compound responses. Chi-square analysis indicates that the tendency to select one syntactic structure over the other was not significantly different from chance in either of these grades or in any other.

The question of response biases toward certain syntactic structures was also considered with regard to the data from Test 3. Table 2.11 provides a breakdown of the number and percentage of compound and phrasal responses for Grades 1 through 3. As with Test 1, chi-square analysis indicates that no grades in Test 3 displayed biases toward one or the other syntactic structure which were significantly greater than chance. Nor did any individual subjects in Test 3 display such a bias.

Table 2.11: Responses Made by Subjects in Test 3 -- Total Number and Percentage of Compound and Phrasal Responses

Grade and No. of Ss	Total Number of Responses	Compound No.	Compound %	Phrase No.	Phrase %
Grade 1 (8Ss)	112	52	46.5	60	53.5
Grade 2 (27Ss)	378	193	51.0	185	49.0
Grade 3 (16Ss)	224	118	52.7	106	47.3

(Grades above Grade 3 showed an approximately equal number of responses for each type.)

Response Biases: Selecting Certain Members of Minimal Pairs

An item analysis was made of the data to determine whether any grades or individuals showed a bias toward selecting one member of any minimal pair over the other member of the pair. Such biases would be reflected by a proportionately large number of correct responses for one member of a pair and a small number of correct responses for the other member of that pair. Chi-square analysis indicated that the difference between members of minimal pairs was significantly greater than chance only with the pair blackboard and black board in the three youngest grades. Subjects in these grades selected the compound more often than they did the phrase ($p<0.05$ in Kindergarten and Pre-1 and $p<0.02$ in Mid-1). Perhaps children selected the picture of a blackboard more often because it is one of the most common objects seen by school children.

Although no other pairs displayed significant biases toward one member, there was a slight general tendency to select items more often than their counterparts. The compounds highchair, blackboard, and hot dog and the phrases red socks, green house, and red head tended to be chosen most often.

The results from Test 3 were also examined to determine whether significant differences existed between members of minimal pairs. Since Test 3 had only one trial for each minimal pair, the data are limited, and it is difficult to draw conclusions with assurance. However, Test 3 showed very little evidence that response biases existed. Chi-square analysis indicated that no grade gave responses that were indicative of any differentiation between members of pairs that was significantly greater than chance.

Also, the general trend toward the selection of certain members of minimal pairs seen in Test 1 does not seem to be supported by the limited data from Test 3. Test 3 not only shows less evidence of biases existing toward the selection of certain members of minimal pairs than does Test 1, but where slight though not significant evidence of biases does exist, the members of pairs which are favored are sometimes opposite to those favored in Test 1. For example, whereas Test 1 subjects in Grade 1 selected the compound blackboard significantly more often than its phrasal counterpart, Test 3 subjects in Grade 1 chose the phrase black board slightly (though not significantly) more frequently. Also, the pair in Test 3 on which subjects made the greatest (but not significant) differentiation between members was redhead-red head for Grade 2. For this pair the general tendency in Test 1 was to select the phrase red head. In Test 3, however, the compound was most often chosen. Thus Test 3 data make the slight response bias trends found in Test 1 appear even less important.

The question of which member of the pair tends to be most often selected may be an individual one. Therefore, the number of individual children whose responses reflected a preference for one member of a minimal pair rather than the other, was determined, as well as the member of the pair preferred in such cases. It was found that 17.5 per cent of the total group of 160 children selected the same member of a minimal pair every time either member of the pair was heard for one or more pairs. Eight per cent of the total group followed this pattern with only one pair of the test, and only 1 per cent selected one member of a pair exclusively on all six minimal pairs. Seven per cent of the total group followed this selection pattern for three or more pairs of the test, i.e. for 50 per cent or more of the items. Table 2.12 shows the number and percentage of children by grade who selected one member of a pair exclusively for one to six pairs. This table indicates that the tendency always to select only one member of a minimal pair occurs mainly with the youngest children. It also shows that a relatively high percentage of children exclusively selected one member of a pair on only one or more of the six pairs but that this percentage, even for the youngest group is still less

Table 2.12: Identification Test 1-- Number and Percentage of Subjects Selecting One Member of a Minimal Pair Exclusively for One to Six Pairs

	Pairs:	1 or more		2 or more		3 or more		4 or more		5 or more		All 6	
Grade	Total No. of Ss	#Ss	%Ss	#Ss	%Ss	#Ss	%Ss	#Ss	%Ss	#Ss	%Ss	#	%
Kinder-garten	(19Ss)	9	47	5	25	4	21	1	5	1	5	0	0
Pre-1	(10Ss)	3	30	1	10	1	10	1	10	1	10	0	0
Mid-1	(30Ss)	3	27	6	20	4	13	2	7	2	7	2	7
Pre-2	(17Ss)	3	18	1	6	1	6	1	6	1	6	0	0
Pre-3	(16Ss)	0	0	0	0	0	0	0	0	0	0	0	0
Pre-4	(20Ss)	2	10	2	10	1	5	1	5	1	5	0	0
Pre-5	(27Ss)	3	11	0	0	0	0	0	0	0	0	0	0
Pre-6	(10Ss)	0	0	0	0	0	0	0	0	0	0	0	0
Pre-7&8	(11Ss)	0	0	0	0	0	0	0	0	0	0	0	0

than half that group. It was also only approximately half as large for the number of subjects who followed this pattern for two or more pairs. The percentage of subjects who did so for more than half the pairs in the test

is fairly low, and only two children out of 160 followed this selection pattern for all six pairs of the test. Therefore, a subject who ignored one member of a minimal pair and selected the other in all cases was more the exception than the rule, even in the youngest grades.

The minimal pairs were analyzed to determine whether this type of response bias pattern was evident to some degree with all six minimal pairs, or whether it occurred only with some pairs and not with others. Table 2.13 indicates the percentage of subjects of the total test group who displayed this response bias for each pair, i.e., the percentage of subjects who polarized their responses to one or the other member of the pair exclusively.

Table 2.13: Identification Test 1 -- Percentage of Subjects Who Exclusively Chose one Member of a Minimal Pair (Calculated for the Entire Test Group of 160 Children)

Pairs	% Ss polarizing response on each pair
Red Sox-red socks	11%
Redhead-red head	8%
Highchair-high chair	7%
Greenhouse-green house	6%
Hot dog-hot dog	6%
Blackboard-black board	5%

From this table we see that there are no great differences in the percentage of subjects who did this for each minimal pair.* Apparently, some subjects demonstrated a bias for one pair, while other subjects did so with a different pair. The total effect was one of small differences between pairs, and the question therefore seems to be more one of individual variation than of a general pattern of preference with such cases. Thus, this individual variation in response pattern seems to explain in part the fact that differences existed between Tests 1 and 3 concerning difficulty of members of pairs.

Although the pattern is not universally consistent, there was a strong general tendency for those subjects who chose only one member of the minimal pair to select the same member. This is indicated by Table 2.14, which shows only responses of those children who demonstrated a response

*The one slight exception to this might be the Red Sox-red socks pair, where the percentage of subjects polarizing their responses was slightly higher than with the other pairs, but this difference was not great.

Table 2.14: Identification Test 1 -- Percentage of Subjects Who Exclusively Chose the Compound or Phrase of the Minimal Pair

	Red Sox (18 Ss)	Greenhouse (10 Ss)	Red Head (13 Ss)	Highchair (11 Ss)	Blackboard (8 Ss)	Hot Dog (10 Ss)
Compound	6%	11%	38%	91%	87%	67%
Phrase	94%	89%	62%	9%	13%	33%

bias to select one member of a minimal pair exclusively. (This group represented 17.5 per cent of the 160 children tested). The table gives the percentage of those children who exclusively chose the compound member of a minimal pair and of those who exclusively chose the phrasal member. As can be seen, with some pairs the differences were less pronounced than with others, but in general they are fairly large. Also, the majority of the children who demonstrated a response bias toward consistent selection of the same member of a minimal pair chose the phrases red socks, green house, and red head, and the compounds highchair, blackboard, and hot dog. As was mentioned above, these are the members of pairs which tended to be preferrred by the test group as a whole in Test 1. This group tendency to prefer certain members of pairs disappeared when these subjects who polarized their responses were removed; thus the 17.5 per cent were apparently responsible for the group tendency noted above.

We might ask why this group of children in Test 1 who did not possess the ability to correctly identify these compounds and phrases distinguished only by stress patterns sometimes polarized their responses and why they tended to favor these particular members of minimal pairs. They may have done so with some pairs by chance. This is a possibility, particularly with those subjects who did so with only one pair; it seems less likely with those who polarized responses with more than one pair. If the children were learning stress patterns item by item rather than by rule, they may have selected the member of the pair which they knew or were beginning to learn and not that which they did not know. Obviously, the biases were not due to structural preference factors, since three of the members most often preferred were compounds and the other three were phrases. Also, as was noted above, there was no strong evidence of existing response biases toward the selection of certain structures. More probable explanations are the degree to which the particular lexical items were familiar to the children, preferences for certain pictures, and the possibility of difficulties of associating test items with their pictorial representations.

The first possibility to be considered is the question of familiarity. The lexical items which were most often selected (i.e., red sócks, green hóuse, red héad, híghchair, blácкboard, and hót dog) were usually those items of the minimal pairs which it might be argued are the most familiar or most often

heard by young children. The compounds Red Sox, greenhouse, and redhead, which were most difficult and least often selected, are lexical items which might possibly (though not obviously) be more unfamiliar than their counterparts. The phrases which were least often chosen, high chair, black board, and hot dog, seem to be more likely candidates as lexical items which are more novel and less often heard by young children than are their compound counterparts. Thus, the lack of familiarity factor could be a possible explanation in these cases. Also, the items most familiar to one child are not necessarily the same for another child, because of differences in background and experience. This would explain why some children who polarized their responses chose the members of minimal pairs that did not correspond to the group's trend. Finally, as mentioned above, the tendencies exhibited by Test 3 subjects did not always agree with the trend in the Test 1 group.

Another explanation for choosing certain members of minimal pairs concerns the nature of the pictures representing each test item. Possibly some pictures were more attention-getting or appealing to these children than others. They may always have selected that picture of the pair which most attracted them. Since all pictures were basically colored line drawings, however, it is not clear why certain pictures would be more appealing than others.

A final possible factor might have been difficulty in associating the phrases heard with the pictorial representations on their answer sheets. Though this is not an obvious factor with most pairs, it may have created some difficulty with the compound Red Sox. It was more difficult to portray this compound with a simple line drawing than it was to portray the other test items. While other items were pictured more or less directly, the Red Sox baseball team was indirectly represented by a picture of a baseball, a bat, and a cap. This may have been too abstract for children to grasp readily. The association factor seems a more plausible explanation for their difficulty with the item Red Sox than does a lack of familiarity. At the beginning of the testing session, before the experimenter had named the items to be tested, the children were asked if they could tell the experimenter the name of Boston's baseball team. Even the youngest children responded correctly and enthusiastically. This is not surprising, since in the Boston area the compound Red Sox is generally heard more frequently than the phrase red socks. Another indication that the primary difficulty was picturing the item Red Sox can be found in the results of the Task Comprehension Test which was given to children in Kindergarten and Grade 1 (for a discussion and description of this test, see above). In the Task Comprehension Test, five children in Kindergarten gave a few incorrect answers. In ranking the errors made by these subjects, Red Sox-red socks was found to be the most frequently missed pair. Whereas the other errors were more or less evenly divided among all pairs, over 30 per cent of all errors made by these five subjects involved the pair Red Sox-red socks.

Therefore, it appears that while familiarity of lexical items may have been a factor contributing to the occasional response bias with members of pairs, in the case of Red Sox the problem of picturing was probably also a factor.

Analysis of the Difficulty of Minimal Pairs

Because there were some children who lacked sufficient linguistic ability to identify these compounds and phrases on the basis of stress, and whose responses were therefore apparently affected by such nonlinguistic factors as degree of familiarity with a lexical item, preference for one picture on the answer sheet, or ease of associating a picture with its corresponding verbal test item, we must ask how much the occasional occurrence of such response biases toward selection of only one member of a minimal pair affected the performance of children on minimal pairs in general and whether there were large differences in the number of correct responses given for minimal pairs as a whole. The data were analyzed to determine whether some minimal pairs were correctly identified more often than others and were thus less difficult than others. The percentage of correct responses made by each grade for each minimal pair was calculated and the pairs were ranked from the least difficult to most difficult, i.e., from those with the highest percentage of correct responses to those with the lowest percentage of correct responses per grade. Table 2.15 shows the pairs and their rankings. It indicates that even among the youngest grades there was

Table 2.15: Identification Test 1 -- Ranking of Minimal Pairs by Grade*

	Réd Sox- Red Sócks	Gréénhouse- Green Hoúse	Híghchair- High Chái r	Rédhead- Red Heád	Blácкboard- Black Boárd	Hót Dog- Hot Dóg
Kinder-garten	1	5	2	4	3	6
Pre-1	3	1	2	5.5	5.5	4
Mid-1	1	2	5	4	6	3
Pre-2	2	4	1	5.5	3	5.5
Pre-3	1	6	2.5	2.5	5	4
Pre-4	1	2	4	5	3	6
Pre-5	2	1	6	3	5	4
Pre-6	3	1.5	5	5	5	1.5
Pre-7 & 8	4.5	1	4.5	2.5	2.5	6
Mean	2.1	2.6	3.6	4.1	4.2	4.4

*1 = most difficult; 6 = least difficult

considerable variation as to which pairs had the greatest number of correct responses and were the least difficult. However, looking at the results from the grades combined, we see that two pairs were generally more difficult than the others: Red Sox-red socks and greenhouse-green house.

The results of Test 3, concerning the ranking of difficulty of minimal pairs, is seen in Table 2.16. These data confirm the results from Test 1 as to the relative difficulty of the Red Sox-red socks and greenhouse-green house pairs, but none of its other pairs correlates with the Test 1 data in terms of ranking of difficulty.

Table 2.16: Identification Test 3 -- Ranking of Minimal Pairs by Grade*

Grade	Réd Sox-Red Sócks	Greénhouse-Green Hoúse	Rédhead-Red Héad	Bláckboard-Black Boárd	Hót Dog-Hot Dóg	Híghchair-High Cháir
1	1.5	1.5	5.5	3	5.5	4
2	3.	1.	2.	4.5	4.5	6
3	3.5	1.5	3.5	5.	1.5	6
4	2.	3.	1.	5.	6.	4
5	3.	4.	1.5	1.5	6.	5
6	1.	5.	3.5	2.	3.5	6
7 & 8	4.5	4.5	4.5	1.5	1.5	4.5
Mean	2.7	2.9	3.1	3.2	4.1	5.1

* 1 = most difficult; 6 = least difficult

More errors were probably made with the Red Sox - red socks pair than with others because of the difficulty of picturing the abstract compound and the resulting tendency to choose the phrase, since more children showed such bias with this pair than with any other. The greenhouse - green house pair may have been more difficult for children not only because of the tendency for some children to display a bias toward choosing the phrasal member of the pair, but also because the two pictures might have been confused. The pictures representing the two members of this minimal pair were more alike than were those of the members of any other minimal pair used in the test. Although they differed in detail, such as presence of flowers or green color, both greenhouse and green house were represented by essentially

the same shape. Hence the similarity of the line drawings may have made it difficult to distinguish between these two items.* Interestingly enough, even the adults found the greenhouse-green house pair the most difficult, which again suggests difficulty in distinguishing the two pictures.

Since these two pairs appear to be generally more difficult than the others, and since this difficulty appears to be caused, at least in part, by nonlinguistic factors, including the picturing of these items, we must ask how poor performance with regard to these pairs affected the general results seen above, such as the developmental trend in the ability to identify these contrasts. It might be argued that nonlinguistic factors confused young children trying to identify members of these two pairs, and therefore the linguistic ability they displayed with other pairs in the test may have been masked. This is not the case, however, because when these two minimal pairs are excluded from the data, the results are essentially the same as when they are included, as Table 2.7 indicates.

Table 2.17. Identification Test 1--Percentage of Correct Responses by Grade (With and Without Greenhouse-Green House, Red Sox-Red Socks)

	Without Greenhouse and Red Sox Pairs	With Greenhouse and Red Sox Pairs
Kindergarten	50%	49%
Pre-1	59%	57%
Mid-1	58%	55%
Pre-2	74%	73%
Pre-3	81%	78%
Pre-4	90%	87%
Pre-5	91%	89%
Pre-6	99%	96%
Pre-7 and 8	97%	99%
Adult	99%	99%

When the two most difficult items are excluded, the total percentage of correct items increases slightly, but the general developmental trend indicated earlier is virtually unchanged. Thus it does not appear that unfamiliarity with any lexical items or difficulty in picturing any particular item altered the data in any significant respect.

* For samples of these two pictures, see the appendix.

Reliability and Correlation of Tests: Test 1 Subjects Repeating in Test 2

In order to determine the reliability of both the subjects and the Identification Test in general, the results from Tests 1 and 2 were compared. (As noted above, Test 2 was identical to Test 1 and was given to 32 of the original test subjects several days after the first testing session.) Subjects were highly consistent in their performance on the two tests. On both tests the performance of the majority of the individual subjects approached roughly the same level of significance (as was determined by chi-square analysis) in terms of percentage of correct versus incorrect responses. Only one of the 32 subjects (a child in Grade 2) who performed at the chance level on the first test improved significantly on the second, moving into the category which produced a number of responses with less than a five per cent probability of occurring by chance. Three subjects who performed at the $p<0.05$ level of significance on the first test performed at the chance level on the second. These three subjects were in Grades 2, 3, and 4. The remaining 28 subjects (87.5 per cent of the repeat group subjects) remained in the same category. Of the latter, 23 scored a significant number of correct versus incorrect responses on both tests, and the remaining five had scores in which the number of correct versus incorrect responses were not significantly greater than chance on either test. Of these five subjects who scored at the chance level on both Test 1 and Test 2, two were in Grade 2, one was in Grade 3, and two were in Grade 5. The last two individuals are most interesting, because 85 per cent of the subjects in Grade 5 scored significantly more correct than incorrect responses, and these two not only had scores which were not significantly greater than chance (i.e., scores under 70 per cent correct responses) in Test 1, but also had scores which were in the same category in Test 2. This demonstrates both the degree of consistency of subjects and the reliable nature of the test.

A Spearman Rank Order Correlation Coefficient was computed to compare the ranks of these subjects on Test 1 and Test 2. It was found that the test-retest ranks showed a positive correlation of 0.81 for 32 subjects, which is significantly greater than chance at the $p<0.001$ level. This high correlation again indicates that subjects were consistent and performed at the same level both times the test was given. The degree of correlation over all subjects is perhaps misleadingly large since a number of subjects were artificially consistent because they had all correct answers or they were performing at the chance level on both tests. To remove this consistency due to "ceiling" effects, the scores of all the very good or chance level subjects were removed. The rho on the resulting 14 subjects was .488, significant at $p<0.05$. As discussed above, this high level of correlation is inconsistent with a nonlinguistic explanation of the differences in performance found on this test for both younger and older children. It seems unlikely that children who performed either very well or very poorly on the basis of factors like fatigue or amount of motivation would be so consistently affected by such factors.

The data on individuals who repeated the test and who made errors on both tests were analyzed to determine whether a correlation existed between the types of error made on Test 1 and on Test 2. Spearman Rank Order Correlation was used to test the similarity of the patterns of error in the two tests. For a given subject, the 12 items were ranked according to the number of errors made on each item (ties were assigned average rank). A similar set of ranks was obtained for errors made on the repeat test. These two sets were then correlated (using the standard Pearson's Product Moment Correlation Coefficient). This gave a Spearman Rank Order Correlation (rho) which could be used as an index of similarity of error pattern across lexical items. The obtained values are listed in Table 2.18.

Table 2.18: Identification Tests 1 and 2 -- Spearman Rank Order Correlation Coefficients for Subjects with Errors on Both Tests

	Subject Number	Rho
Grade 2	2.1	0.3536*
	2.2	0.5176
	2.3	-0.2991
	2.4	0.3014
	2.5	0.3113
	2.6	-0.0360
	2.7	±0.0000
Grade 3	3.1	0.3550
	3.2	0.3536
	3.6	0.0376
	3.7	±0.0000
	3.8	-0.4454
	3.9	0.7704 ($p<0.01$)
Grade 4	4.1	0.4832
Grade 5	5.5	0.3985
	5.6	0.5766
	5.7	0.0182
Adult	A.2	-0.1348

*All rho values not marked as negative are positive. All values are not significantly better than chance unless otherwise indicated.

Because the correlations were weak and there were not a large number of items for each subject, the significance test for rho was not very sensitive. There was only one subject for whom the test was sensitive enough to show that the correlation could not have happened by chance. If we had more

items per person, however, and if we had the same range of correlation as in this table, many of the correlations would become significant. As can be seen from the table, there were many more positive than negative rhos. This would not be expected if the rhos only reflected chance deviations from zero correlation. A Wilcox matched-pairs signed-ranks test was used to determine whether the correlations, paired with theoretical values of zero, had a tendency to be positive. This was done to find out whether the group of children as a whole (17 Ss)displayed significant tendencies to repeat lexical errors. The value of T (with the 15 pairs which were not tied) was 17.0, significant at the level of $p<0.01$. Thus, although there was only one subject whose correlation was strong enough, taken by itself, to be significant,the over-all pattern of resemblance between test and retest errors was significantly greater than would be predicted on the basis of chance. If this is interpreted as a tendency to make errors on unlearned lexical items, and to get better scores on learned lexical items, then the group as a whole gives support to the notion of a weak but significant lexical effect.

To go more deeply into the question of whether a given child tended to make the same errors on the repeat test, the errors made by individual subjects on each item were examined. Unfortunately there was not a large number of tokens for each type, so it was difficult to determine trends. These data do not indicate whether children erred only with certain lexical items or whether errors were more random and not tied to specific lexical items. Some children appeared to err randomly and others showed evidence of knowing some items but not knowing others. If histograms were constructed to show the number of times each error score was obtained for a given item, it would indicate no tendency to group scores around chance and perfect levels; instead responses would be distributed across all categories, demonstrating evidence of gradual learning. The small number of tokens for each item makes statistical testing of the resulting crude histograms difficult. If there is gradual learning, it is difficult to say from these data whether it is a rule or individual lexical items that are learned gradually.

Reliability and Correlation of Tests: Test 1 Subjects Repeating in Test 3

The results of the 21 children who were given both Test 1 and 3 (the partial replication test given one year later) were analyzed and compared in order to consider questions of reliability and developmental improvement. A comparison of the number and percentage of correct responses scored by each of these subjects on the two tests indicates high consistency in performance. Considering the number of correct and incorrect responses scored by each child on each test and the level of significance of each score, it was found that 17 of the 21 subjects (80.9 per cent) had scores on Test 3 which were in the same category in terms of level of significance as their scores on Test 1. Fifteen of these subjects scored more correct than incorrect responses at a level that was significantly greater than chance on both tests, and the other

two children had scores which were not significantly greater than chance on either of the two tests.

The most interesting question to be considered with the comparative data from these two tests is that of the learning and development of individual children; i.e., did the children whose scores were not significantly greater than chance on Test 1 learn of develop sufficient linguistic ability over the period of a year to score a significant number of correct versus incorrect responses on Test 3? Unfortunately, this question cannot be answered with assurance, because of the 21 children who repeated the test one year later, only five had scores on Test 1 which were not significantly greater than chance. Three of these five children did improve one year later and scored significantly more correct than incorrect responses. Two of the five did not improve and had scores that were not greater than chance on the test one year later as well as on the initial test. The three children whose scores improved were older than the two whose did not; the three whose performance improved were in Grades 3, 4, and 5, respectively, during the time of the first test, and the two whose performance remained poor were in Grades 1 and 3, respectively, at the time Test 1 was given. In the case of the First Grader, his age may explain why the year's lapse between tests failed to bring about a significantly improved performance.*

The small number of data again indicate that repeated practice in taking the test does not produce learning effects but rather that improvement in performance on later tests is more the result of developmental linguistic factors.

---

* Only one child of the 21 performed considerably worse on the later test. This child, production subject SB (see Chapter 5), was in the Third Grade at the time of the original test and performed at a level significantly greater than chance on the first test but did not do so on the test given a year later. However, she did perform significantly on the Experimenter Identification Test and other production tests also given a year later. Apparently her failure on the replication test was due to a performance factor like lack of motivation. Since she had also taken the repeat test several days after the main test, the replication test constituted the third time she had taken the test. Under such conditions, it is easier to understand why a lack of motivation could have been sufficiently strong to affect performance so negatively. However, cases where nonlinguistic factors can play such a large role in determining results appear to be rare, and certainly are the exception rather than the rule. This is evidenced by the fact that, in addition to SB, six other children also took all three tests. Unlike SB, five of the six were extremely consistent in their performances on all three tests, and one showed some improvement from test to test. The sixth was in Grade 1, and his score was not significantly better than chance during the initial test. He improved slightly on Test 2 and a year later had reached a significance level of $p<0.02$ on Test 3. The other five scored significantly better than chance on all three tests. Hence, the nonlinguistic factors which apparently affected SB did not appear to affect the six other children tested under similar conditions.

Chapter 3

Preference Experiment with Compounds and Phrases

A test of preferred forms ("forced-choice") was conducted to substantiate the findings from the compound-phrase Identification Test and to determine whether children who had failed with an identification task might succeed with a different testing procedure. This test, which will be referred to as the Preference Test, asked children to listen to compound and phrasal minimal pairs presented in an identical sentential context and then to select the correct sentence, i.e., the one that "sounded best." For example, children heard the two sentences

(A) I put mustard on my hót dog.

(B) I put mustard on my hot dóg.

They were then asked to indicate the correct sentence by selecting either sentence (A) or sentence (B).

This Preference Test differed from the Identification Test in that children had the opportunity to hear and compare the minimal pair stimuli and then to select the one which they judged to be correct, whereas in the Identification Test only one member of the minimal pair was presented and the task was to identify it. The type of response also differed: in the Identification Test children were asked to choose between pictures that represented the minimal pair test items themselves, while in the Preference Test children chose between two letters, A or B, which were labels for the two sentences containing the minimal pair test items. Another obvious but important difference between the two testing methods was that in the Preference Test each compound and phrase was heard in a sentential context, while in the Identification Test the test stimuli were presented in isolation.

In order to verify the results and conclusions drawn from the Identification Test, the procedures and results of the Preference Test will be compared with the Identification Test.

Method

Subjects

The 125 children and 19 adults used as subjects in Identification Test 3 were used as subjects in the Preference Test. (The breakdown of

the number of subjects by grade and sex can be found in Table 2.3.) A number of additional children were given the Preference Test; they will be considered in Chapter 5, since they were the subjects used to gather data on ability to produce the stress patterns of compounds and phrases, and they were given the Preference Test primarily to compare their ability to comprehend with their ability to produce.

Stimuli and Their Presentation

The six minimal pairs of compounds and phrases used as stimuli in the Identification Test were also used as stimuli in the Preference Test. As with Identification Test 3, a total of 13 test items were given in the Preference Test. Each member of the six minimal pairs was presented once, and the first item, hót dog, was repeated at the end of the test. As with Test 3, this additional test item was used only for measuring repetition of errors on a given item and for a comparison with this repeated item in Test 3. The general results will consider only the first 12 items, six compounds and six phrases. The order of these 12 stimuli matched the order of the stimuli given in Test 3, this order also being the order of the first 12 items of Identification Test 1. That is, the items in their sentential contexts were arranged in two sets of six so that one member of each of the six pairs would be presented once per set. The pairs were embedded in the sentences found in Table 3.1:

Table 3.1: Sentence Stimuli with Compounds and Phrases Used in Preference Test

1. (A) I put mustard on my hot dóg.
   (B) I put mustard on my hót dog.

2. (A) We built a book case with a black bóard.
   (B) We built a blácкboard.

3. (A) Father painted our home, and now we have a gréenhouse.
   (B) Father painted our home, and now we have a green hóuse.

4. (A) My favorite baseball team is the Réd Sox.
   (B) My favorite baseball team is the red sócks.

5. (A) The boy with the sunburn had a red héad.
   (B) The boy with the sunburn had a rédhead.

6. (A) The baby ate lunch in his híghchair.
   (B) The baby ate lunch in his high cháir.

7. (A) The teacher wrote on the black bóard.
   (B) The teacher wrote on the blácкboard.

8. (A) The dog in the sun was a hot dóg.
   (B) The dog in the sun was a hót dog.

9. (A) The girl with red hair is a red héad.
   (B) The girl with red hair is a rédhead.

10. (A) We grow flowers in our gréenhouse.
    (B) We grow flowers in our green hóuse.

11. (A) Johnny wore Réd Sox.
    (B) Johnny wore red sócks.

12. (A) The life guard sat above the beach in a híghchair.
    (B) The life guard sat above the beach in a high cháir.

The arrangement of the minimal pairs within each of these sets of test sentences was correlated with the A and B answer choice possibilities. The test design was such that there were an equal number of correct A and B answers and an equal number of correct compound and phrase answers. One half the correct A answers were compounds and the other half were phrases. The same was true of the B answers.

In addition to the 12 sets of test sentences, it was felt that some sentences should be included in the test to give an indication of young children's ability to understand and perform the type of task which the Preference Test involved. (As noted in Chapter 2, a Task Comprehension Test was given to the youngest Identification Test subjects to measure their ability to perform the identification type of task.) However, time, an important consideration in the testing of young children, allowed for only two task comprehension sentences to be used in the Preference Test. The two dummy sentences which were included in the Preference Test, one preceding and one following the test sentences, were the following:

1. (A) The book is on the táble.
   (B) *The book is on the tablé [təbél].

2. (A) *Please close the windów.
   (B) Please close the wíndow.

These dummy sentences were the same type as the regular test sentences with compounds and phrases in that they required the child to choose between two sentences which were identical in all respects except for the different stress patterns given to the test item embedded in each. The difference between the dummy sentences and the regular test sentences was that in the former the lexical item which appeared with two different stress patterns had no minimal counterpart in the language. The words table and window were selected for this test of task comprehension because they are very common words which all children should know. If a child failed to choose correctly

between these two sentences, it is possible that he did not know the stress patterns of these words. But more likely he would fail either because he did not understand the task, or because he was failing to attend to it and to the stimuli for some nonlinguistic reason such as lack of motivation or fatigue.

A stimulus tape was recorded of the experimenter producing the test sentences using the same recording apparatus and procedures that were described in Chapter 2. The testing procedure was similar to that used for the Identification Test. The minimal test phrases were again pronounced for the children and their accompanying pictures from the Identification Test were referred to while the phrases were being pronounced. The children were told that they would again hear these same phrases but that this time they would be in sentences. They were then given a simplified explanation of the task.

As with the Identification Test, the children were given answer sheets on which to record their responses. They heard each of the two sentences they were to choose between pronounced twice. This gave them the opportunity to hear and compare the two and then to hear them again before selecting the correct one. Also as with the Identification Test, children were instructed always to select one sentence even if they were not certain of the correct answer. Omissions, therefore, were again scored as errors.

To be certain that children understood the task, three simple examples were given before the test was begun. The following are the sample sentences used:

Example 1. (A) Give it to the báby.
(B) *Give it to the babý.

Example 2. (A) *We took a bath in the bath túb.
(B) We took a bath in the báthtub.

Example 3. (A) *We write with our pencíls.
(B) We write with our péncils.

The correct answer was given to the children after each example. The examples and instructions were repeated if necessary and any questions were answered. The stimulus tape with the test items was then played without interruption.

The subjects were again tested in small groups; the groups were identical to those that took Identification Test 3 since the Preference Test was given in the same testing session as Test 3.

Results

In the Preference Test, as in the Identification Test, there was no obvious inability to understand, or lack of motivation. The children appeared to respond very positively to the task.

General Results

The Preference Test results support the earlier findings of the Identification Test by indicating that the ability to select the correct minimal pair increases with age. Figure 3.1 shows the percentage of correct responses for the Preference Test according to age and grade. This figure indicates general, gradual improvement beginning with Grade 1, which had 50 percent correct, and rising to Grade 7 and 8, which had 98.5 percent correct. (The scores in Grades 7 and 8 were even slightly higher than that of the adult subjects.)

This developmental trend is the same as that found with Identification Test 3, which was the test given to the same subjects who took the Preference Test. Figure 3.2 compares the general results of these two tests and shows that the group performed in a similar fashion in both. In fact, chi-square analysis indicates that the difference in performance on the two tests was significantly greater than chance only for Grade 4 ($p<0.02$).* There is no obvious reason for this discrepancy in the performance of the adult group, nor is there any good explanation for the slight difference in the performance of Grade 4 on the two tests. As was discussed earlier, however, Grade 4's performance on Identification Test 3 was relatively low in relation to that of Grades 3 and 5, and it was also low in comparison to the performance of Grade 4 on Test 1 (the 30-item Identification Test). Thus, this slight drop in the performance of Grade 4 on Identification Test 3 suggests that nonlinguistic factors like motivation or attention may have had an influence.

The results from the Preference Test also indicate that even the youngest children were successful in selecting the correct option with the dummy sentences (those items included to indicate degree of task comprehension and motivation). Figure 3.3 shows the percentage of correct responses by grade on these sentences contrasted with the percentage of correct responses by grade on the regular test sentences with compounds and phrases. Even the youngest grades scored over 90 per cent correct responses with the dummy sentences, and Grade 4 and above scored 100 per cent correct responses. No child gave an incorrect response for both dummy sentences. In grades 1 through 3 only one, five, and three subjects, respectively, missed one of the dummy sentences. This suggests that few, if any, of the children did not understand the task or were unable to attend to the stimuli and correctly differentiate the stress patterns of the dummy items. The fact that even the youngest grades had a high percentage of correct responses on the dummy sentences while the figures for the actual test sentences were much lower (e.g., a difference of 44 per cent for Grade 1 on the test and dummy stimuli) suggests that the inability to perform the task was due primarily to linguistic factors.

*The difference in Grade 1's performance on the two tests appears to be large, but this is apparently due to the small number of subjects in Grade 1. The value of chi-square is 2.155, which is not significantly greater than chance.

Figure 3.1: Preference Test -- Percentage of Correct Responses According to Grade and Age

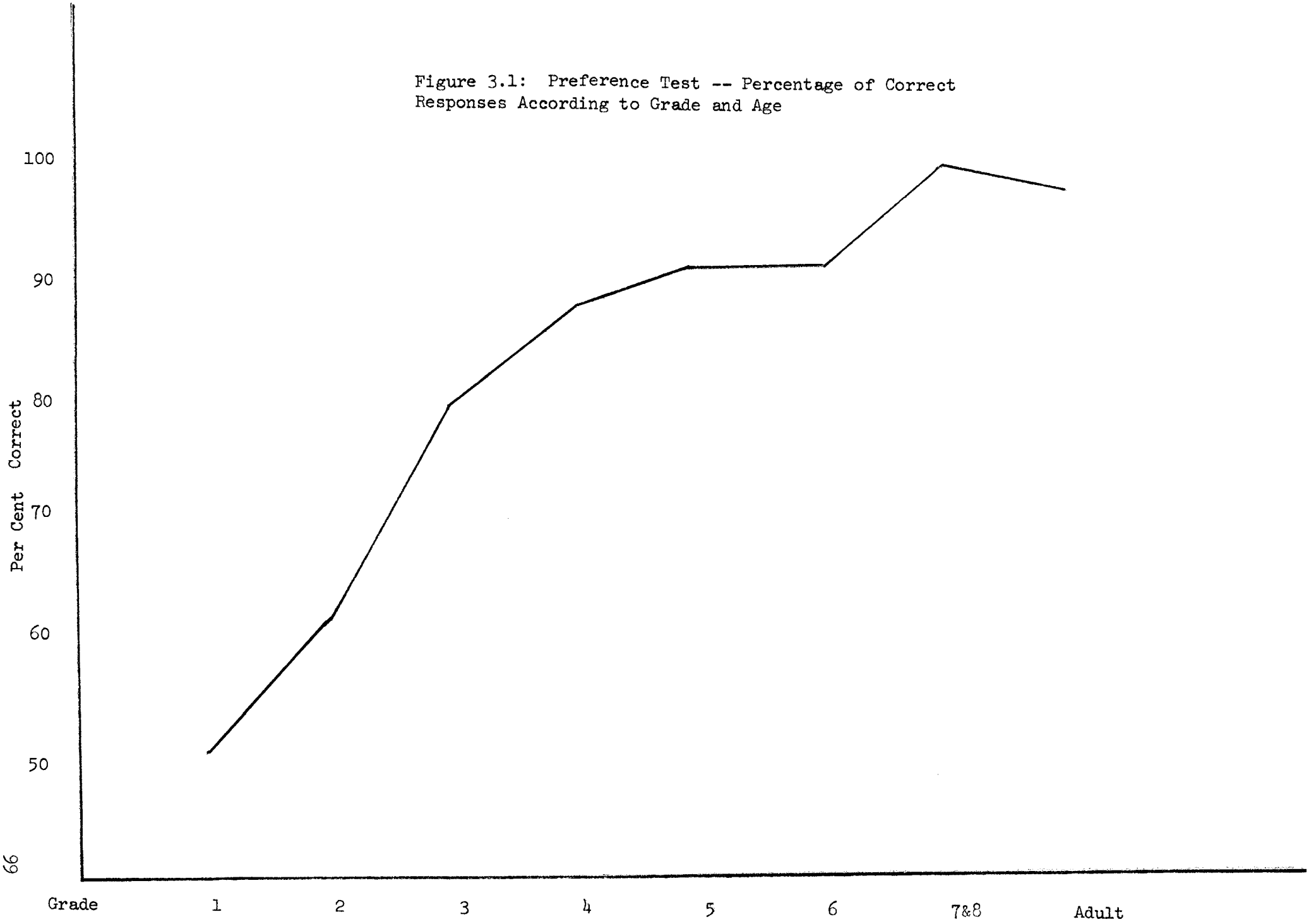


Figure 3.2: Preference Test and Identification Test 3 — — Comparison of Percentage of Correct Responses

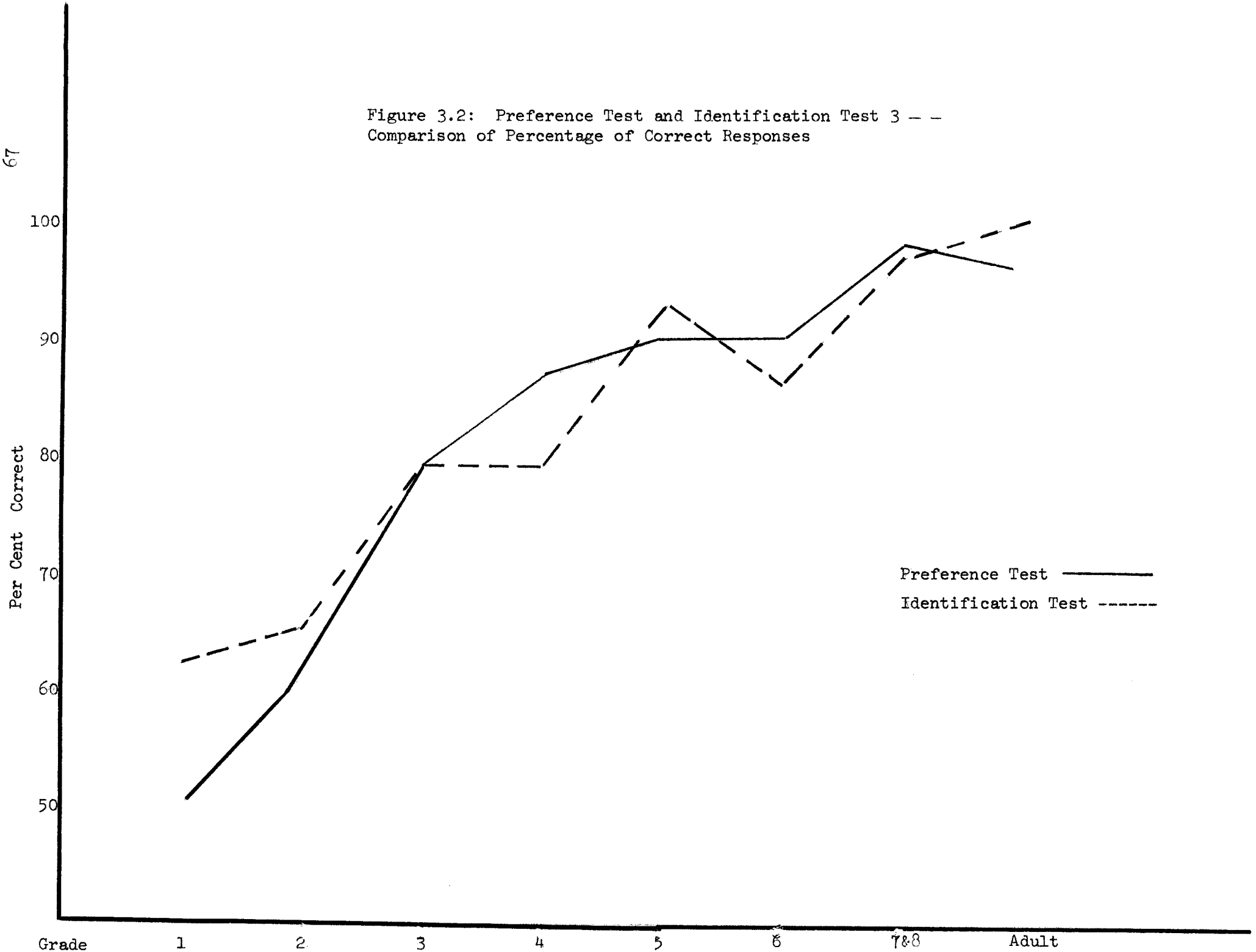
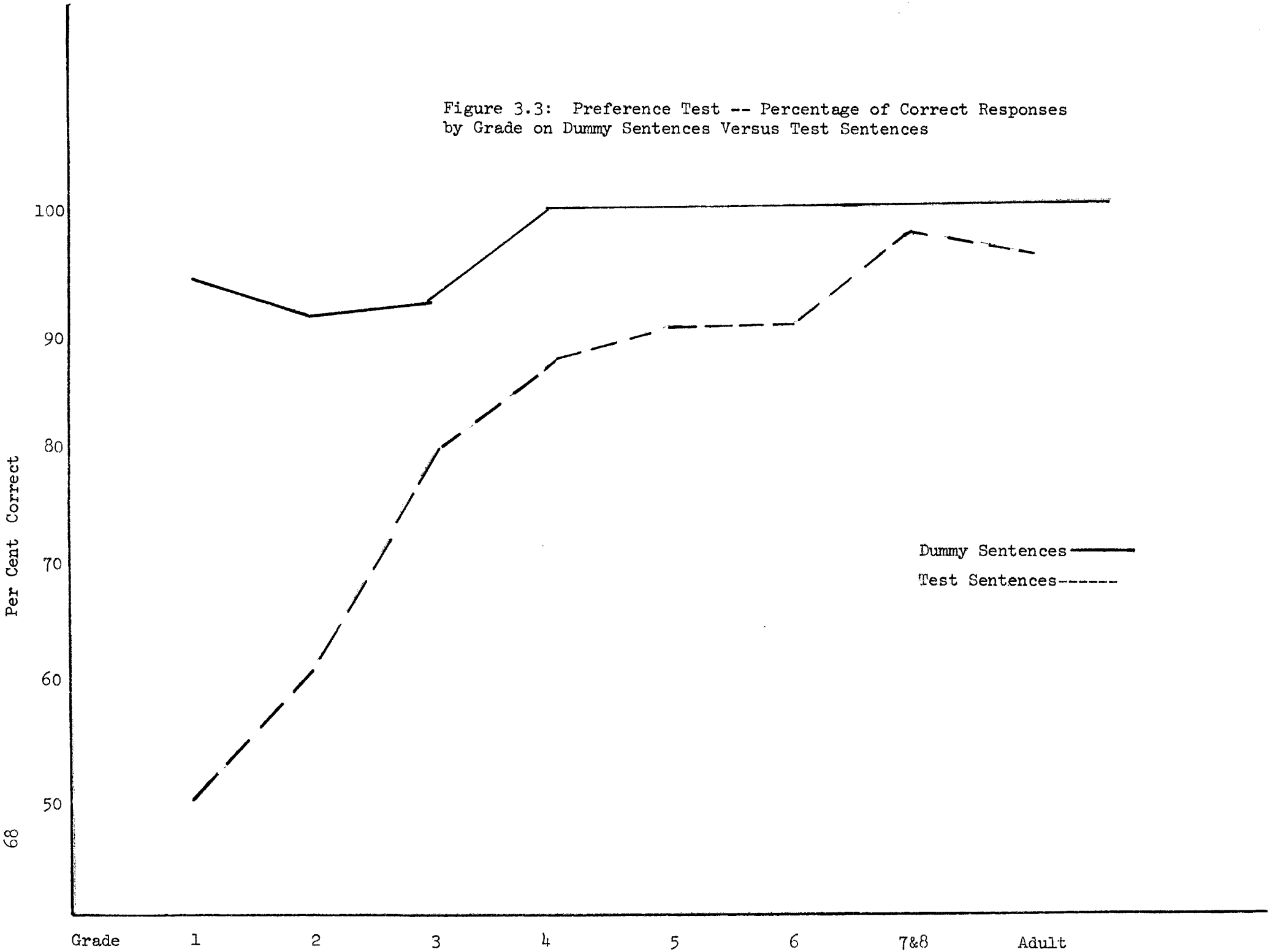


Figure 3.3: Preference Test -- Percentage of Correct Responses by Grade on Dummy Sentences Versus Test Sentences



Frequency Distribution and Individual Variation

A frequency distribution table was compiled to determine the range and breakdown of children's scores within each grade, particularly in relationship to the average scores for each grade plotted in Figure 3.1. Table 3.2. shows that distribution tended to be symmetrical and normal rather than polarized at the two extremes and that in general the majority of the children in each grade had scores which approached the average score for the group:

Table 3.2: Preference Test -- Frequency Distribution of Percentage of Correct Responses of Subjects by Grade

	Number of Subjects Scoring at Each % Interval*						
% Correct Responses	Grade: 1	2	3	4	5	6	7 & 8
	(Mean) (50)	(61.4)	(78.7)	(86.7)	(90.3)	(89.6)	(98.5)
0							
8.3			1				
16.7							
25.0		2					
33.3	1	3					
41.7	3						
50.0	*2	3	1				
58.3	1	6		1	1	1	
66.7		*6	3	2			
75.0		2	1	4	2	2	
83.3	1	2	*3	*5	2	2	
91.7		2	3	5	*6	*5	2
100.0		1	4	8	7	9	*9

[Blank space = 0; interval closest to group mean = *]

* The intervals used in the frequency distribution table were the same as those used in the Identification Tests frequency distribution table rather than those used in Identification Test 1, because the Preference Test had the same number of test items as did Test 3. As previously mentioned, the intervals selected for Test 3 differed from those chosen for Test 1 because of the difference in the number of items in the two tests, Test 1 having 30 items and Test 3 having only 12.

In grades where the group average was above chance but did not reach a high level of accuracy, the group score was more a result of the majority of the children performing at a level consistent with the group average than of approximately half the group performing very well and the other half poorly. This, then, is roughly the same distribution pattern as was seen with the Identification Tests.

Also as with the Identification Tests, the distribution pattern of the Preference Test indicates that there was a certain amount of individual variation within grades. This variation is seen in all except Grade 7 and 8. Although most children performed at a level consistent with the group mean, there were some whose scores differed greatly from that average. A few in the youngest grades performed as well as adults, and a few as late as Grade 6 performed as poorly as the majority of the youngest children.

The amount of variation in the performance of children within the same grade is also apparent in Table 3.3,which shows the percentage of children scoring more or less than 80 per cent correct responses according to grade:

Table 3.3: Preference Test -- Percentage of Subjects Scoring More or Less Than 80% Correct Responses

Grade:	% Ss Less Than 80% Correct (p = ns)	% Ss More Than 80% Correct* (p = <0.02)
1	87.5	12.5
2	81.5	18.5
3	37.5	62.5
4	28.0	72.0
5	16.7	83.3
6	20.0	80.0
7&8	0	100.0
Adults	0	100.0

It indicates a general trend of increasing percentage with increasing age. This continues until Grades 7 and 8, when the 100 per cent level was reached. There was only one dip, which occurred with Grade 6, but it is not statistically significant. The largest jump in percentage occurred between Grades 2 and 3 where the percentage of subjects scoring significantly more correct

*As with the Identification Test 3, a minimum of 80 per cent correct responses (at least 10 correct of 12 items) was required for the probability of chance results to be less than five per cent ($p<0.02$ as was determined by the binomial test). This was due to the small number of items in the test.

than incorrect responses went from 18.5 to 62.5.

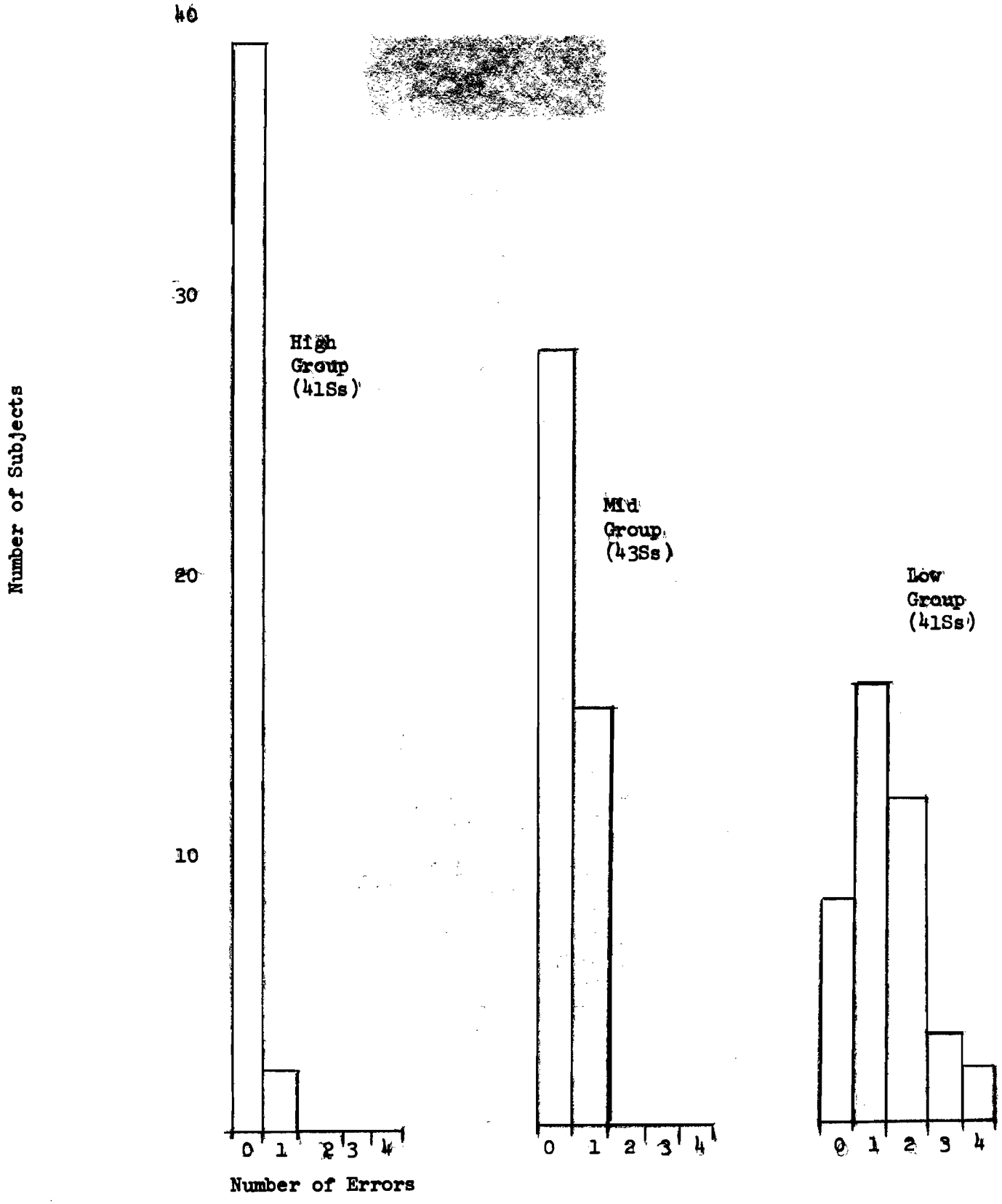
If we compare this table with Table 2.9, which shows the results of the Identification Test 3 taken by the same subjects, we find a good correlation between the two sets of data. The percentages were identical in Grades 1, 3, and 7 and 8, as they were with the adult group, and there was a difference of only 6 per cent or less (the difference of one child) in all other grades except Grade 4. This grade, as previously discussed, demonstrated an unusually low percentage in the Identification Test 3, but in the Preference Test Grade 4's percentage appeared to be roughly at the level which would be predicted on the basis of the scores of Grades 3 and 5 in the Preference Test and the score of Grade 4 in Identification Test 1. The fact that this difference existed between Grade 4's performance on the Preference Test and on Test 3 was indeed atypical and due to nonlinguistic performance factors rather than to linguistic ones. The small dip which occurred with Grade 6 in the Preference Test was also found in Identification Test 3, but it is not significant in either test. The high degree of agreement between the Preference Test and Identification Test 3 provides further support for and confidence in both tests and their results.

Gradual Learning

Identification Test 1 gave evidence that children gradually acquire the skill to differentiate compounds and phrases on the basis of stress: most children in Grades 2 through 5 performed at a level high enough to be significant ($p<0.05$,70% or better) but not at a near perfect level ($p<0.001$, 85% or better). We might also consider this question of gradual learning in terms of the data from the Preference Test. Like Identification Test 3, this test had only 12 items. This number is too small to make comparisons between different levels of significance possible as was done in Test 1. However, the Preference Test, again like Test 3, contained a 13th item, hót dog, which was included to allow for considerations of learning effects and gradual learning. Since both the Preference Test and Test 3 presented two tokens of the stimulus hót dog, each subject taking these two tests responded to hót dog a total of four times in these two tasks. These four items were therefore considered together.

The 125 children were divided into three groups (high, middle, and low) according to each individual's combined total score on the two 12-item tests. Each of these three groups had approximately the same number of subjects (41 subjects were in the low and high groups and 43 in the middle group). Subjects in the high, middle, and low groups scored 96 to 100 per cent correct, 75 to 93 per cent, and 33 to 71 per cent, respectively. Each subject's number of errors (0 to 4) was plotted on one of the histograms in Figure 3.4. A large number of subjects in the high and middle groups scored either 0 or 1 error, which shows that the item had basically been

Figure 3.4: Preference Test and Identification Test 3 -- Total Number of Errors on Four Tokens of Hot Dog by Three Groups of Subjects.



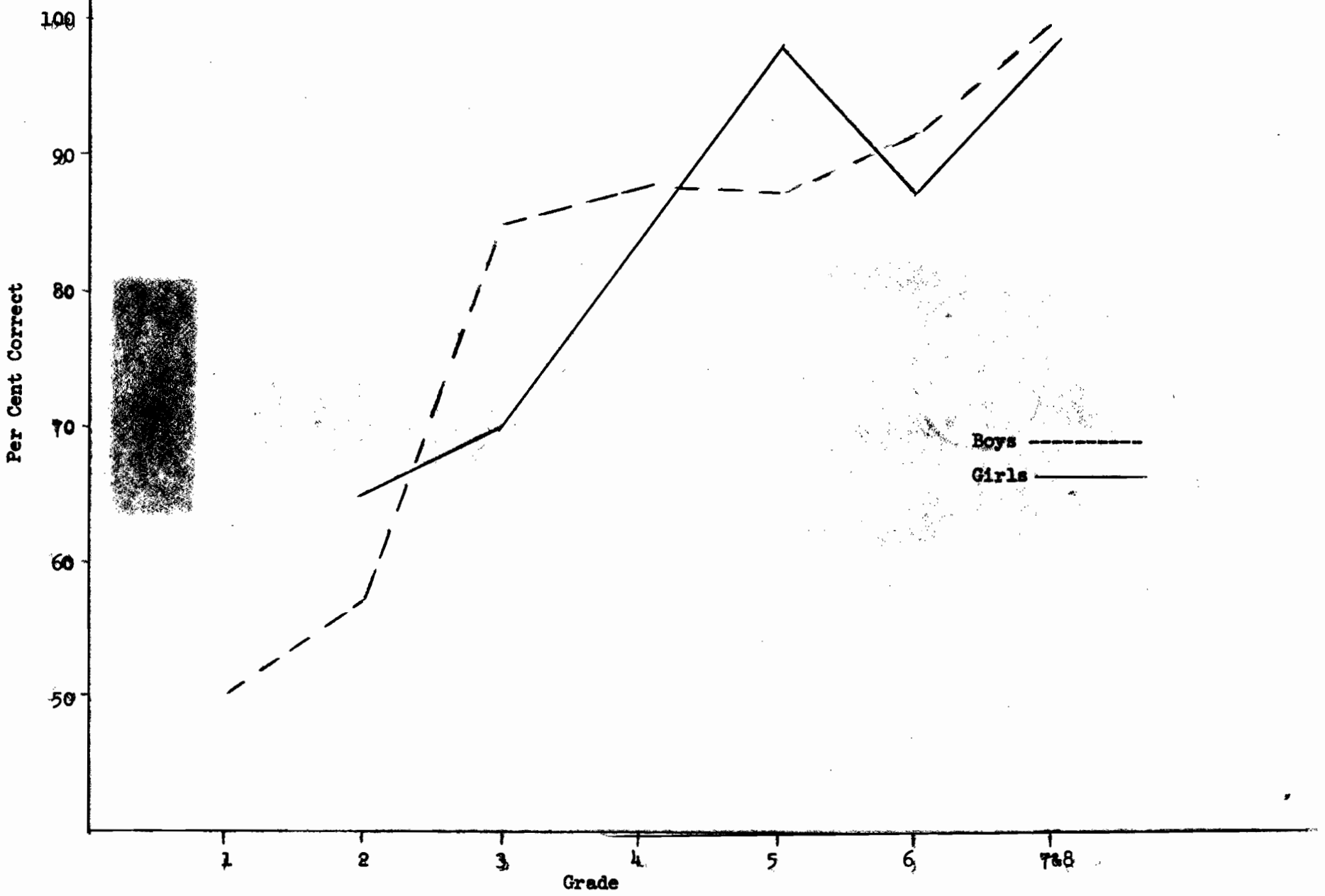
learned by the subjects in those groups. The low group had a wider spread over the possible number of errors. The distribution of the subjects over the number of errors by this group of children is not consistent with all or none learning. Such a notion would predict that errors be centered around the chance score of two where the item is not learned except for extra errors of 0 for those cases where the item is learned perfectly. Instead there was a gradual descent from 0 to 4, suggesting that with some subjects the item was half-learned. The probability of giving correct responses was greater than chance level but far less than 100 per cent. (Future experiments could test this more accurately by administering a larger number of tokens for each item so that the distribution could be resolved more clearly.)

Differences According to Sex

The Preference Test results were analyzed to determine whether any significant differences existed between the performance of boys and that of girls at various ages. The percentage of correct responses according to sex and grade are plotted in Figure 3.5. (No girls were in Grade 1 so no such comparison can be made with that age group.) In Grades 2 through 8 there was no consistent tendency for one sex to perform better than the other. In Grades 2 and 5 girls scored better than boys, but in Grades 3, 4, 6, and 7 and 8 the opposite is true. The Mann-Whitney U test indicates that the only grade where the difference in the performance of girls and boys was significantly greater than chance was Grade 5, where α 0.05 in favor of girls.

Since Identification Test 3 was given to the same subjects used in the Preference Test, a comparison of sex differences between the two tests was made. Figures 2.8 and 3.5, which plot the performance of both sexes in the two tests, indicate little agreement between these tests with regard to the performance of boys versus girls. Comparing first the performance of each sex on the two different tests, we find that in Grades 1 and 2 boys tended to perform better on the Identification Test than they did in the Preference Test. However, in all other grades (except Grade 5) boys tended to perform slightly better on the Preference Test. There was no definite pattern in terms of girls scoring better on one test or the other. In Grades 2 and 4 they did slightly better on the Preference Test, and in Grades 3, 5, and 6 they did better on the Identification Test. In Grade 7 and 8 their performance was the same on the two tests. The results of the two tests agree as to which sex performed better in only one grade of the six -- Grade 5, in which girls performed better than boys on both tests. In all other grades there is disagreement between the two tests as to which sex scored higher, except in Grade 7 and 8, where girls and boys performed equally in the Identification Test, and boys did better than girls in the Preference Test. In the Preference Test, only Grade 5 displayed differences between the scores of girls and boys which were significantly greater than chance. In Identification Test 3 no grades showed

Figure 3.5: Preference Test -- Percentage of Correct Responses According to Sex and Grade.



differences that were significantly greater than chance. Since the Preference Test results showed no tendency for girls to perform consistently better than boys, as the Identification Test did, and since only one grade displayed significant differences between the scores of boys and girls, even further doubt is cast upon the importance of sex as a factor affecting the acquisition of these stress patterns.

Learning Effects

The role of learning effects is difficult to evaluate with the Preference Test because of the small number of items it contained. There was only one trial for each type of stimulus with the exception of the first item, which was repeated after each of the other 12 stimuli had been presented in order to give at least some indication of possible learning effects. (This was also done with Identification Test 3). A comparison of the number of errors made by each grade on the first and second trials of this item, hót dog, showed no significant differences, thus agreeing with the Test 3 results. In no grade did the number of errors on the second trial decline; hence no grade showed even a slight learning effect. (There were also no indications of fatigue effects, since the differences between the number of errors on the first and second trial in each grade was either minimal or nonexistent.)

Another indication of the presence of learning effects might be found by comparing the number of errors made on the first six sentence pairs with those made on the second six. Such a comparison is, of course, complicated by the fact that the sentences were not identical in the first and second groups. Nevertheless, there is reason to assume that the degree of difficulty of the sentences in the first and second halves of the test was roughly equivalent, since they both presented the same number of compound and phrase correct responses, the same number of A and B correct answers, and one member of all six minimal pairs. However, when this comparison was made, it was found that the number of errors in the first and second halves of the test was not significantly different for any grade or for the test group as a whole. Hence, no obvious learning effect and no fatigue or motivational factor is involved. This agrees generally with the findings of Identification Test 1 where there was no evidence that these factors were significant in any grade.

Response Biases: Selection of Letters on Answer Sheets

The results were examined to determine whether there was a response bias toward selecting one letter rather than the other on the answer sheet. Because the order of the two letters children chose between, A and B, was alphabetical in all cases, a possible response bias toward selecting one letter might also be a question of the positioning of the letters on the

answer sheet, since the A always appeared to the left of the B. Examination disclosed no evidence of such biases with any age group. A few more A answers were given by Grades 1 and 2 and more B answers by all other grades, but the differences were not significant for any grade. In the entire test group of 125 children, only two individuals displayed a bias toward selecting only one of the two letters. One child in Grade 3 selected the letter A exclusively, and one child in Grade 6 chose A for five of the six pairs.

Response Biases: Type of Syntactic Structure

The data were analyzed for evidence of a bias toward selecting compounds more often than phrases, or vice-versa. Table 3.5 shows the number and percentage of responses for each structural type. The table indicates

Table 3.5: Preference Test -- Total Number and Percentage of Compound and Phrasal Responses

Grade and No. of Ss		Total No. Responses	Noun Compound No.	Noun Compound %	Noun Phrase No.	Noun Phrase %
1	(8)	96	54	56.3	42	43.8
2	(27)	324	184	56.8	140	43.2
3	(16)	192	103	53.6	89	46.4
4	(25)	300	158	52.7	142	47.3
5	(18)	216	117	54.2	99	45.8
6	(20)	240	133	55.4	107	44.6
7 & 8	(11)	132	67	50.8	65	49.2
Adults	(19)	228	120	52.6	108	47.4
TOTAL	(144)	1728	936	54.2	792	45.8

that when subjects were presented with a choice between the compound and the phrasal member of minimal pairs, all age groups displayed a tendency to select the compound. However, chi-square analysis shows that this tendency was not significantly greater than chance for any grade except Grade 2 of the individual groups, where $p<0.05$. But because all the groups displayed this tendency and because it approached significance in most groups, the group of 125 children was considered as a whole. We then find that the tendency of the entire group to choose the compound was significantly greater than chance at the $p<0.001$ level of significance. Thus, there did appear to be a general tendency to select the compound stress pattern more often than the phrasal stress pattern.

Only two children showed a strong tendency to select one structural type over the other. One child in Grade 1 and one child in Grade 6 selected compounds for 11 of their 12 responses on the Preference Test. Strangely

enough, these two children both performed well on the other tests with compounds and phrases (the Identification Test and the Semantic/Acoustic Test). Thus, their poor scores and biases in the Preference Test were apparently not due to lack of linguistic knowledge but rather a matter of their preferring the compound stress pattern to the phrase stress pattern under the conditions of this test.

Why this tendency to select compounds more often than phrases should exist is not obvious. It was not found in the Identification Test. The fact that children were selecting between pictures to identify the item they heard in the Identification Test while in the Preference Test they were asked to indicate the sentence which "sounded the best" seems to imply a preference for the compound stress pattern. This bias may be a reflection of the frequency of appearance of compounds with children and of the fact that compounding is still a productive process in the language. Young children may be more familiar with this stress pattern than the other because it may be acquired earlier, or because it may be preferred to the other in their own productions of these minimal pair contrasts. Our knowledge of the early acquisition of these and other stress patterns is too sparse to be helpful in these speculations, but the production data for these contrasts, which is discussed in Chapter 5, provide some additional information.

One other possible explanation for this general tendency to prefer compounds must also be considered. In some cases the test item with primary stress on the first syllable may have been selected not because it indicated a compound, but because it indicated a phrase whose primary stress had been shifted to the first syllable by contrastive stress. In constructing the test sentences an attempt was made to limit the possibility of contrastive stress interference; however, this proved to be difficult. Some test sentences may allow contrastive stress to give a second possible reading. Perhaps the most likely sentence in which this may be the case is, "Father painted our home and now we have a {greenhouse / green house}." The now could signal contrastive stress shift (i.e., a gréen house, not a brówn house). This tendency would have been even stronger if the now had received heavy stress, which it did not. Even though some stimuli are possibly ambiguous, one reading is more probable than the other. The fact that the general results from this test correlate so well with those of the Identification Test suggests that any possible contrastive stress interference and stimuli ambiguity did not seriously affect the results. The context tended to disambiguate the stimuli for the children, and in these potentially ambiguous sentences most subjects did not interpret them as ambiguous but gave them the intended reading.

Response Biases: Selecting Certain Members of Minimal Pairs

An item analysis was considered as a way of finding out whether there was any bias toward the selection of one member of a minimal pair more often than the other. Because the Preference Test had a small number of items presented only once (with the exception of one item), and because the

sentential contexts for minimal pairs may have varied in difficulty, thus further complicating the picture, the data offer only limited clues with respect to this question. However, these limited results suggest possible trends. Although there was a tendency for groups to have more correct responses with the compound member, as we would expect from the response bias data on structural preference, chi-square analysis indicates that the differences in the number of correct responses between members of minimal pairs are not significantly greater than chance with any pair. When we consider the entire test group of 125 children as a whole, chi-square analysis again indicates that differences in the number of correct responses between members of minimal pairs are not significantly greater than chance for any pairs.

It is interesting that in Identification Test 1 a slight tendency existed for certain children to select one member of a minimal pair more often than the other. The minimal pairs most often selected in those cases were the compounds hot dog, blackboard, and highchair, and the phrases red socks, red head, and green house. It was suggested that the main difficulty may have been associating the phrases heard with the pictures on the answer sheets, or lack of familiarity with the items. The fact that the Preference Test indicated a slight structural preference and that it displayed no evidence of the biases seen in the Identification Test with members of minimal pairs suggests that the biases seen with a few children in the Identification Test were connected with the type of task. The children who displayed biases in the Identification Test probably chose the more familiar or appealing picture, or the one which was less confusing. In the Preference Test, where no pictures were used, these biases did not appear. Thus, the biases appearing with a small number of children in the Identification Test were most probably connected solely to selection of certain pictures rather than to certain structures or phrases, whereas those in the Preference Test were connected to certain structures rather than to certain pictures of items or letters.

Analysis of the Difficulty of Minimal Pairs

Since the Identification Test results showed that two minimal pairs were consistently incorrectly identified, a ranking of the difficulty of minimal pairs in the Preference Test results was compiled to see to what degree the rankings were correlated in the two types of tests. Table 3.6 lists these rankings for the Preference Test. It indicates that the most difficult pairs in the Preference Test were the blackboard-black board and the redhead-red head pairs. If we compare the two tables with the rankings for difficulty of pairs in Identification Test 1 and Identification Test 3 (found in Tables 2.17 and 2.18, respectively), we see almost no agreement in the results of these three tests with respect to this question. For example, in both of these Identification Tests the pair Red Sox-red socks

Table 3.6: Preference Test -- Ranking of Minimal Pairs by Grade - Most Difficult to Least Difficult*

Grade	Blackboard-Black Board	Redhead-Red Head	Highchair-High Chair	Greenhouse-Green House	Hot Dog-Hot Dog	Red Sox-Red Socks
1	2	1	3	4	6	5
2	1	2	4	3	6	5
3	1.5	1.5	3.5	5	6	3.5
4	2	5	3.5	3.5	1	6
5	2.5	1	4.5	2.5	6	4.5
6	1	2.5	2.5	4.5	4.5	6
7&8	1	2	4.5	4.5	4.5	4.5
Mean	1.6	2.1	3.6	3.9	4.9	4.9

* 1 = most difficult; 6 = least difficult

was the most difficult, yet in the Preference Test the fewest errors were made with this pair and it seemed the easiest. Also, the pair greenhouse-green house, which was the second most difficult in the Identification Tests, ranked fourth in difficulty in the Preference Test. The two pairs which were most difficult in the Preference Test were also found to be easier in the two Identification Tests, their rankings being third and fourth in one Identification Test and fourth and fifth in the other.

Why more errors were made on some pairs than on others in the Preference Test is not clear. As mentioned in the section "Response Biases: Type of Syntactic Structure," it may be that the sentential contexts for some pairs were ambiguous in terms of limiting what the correct answers should be. It is also possible that some items were more familiar than others. It is perhaps more likely however, that the rankings in this Preference Test are due to chance and would not be repeated in a replication test. The important point, however, is the lack of general agreement of the Preference Test results with the Identification Tests' findings that the pairs Red Sox-red socks and greenhouse-green house had the greatest number of errors and were consistently the most difficult of the six pairs. This lack of agreement gives support to the explanation that these differences in pairs in the Identification Test resulted from problems in picturing the items. Since the Preference Test used no pictures and thus had no such problems, these pairs were not consistently more difficult than the others in that test.

Reliability and Correlation with Other Tests

As we have seen, evidence of the general reliability of the Preference Test has been indicated throughout the discussion of results by its close correlation with the results of Identification Test 3. The similarity of results seen above in Figure 3.2 indicates that the Preference Test offered a reliable second method of eliciting knowledge concerning the stress patterns tested.

A more direct indication of reliability and consistency is the degree of correlation of each subject's performance on the two tests. Did a child who performed well on the Identification Test also perform well on the Preference Test, and vice versa; or was there little correlation in performance of the two different tasks? Table 3.7 indicates the Pearson Product Moment Correlation Coefficient r and its associated level of significance for each grade on Identification Test 3 and the Preference Test. As we would expect, the degree of correlation in Grade 1 is low because these children had not acquired the linguistic skill being tested and thus performed randomly. The later grades showed a high degree of positive correlation, indicating that ability to do well on one test is closely related to the ability to do well on the other test. That the correlation of Grade 3 was statistically not significant may be a reflection of the fact that fewer subjects were in this grade than in either of the grades immediately before and after it.

Table 3.7: Pearson Product Moment Correlation Coefficient r for Identification Test 3 and Preference Test

Grade	Number of Subjects	r	Level of Significance
1	(8)	0.188	ns
2	(27)	0.451	$p<0.02$
3	(16)	0.335	ns
4	(25)	0.757	$p<0.01$
5	(18)	0.566	$p<0.02$
6	(20)	0.481	$p<0.05$
7 & 8	(11)	0.767	$p<0.01$
TOTAL	(125)	0.635	$p<0.01$

The positive correlation is reasonably high and if there were a larger number of subjects continuing in the same direction, it would probably be significant. The correlation for the entire group of test subjects is also very high.

Another way of examining the question of consistency of individual subjects on both tests is to determine the percentage of subjects who scored at roughly the same level on both tests. If we divide subjects on each test into two groups -- those who performed at a level of significance of $p<0.02$

(10 correct responses of 12 items) and those who performed at a level that it not significantly better than chance -- we find that 81 per cent of all subjects were in the same category on both tests. Of this 81 per cent roughly two-thirds had significantly more correct than incorrect responses on both tests, and one-third had scores which were not greater than chance on either test. Table 3.8 gives the breakdown by grade of the percentage of subjects who scored at roughly the same level on both tests.

Table 3.8: Percentage of Subjects by Grade Who Either Did or Did Not Score a Number of Correct Versus Incorrect Responses That Was Significantly Greater than Chance on Both the Preference Test and Identification Test 3

Grade	%Ss scoring at the level $p<0.02$ on both tests	% Ss scoring at the level of $p<$ns on both tests
1	0	75
2	11	70
3	50	25
4	48	24
5	78	6
6	75	10
7 & 8	100	0

It is interesting to note that even in the higher grades a number of subjects did not perform significantly better than chance on either test.

We might also ask whether a tendency existed for children to do better on either the Identification Test or the Preference Test. Of the 19 per cent of the total group who displayed inconsistent performance on the two tests, 54 per cent did better on the Preference Test and 46 per cent did better on the Identification Test. Considering the number of errors made by each child, of the entire test group 35 per cent did better on the Preference Test, 32 per cent did better on the Identification Test, and 33 per cent had identical scores on both tests. However in Grade 4 there was a strong tendency to score better on the Preference Test. Of the 25 subjects in this grade, 14 did better on the Preference Test while only 2 did better on the Identification Test. (The other 9 did equally well on both.) Chi-square analysis shows this to be significantly greater than chance at the $p<0.02$ level. No other grade showed a significant tendency. These results suggest that a preference task is slightly less difficult than an identification task.

In order to determine whether there was a correlation between the specific lexical items missed by children on Identification Test 3 and the Preference Test, chi-square analysis (two by two contingency table) was

used to test for an association between the response on the first test and the response on the second for each lexical item. Because of the possibility that the responses of a given child are not statistically independent of one another, as is required by the chi-square test, subjects whose overall accuracy level was approximately the same were grouped together in order to correct for bias introduced, for example, by some subjects doing very well on all items and other subjects doing poorly on all items. While this is not a complete solution to the independence problem, it was the best that could be done with the data at hand. The results should be viewed with caution, however, because of this incompletely resolved difficulty.

In the groupings of subjects who scored 92 to 83 per cent correct (1 or 2 errors) on both tests, the chi-square value was not significant. The similar errors were slightly higher than expected and the unlike errors were lower than expected, however. For subjects who scored 75 to 67 per cent correct (3 or 4 errors) on both tests, again chi-square was not significant. The subjects who scored 58 to 50 per cent correct (5 to 6 errors) had a chi-square value significant at the 0.05 level. In this case, however, the errors were significantly unlike each other rather than significantly alike. There is no obvious reason for this opposite significance. Of course, it may just have been due to an unlikely chance-sampling fluctuation. But on the other hand, it may suggest that other factors are operating here which we cannot fully explain, and this should remind us to view these particular results with caution.

Chapter 4

Semantic/Acoustic Experiment with Compounds and Phrases

In the previous tests, the failure of the youngest children to differentiate correctly between compounds and phrases on the basis of their distinctive stress patterns and the relatively late acquisition of the linguistic ability to use stress patterns suggests that in a normal language situation, either these children are unable to make the proper semantic interpretation of these compounds and phrases because they do not yet know the syntactic constructions or lexical items involved; or they are able to disambiguate these compounds and phrases but do so on the basis of the semantic and syntactic information of the sentential context rather than the stress patterns. Since the children's reaction to the test stimuli indicated that they did know these lexical items, the latter possibility seems more plausible for a number of reasons.

In normal language use, even adults may not always produce the distinction between these pairs, owing to various performance factor limitations. Also, contrastive or emphatic stress can cause the normal stress pattern of a phrase (primary stress falling on the second element) to become that of a compound (primary stress appearing on the first element) -- e.g., "It was a lów chair, not a hígh chair." General studies in speech perception show that listeners normally do not depend only upon phonetic information to interpret sentences. The message is interpreted by using various kinds of linguistic knowledge -- semantic and syntactic as well as phonetic. In fact, the phonetic information is not necessarily the most important*. Generally we do not look for phonetic detail as much as for the over-all meaning of the sentence. Thus, if the phonetic distinction differentiating these minimal pairs were missing because of performance factors or contrastive stress shift, listeners would use information from the sentential context to disambiguate them.

Although adults as well as children use semantic and syntactic cues to disambiguate these minimal pair contrasts, the Identification Test showed that adults and older children have the ability to use only the acoustic cues when semantic and syntactic cues are not available. However, most younger children do not have the linguistic ability to differentiate semantically between these pairs on the basis of acoustic cues only; if they

*For examples of speech perception being influenced by a speaker's knowledge of the language as well as by the acoustic cues of the speech signal, see Lieberman (1965) and Garrett, Fodor, and Bever (1966). Selezneva (as reported in Slobin, 1966) reports an experiment in Russian indicating the effect of rules for stress placement on the perception of stress.

disambiguate them at all, they must use semantic and syntactic cues. This would mean that such young children may still correctly understand the meaning of sentences and could disambiguate these minimal pairs even though they failed to recognize and use the phonetic stress patterns associated with each member of the pair consistently.

If these assumptions are correct, when we construct a sentence in which the semantic and syntactic information of the sentence predicted the occurrence of one member of a minimal pair but in which the other member of that pair was actually pronounced (e.g.,"It was very warm yesterday, and when I went for a walk, I saw a hót dog"), there would be a difference in the possible responses of young children and of older children and adults. Adults and most older children could presumbably interpret the member of the minimal pair of such a sentence as either being the one which agreed with the sentential context or the one which was actually pronounced, since they have a well-established linguistic ability to use the acoustic cues as well as the semantic and syntactic cues of the sentence. Their responses would thus be determined by the instructions given. On the other hand, young children who do not have the linguistic ability to use acoustic cues would always respond with the member of the pair which agreed with the sentential context, irrespective of how it was pronounced. Children who fall between these two extremes, being in the process of acquiring the linguistic ability to distinguish these pairs by phonetic stress pattern but not having completely mastered that skill, might respond with some confusion to the sentence, since they would find the two types of cue in disagreement. Without complete confidence or ability to distinguish all pairs on the basis of acoustic cues alone,they might give somewhat random responses; i.e., they might answer according to acoustic cues in some cases and to semantic and syntactic cues in others.

The following experiment, which will be referred to as the Semantic/Acoustic Test, was conducted to test these assumptions and predictions. The acoustic cues signaling the phonetic stress patterns of the minimal pairs to be tested contradict the semantic interpretation which would normally be made on the basis of the semantic and syntactic information of the sentential context in which these pairs appear. (Children whose responses to these minimal-pair test items were based on semantic and syntactic information will be said to base their responses on semantic cues, and children whose responses were based only on the phonetic stress patterns of the pairs will be said to base their responses on acoustic cues.)

Method

Subjects

The same 125 children and 19 adults who served as subjects for Identification Test 3 and the Preference Test were used as subjects in this test.

(See Table 2.3, Chapter 2). Also as with the other tests, this one was administered to an additional group of children to allow for comparison of comprehension and production skills (see Chapter 5).

Stimuli and Their Presentation

The same six minimal pairs of compounds and phrases, differentiated only with respect to stress, were used as stimuli in the Semantic/Acoustic Test. (See Table 2.5 for a listing of these pairs.) In this test these contrasting pairs were embedded in sentences which were constructed so that the semantic cues of the sentential context opposed the acoustic cues of the embedded member of the minimal pair. That is, the semantic interpretation suggested by the sentential context would indicate that one member of the minimal pair was correct, but the other member of the pair would be pronounced instead. These sentences are listed in Table 4.1.

Table 4.1: Sentence Stimuli Used in Semantic/Acoustic Test

1. It was very warm yesterday, and when I went for a walk, I saw a hót dog.
2. Sally wanted to practice writing her name, so she looked for a black boárd.
3. We didn't want the flowers to freeze, so we put them in the green hóuse.
4. Before I put on my shoes, I saw the Réd Sox.
5. When Stevie was in the park yesterday, he saw a blond and he saw a red heád.
6. The baby was hungry, so his mother put him in the high cháir.
7. The men were building a house and they needed a bláckboard.
8. Billy was very hungry; he wished he could find a hot dóg.
9. The baby spent the whole day in the sun without his hat. Then the baby looked in the mirror and saw a rédhead.
10. Father finished painting yesterday. Now we have a gréenhouse.
11. It rained so hard yesterday that we couldn't see the red sócks.
12. The king sat on the top of the steps in his híghchair.

These test sentences were constructed in such a way that they would not be totally ambiguous, with one member of the pair as likely to appear as the other (e.g., It's a $\left\{\begin{array}{l}\text{blackboard}\\ \text{black board}\end{array}\right\}$.)

Such a test sentence would only elicit an arbitrary decision or perhaps some kind of preference for one member of the pair over the other (the basis of which would be difficult to determine), and it would not test the claims we are concerned with here. On the other hand, these sentences were not constructed so that only one semantic interpretation was possible. Although the sentence frame would signal one interpretation with one member of the pair as the most likely on the basis of semantic and syntactic information, a second possible interpretation when the other member of the pair was pronounced was also possible. This second sentence reading was usually somewhat unlikely, but it was nonetheless feasible. For example, the compound which is pronounced in the sentence, "The king sat on the top of the stairs in his <u>high</u>chair" gives a somewhat humorous reading, but one that is possible if the king were only two years old. This is important because the sometimes impoverished nature of the speech signal in normal speech requires even adult listeners to depend heavily on semantic and syntactic cues in speech perception. If a second semantic reading with the member of the minimal pair which had actually been pronounced were impossible, then even adults would likely ignore the acoustic cues or think that they had not heard the test item correctly.

This test is similar to the Preference Test (see Chapter 3) in that both use stress patterns which contradict the context of the sentences in which they appear. However, in the Preference Test, both the right and wrong patterns were given; the task was to select between the two. In this test the correct stress pattern was not given, and the task was to identify the compound or phrase from the stress pattern when it was in conflict with the context of the sentence in which it is embedded. In the Preference Test subjects were being tested to determine whether they could select the correctly stressed compound or phrase within a sentential context. In this test, on the other hand, we are interested in determining what cues are most important in making the selection -- semantic or acoustic.

A total of 14 test sentences were used in the Semantic/Acoustic Test. The order of the pairs in this test matched the order of the pairs in both the Identification and Preference Tests. The sentences listed in Table 4.1 presented the 12 members of the minimal pair test items, each member of every pair being presented once. The first and second test sentences were presented a second time after all 12 test sentences had been given, bringing the total number of test sentences to 14. The main reason for repeating these sentences was to allow for a comparison in performance at the beginning and end of the test. Such a comparison seemed particularly important in this Semantic/Acoustic Test since it is possible that a subject may become aware of the purpose of the test after hearing a number of the test items and then change his behavior accordingly when the same sentences are repeated. Only two sentences were repeated at the end of the test, mainly because of time considerations. However, it was felt that those two sentences would be sufficient to gain a general idea concerning subjects "catching

on" and changing their behavior.

As with the Identification and Preference Tests, a stimulus tape was recorded of the experimenter producing the test sentences; the recording procedures and apparatus described in Chapter 2 were employed. Before this stimulus tape was played, the experimenter explained the task to the children. Since the children had completed Identification Test 3 immediately before this Semantic/Acoustic Test was given, it was not necessary to explain and describe the six minimal pairs again in detail, but the experimenter did review what the pairs were by again pronouncing each member of the minimal pair while showing the picture of that item. The children were then told that they would again hear the words they had just heard in the Identification Test, but that this time they would hear them used in sentences. Each sentence was pronounced only once before children were asked to make judgments. The same answer sheets with pictures of the minimal pairs which children had used for recording their responses in the Identification Test were used in this test. Again the order of the pictures on the answer sheets corresponded to the order of stimuli on the test tape. Children were then told in simple terms to listen to the test items which were presented in the sentences, to decide which member of the minimal pair was being pronounced in each sentence, and to mark the picture on the answer sheet which corresponded to the item which was pronounced. The experimenter emphasized several times at the beginning of the test as well as at various times during the test itself that children should always mark the picture of the word which was pronounced, even if the sentence itself sounded strange or funny to them. Since this instruction was crucial to the experiment, the point was made several times to insure that all the children understood the directions. The experimenter gave children an opportunity to ask questions at this point in order to again make certain that there was no confusion concerning the task. Two example sentences -- Nos. 3 and 10 of the test sentences -- were then played for the children. Because actual test sentences were played as examples and because of the nature of the task, the correct answers were not given to these two sentences, a feature which differed from the other tests given. The purpose in giving these examples was to accustom the children to hearing the kind of test stimuli which the Semantic/Acoustic Test involved, to give them a feeling for how long the time interval was between test sentences so that they would know how long they had to make judgments and mark their responses on their answer sheets, and to ensure that the instructions were clearly understood. The stimulus tape was played to the children with only an occasional interruption by the experimenter reminding them always to "Choose the word which you heard me say, even if the sentence sounds strange or funny."

The children were again tested in small groups, as were those who took Identification Test 3 and the Preference Test, since all these tests were administered during the same session.

Results

The Semantic/Acoustic Test was perhaps the most enjoyable and satisfying for the children of all the tests given. The youngest felt it was easy and were confident that they were making the correct choices, a feeling they did not have with other tests in which decisions were more difficult for them.

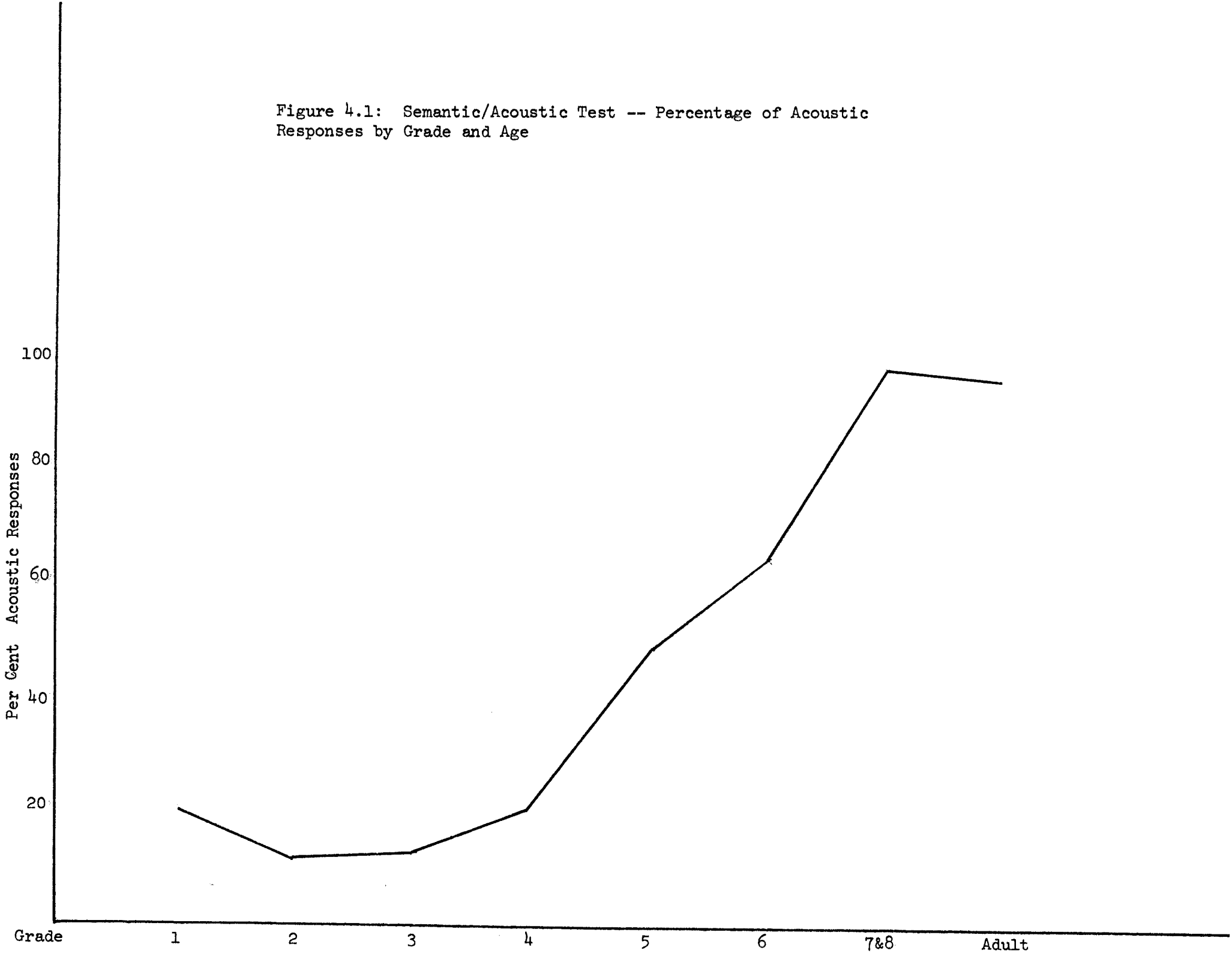
General Results

The data from the Semantic/Acoustic Test confirm the assumptions and predictions made on the basis of the Identification Test concerning developmental differences in the ability to attend to the acoustic cues distinguishing these compounds from phrases in the presence of opposing sentential semantic cues. Figure 4.1 indicates the percentage of acoustic responses plotted against age and grade. There was a general tendency for the ability to base responses on the acoustic cues signaling the phonetic stress patterns of these minimal pairs to increase with age.* In Grades 1 through 5 the percentage of acoustic responses is below 50. The largest jump occurred between Grades 6 and 7 and 8, where the percentage of acoustic responses rose from 63 to 97. Chi-square analysis indicates that the only grade where the number of acoustic responses was not significantly different from chance was Grade 5, with 48 per cent. Grade 6 scored acoustic responses significant at $p<0.01$ and all other grades had scores which are significant at the $p<0.001$ level. Unlike Grades 6 through 8, the significance of the scores in Grades 1 through 4 indicates a very low number of acoustic responses and thus a highly significant number of semantic-type responses. Thus, the predicted tendency for the youngest children to use semantic cues, the oldest to use acoustic cues, and those in between to use both and thus score more randomly is apparent in this figure.

If we compare Figure 4.1 with Figure 3.2,which shows the percentage of correct responses by the same test subjects in Identification Test 3 and the Preference Test, we find excellent agreement as to when full proficiency was reached in all three tests. Acoustic responses reached 97 per cent in Grade 7 and 8 in the Semantic/Acoustic Test. It was also not until these grades that the percentage of correct responses reached over 95 per cent in Identification Test 3 and the Preference Test. On the other hand, in the Semantic/Acoustic Test the percentage of acoustic responses in Grades 2 through 5 and in Grade 6 to some extent is lower than would be expected on the basis of performance in these other tests. This suggests that only when children have sufficient ability with stress patterns to obtain a near perfect score in identification and preference tasks are they able to consistently ignore the opposing sentential contexts and attend to the stress

*The only exceptions to this developmental trend were in Grade 1 and the Adult Group. Grade 1's performance appeared to be relatively higher because of the smaller number of subjects in that group. The dip with the adult group is also not significant.

Figure 4.1: Semantic/Acoustic Test -- Percentage of Acoustic Responses by Grade and Age



patterns. The data from Identification Test 3 and the Preference Test thus verify that the lack of ability with these stress contrasts is the crucial factor in young children's finding it difficult to ignore the semantic and syntactic information of the sentential context.

Frequency Distribution and Individual Variation

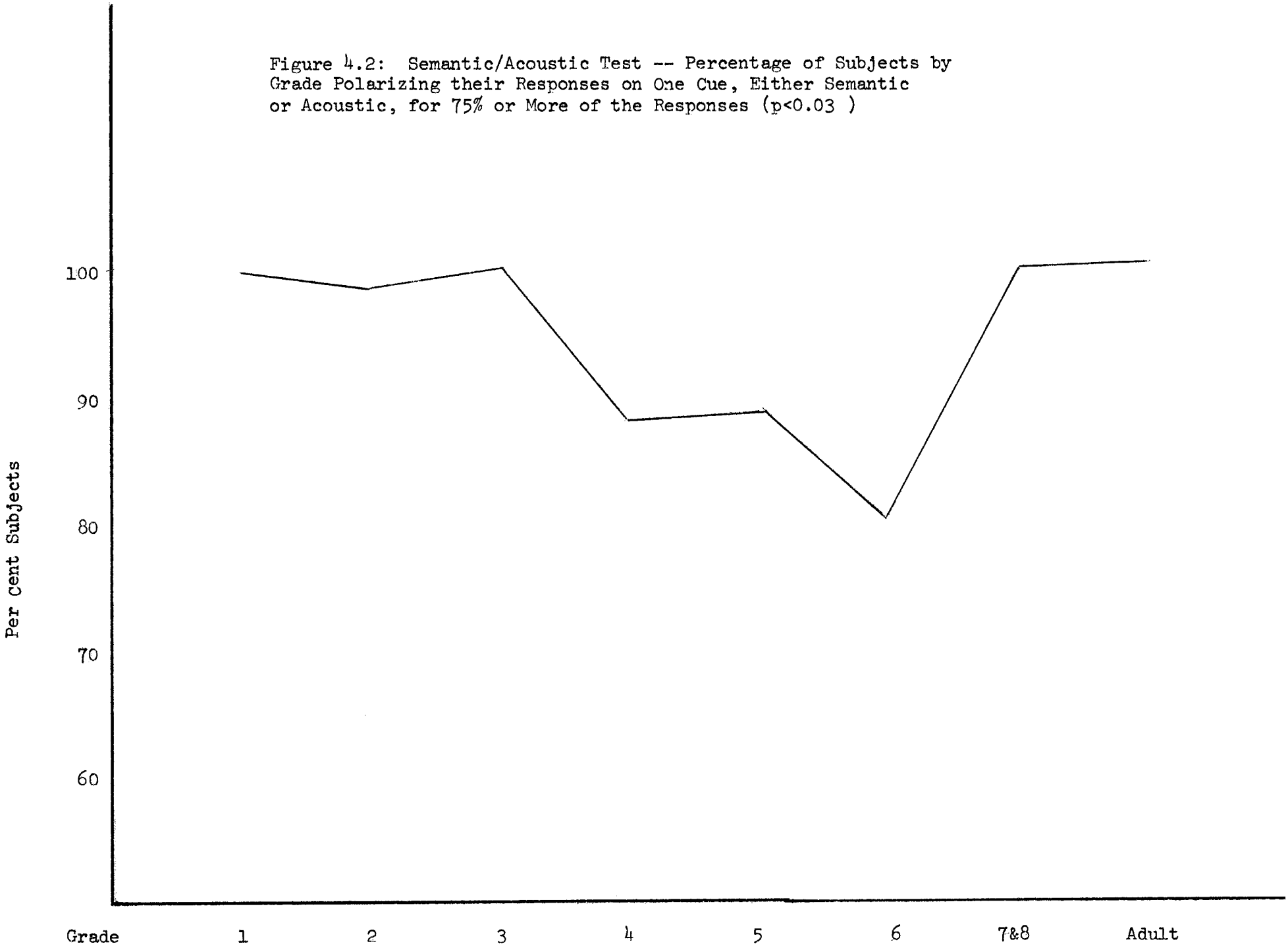
In order to determine whether the group averages of acoustic responses seen in Figure 4.1 resulted from the majority of the children having scores which approximated the group mean, or whether these averages resulted from some children basing the majority of their responses on one type of cue and other children basing them on the other, the frequency distribution seen in Table 4.2 was compiled. It indicates that unlike the distribution in other tests, the distribution in this Semantic/Acoustic Test was not normal and symmetrical. Instead, there is an "all or none" tendency; the majority of the subjects in every grade tended to base almost all their responses on one type of cue.

Table 4.2: Semantic/Acoustic Test -- Frequency Distribution of Subjects and their Response Types

Responses		#Ss Scoring at Each Response Type Interval							
#Acoustic	#Semantic	Grade 1	2	3	4	5	6	7&8	Adult
14	0			1	1	4	7	8	17
13	1				1	3	3	2	1
12	2							1	
11	3	1	1		1				
10	4		1		1		2		
9	5						2		
8	6				2	2			
7	7								
6	8								
5	9								
4	10								
3	11	2				1			
2	12	2	5	2		1			1
1	13	1	10	8	6	5	2		
0	14	2	10	5	13	2	4		

The statistical significance of these trends is indicated by Figure 4.2, which shows the percentage of children by grade who displayed this polarization tendency with 75 per cent or more of the total number of test stimuli, the probability of this high degree of polarization occurring by chance being $p<0.03$, as was determined by the binominal test. In all grades except Grades

Figure 4.2: Semantic/Acoustic Test -- Percentage of Subjects by Grade Polarizing their Responses on One Cue, Either Semantic or Acoustic, for 75% or More of the Responses ($p<0.03$)



4 through 6 close to 100 per cent of the subjects polarized their responses on a significant number of test items, this pattern being followed by exactly 100 per cent of the subjects in Grades 1, 3, 7 and 8 and the adult group, and by 97 per cent of the subjects (26 of 27 children) in Grade 2. Thus the subjects in these grades were apparently very confident and consistent in making their responses in this test. A smaller percentage of subjects in Grades 4 through 6 displayed this consistency and confidence, but still over 80 per cent of them polarized their responses on a significant number of items. These percentages in all grades again bear out the predictions concerning the confident use of one cue in the youngest and oldest age groups and the more random use of one or both cues in the intermediate age groups.

The prediction concerning the cues on which each age group would base their responses was verified, as indicated in Figure 4.3. It shows the percentage of subjects whose responses in over 75 per cent of the sentence stimuli (where $p<0.03$) indicated attendance to either the sentential context or the acoustic cues which signaled the phonetic stress pattern of the minimal pairs. Although there is a very small amount of individual variation, an extremely high percentage of the children in Grades 1 through 3 polarized on semantic cues and ignored the acoustic cues; on the other hand, subjects in Grade 7 & 8 and adults were able to follow instructions and attend to acoustic cues. Grade 4 had a higher percentage of children scoring a significant number of responses on the basis of semantic cues rather than acoustic cues, but the difference was less marked with this group; also there was more individual variation here than in younger groups. The crossover in importance of cues occurred between Grades 5 and 6, the age period where fewer subjects polarized their responses on a significant number of test items. It thus again appears that most children in these intermediate groups attended to acoustic as well as semantic cues, but they were not as consistent in basing their responses on one or the other cue as were the younger and older subjects.

Differences According to Sex

The data were analyzed to determine whether there were any significant differences at any age according to sex. Figure 4.4 gives the percentage of acoustic responses of boys and girls plotted against grade. Again, since no girls were in the Grade 1 group the comparison is incomplete, but in all other Grades except 2 and 7 & 8 the girls tended to give more acoustic responses than the boys. The Mann-Whitney U test indicates that this difference in performance between boys and girls was significantly greater than chance only in Grade 5 where $\alpha<0.05$. Large jumps in percentages between grades occurred with both boys' and girls' groups, but they occurred between different grades. The biggest jump in performance for girls was between Grades 4 and 5, and for boys between Grades 6 and 7 & 8.

Figure 4.3: Semantic/Acoustic Test -- Percentage of Subjects Scoring 75% or More Semantic Responses ($p<0.03$) or 75% or More Acoustic Responses ($p<0.03$)

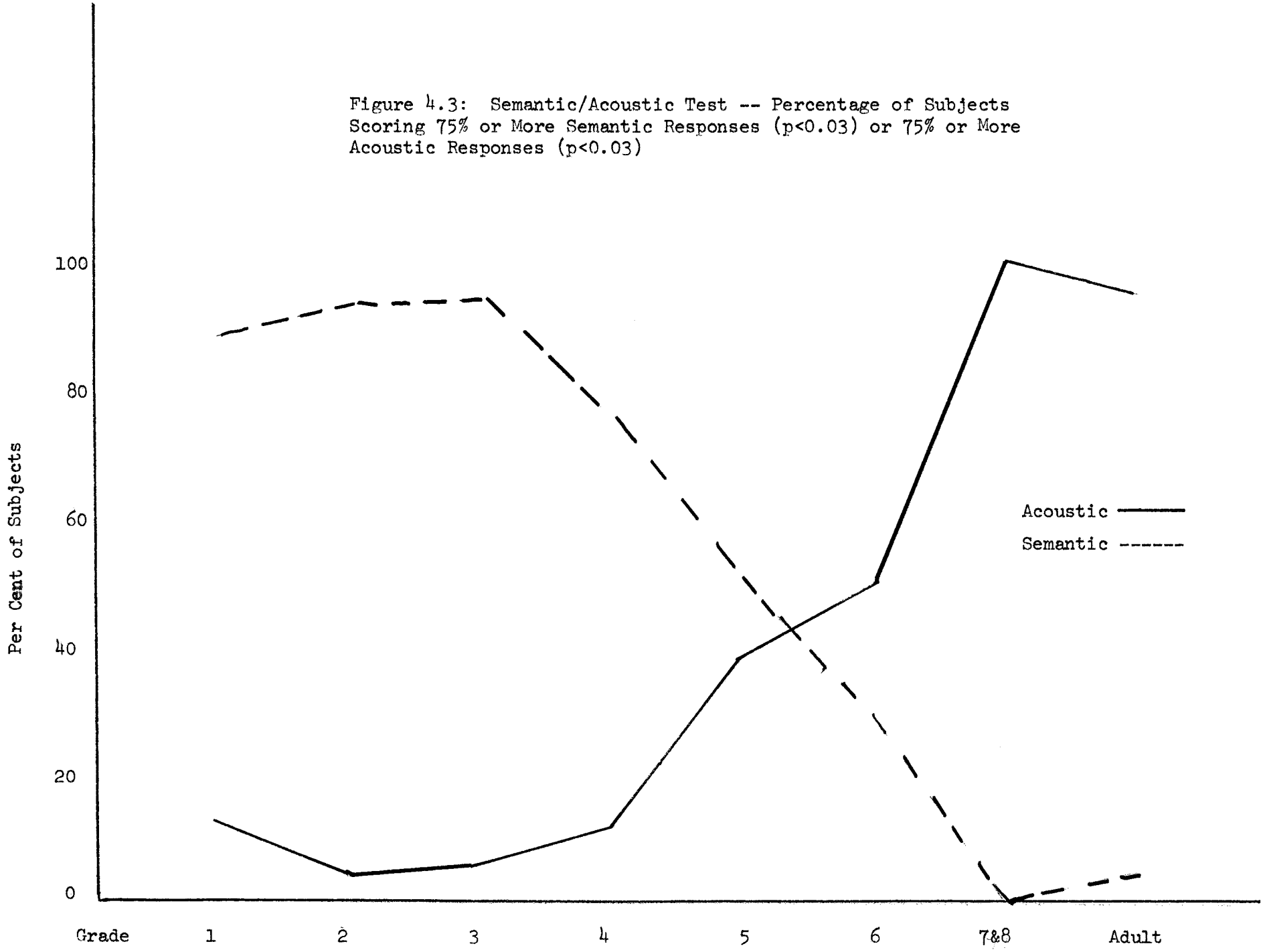


Figure 4.4: Semantic/Acoustic Test -- Percentage of Acoustic Responses Plotted Against Grade According to Sex

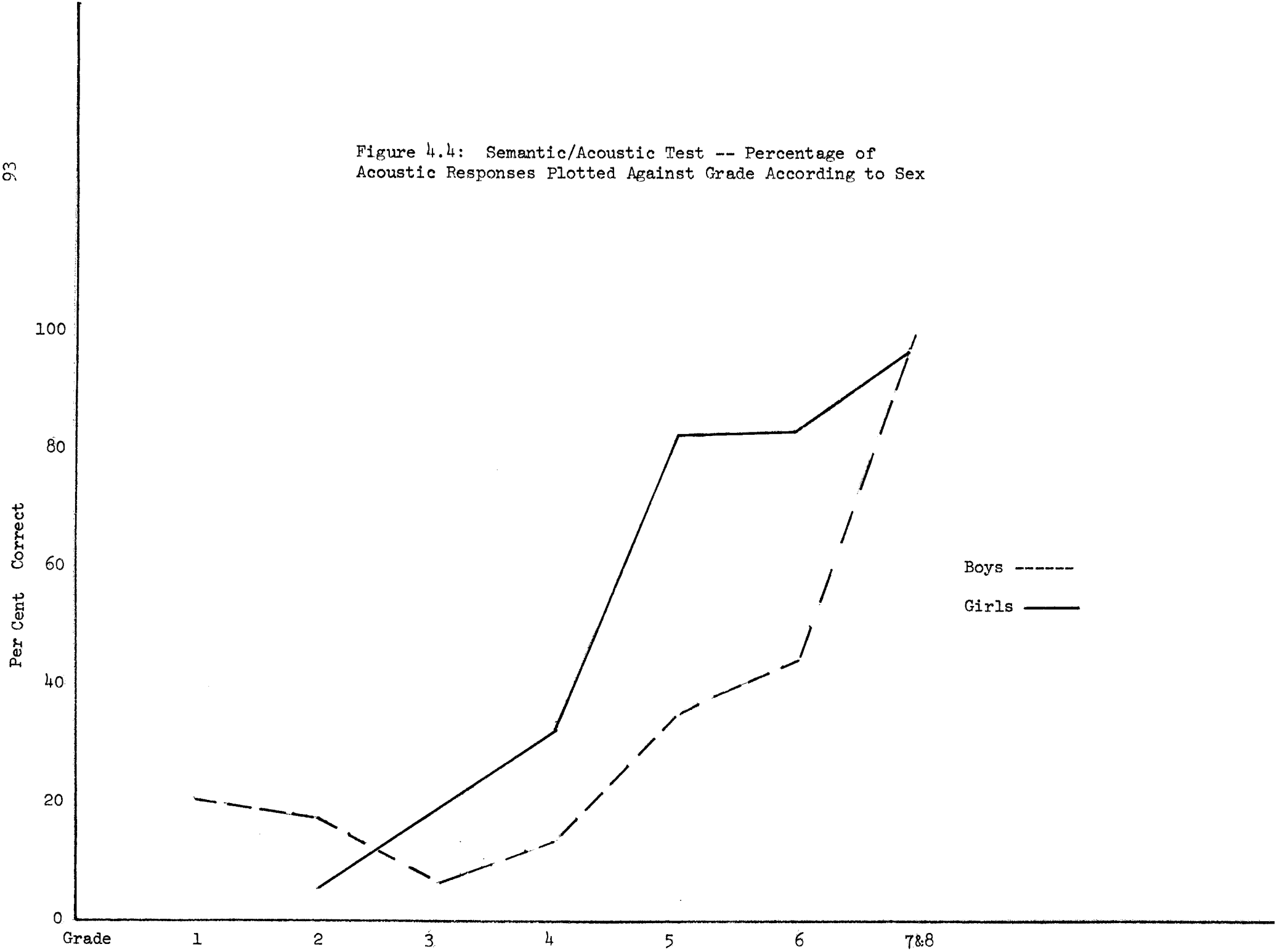
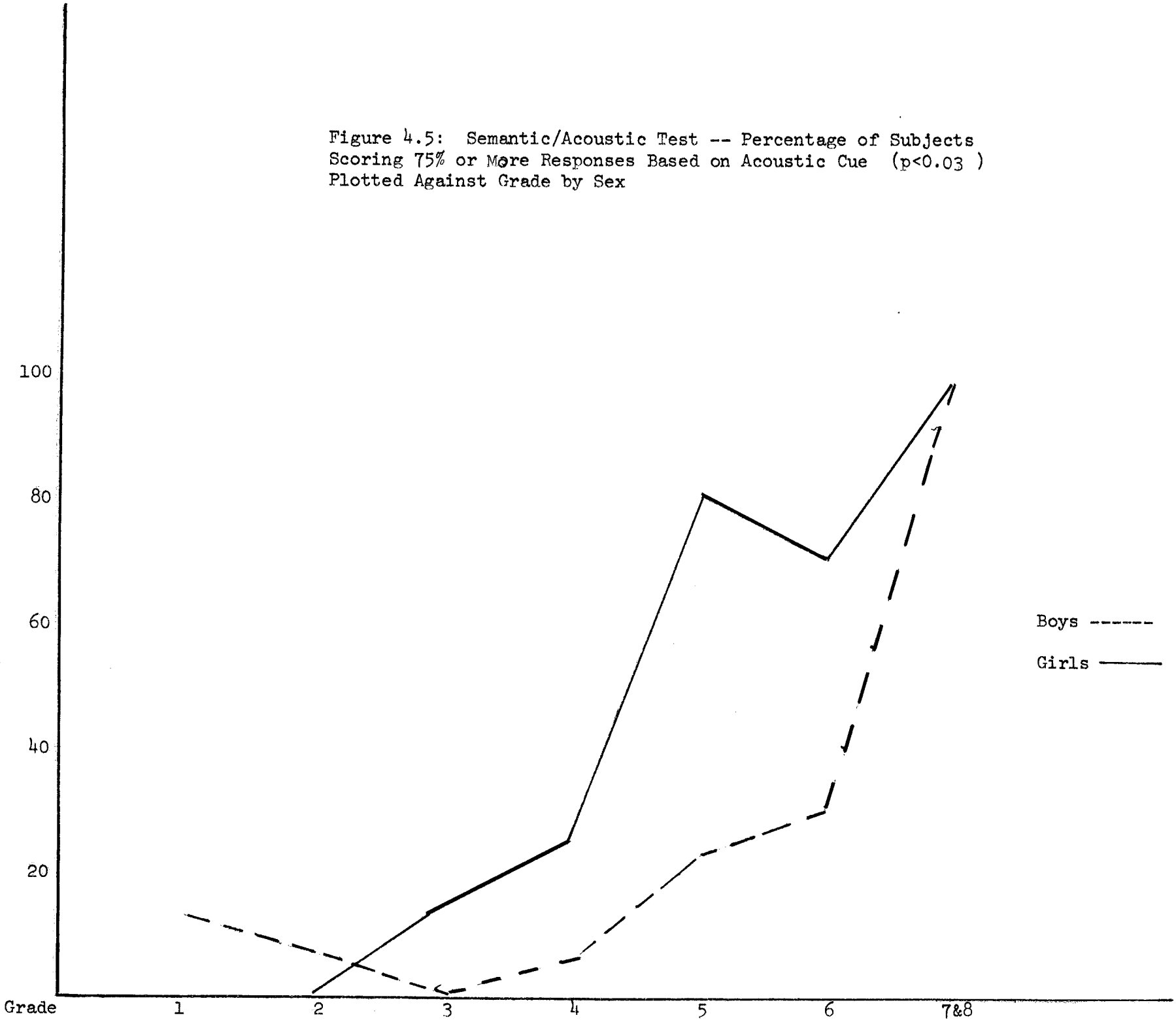


Figure 4.5: Semantic/Acoustic Test -- Percentage of Subjects Scoring 75% or More Responses Based on Acoustic Cue ($p<0.03$) Plotted Against Grade by Sex



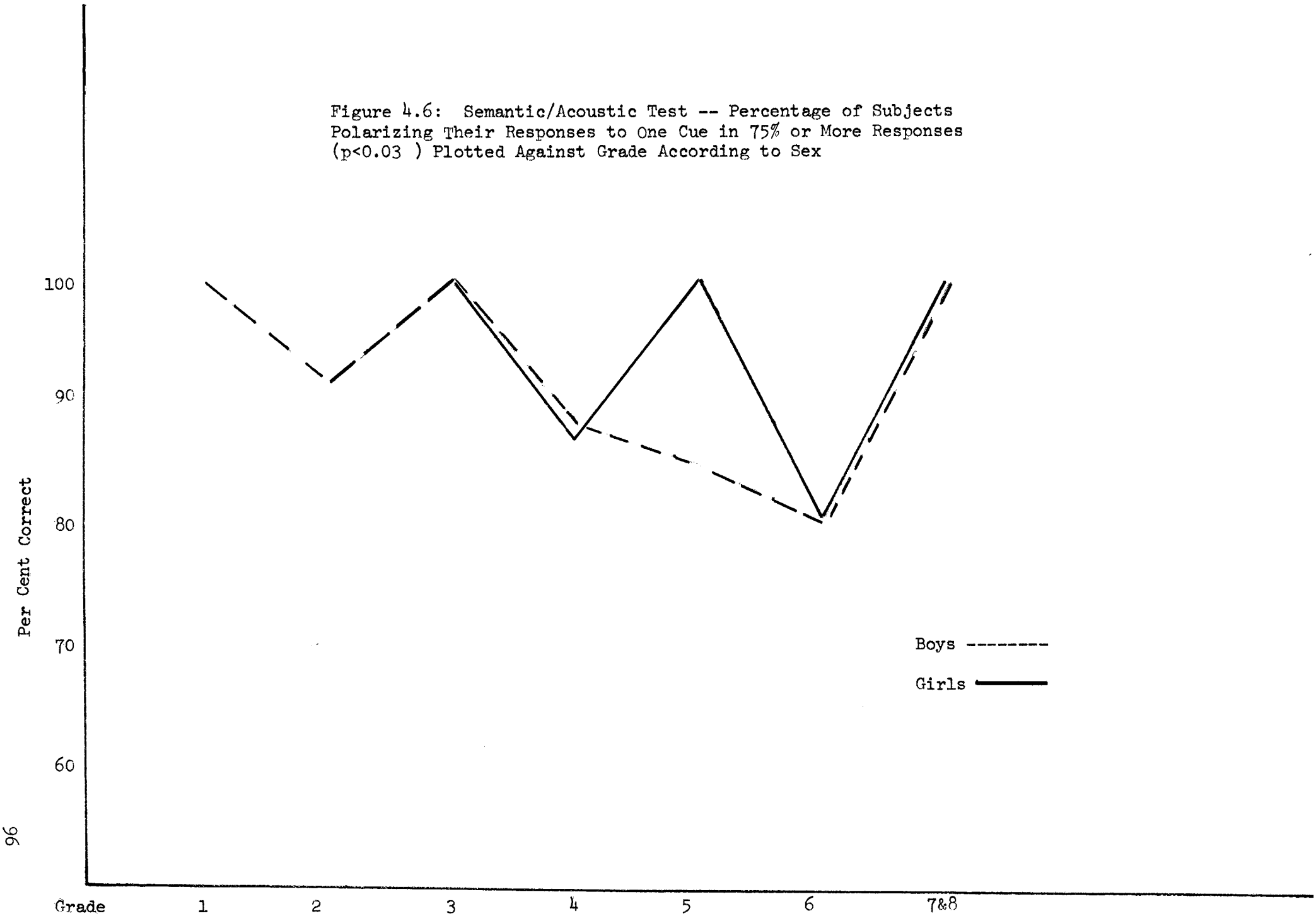
We see the same general tendency for girls to give more acoustic responses in Figure 4.5, which shows the percentage of subjects basing 75 per cent or more of their responses on acoustic cues ($p<0.03$). More girls used acoustic cues for a significant number of their responses than did boys in all grades except Grade 2. Again, the most marked difference between the performance of boys and girls occurred at Grades 5 and 6, two of the three grades in which there was the least amount of polarization to one or the other cue. The fact that a large difference occurred between the performance of boys and girls at those ages could indicate that sex differences are a factor in this polarization tendency at those ages. Figure 4.6, which displays the percentage of subjects who polarized to one cue in a significant number of their responses, shows that there are some differences in the performance of boys and girls in Grade 5, but not in Grades 4 or 6. (The only other difference in this figure is seen in Grade 2, but there it is due to the smaller number of subjects and was accounted for by only one child.)

The tendency of girls in Grades 3 through 6 to perform better than boys is somewhat stronger in this test than in the previous ones. Girls apparently have more confidence or ability to attend to acoustic cues and ignore sentential context earlier than boys. Boys at these ages may have acquired the linguistic skill to use these stress patterns at approximately the same time, but they may be less adept at using this skill, and hence are more reluctant to ignore sentential context. The correlation data given below in Table 4.4 indicate that a certain percentage of subjects who gave a significant number of correct responses on either the Identification or Preference Test or on both still gave a significant number of responses based on sentential context in the Semantic/Acoustic Test. Looking at these data with regard to sex, we find that the majority of these subjects who had the linguistic knowledge to use acoustic cues but nonetheless used semantic information were boys. Thus, these differences according to sex seem to play a role in this type of task.

Learning Effects

The data were analyzed to determine whether learning had occurred during the administration of the test, or whether the opposite effect of fatigue or lack of motivation played an important role. The main evidence examined was the number of acoustic cue responses on the first and second trials of the two sentences played at the beginning and at the end of the test. To supplement this, a comparison was made between the number of children who gave more acoustic responses on the first half of the test and those who did so on the second half. This comparison had limitations since the two parts of the test did not present identical sentence stimuli, but it could be indicative of learning effects.

Figure 4.6: Semantic/Acoustic Test -- Percentage of Subjects Polarizing Their Responses to One Cue in 75% or More Responses ($p<0.03$) Plotted Against Grade According to Sex



Chi-square analysis indicated that on the basis of these two criteria there were no significant learning or fatigue effects for any grade, for Grades 4 through 6 together, or for the test group as a whole. The number of children who gave more acoustic responses for the two repeated sentences was slightly higher on the second trial than on the first; conversely, the number of children who gave more acoustic responses on the second half of the test was slightly lower than on the first. In both cases, however, the differences were insignificant.

The same two criteria were used with the scores of individual subjects to see if any individual displayed evidence of learning during the test. Only two of the 125 children and 19 adults consistently showed that learning may have taken place. One Fifth Grader used semantic cues to respond to the first six sentence stimuli, and one adult used semantic cues for the first twelve sentence stimuli and then switched to acoustic cues for the last two stimuli. All other subjects in the test group who fluctuated between dependence on semantic and acoustic cues with any stimuli did so at random in the test, and thus displayed no evidence of learning or fatigue.

6. Response Biases

Since 93 per cent of the total test group of 125 children and 19 adults based a significant number of their responses on only one cue, response biases could not have affected a large number of subjects. However, because (with the exception of one Second Grader) the remaining 7 per cent who did not show this significant polarization were in Grades 4 through 6, some response biases could have affected children at these ages, since they responded with some confusion to both acoustic and semantic cues.

The data were thus analyzed to determine whether these three grades displayed biases with respect to positioning of pictures on the answer sheet. More responses were made with the picture of the left member of the pair than the right. However, chi-square analysis indicates this difference is not significantly different from chance.

The possibility of response biases according to syntactic structure were also considered with Grades 4 through 6. Although slightly more compound answers were given than phrasal answers, chi-square analysis indicated that the difference was not significant. Differences in difficulty of members of minimal pairs or greater difficulty of some pairs themselves in relation to other pairs were examined as another possible response bias. But there was no evidence of significant bias occurring in either case. Thus it appears that with Grades 4 through 6 the only response bias was that regarding positioning of pictures on the answer sheets.

Reliability and Test Correlation

Since the Semantic/Acoustic Test was not given a second time to these subjects, it is not possible to measure the reliability of the test and the subjects by comparing such performances. However, from the data discussed in connection with learning effects and the general results, it is apparent that most subjects' behavior was consistent throughout the test. Generally, if a subject attended to only one cue at the beginning of the test, he continued to attend to that cue throughout the test. Likewise, if he performed more randomly and fluctuated between the two cues at the beginning of the test, he continued to fluctuate at random for the remainder of the test. Also, responses on the two repeated sentences indicated consistent behavior at the beginning and end of the test.

The question of consistency in performance of individual subjects on Identification Test 3, the Preference Test, and the Semantic/Acoustic Test was also considered. A comparison was made of performance on the three tests in terms of the level of significance scored by each individual on each test, and the results of this comparison are summarized below. In Table 4.3 scores of individual subjects on the Semantic/Acoustic Test are divided into groups according to whether subjects used acoustic cues, semantic cues, or no one cue in responding to over 75 per cent of the stimuli ($p<0.03$ by the binominal test). The scores of subjects in each group on Identification Test 3 and the Preference Test are then categorized as to whether they were significantly different from chance on both, one, or neither of these tests.

Table 4.3: Breakdown of Subjects in Semantic/Acoustic Test in Terms of Significance and Performance of These Groups on Identification Test 3 and Preference Test

Performance of Subjects on Semantic/Acoustic Test: Of the Ss scoring:	Corresponding Scores of These Ss on Other Tests -% Ss scoring in each division:		
	Both Tests $p<0.03$	One Test $p<0.03$	Neither Test $p<0.03$
Acoustic Cue $p<0.03$	96%	2%	2%
Semantic Cue $p<0.03$	32%	23%	45%
Neither Cue $p<0.03$	60%	40%	0%

These data indicate that subjects who significantly based their responses on acoustic cues also performed well on the other two tests; 96 per cent of these subjects scored significantly on both other tests and another subject (2 per cent) scored significantly on one other test. Only one subject (2 per cent) who based responses on acoustic cues did not score significantly on either additional test. With the exception of this one subject, then, the results

are as expected. Subjects who had sufficient linguistic knowledge of these stress patterns to use acoustic cues in the Semantic/Acoustic Test also had the knowledge to perform well on the other tests.

This table also indicates that, taken as a group, the subjects who did not use one cue in the Semantic/Acoustic Test for a significant number of responses performed fairly well on the other tests, 60 per cent of them scoring significantly on both tests and 40 per cent scoring significantly on one. None of these subjects displayed a significant lack of knowledge of these stress patterns in either additional test. These subjects, with the exception of one child, were in Grades 4 through 6, and were thus in the stage of mastering their linguistic ability with these stress patterns; but they did not have complete confidence or competence with these patterns, as is evidenced by their generally good but not perfect performance on the other two tests. The fact that they did not utilize acoustic cues in all cases on the Semantic/Acoustic Test thus correlates with these other facts.*

As Table 4.3 also indicates, the subjects who based their responses on semantic cues were less predictable in terms of behavior on the Identification Test and the Preference Test than were those basing responses on acoustic cues or on both. Some of them obviously had enough linguistic skill with these stress patterns to score significantly on one or both of these other tests, but for some reason they chose to disregard this knowledge in favor of the perhaps more persuasive semantic information. However, a very large percentage of them had scores that were not significantly different from chance on the other tests, which explains their dependence on semantic cues.

Another way of comparing the performance of individual subjects on the three tests is seen in Table 4.4. This table divides subjects into groups according to their performance on Identification Test 3 and the Preference Test in terms of whether or not scores were significantly greater than chance on one or both tests. It then considers the performance of these various groups of subjects on the Semantic/Acoustic Test, listing whether they used acoustic cues, semantic cues, or no one cue in over 75 per cent of their responses ($p<0.03$).

*In all scores of the children using both acoustic and semantic cues, slightly more responses were based on acoustic than semantic cues.

Table 4.4: Breakdown of Subjects in Identification Test 3 and Preference Test in Terms of Significance and Performance of These Groups on Semantic/Acoustic Test

Performance of Ss on Test 3 & Preference Tests: % of Ss Scoring:		Corresponding Scores of These Ss on Semantic/Acoustic Test: % Ss Scoring in Each Category:		
		Acoustic Cue p<0.03	Semantic Cue p<0.03	Neither Cue p<0.03
Both Tests	p<0.02	63%	33%	5%
One Test	p<0.02	4%	73%	23%
Neither Test	p<0.02	3%	97%	0%

These results show that 97 per cent of the subjects who failed to score significantly more correct than incorrect responses on the Identification and Preference Tests (that is, every subject except one) based a significant number of their responses in the Semantic/Acoustic Test on semantic cues. This again verifies the prediction that subjects who lack linguistic ability with stress patterns will not perform randomly, but will use information from the sentential context to disambiguate these minimal pairs.

On the other hand, subjects performing significantly on both the Identification and Preference Tests showed a greater tendency to use acoustic cues than semantic ones, the figures being 63 per cent and 33 per cent, respectively. These subjects seemed to have possessed the necessary linguistic skill to use acoustic cues, and the majority of them did so, but for some reason in the case of 33 per cent of them they apparently chose to respond to the information of the sentential context.

Subjects who scored significantly on only one of the tests tended to use semantic cues more often than acoustic, and a fairly sizable percentage of these subjects vacillated between the two types of cue. This is again as one would expect; those subjects who had the linguistic ability to score a significant number of responses on one test but not the ability to do so consistently also had the linguistic knowledge to attend to acoustic cues but not to do so in all cases. Therefore they fluctuated between semantic and acoustic cues, or depended on semantic cues entirely. Subjects' behavior was so consistent in most cases that it is possible to make certain generalizations and predictions concerning performance on the Identification and Preference Tests on the basis of the Semantic/Acoustic Test, and vice versa.

The Effect of Test Instructions

In this experiment the older children and adults had the ability to ignore semantic cues and attend to acoustic ones, and did so in response to the experimenter's instructions. As noted earlier, in normal language use listeners generally look for the over-all meaning of a sentence rather than

attend to phonetic detail. Thus if listeners heard these test sentences in a nontesting situation where they were not instructed to attend to acoustic cues, they would probably attend to the semantic ones. Although this possibility was not studied, the effect of the test instructions on subjects' responses was examined. That is, the question of what cues an adult would listen to in these sentences if no instructions were given to slant the responses in the direction of the acoustic cues was considered.

An additional 10 adults were tested who were not part of the original testing group.* All 10 were students residing in the Boston area. Their ages ranged from 20 to 30. Two were tested individually and the remaining 8 were tested in two groups of 4.

The stimulus tape, answer sheets, and all testing procedures were the same as in the main Semantic/Acoustic Test with the exception of the instructions. Instead of telling subjects to answer according to the way the test item was pronounced, they were told only to mark the compound or phrase they thought correct. The tape was then played without interruption.

Results indicated that just as in the main Semantic/Acoustic Test these subjects tended to select either one cue or the other and to use that cue consistently. Eight of the 10 polarized their responses to one cue for over 75 per cent of the stimuli ($p<0.03$ by the binomial test). Of these eight, five attended to acoustic cues and three used semantic cues. The two subjects who did not polarize their responses for over 75 percent of the stimuli attended more to acoustic than to semantic cues. They reported finding the task ambiguous and somewhat baffling. All other subjects except one also reported that they realized the conflict and consciously selected one or the other type of response. Hence, even though these test sentences were heard under testing conditions rather than in a normal speech situation (where there would be an even stronger tendency to ignore the conflicting stress pattern in favor of the over-all sentential meaning) there were a number of adults who nevertheless attended to semantic cues.

*Six of these subjects were students at M.I.T. and were tested with the aid of Sam Pestin of the M.I.T. Psychology Department.

Chapter 5

Production Experiments with Compounds and Phrases

The experiments in Chapters 2, 3, and 4 have dealt with the comprehension aspect of children's ability with stress in compounds and phrases. To more fully understand the acquisition of these stress contrasts, the ability to produce them was also examined.

Various aspects of production can be studied, one of which is the ability to produce the test utterances spontaneously in natural speech without the aid of test pictures or elicitation procedures. However, this can only be studied under intensive observational conditions. Such a study would involve the investigator's keeping complete records of all a child's utterances, a tedious and time-consuming, if not impossible, feat. Therefore this type of testing was not used for purposes of the present study, though the experimenter always watched for any spontaneous utterances of the test items heard in preliminary observations and conversations with the children.

Two other aspects of production which could be experimentally examined were investigated, however. The first was the ability to reproduce test stimuli modeled verbally by the experimenter. This type of production is hereafter referred to as "imitation." The second was the ability to produce test stimuli in response to pictures. In this case the child was asked to label pictures, thus testing his understanding of the stress contrast differences. This ability will hereafter be referred to as "production." Unlike this latter, imitation does not necessarily involve the ability to understand the phonological stress differences and the semantic referents for each item; it can consist merely of the repetition of the parameters in the acoustic signal which indicate stress.

Among the questions to be considered in production experiments are the following: Are children able to produce correct stress patterns of these pairs distinctively enough so that an adult listener would be able to identify the stimuli being named? Are they able to produce them sufficiently clearly and correctly so that at a later time they themselves can identify their own productions? One of the more interesting questions examined is the possibility that a child who failed in the various comprehension tasks did so not because he had no way to distinguish between the minimal pair test items in isolation, but because he followed a rule which did not match the adult rule or the dialect or idiolect of the experimenter.

A child who fails in the comprehension tests may be able to produce the pairs distinctively using his own rule. When this is the case, an adult listener may not be able to identify the child's productions correctly, but the child himself can do so. Furthermore, he can more easily and correctly identify his own productions than he can the experimenter's. To examine this possibility, the Experimenter Identification Test was first given to each child. This test was identicial to the Identification Test in Chapter 2 except that it was given to each child individually. Then each child was given the Production Test; he was asked to produce a list of stimuli which matched that of the original Identification Test stimulus tape produced by the experimenter. Later he was given the Self-Identification Test, in which he was asked to identify his own test utterances under conditions identical to those under which he took the Experimenter's Identification Test. This series of tasks not only permitted the study of a child's productions and the possibility of his creating unique rules to distinguish between these compounds and phrases; it also provided data for the second area of interest: the relationship between comprehension and production of these stress patterns.

In child language research the general assumption has been that children learn to comprehend language before they are able to produce it. Little had been done to substantiate this assumption until Fraser, Bellugi, and Brown (1963) experimentally studied the acquisition of 10 grammatical contrasts with three-year-olds. They found that these children were able to imitate the test contrasts before they could understand them, but that they could correctly produce these grammatical contrasts in response to pictures only after they could understand them. It is probable, but not certain, that the same order of acquisition and the same relationship between imitation, comprehension, and production exist with all aspects of language. Thus, this relationship should be studied together with the acquisition of stress contrasts and other aspects of language.

In addition to the Production Test in which children were given a list of stimuli to pronounce, the Pair Production Test was administered to test production ability. In this test they were asked to produce both members of each minimal pair, one after the other, in order to determine, under conditions in which differences would be maximized, whether they could correctly produce these differences. Children were also then asked to imitate test stimuli (the Imitation Test).

METHOD

Subjects

Twenty-five children were used as subjects in the various production tests. They were in Grades 1 through 4, and their ages ranged from five to nine years. Testing was done with children of these ages because it was felt that questions concerning production would be most interesting with

children who had either imperfect or no ability to comprehend these stress contrasts. Fourteen of the children were members of the M.I.T. Day Camp and had been given Identification Test 3, the Preference Test, the Semantic/Acoustic Test and the Noun/Verb Test. (See Chapter 2 for descriptions of subjects' background.) The remaining eleven were from a church organization for children and were basically middle-class with some variation in socio-economic background, as was evident from the professions of their parents and the neighborhoods in which they lived. As were the subjects in the comprehension tests, these children were native speakers of English and resided in the Boston area. Table 5.1 shows the number, grade, and sex of the children used in the production studies.

Table 5.1: Number of Subjects for Production Study by Grade and Sex

	Boys	Girls	Total
Grade 1	4	3	7
Grade 2	3	5	8
Grade 3	3	4	7
Grade 4	2	1	3
Total	12	13	25

In order to make the labeling of subjects consistent with that used in the comprehension tests, children who had completed Kindergarten but had not yet entered Grade 1 are referred to as First Graders, etc. Children were tested throughout the summer of 1970.*

Materials and Testing Procedure

Production tests were administered on two different days. On the first day both an identification test of the experimenter's productions (the Experimenter Identification Test) and the Production Test itself were given. The identification test of the child's own productions (the Self-Identification Test), the Imitation Test, and the Pair Production Test were given on the second day.** The time period between the first and second testing sessions varied between two days and one week. Each session lasted approximately 20 to 25 minutes.

* An additional 19 First and Second Graders from the M.I.T. Day Camp were given only the Production and Pair Production Tests. Eight of these subjects were boys and 11 were girls.

** The Contrastive Stress Test discussed in Chapter 6 was also given to some children at the end of the final testing session.

The experimenter tested all children individually in a small, quiet room. Wollensak T-1500 tape recorders were used to play the stimulus tape for the Imitation, Experimenter Identification, and Self-Identification Tests, and to record for the Production, Imitation, and Pair Production Tests (see Chapter 2 for tape recorder specifications). The six compound and phrasal minimal pairs used as stimuli in the comprehension tests were again used as stimuli in these tests. The pictures used to illustrate each member of minimal pairs used in comprehension tests were again used in all tests of comprehension and production except the Imitation Test. In the comprehension tests, however, instead of being told to mark pictures on the answer sheets, the children were shown one pair at a time and responded to each stimulus by pointing to one of the two pictures. The experimenter then recorded all responses. In the production tasks, the experimenter simply gave each child the picture or pictures to name, one at a time. The experimenter got to know the children individually before the testing sessions and talked informally with each one before any test was given.

The Experimenter Identification Test, given in the first testing session, used the original stimuli recorded for the three identification tests. Thirty stimuli (the same number used in Test 1) were given. However, if a child made no errors on the first half of this test he was not given the last half, since it was obvious he could perform the task. Stimuli were presented in the same order as in Identification Tests 1 and 3. As with the latter, each stimulus was presented only once. This was done because the Experimenter Identification Test was to be compared with the Production and Self-Identification Tests and pilot studies indicated that it was confusing and unnatural for children to repeat each stimulus. The presentation of test stimuli and the testing procedures were the same as for Identification Test 1 (except that children were tested individually rather than in groups, and indicated their responses by pointing to pictures rather than filling in answer sheets).

The Production Test was given immediately after the Experimenter Identification Test. Each child was asked to record a tape like the one he had just heard. He was told to pronounce all words as carefully as he could, since he would be asked to listen to them in a few days and to identify each test item just as he had done in the Experimenter Identification Test.

Every child practiced using the tape recorder before the Production Test was recorded. The tape was then made by having the experimenter pronounce the item number and show the child the picture he was to name. The number and order of items was identical to that of the Experimenter Identification Test. Care was taken to leave approximately the same time interval between items as had been used in the experimenter's stimulus tape and to make the child's tape as similar as possible to it.

In the Self-Identification Test, administered a minimum of several days later, each child was given the same instructions as for the Experimenter

Identification Test. Also, test items were reviewed and renamed. The child then heard the tape of his own productions. As in the Experimenter Identification Test, he identified each test item by pointing to one of the two pictures of each member of minimal pairs and the experimenter recorded his responses.

The Imitation Test was then given. The stimulus tape for the Experimenter Identification Test was played and each child was asked to listen and then copy the model utterances as exactly as he could. Imitations were recorded on a second tape recorder. No pictures were used, and the child did not have to give any indication of understanding what he was saying.

The Pair Production Test followed. Each child was shown pictures of the minimal pairs, one by one. He was asked to name the two members, one immediately after the other, and to maximize the differences between them. All six pairs were presented twice, in such a way that the same member of a given pair would be pronounced both before and after the other member. For example, both sequences bláckboard - black bóard and black bóard - bláckboard were given. The pairs were presented at random. By this time the child was very familiar with the stimuli and their pictures but nevertheless the experimenter briefly reviewed the stimuli and their associated pictures before testing started. Also, several examples were modeled by the experimenter to give the child some practice and make sure he understood the task. The test pairs were then recorded.

Most First Graders and a few Second and Third Graders were given a second Experimenter Identification Test and a Production Test, in a third session. This was done mainly to see if learning had taken place over the testing session. It also gave children a second opportunity to produce utterances correctly.

Since 11 of the children tested in these production studies had not been part of the original group given the comprehension tests, Identification Test 3 and the Preference, Semantic/Acoustic, and Noun/Verb Tests were administered to them. They were tested in two groups rather than individually, in order to provide data from a group testing situation which could be compared with their performance when tested individually. This was done to verify the pilot study findings that testing young children in small groups did not hinder their performance. All materials and procedures were the same as outlined in Chapters 2, 3, 4, and 6.

RESULTS

The results of the 25 subjects for all compound and phrasal production and comprehension tests given individually and for the comprehension tests given in small groups are listed in Table 5.2. Since the raw scores or percentage scores for the various tests do not give a good basis for comparison, the results are reported in terms of statistical significance. For all tests except the Semantic/Acoustic Test, any subject whose number of correct answers was not high enough for the probability of chance results to be less than 5 per cent (as was determined by either chi-square analysis or the

Table 5.2

Scores of Production Test Subjects on All Production and Comprehension Tests
(Scores Stated in Terms of Statistical Significance)

				Administered to Individuals					Administered to Groups		
GRADE	SUBJECT	AGE	SEX	EXPERIMENTER ID TEST	SELF ID TEST	PRODUCTION TEST	PAIR PRODUCTION TEST	IMITATION TEST	ID TEST	PREFERENCE TEST	SEMANTIC/ACOUSTIC TEST
1	NU	5.10	M	ns	ns	ns	ns	.05	ns	.05	Semantic
	JT	5.11	M	ns	ns	ns	ns	.05	ns	ns	Semantic
	BK	6.3	M	ns	ns	ns	ns	.05	ns	ns	ns
	JM	6.4	M	ns	ns	ns	ns	.05	ns	.05	Semantic
	KB	5.8	F	ns	ns	ns	ns	.05	ns	ns	Semantic
	DC	5.10	F	ns	ns	ns	ns	.05	ns	ns	Semantic
	JF	5.11	F	ns	ns	ns	ns	.05	ns	ns	Semantic
2	SR	7.3	M	ns	ns	ns	ns	.05	ns	ns	Semantic
	FC	7.5	M	ns	ns	ns	ns	.05	ns	ns	Semantic
	MM	7.5	M	ns	ns	ns	ns	.05	ns	ns	Semantic
	PB	6.9	F	ns	ns	ns	ns	.05	ns	.05	Semantic
	LH	7.0	F	ns	ns	ns	ns	.05	ns	ns	Semantic
	AB	7.1	F	ns	ns	ns	ns	.05	ns	ns	Semantic
	JS	7.2	F	.05	ns	ns	ns	.05	ns	.05	Semantic
	RB	7.3	F	.05	ns	ns	.05	.05	*--	--	--
3	DB	8.1	M	.05	ns	ns	ns	.05	.05	.05	Semantic
	ML	8.3	M	.05	.05	.05	.05	.05	.05	.05	Semantic
	JS	8.4	M	.05	ns	ns	ns	.05	.05	.05	Semantic
	MU	7.4	F	.05	.05	.05	.05	.05	.05	.05	Acoustic
	CW	7.9	F	ns	ns	ns	.05	.05	.05	.05	Semantic
	AM	7.10	F	.05	.05	.05	.05	.05	.05	.05	Acoustic
	LS	8.3	F	ns	ns	ns	ns	.05	ns	ns	Semantic
4	DH	8.3	M	.05	ns	ns	ns	.05	.05	.05	ns
	DS	8.10	M	.05	.05	.05	.05	.05	ns	.05	ns
	SB	8.8	F	.05	.05	.05	.05	.05	ns	ns	Semantic

* Subject RB was unavailable for testing with groups. The Semantic/Acoustic Test was administered individually and she responded to semantic cues with all items.

binomial test, depending on the particular test) had his score reported as nonsignificant (ns); anyone whose number of correct answers was high enough to be significant at this level of probability or better had his score listed as $p<0.05$. On the Semantic/Acoustic Test, semantic or acoustic indicates cases where either one was used in a number of responses significantly greater than chance; otherwise ns is the notation.

The Production Test

Two adults listened to each child's Production Test tape and judged each item as a compound or a phrase. These judgments were then compared with the list of stimuli which the child was attempting to produce. Those which did not correspond to the stimuli the child was told to produce were counted as errors. Table 5.2 lists the subjects in the Production Test who had a significant number of productions judged to be correct. Only five of the twenty-five subjects were successful in this task. No child in Grade 1 or 2 produced a number of correct reponses that could be classified as significantly better than chance; in Grades 3 and 4 approximately half the children were able to do so.

As can be seen from Table 5.3, the 20 children whose number of correct responses was judged not significantly different from chance were consistent in their answers; 16 of them produced the compound stress pattern for a significant number of items. In fact, most of these children produced over

Table 5.3: Production Test -- Response Biases of Subjects Whose Number of Answers Judged to Be Correct Was Not Significantly Greater than Chance

Grade	Total #Ss scoring ns	#Ss producing compound stress pattern responses ($p.< 0.05$)	#Ss producing phrasal stress pattern responses ($p <0.05$)	#Ss producing both patterns ($p<$ ns)
1	7	6	0	1
2	8	7	0	1
3	4	2	0	2
4	1	1	0	0
Totals	20	16	0	4

90 per cent of the stimuli as compounds. No subject produced a high number of phrasal stress patterns.* The remaining four subjects produced both stress patterns but did so at random. Three of the four had a higher (though not significant) number of compound stress responses. The tendency to produce these pairs with primary stress on the first syllable is therefore very strong. This may relate to the fact that the compound stress pattern was more often selected than the phrasal one in the Preference Test (see Chapter 3). If a child produced only this pattern he would probably prefer it in a comprehension test, or vice versa.

Why the compound stress pattern should be produced most often is not obvious. Livant (1962) also reports that children had difficulty producing the distinction between bluebird house and blue birdhouse. He notes they cannot reliably produce a three-word compound with primary stress on the second element; instead they produce it on the first. His only explanation for children's tendency to place stress on the first element is that it might have been due to a set where children had already had the 132 pattern before encountering the 213 pattern. Since some children who always produce stress on the first element in the present study heard the phrasal pattern prior to the compound pattern, this explanation appears untenable. As discussed in Chapter 3, the tendency could relate to early acquisition frequency, and productivity of the compound structure. It may also be related to contrastive stress rule interference. Contrastive stress rule application can shift stress to the first syllable of a phrase, making it homophonous with the compound (e.g., it's a gréen house, not a réd one). However, it is extremely unlikely that contrastive stress would shift stress to the second syllable of a compound , making it homophonous with the phrase; commonly, contrastive stress does not affect the stress position in compounds (e.g., it's a gréenhouse, not a garáge). The fact that contrastive stress frequently shifts phrasal stress to the first syllable while the compound generally retains its first-syllable emphasis may contribute to a child's preference for the compound stress pattern.

One additional possibility that cannot be ruled out is that when children produced test items with primary stress on the first syllable, they were using contrastive stress rather than compound stress. For example, if the child were thinking of a high chair as a chair that is high, as opposed to a chair that is low, he might very well produce the stress pattern high chair. Therefore at this point it cannot be clearly determined why children produced stress on the first syllable of these test items.

---

* None of the additional 19 First and Second Graders given the Production Test had a significant number of responses judged to be correct. Again, more children displayed biases toward the compound stress pattern than the phrasal one, 12 of the 19 producing only compounds. Also 6 children used both patterns but produced more compounds than phrases. Unlike the main test, however, in this sample one child produced only the phrasal stress pattern.

The question of learning effects was also considered, since 12 of the children repeated the Production Test after having completed the Self-Identification, Pair Production, and Imitation Tests. Only one child, Third Grader CW, showed marked improvement and performed significantly on the second trial after performing randomly on the first. She had produced both compound and phrasal stress patterns on the first test, but did so more accurately on the repeat test.*

Although the remaining 11 children had scores not significantly different from chance on both tests, 3 showed other evidences of becoming more aware of the stress contrast differences after practice on tests which intervened between the first and second Production Tests. Second Graders JS, FC, and SR produced all stimuli with the compound stress pattern on the first trial. By the second trial they were aware that some difference existed, but they were unsure of how to signal it systematically. They used both stress patterns on the second trial, but at random. They also experimented with ways to produce the differences; for example, SR occasionally tried to indicate a difference between members of pairs by speaking in a low monotone voice on both syllables of some test items and in a falsetto on others. Of the remaining eight children who repeated the test, seven did not alter their behavior the second time. The eighth child, Second Grader RB, produced both compound and phrasal stress patterns on the first production test, but used all compounds on the second. Thus a few children appeared to have benefited from having performed several different production tasks, but they were in the minority, and the degree of learning was not great.

Results of the Production Test were also analyzed for differences according to sex. As Table 5.4 indicates, there were no differences in

Table 5.4: Number of Subjects Performing Significantly on Production Test According to Sex and Grade

	#Ss Scoring Significantly	
Grade	Boys	Girls
1	0 of 4	0 of 3
2	0 of 3	0 of 5
3	1 of 3	2 of 4
4	1 of 2	1 of 1
Totals	2 of 12	3 of 13

* She also performed significantly on her second trial of the Experimenter Identification Test.

Grades 1, 2, and 4. In Grade 3 there was only a difference of one.* Although the trend for girls to perform better than boys may have been stronger if a larger number of subjects were considered, the differences do not appear great.

The Pair Production Test

Results from the Pair Production Test were analyzed to determine whether children who failed in the Production Test might have more success if they produced test stimuli in pairs rather than one at a time. Producing the stimuli as minimal pairs gave them more opportunity to contrast the stress differences between members.

As with the Production Test, two adults listened to each child's performance and judged each member of the pair as a compound or a phrase. The judgments were matched with the list of stimuli, and the pairs that did not agree were marked incorrect. Table 5.2 lists these judgments. Of the 25 children, 7 produced a number of correct pairs significantly different from chance (as was determined by the binomial test). With the exception of Second Grader RB, all were in Grades 3 and 4.

All children who produced a significant number of responses judged as correct in the Production Test also had significant scores in the Pair Production Test. The reverse was not true, however; two children produced a significant number of responses correctly on the Pair Production Test but not on the Production Test. This suggests that it was easier to produce these stimuli correctly in pairs than in isolation.

Just as in the Production Test, some children knew there was a difference between members of pairs but could not consistently and correctly signal that difference. For example, some used the same stress pattern on both members of a pair but said the entire first member very rapidly and the second extremely slowly. A number of others were able to produce the two stress patterns correctly but always produced the compound and then the phrasal patterns, regardless of the order in which the pictures they were identifying appeared.**

---

* One additional girl, CW, performed significantly on the second trial of the Production Test, however.

** Of the additional 19 First and Second Graders who took the Pair Production Test, 18 were unable to produce the pairs correctly. One First Grade boy was able to do so with all but one pair. He was unsuccessful in the Production Test, however. All children who were unable to produce the differences correctly produced the compound stress pattern more often than the phrasal one, with the exception of the one girl who had produced all stimuli with phrasal stress in the Production Test; she also used only the phrasal stress pattern in this Pair Production Test.

After the production tasks had been completed, the experimenter questioned each child about the nature of the differences between minimal pairs. Many of the youngest children either gave no clear answer or were unable to say what the differences were. Of the children who did report a difference, most said that it was related to duration. They felt that compounds are said faster than phrases. Some mentioned differences in amplitudes, and a few noted differences of fundamental frequency. These subjective impressions of what acoustic correlates of stress are important may have no relation to what children actually use in stress production and perception, but they were nonetheless interesting.*

The Imitation Test

Children's imitations were judged by two listeners as correct or incorrect on the basis of the model utterance. The results of this test are shown in Table 5.2. All children tested were able to mimic the stress contrasts successfully. Some needed a little practice, but once they understood the task they imitated the stimuli correctly. The fact that they could imitate showed that they were able to perceive the stress patterns in the speech signal; however, given the fact that even infants are sensitive to the acoustic correlates of stress, it would be most surprising if older children could not perceive it. Where they failed was in the comprehension of stress patterns, i.e., they were unable to use stress at a phonological level as well as at a phonetic one. Since even the youngest children who failed in the production and comprehension tasks were able to imitate correctly, these findings are in agreement with those of Fraser, Bellugi, and Brown (1963): children can imitate contrasts before they can comprehend or produce them.

The Self-Identification Test

Results of the Self-Identification Test showed that only five children were able to identify correctly a significantly-better-than-chance number of their own productions, as Table 5.2 indicates. They were all in Grades 3 and 4, and all five had had significant scores on the Production Test, indicating that the adult listeners could also identify their productions. On the other hand, not one of the children who was unable to identify his own productions had had a significant score on the Production Test; thus, there was good correlation between the adult listeners' judgments and the children's identifications. If the children could identify their own productions, the adults could too; if the children failed, so did the adults.

* Some perception experiments with synthetic speech were conducted with children who had been given Identification Test 1, in order to examine what acoustic correlates of stress are important to children. Acoustic analysis was also made of children's productions of stress. These findings will be reported elsewhere at a later time.

Data from the Experimenter and Self-Identification Tests were compared to determine whether any child had more success in identifying his own productions than the experimenter's. Table 5.5 indicates the total number of children in each grade categorized according to the statistical significance of their scores on these tests. It shows that no child was more

Table 5.5: Number of Subjects with Significant or Nonsignificant Scores on Experimenter and Self-Identification Tests According to Grade

Grade	$p<0.05$ on Neither Test	$p<0.05$ on Both Tests	Exp. = $p<0.05$ Self = ns	Exp. = ns Self = $p<0.05$
1	7	0	0	0
2	6	0	2	0
3	2	3	2	0
4	0	2	1	0
Totals	15	5	5	0

successful in identifying his own productions, suggesting that none had created a unique rule for distinguishing between these compounds and phrases. Also, apparently none failed because the experimenter had different dialect or idiolect. In fact, the data contained in Table 5.5 indicate that five of the children had more success with the experimenter's productions than they did with their own. If the second trial of the Experimenter Identification Test is considered, two more children are added to this group; they had success with the second trial of the Experimenter Identification Test but failed on the Self-Identification Test. Since approximately half these seven children were girls and the other half boys, sex differences do not appear to be important here.

Comparison Between Experimenter Identification Test and Identification Test 3

Results from the Experimenter Identification Test and Identification Test 3 were compared in order to verify the pilot study findings that no significant differences existed between children's performances individually and in small groups. The only difference between these two tests (other than the individual versus small group testing condition) was that the number of items differed. The Experimenter Identification Test had 30 items, all of which were given to all children except those who made no errors on the first 13 items; Test 3 had only the first 13 of these items. Given the high correlation of performance between Tests 1 and 3 on the one hand, and the first 12 items and the entire 30 items on Test 1 on the other (see Figure 2.3,

Chapter 2), this difference does not appear important. Table 5.6 shows the number of children in each grade who scored significantly more correct than incorrect responses on either or both of these tests. The table shows that the scores in both tests of 20 of the 24 children fall into the same category of significance, and four did not. Three of the latter did better on the test given individually than on the one administered to the group, and the fourth performed better under group testing conditions. Hence, although a few children seem to have been able to perform better under individual testing conditions, the performance of the majority did

Table 5.6: Number of Subjects Scoring Significant Responses on Either or Both the Experimenter Identification Test and Identification Test 3

Grade	Total Ss Considered*	#Ss Scoring: $p<0.05$ on Both	$p<0.05$ on Neither	$p<0.05$ Exp. $p<ns$ on Test 3	$p<ns$ on Exp. $p<0.05$ on Test 3
1	7	0	7	0	0
2	7	0	6	1	0
3	7	5	1	0	1
4	3	1	0	2	0
Totals	24	6	14	3	1

*One child in Grade 2 did not take Test 3, and therefore was excluded from this comparison.

not appear to be affected by this difference. It would seem reasonable to assume that the children who had the greatest difficulty in a group testing situation would be the youngest, since they have had less experience in group testing. Yet, interestingly enough, none of the children who had just finished Kindergarten and were about to enter Grade 1 did any better when tested individually on the first trial than they did in small groups. This again suggests that the differences in testing conditions were not an important factor in comprehension.

Each child's results on the Experimenter Identification Test and Test 3 were compared, by item, in an attempt to determine whether children erred repeatedly on the same items. Such repeated errors on only certain items would argue against use of a stress rule rather than learning stress item by item. However, the data were not clear with regard to this question.

Some children repeatedly missed only certain lexical items, while others appeared to "half know" all items and to err more at random. Therefore no conclusions could be drawn in this connection.

Comparison of Identification Test 3, Preference Test, and Semantic/Acoustic Test with Production Subjects

In Grades 1 and 2, four of the fifteen children scored significantly more correct than incorrect responses on the Preference Test but not on Identification Test 3. Fourth Grader DS, and an additional Second Grader, JS, did the same, although both had significant scores on the Experimenter Identification Test. The reverse situation -- a significant score on Identification Test 3 and a random score on the Preference Test -- did not occur with any of these children.

In Chapter 3 it was shown that children tended to perform slightly better with a preference (forced-choice) task than an identification task, and the children to whom the production tests were given buttressed this conclusion. In general they appeared to be successful in selecting the correct stress pattern from right and wrong options earlier than they were in identifying an item as a compound or a phrase on the basis of stress.

Only two children deviated slightly from this pattern. They failed on the Preference Test and also on Identification Test 3, as would be expected; however, they succeeded with the Experimenter Identification Test. One First Grader, JT, scored significantly more correct than incorrect responses the second time he took the Experimenter Identification Test, but he failed to do so on any other test. However, this success is suspicious because of his failure on the first Experimenter Identification Text, Identification Test 3, and all other tests. Fourth Grader SB also scored significantly on the Experimenter Identification Test and randomly on the Preference Test. However, the fact that she failed on the Preference Test and Identification Test 3 was unusual for a number of reasons: first, she displayed an ability to identify these compounds and phrases by giving significantly more correct than incorrect responses on the Experimenter Identification Test and on the production tasks and Self-Identification Test. Second, she performed on a level significantly different from chance on Identification Tests 1 and 2, which she had taken on a year earlier. Finally, she performed randomly on all the tests (including the Noun/Verb Test) she took at the same time as she did Test 3 and the Preference Test. Her performance on this day was atypical, and was undoubtedly not a reflection of lack of skill with these stress contrasts, but rather attributable to performance factors.

The results of the Semantic/Acoustic Test seen in Table 5.2 indicate that only two of the 25 children, both Third Grade girls, attended to acoustic cues in a significant number of their responses. They both had significant

scores on all other tests, which supports the suggestion that children are not able to ignore semantic cues and attend to acoustic ones until they have completely mastered the ability to use stress in these pairs (see Chapter 4). Third Grader ML, the only other subject who had significant scores on all other tests, attended to semantic rather than acoustic cues. The reason for this is not clear. Although his scores were significant on the other tests he did make a few errors, which suggests that his level of competence was either newly acquired or not yet perfect enough to allow him to ignore the semantic cues with confidence.

All other subjects except one First Grader and two Fourth Graders attended to semantic cues in a significant number of their responses. The younger child, BK, did not score well on any other tests with compounds and phrases, so his fluctuation between semantic and acoustic cues may have been a reflection of performance factors.* The two older boys showed evidence of being in the process of mastering this ability with regard to stress contrasts. They scored significantly better than chance on some of the other tests; in fact, one of them scored significantly on all tests except one. As was seen in Chapter 4, many children who were in the process of perfecting ability with stress contrasts were able to attend to acoustic cues but were not confident enough to ignore the semantic ones consistently, and thus they fluctuated between the two. These two boys appear to be in this stage.

A Comparison of the Production and Comprehension Tests

In comparing the results from the Production Test with those of the Experimenter Identification Test, we find no instances where children had more success with the former than they did with the latter. However, the reverse was true of one subject in Grade 4, and with two each in Grades 2 and 3. The fact that these five subjects had more success in comprehending stimuli than in producing them, is evidence that ability to comprehend precedes that of production. However, there was no evidence of the opposite tendency.

If Identification Test 3 and the Preference Test are also considered, there is further evidence that the order of acquisition is comprehension first and production second. Again, a number of children had significant scores on one or both these tests, but performed randomly on the Production Test; and again performance was rarely the other way around. Only one child, Fourth Grader SB, performed significantly on the Production and Self-Identification Tests and randomly on the Preference Test and Identification Test 3. However, her failure on the latter two tests was atypical, as explained earlier, and was probably due to performance factors rather than lack of linguistic ability, since she succeeded with the Experimenter Identification Test, Identification Tests 1 and 2, and the production and self-identification tasks.

---

* However, BK did score a significant number of correct responses on the Noun/Verb Test, as will be seen in Chapter 6.

The generalization to be drawn from these various comprehension and production tasks is that the order of mastery is imitation, comprehension, and finally production, as Fraser, Bellugi, and Brown (1963) have suggested. However, the types of task used to test skills of comprehension and production vary in difficulty; and children are able to perform some before others. An examination of the results of all comprehension and production tests, collated in Table 5.2, indicates the order in which the ability to perform certain types of tasks as well as comprehension and production skills is acquired. The scores of these 25 children suggest the following generalization with regard to order of acquisition.

First, even the youngest children are able to imitate stress patterns without difficulty. They are able to perceive patterns at a phonetic level and can mimic them. When they reach the age where they begin to develop comprehension abilities, the first type of task they are able to perform well is a preference one. Given the right and wrong stress pattern to compare, they are able to select the correct one -- i.e., the one which sounds best. Then, as they gain further ability with these contrasts, they begin to be successful in correctly identifying stimuli in isolation.

At a later point they start to develop production skills. They are first successful in producing the stress patterns of stimuli correctly when they say one member of a minimal pair immediately after the other, since they are able to contrast them. Then they develop the ability to produce stimuli correctly in isolation. When they can do this, they can correctly identify their own productions, since the ability to identify was mastered somewhat earlier.

The final skill appears to be the ability to identify stimuli by stress patterns when they are semantically contradicted by the sentential context. The ability to ignore semantic cues and attend to acoustic ones was the last task of those studied to be mastered. Thus, although the general order of mastery is imitation, then comprehension, and finally production, the type of task varies in difficulty and can even influence this general order; the Semantic/Acoustic Comprehension Test was more difficult for the children than were the production tasks, and it was not until they had mastered both comprehension and production skills that they could perform this type of comprehension task.

Chapter 6

Experiments with Contrasting Stress Patterns in Other Constructions

The experiments presented in the previous chapters with compounds and phrases indicate that linguistic ability with the stress patterns which distinguish these structures is acquired mainly as a function of age, and that this ability is not mastered until relatively late in the language-acquisition process. An important question raised by these experiments is the relationship of these findings to stress patterns in other types of contrasting constructions. As noted in Chapter 1, children use stress very early, but we do not know in detail how ability with stress is acquired in constructions where it can be predicted by rule and signal differences in meaning.

In light of the findings with regard to stress as it distinguishes compounds and phrases, we might ask whether the ability to use stress with other contrasting structures is also acquired in the same way. Is ability with stress in other structures also acquired as a function of age, and is it mastered relatively late? If the time of mastery is not the same, what is the relative order of acquisition for ability with stress in compounds and phrases in other contrasting constructions?

In order to examine these questions, a number of additional formal and informal experiments were undertaken. They were intended to give only preliminary indications, and more extensive testing will be required before the results can be considered conclusive. Two additional structures studied were noun and verb pairs distinguished by differences of both stress and vowel quality, and sentences with emphatic or contrastive stress.*

I. Comprehension Experiment with Stress as it Distinguishes Nouns From Verbs

The main experiment conducted to explore some of these questions was a comprehension experiment with contrasting noun and verb pairs. (This experiment will hereafter be referred to as the Noun/Verb Test.) The stimuli used were noun and verb pairs such as <u>récord</u> and <u>recórd</u>, which differ with respect to stress patterns, primary stress falling on the first syllable

*A series of experiments was also conducted with structures involving contrastive stress and pronominal reference in coordinate sentences and sentences with subordinate clauses; however, these data will not be discussed here, but will be considered elsewhere at a future time.

in the noun and on the second syllable in the verb. Unlike the compound and phrase pairs tested, these noun and verb pairs are not minimal with respect to stress since they have an additional vowel-quality difference. The stressed vowel in a given member of a pair is reduced in quality when it appears unstressed in the other member of that pair.

The contrasting stress and vowel quality patterns differentiating such noun and verb pairs, which have often been described, have recently been accounted for by Chomsky and Halle (1968) in terms of formal rules: vowel reduction rules, and the Main Stress Rule. In order to account for the relationship between the nouns and verbs of these pairs, the Main Stress Rule assumes that the verb is the underlying form and that the noun is transformationally derived from the verb. (Chomsky and Halle give no syntactic, morphologic, or phonetic reasons for assuming that the noun is derived from the verb rather than vice versa.*) According to this rule, the noun differs from the verb in that it goes through one more cycle of the rules.

Whether the rules of Chomsky and Halle (1968) constitute a correct linguistic and/or psychological description of the stress- and vowel-reduction patterns of these contrasting noun and verb pairs is not the primary issue here. We are concerned with children's knowledge of these stress and vowel reduction patterns at various ages, and with the relationship of this knowledge to ability with stress in compounds and phrases. It is hoped that such information will suggest additional questions and studies.

The relationship between ability with stress as it differentiates these nouns from verbs and ability with stress as it differentiates compounds from phrases is not obvious, particularly in terms of the age at which the two types of stress patterns are mastered. A second cue of vowel reduction works with the stress cue to signal the meaning difference; hence ability with stress in these nouns and verbs might be mastered at an earlier age than ability with stress in compounds and phrases, which are phonetically different only with respect to stress and which have no additional vowel-reduction cue. Also, the syntactic distinction of noun versus verb is perhaps more basic and obvious than that of two constructions involving nouns, and this might also contribute to earlier learning of the stress patterns. On the other hand, children may be able to use word order and inflectional information to correctly distinguish between nouns and verbs without recognizing differences in stress and vowel reduction. In addition, the fact that most of the noun and verb pairs are based on a Romance vocabulary which is generally more sophisticated and less common to a preliterate child would perhaps suggest that children do not become aware of these vocabulary items until relatively late in the acquisition process; thus they also may not acquire skill in distinguishing the pairs on the basis of stress and vowel reduction until relatively late.

---

*This assumption has been criticized by Ross (1969), who points out that some verbs, such as to police, to voice, and to machine, "feel"as though they have been derived from nouns rather than vice versa. Ross also discusses other problems with the Main Stress Rule and proposes alternative solutions.

In order to make the comparison between these different structures as reliable as possible, the Noun/Verb Test was conducted with the same subjects and the same experimental procedures used in the Preference Test with compounds and phrases. Also, since testing with compounds and phrases was done with real words rather than nonsense words, real words were also used in testing with nouns and verbs. The same type of task used in the Preference Test was selected for testing with nouns and verbs because it did not necessitate picturing each item, something that was difficult and may have been confusing to the children, given the vocabulary items involved. Results of the Noun/Verb Test were therefore compared with those of the Preference Test with compounds and phrases. The analysis of the Noun/Verb Test is somewhat less detailed than that of the Preference Test because of the limitations imposed by the smaller number of noun and verb pairs used in testing and the preliminary nature of the experiment.

METHOD

Subjects

The 125 children and 19 adults used as subjects in this experiment were the same as those used in the Preference Test, Identification Test 3, and the Semantic/Acoustic Test. (See Table 2.3 for a breakdown of the number of subjects by grade and sex.)

Stimuli and Their Presentation

Three pairs of contrasting nouns and verbs were used as stimuli in this experiment. They were embedded in the sentences listed in Table 6.1.

Table 6.1: Sentence Stimuli Used in the Noun/Verb Test

1. (A) Let's recórd his voice.
   (B) *Let's récord his voice.

2. (A) *Jane gave Billy a presént.
   (B) Jane gave Billy a présent.

3. (A) When will the director condúct the orchestra?
   (B) *When will the director cónduct the orchestra?

4. (A) The récord is on the shelf.
   (B) *The recórd is on the shelf.

5. (A) *When will John préśent a speech?
   (B) When will John presént a speech?

6. (A) *Jimmy was very noisy; the teacher said his conduct was terrible.
   (B) Jimmy was very noisy; the teacher said his cónduct was terrible.

As noted earlier, because noun and verb pairs which are distinguished only by stress and vowel reduction are generally not common in the vocabulary of very young children, it was difficult to find suitable pairs for testing. In addition to the three pairs listed above, several others were used as stimuli in pilot studies for this experiment. However, since most children were not familiar with these additional pairs, they were excluded from further testing.

Because the test with noun and verb pairs was given in conjunction with the Preference Test with compounds and phrases, the basic test design and procedures employed were the same. (See Chapter 3 for details of the basic procedure.) The dummy items described in Chapter 3 preceded and followed the presentation of the Noun/Verb Test sentences.

One difference between procedures for the Preference Test with compounds and phrases and for the Noun/Verb Test is that the stimuli of the former were pronounced in isolation and defined before the test was administered, but the items of the latter were not. The first time children heard the noun and verb pairs was during the test itself, when they were asked to listen to the A and B sentences and select the one which "sounded the best."

RESULTS

General Results

The results of the Noun/Verb Test indicate that the ability to select the correct noun and verb in sentences with these pairs is mastered mainly as a function of age. Figure 6.1 displays the percentage of correct responses plotted according to age and grade. A steady developmental increase occurs in the percentage of correct responses, starting with Grade 1's 67 per cent and rising to Grade 7 & 8's 100 per cent*

---

*The Noun/Verb Test was given to 24 of the children who took the production tests described in Chapter 5. Of the 24, 18 scored significantly more correct than incorrect responses. Those who did not have significant scores were First Graders JF and JM; Second Graders FC, MM, and SR; and Fourth Grader SB.

Figure 6.1: Noun/Verb Test --Percentage of Correct Responses Plotted Against Grade and Age

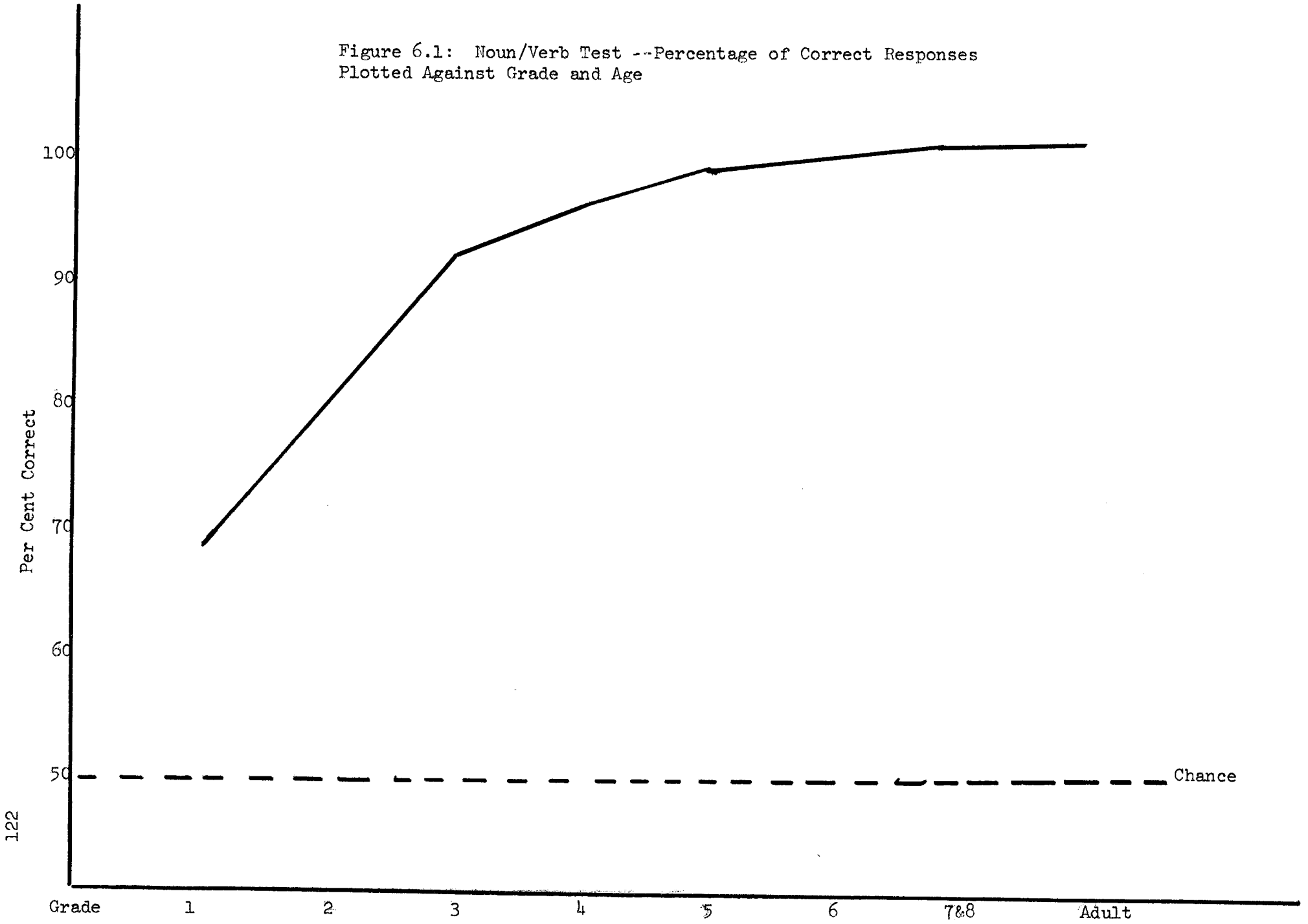


Figure 6.2: Noun/Verb Test & Preference Test with Noun Compounds and Noun Phrases --Comparison of Percentage of Correct Responses

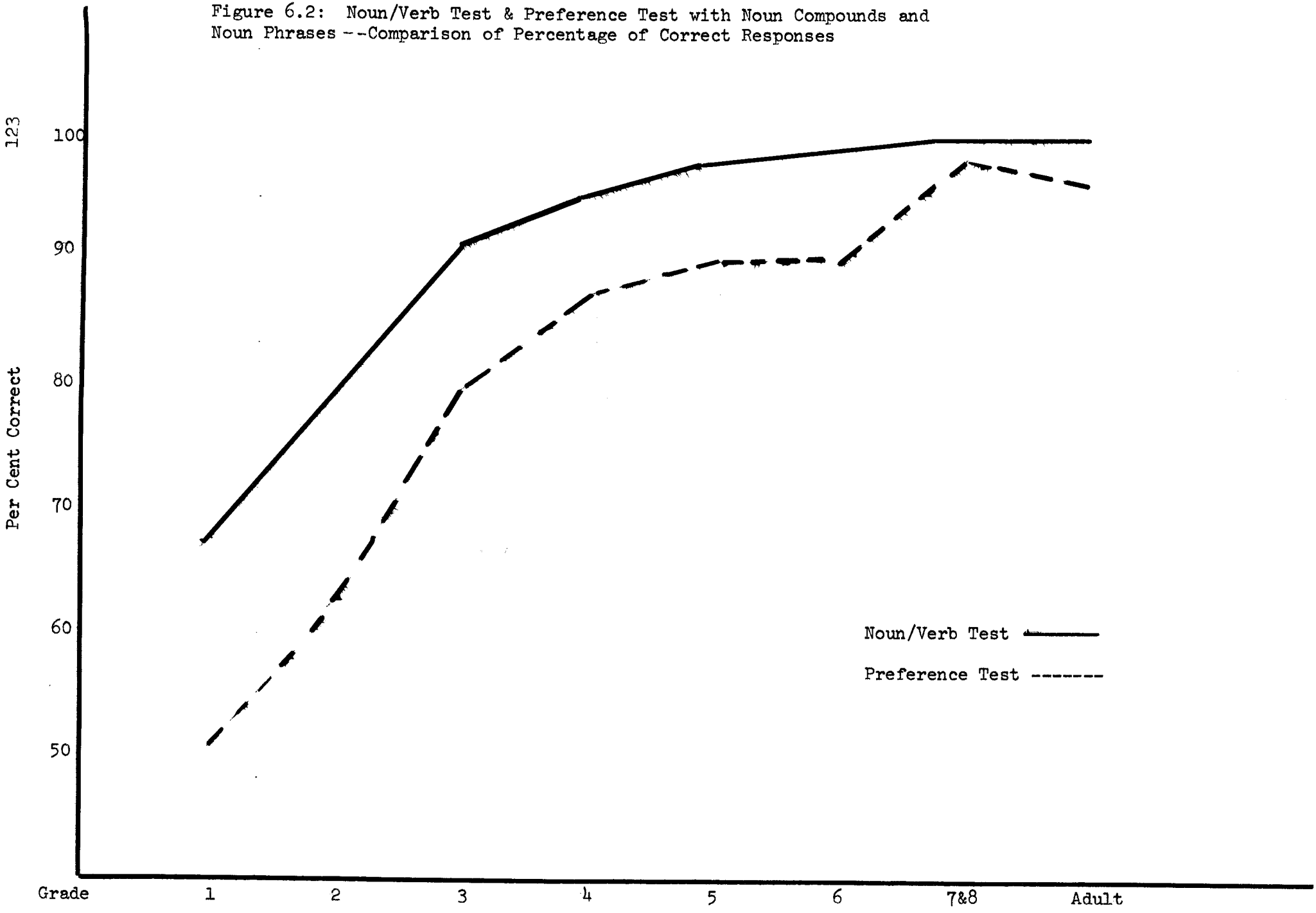


Figure 6.2 compares the results of subjects' performance on the Noun/Verb Test with their performance on the Preference Test. It indicates that children reach a high level of accuracy with these nouns and verbs much earlier than they do with the compounds and phrases tested. For example, children had over 90 per cent correct responses by Grade 3 with the Noun/Verb Test, but they did not reach that level of accuracy with compounds and phrases until Grade 5.

Earlier acquisition of ability with nouns and verbs is also suggested by the fact that in every grade, subjects tended to perform better on the Noun/Verb Test than they did with compounds and phrases. Chi-square analysis indicates that the tendency of individuals to perform better on the Noun/Verb Test than on the Preference Test is significantly greater than chance at the $p<0.05$ level or higher in Grades 2, 4, 5, and 6.

Although performance was generally much higher on the Noun/Verb Test than on the Preference Test, it did not reach the high level of accuracy that was attained on the dummy sentences in this test. A comparison of Figure 3.3 with Figure 6.1 shows that the differences between performance on the dummy sentences and that on the noun and verb pairs was very small in Grades 3 and above (since accuracy is very high with both), but in Grades 1 and 2 the difference is fairly large. Although the children in Grades 1 and 2 showed a high degree of linguistic ability with these noun and verb pairs, they did not have sufficient linguistic knowledge or familiarity with these pairs to perform as well with stress in these constructions as they did with stress in the dummy sentences.

Frequency Distribution and Individual Variation

In order to determine the range and breakdown of the children's scores within each grade, the frequency distribution found in Table 6.2 was calculated. Distribution in this test tended to be normal and symmetrical, and thus was similar to the distribution seen with compounds and phrases in the Preference and Identification Tests. Instead of tending to score either all correct or none correct as with the Semantic/Acoustic Test, children in each grade generally had scores that were close to the average scores for the grades recorded in Figure 6.1.

As with the Preference and the Identification Tests with compounds and phrases, some individual variation was seen within grades, particularly among the younger children. However, the amount of variation was relatively small. Also, since linguistic ability with the noun and verb pairs tended to be acquired earlier than ability with compounds and phrases, the number of children scoring all six items correct was generally higher in each grade for noun and verb pairs than for compounds and phrases. Table 6.3 lists

Table 6.2: Noun/Verb Test - Frequency Distribution of Per Cent of Correct Responses of Subjects by Grade

% Correct Responses	No. Correct Responses	Number of Subjects Scoring at Each Interval						
	Grade:	1	2	3	4	5	6	7 & 8
	% (Mean)	(67)	(79)	(91)	(95)	(98)	(99)	(100)
0.0	0		1					
16.7	1	1						
33.3	2	1						
50.0	3		5	2	1			
66.7	4	*2	3		1			
83.3	5	4	*7	*3	2	2	1	
100.0	6		11	11	*21	*16	*19	*11

[Blank space = 0; interval closest to group mean = *]

Table 6.3: Noun/Verb Test: Percentage of Subjects Scoring 100 Per Cent Correct Responses

Grade	% Ss
1	0
2	41
3	69
4	84
5	89
6	95
7 & 8	100
Adult	100

these data in terms of the percentage of children scoring 100 per cent correct for each grade.* If we compare this table with the frequency distribution data for the Preference Test seen in Table 3.3 (which shows the percentage of subjects scoring at least 80 per cent, i.e. 10 out of 12 items correct), we see that the most marked difference occurred with Grade 2; 41 per cent of the children scored 100 per cent correct responses ($p<0.02$) in the Noun/Verb Test, and only 18.5 per cent scored over 80 per cent correct ($p<0.02$) in the Preference Test. This is a particular illustration of the fact that the average scores of grades as a whole were much higher with nouns and verbs than with compounds and phrases, and also that more individuals performed at a high level of accuracy with the former construction than with the latter.

Differences According to Sex

Results from the Noun/Verb Test were analyzed to determine whether any significant differences occurred between the performance of boys and that of girls at any age group. Figure 6.3 plots the percentage of correct responses for both boys and girls at all age groups except Grade 1, where no female subjects were in the sample. This comparison shows that both boys and girls performed at a relatively high level even in the youngest grades, and that there were no marked differences in their performance at any of these ages. The biggest difference is seen with Grade 2, where girls performed better than boys, but the Mann-Whitney U indicates that this difference was not significantly greater than chance in this or any other grade. In fact, the Noun/Verb Test displayed fewer differences according to sex at various ages than did any of the preceding tests.

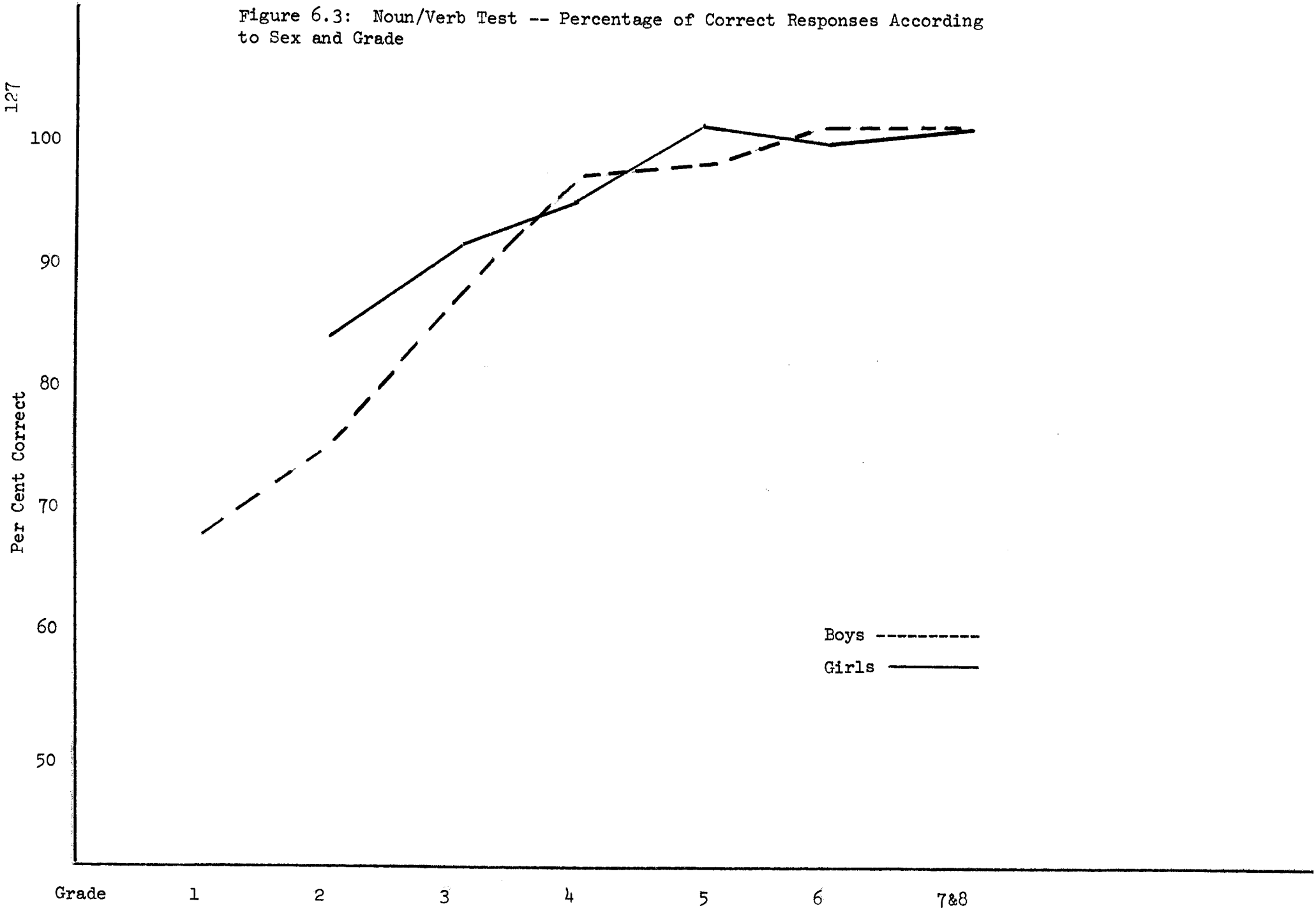
Response Biases

Because the number of test items in the Noun/Verb Test was so small it is difficult to determine whether any significant response biases were present. Nevertheless, the nonlinguistic factors which potentially affected responses of children who did not have the ability to base their responses on linguistic information are briefly considered here, since they may be suggestive of response bias trends.

In Grades 1 through 4 (those with the greatest number of errors) more subjects marked B on the answer sheet than A. (B answers were always located to the right of the A answers on the sheets.) However, only one subject in the whole group of 125 children responded with the same letter to all stimuli,

*Since the Noun/Verb Test had only six test items, it was necessary for children to score all items correct in order to have less than a five per cent probability of chance results, as was determined by the binomial test.

Figure 6.3: Noun/Verb Test -- Percentage of Correct Responses According to Sex and Grade



and his bias was not in the direction suggested, since he marked all A's.

The data also indicated a potential bias toward selecting verbs rather than nouns, since in Grades 1 through 4 more errors were made with the latter than with the former. However, only one child chose the verb for all six stimuli.

The members of the minimal pairs which tended to be missed most often were the nouns conduct and record. Virtually no differences were found in the number of errors made with the noun and verb of the present-present pair. Whether the greater number of errors were due to the vocabulary items themselves, to the noun or verb structure and stress pattern, to the test sentences in which the pairs were embedded, or to chance cannot be determined because of the limited data.

Considering the pairs themselves, we find no consistent ranking of difficulty of pairs within grades. Only Grades 1 and 2 gave similar rankings, with the order of difficulty from most to least being record, conduct, and present. As with the members of pairs, the reason why the greater difficulty was experienced with one pair rather than another cannot be determined from these limited data.

Reliability and Correlation with Other Tests

Although the Noun/Verb Test was limited in terms of the number of test items presented, its general results were consistent so far as continuous improvement with each successive grade and general agreement with the Preference Test results were concerned, as was shown in Figure 6.2. There also appeared to be good correlation on an individual subject basis when the performance of any individual child was compared for the two tests. If the subjects are divided into two groups for each test--those who performed at a level of significance greater than $p<0.02$ as was determined by the binomial test and those who performed at the chance level--we find the majority of subjects in every grade performed at the same level of significance on both tests, as can be seen from Table 6.4.

Table 6.4: Noun/Verb Test Versus Preference Test with Compounds and Phrases - Percentage of Subjects Scoring at Same Level of Significance on Both Tests (i.e., either $p<0.02$ on both tests or $p<ns$ on both tests)

Grade	% Ss
1	87.5
2	70.4
3	81.3
4	68.0
5	83.6
6	75.0
7 & 8	100.0
Adults	100.0

That is, these subjects either scored significantly more correct than incorrect answers on both tests, or did randomly on both tests. The scores of over 70 per cent of the subjects agreed in this manner in every grade except Grade 4, where only 68 per cent of the subjects displayed such consistency. Taking the group of 125 children as a whole, we find that an average of 78 per cent of the subjects performed consistently on both tests in terms of levels of significance reached.

Table 6.5 gives further data on the correlation of individual children's performances on the Noun/Verb Test and the Preference Test with compounds and phrases. It shows the breakdown by grade of the number and percentage

Table 6.5: Noun/Verb Test Versus Preference Test with Compounds and Phrases - Number and Percentage of Subjects Scoring Randomly or Significantly Better Than Chance on the Two Tests

		No. and % Ss Having Scores Significant at $p<0.02$							
No. of Ss	Grade	Both Tests		Neither Test		Noun/Verb Test Only		Preference Test Only	
		No.	%	No.	%	No.	%	No.	%
(8)	1	0	0	7	87.5	0	0	1	12.5
(27)	2	4	14.8	15	55.6	7	25.9	1	3.7
(16)	3	8	50.0	5	31.3	1	6.3	2	12.5
(25)	4	16	64.0	1	4.0	6	24.0	2	8.0
(18)	5	14	77.8	1	5.6	2	11.1	1	5.6
(20)	6	15	75.0	0	0	4	20.0	1	5.0
(11)	7 & 8	11	100.0	0	0	0	0	0	0
(19)	Adults	19	100.0	0	0	0	0	0	0

of subjects scoring either randomly or significantly on one or both of these tests. As might be expected, the number and percentage of children who scored significantly more correct than incorrect responses on both tests increased with age; it was higher than the number and percentage of subjects scoring randomly in both tests in every grade except Grade 1. In fact, of the 78 per cent of the group performing consistently on the two tests in terms of the levels of significance of their scores, 55 per cent scored significantly and only 23 per cent randomly on both these tests.

Looking at the other 22 per cent of the entire group who did not perform with significant linguistic ability on both tests, we see that more of these subjects performed well on the Noun/Verb Test than on the Preference Test; 16 per cent performed significantly on the Noun/Verb Test but randomly on the Preference Test with compounds and phrases, and 6 per cent did the opposite. Hence, the data indicate, in general, that if a child performs well on the Preference Test with compounds and phrases, he will also perform well on the Noun/Verb Test, but that the converse is not necessarily true.

A Pearson Product Moment Correlation Coefficient r was calculated to determine the degree of correlation between scores on the Noun/Verb Test and the Preference Test. The results are shown in Table 6.6.

Table 6.6: Pearson Product-Moment Correlation Coefficient r for Noun/Verb Test Versus Preference Test

Grade	Number of Subjects	r	Level of Significance
1	8	0.109	ns
2	27	0.416	$p<0.05$
3	16	0.741	$p<0.01$
4	25	0.291	ns
5	18	0.087	ns
6	20	-0.183	ns
7 & 8	11	0.000	ns
Total	125	0.599	$p<0.01$

Considering the test group as a whole the degree of correlation is very high, indicating that ability to perform the task in the Noun/Verb Test is related to ability to perform the task in the Preference Test. However, when one considers the correlation within each individual grade group, certain differences become apparent. As one would expect, in Grade 1, where most subjects performed randomly on both tests, there is little statistical correlation between scores on one test and those on the other.

In Grades 2 and 3 there is a close correlation between ability to perform well on both tests. In these grades, those who had the ability to distinguish nouns from verbs on the basis of stress and vowel reduction were also able to distinguish between the compound and phrasal minimal pairs; likewise, those who did poorly on the Noun/Verb Test did not do well on the Preference Test. In Grades 4 and above, almost all children had learned to recognize the noun-verb distinctions, but at the same time a number of them still lacked the linguistic sophistication to distinguish the compound and phrasal minimal pairs on the basis of stress patterns. Hence, there was not a high statistical correlation between the two tests in these grades. The over-all correlation, however, was highly significant.

II. The Contrastive Stress Test

The ability to disambiguate compounds and phrases on the basis of stress has been shown to be acquired relatively late in the language acquisition process. One possible reason for this is the interference of contrastive stress. As mentioned earlier, contrastive stress shift can cause the stress pattern of a phrase to become homophonous with that of the compound. Such a shifting of stress might confuse the child or make the stress patterns of compounds and phrases appear optional or unimportant, and hence contribute to later learning.

A number of studies have been made which indicate that children learn contrastive or emphatic stress very early. (See Chapter 1 for a review of this research.) According to the results of these studies, children generally have ability with most uses of contrastive stress by the age of six at the very latest, and probably much earlier. The earliest examples cited were about two years of age. This suggests that the children dealt with in the present study should all have been able to use contrastive stress since the youngest children tested were at least five years old. However, since contrastive stress plays a potentially important role in the acquisition of compound and phrasal stress, it was felt that some inquiry should be made into the question of whether the children in this study did indeed have competence with contrastive stress. For this reason, the following somewhat informal test, the Contrastive Stress Test, was conducted.

METHOD

Subjects

Twelve of the children who were subjects in the production studies were used in this test. (For details concerning subjects, see Chapter 5.) Table 6.7 lists these subjects by initials, age, sex, and grade.

Table 6.7: Subjects in the Contrastive Stress Test Listed by Initials, Grade, Sex, and Age

Grade	Subject's Initials	Sex	Age
1	NU	Male	5.10
	JM	Male	6.4
	JF	Female	5.11
2	SR	Male	7.3
	MM	Male	7.5
	FC	Male	7.5
	LH	Female	7.0
	JS	Female	7.2
3	ML	Male	8.3
	MU	Female	7.4
4	DS	Male	8.10
	SB	Female	8.8

Stimuli and Testing Procedures

Questions and test phrases were constructed to elicit answers using the six phrases tested in other experiments in this study. These test stimuli and the answers elicited are listed in Table 6.8. The phrasal answers

Table 6.8: Question and Answer Frames and Stimuli Used in Contrastive Stress Test

Question Frames

Experimenter:	Are these _____?	Child:	No, they're _____.
	red shóes blúe socks		red sócks réd socks
Experimenter:	Is this a _____?	Child:	No, it's a _____.
	green cár réd house		green hóuse gréen house
	black hámmer gréen board		black bóard bláck board
	high táble lów chair		high cháir hígh chair
	red árm blúe head		red héad réd head
	hot cát cóld dog		hot dóg hót dog

call for contrastive stress to appear on one or the other syllable. For example, if the experimenter pointed to a picture of a house colored green and asked the child, "Is this a green cár?" the child should answer, "No, it's a green hóuse." Likewise, if the experimenter pointed to the same picture and said, "Is this a réd house?" the response should be, "No, its a gréen house." In the first sentence, the phrase would have contrastive stress on the syllable where stress would normally appear. In the latter sentence, the stress would fall on the first syllable, making the phrase homophonous with the compound greenhouse.

Compound members of minimal pairs were not used in the study, since eliciting compounds would not indicate whether contrastive stress or compound stress was being used. For example, if the experimenter pointed to a blackboard and asked, "Is this a táble?" the child should answer, "No, it's a blácкboard." With such an answer there is no way to be certain whether contrastive or compound stress is being produced.

The 12 test phrases were embedded in the test frames, "Is this a_____?" or "Are these _____?" All 12 phrases were presented at least once. If the child made no errors, phrases were not presented again. If there were any errors, six of the phrases were presented a second time. The order of presentation was random. Each question was asked only once before the child was expected to answer unless the child had difficulty in responding, in which case the question was repeated.

This Contrastive Stress Test was given to subjects individually in conjunction with other production tests. However, in an attempt to prevent confusion with the compound and phrasal stress rules and the contrastive stress rule, it was given at the end of the second day of testing, after the conclusion of all other production tests and after some informal discussion which followed the last production task. The experimenter presented the questions to the children orally, pointing to the pictures of phrases used in the Identification Test and Semantic/Acoustic Test. (These pictures are found in the Appendix.) The entire test was recorded on a Wollensak T-1500 tape recorder. (See Chapter 2 for tape recorder specifications.)

RESULTS

As with the production tests, two adults listened to each child's responses and judged them as correct or incorrect. Results showed that 10 of the 12 children tested displayed the ability to use contrastive stress correctly. Six of them did so for all twelve phrases. The other four had only one to three errors in eighteen phrases. JF, a First Grader, missed the first two phrases but then,with the exception of only one item, gave all remaining phrases correctly. Second Graders JS, MM, and SR missed only one or two items.

Two of the four subjects corrected themselves in one case when they made an error; all four gave the impression of having missed a few items because of carelessness rather than lack of ability. All errors were with phrases that called for stress on the second element; the tendency to put stress on the first syllable appeared to be so strong that it was as though the children had to consciously attempt not to use that pattern.

The remaining two subjects who made a larger number of errors were First Grader JM and Second Grader LH. They made from 10 to 11 errors on 18 phrases. Again, in all cases the errors were with phrases that should have had stress on the second element. It was difficult to tell whether or not these two children had competence with contrastive stress, since they produced some answers correctly but made so many errors. Their failure may have been due to performance factors since this task was given at the end of a 20 to 25 minute testing session; they may have been too tired to produce the correct answers. Some support for this explanation can be found in the fact that although their error was always to put primary stress on the first syllable, it was seldom a strong, emphatic stress, and sometimes it was only slightly stronger than that on the second syllable. The correctly stressed phrases had stronger primary stress. Therefore the errors may have been due to the fact that these children did not always make the effort to respond correctly.

A second possible explanation is related to the fact that in the various production tests these two children were unable to correctly produce the distinction between the compound and phrasal minimal pairs. In the Production Test they both produced over 90 per cent of the stimuli with a compound stress pattern; in the Pair Production Test, both attempted to make a difference between the compound and phrase but were not successful in doing so correctly. Since they had almost always labeled these pictures of phrases with a compound stress pattern in the various tests, they may have been unable to pay attention to the Contrastive Stress Test questions and to break away from using the compound label. It is possible that they would have produced contrastive stress more successfully if they had been presented with different phrases and pictures, and if testing had been done when they were not tired and had not been given other tasks involving different (and potentially confusing)stress rules.

In spite of the inability of these two children to perform the Compound Stress Test task, it seems reasonable to assume that in general the contrastive stress rule has been acquired and is used correctly by the time children reach the age of five or six. With a few possible exceptions, then, it can be assumed that the children tested in these comprehension and production studies with compounds and phrases had all acquired rules for contrastive stress.

Chapter 7

Summary and Discussion of Test Results

The main purpose of this study was to gain information about the acquisition of stress that marks grammatical distinctions and that can be predicted by rule. Since earlier studies on the acquisition of stress in English have generally centered around the first few years of life, they do not deal with complex stress rules such as those which distinguish compounds from phrases. Stress data on children's language in the first few years at best deal only with use of lexical stress. Such stress phenomena may be only phonetic, probably do not mark meaning differences, and are not necessarily predictable by rule. Also, studies of stress in older children have focused on aspects of the use of contrastive or emphatic stress rather than unemphatic, or noncontrastive, stress. The present study was made in an effort to determine the acquisition of unemphatic stress patterns which signal differences of meaning. Children's knowledge of the stress patterns that distinguish compounds and phrases was tested; to a lesser extent, knowledge of the stress patterns which distinguish certain nouns and verbs was examined. Use of contrastive stress was also briefly studied, for comparative purposes.

The results obtained from all the experiments on stress comprehension and production were in basic agreement. This fact in itself is a good indication that the different tests offered reliable measures of children's competence with these stress contrasts. The results of the tests on comprehension of the stress patterns in compounds and phrases will be compared first: Identification Test 1; Identification Test 2 (an exact repetition of Test 1); Identification Test 3 (a partial replication of Test 1); the Preference Test; and the Semantic/Acoustic Test. The results of the comprehension tests with stress in noun and verb pairs will then be summarized and compared. Finally, production data of both noncontrastive(i.e., unemphatic)and contrastive stress with compounds and phrases will be considered. A discussion of the results of all these tests follows.

Summary of Results of Comprehension Tests with Compounds and Phrases

The main conclusion to be drawn from all comprehension tests with compounds and phrases is that the ability to identify or select compounds and phrases correctly and consistently on the basis of their distinguishing stress patterns is learned mainly as a function of age. Although a certain awareness of stress can be observed very early in life, ability with stress in compounds and phrases is not acquired until relatively late in the language acquisition process. Most of the basic syntactic and phonological distinctions are learned by the time a child has reached the age of six. However,

a child at five does not yet seem to show linguistic ability with these stress contrasts; one at twelve does; and the older a child under twelve is, the more likely he is to have acquired it. The Semantic/Acoustic Test showed that the ability to ignore the conflicting sentential context and to correctly identify the compounds and phrases on the basis of stress alone is acquired later than the ability to perform other tasks. The percentage of correct responses made by children in Grades 1 - 6 was considerably lower in this test than in any other. Presumably this was because children who did not have a firm and well-established knowledge of these stress patterns were not able to consistently ignore the tendency to use contextual information to assign a semantic reading to these compounds and phrases; children in Grade 7 & 8 always used a significant number of acoustic cues, and they were the only grade to have near-perfect scores in the other comprehension tests.

Another general finding of these experiments is that a certain amount of individual variation exists at all levels. A First Grader may perform these tasks as well as adults, and a Fifth Grader may do as poorly as the youngest children tested. The reasons for this variation cannot be determined on the basis of these data. Children in the study varied in I.Q., socioeconomic status, educational background, etc., and factors such as these could be partially responsible for the differences in performance. Gleitman and Gleitman (1970) demonstrated that not all speakers of English necessarily have identical grammars. In testing adults' ability to paraphrase noun compounds, they found evidence that secretaries did not follow the same grammatical rules as did graduate students. In the present study, however, no evidence was found of adults' differing in ability to correctly select and identify two-word compounds and phrases by means of stress patterns. Nonetheless, nonlinguistic factors such as socioeconomic status could contribute to differences in the age at which this ability is acquired, a phenonemenon which has been noticed with other apsects of speech and grammatical development (c.f. Templing, 1957; Lenneberg, 1965; McCaffrey, 1969, 1970.)

Very little evidence of significant learning effects was found in any of the tests. No grade improved significantly on the second trial of any item or items. A small, though not significant, tendency existed for the younger grades to improve in the second half of Identification Test 1, and only one subject who took Identification Test 2, a Second Grader, improved significantly on the second trial. Identification Test 3 and the Preference and Semantic/Acoustic Tests also showed no significant learning effects. In the latter, only two of the 125 children made a major switch from dependence on semantic cues to dependence on acoustic cues and showed any evidence of "learning." Thus, additional practice with the task was apparently not an important factor;in general, children who initially failed in the task did not improve with practice.

There is some question as to the importance of sex as a factor in the acquisition of compound and phrasal stress patterns. In Identification Test 1, the girls tended to score better than the boys in the grades up to Grade 5. In Identification Test 3,girls also tended to perform better than boys in Grades 3 through 6, but in Grade 2 boys performed better than girls. (There were no girls in the Grade 1 sample who took Test 3.) However, in Test 1 only Grade 4 showed significant differences between the performance of boys and girls, and in Test 3 none of the differences was significantly better than chance. The Preference Test results indicated that the difference between the performance of boys and that of girls was significant only in Grade 5, where girls performed better. In this test, boys performed better (though not significantly so) in all grades except Grades 2 and 5. (Again, no girls were tested in Grade 1.) Thus further doubt is cast upon sex differences as an important factor here. On the other hand, the Semantic/Acoustic Test results indicated that boys in Grades 3 through 6 generally were slower than girls in learning to ignore the sentential context and utilize the acoustic cues. Therefore, the importance of sex as a general factor in ability with these stress patterns is not clear. It is not unusual to find differences between girls and boys with regard to some aspects of language development. (See McCarthy, 1954, for references.) Generally, however, the differences according to sex that have been noted occur with aspects of speech articulation and vocabulary rather than with acquisition of grammatical rules. For example, Templin (1966) notes a difference in the performance of boys and girls in speech articulation studies, but Berko (1958) found no sex differences in her study of English morphology.

In all these tests little evidence that response biases affect the general results was found. In no case was the existence of a bias toward selecting certain syntactic structures or toward selecting pictures or letters in certain positions on the answer sheet statistically significant. However, in the Preference Test a tendency existed to select the compound stress pattern more often than the phrasal one. (This bias agreed with the production data findings that most children who did not display linguistic ability with these stress patterns in comprehension tests tended to produce primary stress on the first syllable of all test stimuli in the production tasks.)

The difficulty with certain members of minimal pairs did not agree in all tests. In Identification Test 1, the Kindergarten and two First Grade groups found the phrase <u>black bóard</u> to be significantly more difficult than its compound counterpart. (These were the only significant differences between members of pairs in any grade or any test, however.) The Test 1 group as a whole displayed a slight (though not significant) tendency to select the compounds <u>hót dog</u>, <u>blácкboard</u>, and <u>híghchair</u>, and the phrases <u>red sócks</u>, <u>red héad</u>, and <u>green house</u>. Apparently, however, this was because a small group of children polarized their responses to one member of a pair. Identification

Test 3 did not show the same biases. The only two pairs with which the subjects showed a tendency to select one member of the pair more often than the other were redhead and greenhouse; and in both cases it was the compound that was preferred rather than the phrase, as in Test 1. In the Preference Test, the compounds gréenhouse and blackboard tended to be selected more often than their counterparts. Thus the bias toward selecting one member of a minimal pair appear to be more individual than general, if it exists.

Likewise, the tendency to find certain minimal pairs more difficult than others was more individual than general. However, in general the pairs Red Sox - red socks and greenhouse - green house appeared to be more difficult in both Identification Tests. This same ranking was not apparent in the Preference Test, however. The conclusion was drawn that the problem of picturing these items was responsible for the difficulty in the Identification Tests.

Since the same children were used as subjects in Identification Test 3 and the Preference and Semantic/Acoustic Tests, a correlation can be made of individual performance on these tests. In general, a child who performed well on the Identification Test was likely to do well on the Preference Test; the reverse was usually, though not always, true. If a subject used acoustic cues for the Semantic/Acoustic Test in a significant number of his responses, then in general he had sufficient linguistic ability to score a significant number of correct versus incorrect answers on both the Identification and Preference Tests. If he fluctuated randomly between dependence on acoustic and semantic cues, he apparently had some knowledge of these stress patterns and may have performed significantly on one or both of the other tests. If he used only semantic cues in his responses, he may or may have not had the linguistic ability to give significantly more correct than incorrect answers on the other tests. If he performed well on the additional tests, he apparently had the necessary linguistic knowledge to attend to acoustic cues; his use of semantic ones may have reflected either that he had not paid attention to the instructions, or that his knowledge was insufficiently well established to permit him to ignore conflicting semantic cues. Therefore he followed the strong natural tendency of listeners to depend on semantic cues and exclude contradicting acoustic ones. As noted before, in normal language use listeners interpret the message by using syntactic and semantic, as well as phonetic knowledge; they generally look for the over-all meaning of the sentence, not for phonetic detail.

Summary of Noun/Verb Test

The results of the Noun/Verb Test suggest that the ability to select the correct member of these contrasting pairs is mastered gradually and mainly as a function of age, as was the case with contrasting compound and phrase pairs. Very few children differed greatly in performance from the

average performance of the other children in the same grade, which suggests that age is a more crucial factor here than differences in experience or background. Children in Grades 1 and 2 displayed some ability with this task, and those in Grades 3 and above displayed near perfect ability.

Differences according to sex did not appear to be an important factor in the acquisition of this knowledge. In fact, there was less evidence of differences between the performances of boys and girls of various ages in this test than in any of the tests with compounds and phrases. The existence of response biases was not easy to establish because of the small number of items in the test.

Ability with stress in the noun and verb pairs appears to be acquired earlier than ability with stress in compounds and phrases. Children who performed well on the latter test also tended to perform well on the Noun/Verb Test, but the converse was not always true. Exactly how early stress ability with nouns and verbs is acquired cannot be answered on the basis of these data, since the youngest children tested already displayed a certain degree of competence with these structures. This was not true with compounds and phrases since most children in the youngest grade did not score a significantly-greater-than-chance number of correct responses. However, ability with nouns and verbs appears to be acquired later than knowledge of stress in the items table and window, since the youngest children performed better with the latter.

Summary of Production Data

The judgments of two adult listeners indicated that the majority of the children, particularly the youngest, failed to produce compound and phrasal stress correctly; in most productions of test stimuli primary stress was placed on the first syllable. This tendency to produce stimuli with the compound stress pattern was most evident among the younger children who had not succeeded in the earlier comprehension tasks. Children appeared to have earlier success in correctly producing both members of a minimal pair in succession, emphasizing the differences (the Pair Production Test), than in producing each member randomly in isolation (the Production Test). Although the youngest failed to produce these stimuli consistently and correctly, children of all ages were able to imitate the stress patterns of the compounds and phrases correctly. Also, children who were successful in producing the test contrasts correctly were also successful in comprehending them, but the reverse was not always true.

Comprehension tests of self-produced compounds and phrases indicated that children did not perform better with their own productions of test stimuli than with the experimenter's. Children who failed to perform well in comprehension tests which required identification of the experimenter's productions of test stimuli performed at the same level of accuracy, or had even more

errors, when attempting to identify their own productions of these stimuli. Therefore no evidence existed that such children failed with the experimenter's production because they had unique rules for signaling the phonetic distinctions between compounds and phrases.

Tests with contrastive stress in phrases indicated that in general even the youngest children were able to produce this kind of stress. On several occasions there was evidence of a strong tendency to produce phrases as compounds when the children did not make the effort to produce the contrastive stress. Most of these children were able to correct themselves either immediately or when the stimulus sentences were repeated.

Some Possible Explanations

The results of all the tests suggest that age is the most important factor in the ability to correctly and consistently identify or select the proper compound or phrase on the basis of stress. Since children generally do not begin to acquire ability with these stress distinctions until after the age of five, these data support Brown's (1969) findings that in the two-word stage children do not consistently use suprasegmentals to mark semantic relations; furthermore, they suggest the need to re-evaluate the claims that two-year-olds use stress contrastively to mark grammatical relations. (For example, Miller and Ervin, 1964, and Ervin, 1966, cite examples like "Chrísty room" and"Christy róom", where stress signals possessive versus locative; Weir, 1962, reports that stress distinguishes the compound from the phrase in "Twófeet" versus "twó féet." See Chapter 1 for discussion of and alternative explanations for these data.)

Several possible reasons can be suggested as to why older children should perform better with these tasks. First, it may be that nonlinguistic factors like degree of mechanical ability or length of attention span are responsible for the differences. Perhaps the older a child is, the better he can perform these mechanical tasks. Although such factors as these undoubtedly account for some of the errors made in the tests, it is unlikely that they are the major determinants of the developmental trend that was observed. One finding which supports this conclusion is that the Task Comprehension Test indicated that even the youngest children were able to perform the task used in the Identification Test successfully.

Since the testing conditions in this test were the same as in the Identification Test, the only reason why the identification task was more difficult for young children appears to be a linguistic one. Another argument against the importance of such nonlinguistic factors (other than age) is that all children performed at a high level of accuracy with the dummy items used in the Preference Test. Since these items were interspersed with regular test items at the beginning and end of the task, it is unlikely that factors like a short attention span and mechanical difficulties could have influenced only the test items and not the dummy ones. A third argument against the importance of such nonlinguistic factors is the high degree of consistency in the responses of the youngest children in the Semantic/Acoustic Test.

Children had little difficulty in attending to the same type of cue and in responding to it accurately. Also, the consistency which all children demonstrated in the different test tasks suggests the absence of important nonlinguistic factors of this type. Since the errors made on the second half of Identification Test 1 were no greater than on the first, a short attention span was not the problem with the youngest children. One may also suggest that the performance of all children, and particularly the youngest,was adversely affected by the fact that they were tested in small groups rather than individually. This objection is answered by the comparative data on children tested under both conditions; these data give no evidence of any important differences with most of the children. Thus, although it is likely that nonlinguistic performance factors such as mechanical ability or attention span may have been responsible for failure in some cases, it is obvious that the influence of such factors was quite small.

A second hypothesis that might be proposed to explain the superior performance of older children with compound and phrasal stress patterns is that they may have had more of the test items in their vocabularies than the younger children had. The main evidence against vocabulary being the key factor comes from the Semantic/Acoustic Test. Since the youngest children invariably selected the compound or phrase which agreed semantically with the sentential context, it is clear they had a fairly good concept of the meanings of the test stimuli.

The most obvious explanation for the developmental trend which resulted is that it is a reflection of differences in linguistic ability with the stress contrasts tested. However, it is not obvious exactly what this ability reflects. It has been noted that children of five do not have sufficient ability with these stress patterns to perform the tasks well; at age twelve they do. Between these age extremes, children gradually master this skill. But exactly what are they acquiring and how do they do it?

As noted in Chapter 1, at least two possible hypotheses can be posited to explain how children acquire this skill. The first is based on Chomsky and Halle (1968). According to their view, lexical items such as compound nouns learned by the child are not marked for stress; instead, he hypothesizes stress rules which, with the aid of the innate principle of the transformational cycle, assign stress contours on the basis of syntactic information. Lexical, phrasal, and sentence stress are all assigned by stress rules. Stress is assigned to compounds by the Compound Rule and to phrases by the Nuclear Stress Rule, making the perception of stress contours and presumably the acquisition of stress rules heavily dependent upon a speaker's knowledge of the grammatical structure of the language. This grammatical knowledge is therefore a basic prerequisite to the child's learning of these stress rules. According to this hypothesis, a child's ability to distinguish these compounds and phrases on the basis of stress depends on whether or not he has learned the rules which assign stress.

An alternative hypothesis is that children do not learn or operate with stress rules; rather, they learn these lexical items with associated stress patterns and then store them in memory. The compound is learned as a single lexical item with primary stress on the first syllable. The phrase has primary stress on the second syllable as a result of a general sentence intonation rule which assigns stress to the nuclear syllable of the sentence. This rule would assign stress to the last stressed syllable in sentences, and in the case of noun phrases, to the last item of the phrase. Under this hypothesis, children are able to differentiate between pairs such as gréenhouse and green hóuse on the basis of stress first, when they have learned that it can be used to make lexical distinctions, and second, when they have learned the lexical items and the associated stress pattern for each item with its different meaning. A child who associates only one meaning with each of these pairs would not attend to stress differences, and would thus fail in tests with pairs he had not yet learned. Additionally, a child who associated two meanings but did not know that stress could be used to make lexical distinctions would also fail. The possibility of compounds being learned as lexical items instead of being generated by rule is not impossible, particularly in light of the findings of Berko (1958) and Livant (1962), which indicated that some compounds with set or "frozen" meanings were first learned as single nouns; young children failed to recognize the constructional features of these compounds. Since these researchers also found evidence of compounding rules as children produced unique compounds, and since evidence of rule existence has been demonstrated with adults (c. f. Gleitman and Gleitman, 1970),rules must be acquired and used at some time for novel compounds; however, we do not know how general the rules for compounding are. It is not essential that children have rules for the compounds studied here. For example, young children may not realize that hót dog is a compound composed of the two related elements hot plus dog, just as they are not confused by the fact that window could be a compound composed of the two elements win plus dough (e.g., "Try the wíndow" versus "Try to wín dough" (i.e, not lose it).

It is also possible, and even probable, that some variation of both hypotheses are correct but account for different stages of the acquisition process; a child probably learns some or all of these compounds and phrases as lexical items in the beginning and stores these stress patterns in memory until some later time when he learns the rules; then stress in all these items could become predictable by rule and no longer need to be stored. However, it is possible that even after rules for stress assignment are learned, some compound stress patterns remain stored in memory while others have stress assigned by rule. For example, stress in blácкboard might become predicted by rule, as a child becomes aware that it is a compound composed of the related elements black and board. But stress in hót dog may never be assigned by the Compound Rule; it may remain a memorized stress pattern for a single lexical item. In view of the Gleitmans' (1970) findings,

it is even possible that not all speakers are the same with respect to this question; some may use rules to assign stress to compounds, others may not. (Ability with the noun-verb distinction could also reflect either learning of a stress assignment rule or memorizing and storing lexical stress patterns. Since single items are involved, it is especially possible that stress patterns could be memorized and stored rather than being assigned by rule.)

Although there is no obvious way to discover whether lexical items with memorized stress patterns are no longer stored in memory after stress rules are acquired, it should be possible to find out whether children are at the stage where they are learning rules and generalizing to new items, or whether they are learning stress patterns item by item. These two hypotheses make different predictions about what kinds of error a child in the process of mastering this skill should make. If the child is in the stage of learning rules, any errors he makes should be random and evenly spread over all lexical items. This is because such errors should mainly reflect performance factors. On the other hand, if the child is learning lexical items, until he has learned all items he should tend to make errors on those which he has not yet learned and to identify correctly those which he already knows. (Again, a certain number of his errors could be attributed to performance factors and would have to be ignored.)

Although these hypotheses seem perfectly testable, unfortunately the data from these experiments do not seem to point to one or the other of them. A comparison of the errors made on Identification Test 3 and on the Preference Test by subjects who had scores of approximately the same range of accuracy indicated no significant positive correlation on errors made in the two tests (as was determined by chi-square), which would provide support for a rule hypothesis. (In one instance, in fact, there was a negative correlation.) On the other hand, the Spearman rank-order correlation coefficients (rhos) for the error patterns of subjects who took both Identification Test 1 and Identification Test 2 (the repeat test) indicated a weak but consistent lexical item effect. If errors were totally random, we would expect as many negative correlations as positive correlations, but there were far more positive than negative ones. And, in fact, the Wilcox matched-pairs signed ranks test with the rhos of these repeat group subjects matched to theoretical values of zero indicated that the tendency for subjects to have positive correlations was significant at $p<0.01$. Thus, there is no firm basis on which to decide between these two hypotheses. The same ambiguity was evident with the examination of individual error patterns on the two tests. Some children showed a clear tendency to miss certain items and not others, while others made errors randomly and performed at the same level of competence on all items. It is very hard to tell whether these patterns are due to chance or whether they reflect genuinely different processes on the part of individuals.

The results of all the experiments in this study with both compounds and phrases as well as with noun and verb pairs indicate that the ability to correctly identify or select items on the basis of stress patterns is acquired gradually. For example, it was fairly clear from the gap between those subjects having 70 per cent correct responses and those having 85 per cent correct in Identification Test 1 (see Figure 2.5) that some sort of gradual learning was taking place. Also, the distribution of error scores (i.e., per cent correct on a given lexical item) of individuals taking both Identification Tests 1 and 2 seemed inconsistent with an "all or none" learning hypothesis; the same seemed true of the analysis of a larger number of responses to hót dog on Identification Test 3 and the Preference Test (see Table 3.4).

Two different explanations can be offered for this gradual learning phenomenon under the two hypotheses. Under the lexical learning hypothesis, if lexical items and their associated stress patterns are being learned, it is reasonable that this would be a gradual process. Also, as errors occur in the process of learning, the item missed may not yet have been learned semantically (though this seems unlikely because of the Semantic/Acoustic Test results); more probably,however, its associated stress pattern and the fact that it can signal a meaning difference has not yet been learned. Also, it could be missed because the stress pattern has been very recently associated with the lexical item, and thus the item and its stress pattern might not be as well established as those learned earlier. The possibility of forgetting or deleting newly acquired items is so common that it has been suggested as a possible universal of language acquisition (Slobin, 1970a): "It is categories which have been most recently acquired which tend to be deleted from speech under pressure of performance limitations" (p. 17).

In order to explain the gradual learning of rules, we would have to say that although the rules were acquired fairly early,they were not perfected and used consistently until considerably later. A child who can disambiguate the minimal pairs used in these experiments 70 per cent of the time would have to have internalized the rules in order to do so well. Yet since he still has 30 per cent incorrect responses, he is not applying them consistently. It is possible that the rules the child has differ from those of the adult in that they are optional rather than obligatory. But if they are optional, what are the conditions under which they do or do not apply? The data give no answers as to what these conditions might be or as to why the rules were not applied in all cases. Possibly the rule is the same and children possess competence with these rules but fail in performance. In normal language use, adults make many performance errors, and children must also. But can all errors be attributed to performance factors? Since the number of errors made with dummy items and with the Task Comprehension Test was considerably lower than in the case of the regular test items and the regular test, additional factors must be involved. Usually when one thinks of rule-governed behavior, it is more in "all or nothing" terms. It has been assumed by some that a rule, if learned, should be applied always and in all cases

rather than just part of the time. It should be acquired suddenly, not gradually. This may not be the way all rules are learned, however. Brown (1970) and Blount (1969), in observational studies, both note that children learn grammatical inflections gradually rather than instantaneously. Brown reports that with Adam, Eve, and Sarah, none of their grammatical morphemes were acquired suddenly, going from total absence to reliable presence: "There is always a considerable period, varying in length with the particular morpheme, in which production-where-required is probabilistic" (p. 16). Also, the proportion of times a grammatical morpheme is present gradually rises with time. He considered the morpheme fully acquired when it was present in 90 per cent of obligatory contexts in three successive 2-hour samples. In spite of the assumption that rule acquisition should be sudden, he claims that children are, nonetheless, learning rules:

> The variable character of the grammatical morphemes suggests a kind of rule learning that proceeds on a rather molecular level and gradually rather than abruptly. In certain respects, then, it is like operant conditioning or habit acquisition, and yet there is no doubt that what is learned is a set of rules since the responses generalize in just the ways that the rules describe. Very likely this is only one of several, or even of many, kinds of learning involved in the total process of acquiring a first language (p. 239).

Fischer (1971) also notes that some children in her study gave evidence of knowing dative constructions, but they had not yet effectively mastered them: "Children know the rule, but something prevents them from using it very well" (p. 148). Since this gradual learning or gradual mastery phenomenon appears to be present with other kinds of rule-governed behavior, its occurrence with stress rules may not be a serious objection to the rule-learning hypothesis.

Although these experimental data do not establish with certainly exactly what is being learned, they clearly indicate that children do not become proficient at distinguishing compounds and phrases on the basis of stress until relatively late in the language-learning process. There are several reasons for this, and, perhaps, for why it might be gradual.

One reason is that the phonetic patterns which differentiate compounds and phrases may not always be realized in the acoustic signal. This is true even with adults, whose speech serves as a model for the child. Adults may not always make the distinction in production because of various performance factors, and because the semantic content of the sentence usually carries enough information to make comprehension possible without the help of this acoustic distinction.

A second reason they may be learned late is that there are numerous exceptions to the rule that primary stress is on the first syllable in compounds, and on the second in phrases. One such example is Fifth Street versus Fifth Avenue which have different stress patterns but no syntactic justification

for calling one a compound and one a phrase.* Li (1970) notes many examples such as this. If rules are being learned, a large number of such exceptions might obscure the generalization and make it necessary for the child to have a fairly large sample of compounds and phrases before he could learn the rules. If lexical items with associated stress patterns are being learned, however, this should not affect the time it takes children to learn that stress can be used to signal different meanings; furthermore, these exceptional items would be treated in the same manner as other compounds and phrases.

Another possible reason for a child's late acquisition of the ability to use stress patterns to distinguish compounds and phrases is that this might be connected with learning to spell, and spelling is learned fairly late. Since preliterate children do not use compound and phrasal stress differences, and since the stages at which children begin to use them occur when they have learned how to read and spell, it may be that there is a connection between learning these kinds of skills. Certain kinds of linguistic knowledge could be connected with the ability to spell. For example, Moskowitz (forthcoming) claims that children have no knowledge of vowel shift rules except insofar as they know about long and short vowels in spelling. In the case of these stress distinctions, however, it seems unlikely that spelling would have any bearing on the question because stress is not marked in English orthography. Since compounds are very often (though not always) spelled as one word while phrases are written as two, presumably spelling could call attention to the fact that two separate referents are involved, which would aid a child who had failed to notice that the combination of these two words could be either a compound or a phrase. But orthography gives no simple clues as to which syllable receives primary stress, and thus it could not provide information concerning the stress patterns associated with compounds and phrases. Also, Read (1971 and personal communication), in a study of unique phonetic spellings made by preschool children, found no evidence of children creating ways to indicate stress or intonation in spelling. Spelling, therefore, does not appear to be a crucial factor in learning these stress patterns.

Another reason this ability could be acquired late and, particularly, in a gradual manner is that children may have a bias toward selecting certain lexical items more often than others. Such biases could be due to either familiarity or structural preferences. If a rule is involved, biases could at first prevent the stress rule from being applied consistently; they then could later be given up item by item. According to a lexical hypothesis, such biases would determine the order in which the stress patterns of lexical items would be memorized and stored. Since only 17.5 per cent of the Identification Test 1 group showed a tendency to polarize their responses and to select only certain members of minimal pairs for even one pair, some support is given to a possible familiarity bias explanation, but it is not strong. On the other hand, most children displayed a bias toward producing

---

*Chomsky and Halle (1968) handle examples such as these with readjustment rules. Surface structure modification occurs so that Fifth Avenue is not dominated by the node "noun."

stress on the first syllable, and some tendency existed to select this pattern more often in the Preference Test. As noted earlier, the source of this bias is not clear. It could either reflect preference for the compound structure or it could result from contrastive stress interference.

Possible contrastive stress interference is probably the most important reason for children's late and gradual acquisition of the ability to use stress to distinguish compounds and phrases. As discussed earlier, contrastive stress shift negates the phonetic stress differences between compounds and phrases by making the phrase and the compound homophonous. This homophony could confuse the child and make the learning of the compound and phrasal stress distinctions more difficult. If rules are being learned, its interference might prevent consistent application of the more recently learned compound and phrasal stress rules. Or it might make these later rules look as though they apply randomly or with optional conditions. With either the rule or lexical-learning hypothesis, contrastive stress shift would create conflict that could help to explain late learning. If a child learns contrastive stress early, as the studies reviewed in Chapter 1 suggest, and if he uses this type of stress to mark grammatical relations like topic and comment, then it is not unreasonable to suggest that this might interfere with the learning of stress distinctions in compounds and phrases.

Support for this conclusion is found in the data of Hornby (1970, 1971). As noted in Chapter 1, he studied contrastive stress and other surface structure devices as they mark the topic-comment relation. He gave children sentences similar to the following:

1. The Indian is building the igloo. (active)
2. The igloo is being built by the Indian. (passive)
3. It is the Indian who is building the igloo. (cleft agent)
4. It is the igloo that the Indian is building. (cleft object)
5. The one who is building the igloo is the Indian. (pseudo-cleft agent)
6. What the Indian is building is the igloo. (pseudo-cleft object)
7. The Índian is building the igloo. (stressed agent)

For each stimulus, two pictures were given to children. With sentences such as these, one picture would depict an Indian building a teepee and the other an Eskimo building an igloo. Children were told that they would hear something about one of the pictures but that something would be wrong with what was said. They were asked to select the picture which the sentence was about and then to tell the experimenter what should have been said about the picture. His results showed that contrastive stress was the primary device used by First, Third, and Fifth Graders to signal the comment, but it was not equally important to children at these different ages. Production task results indi-

cated that the tendency to employ contrastive stress to mark the comment declined as a function of age, whereas the frequency of correct passive, cleft, and pseudo-cleft increased with age. The mean number of sentences in which the topic-comment relation was correctly marked by stress in children's productions was approximately 16.5, 13.5, and 12.0 for Grades 1, 3, and 5, respectively. Similar results were found in comprehension tasks, since selection of topic for contrastively stressed sentences declined in frequency between the First and Third Grades. However, between Grades 3 and 5 there was an increase in the selection of the topic in sentences with contrastive stress. He concludes that the change in pattern of response from First to Third Grade is due to the emergence of the role of temporal order as the major determinant of the topic-comment distinction. From Third to Fifth Grade the general importance of word order declined.

Hornby's finding that after the First Grade the use of contrastive stress to mark the comment declines in importance as use of other devices for marking it increases corresponds well with data of this study of stress in compounds and phrases. It supports the hypothesis that learning of stress in compounds and phrases is perhaps hindered by the contrastive stress rule, since it is at the time when the importance of contrastive stress to mark the topic-comment relation declines that the ability to use stress in compounds and phrases begins to increase. Hornby found in both comprehension and production studies that children depended heavily on contrastive stress in Grade 1; we found that in Grade 1 children performed very poorly in distinguishing compounds and phrases on the basis of stress. In Grade 3 Hornby discovered that in comprehension tests word order emerged as the major determinant of the topic-comment distinction; we found that in Grade 3 most children show statistically significant but not perfect ability to use stress to distinguish compounds and phrases. In Grade 5 Hornby found that word order declined in importance in marking topic-comment relations and contrastive stress increased slightly in importance, but in production, contrastive stress was less important than in Grades 1 or 3; we found that in Grade 5 children performed at a very high level of accuracy with compounds and phrases. Thus Hornby's data add support to the hypothesis that contrastive stress can affect the learning of stress in compounds and phrases, since the time when the child's focus of attention is turned away from contrastive stress in topic-comment relations is the time when he begins to acquire the ability to use unemphatic stress to distinguish compounds and phrases.

Among the reasons why a young child may attend to contrastive stress more than he does to stress distinction between compounds and phrases is that at early stages the child depends heavily upon the semantic interpretation he gets from lexical and syntactic information, and contrastive stress is closer to semantics than is unemphatic stress; that is, it is more directly related to superficial meaning without involving deep grammatical analysis. The fact that semantic information is crucial to young children was apparent in the Semantic/Acoustic Test; younger children were locked in on semantic cues and the overwhelming majority used them, ignoring the stress patterns of compounds and phrases. Other researchers have also noticed that children in the process of grammatical development are heavily influenced by semantics (e.g., Bates, 1969; Fischer, 1971).

The Order of Acquisition

The interaction of contrastive stress and the learning of stress distinctions in compounds and phrases suggests interesting questions concerning the order of acquisition of stress. On the basis of the results of these comprehension and production experiments and the information on stress acquisition available from other studies, the following hypothesis concerning the learning of stress can be postulated.

The first occurrences of what we call stress in adult language appear very early in infant productions. Infants are also responsive to the acoustic correlates of stress. However, children's use of stress in the prelexical, babbling stage is a phonetic one. It is not predictable by grammatical rule, and it does not mark grammatical distinctions or meaning differences. In this study even the youngest children were able to imitate stress patterns of compounds and phrases correctly. This ability to perceive and produce stress in imitation is much like that of a child in its first year; stress is used in a phonetic sense but not in any phonological way. When the child begins to learn lexical items, he probably imitates and memorizes the stress patterns along with other phonetic properties. Such stress patterns are therefore still unpredictable by rule.

Most likely, the first rule of stress that a child uses is that of contrastive stress, a rule that has not yet been described in a generative grammar. Exactly how early this rule is learned cannot be determined at this point, but it probably appears at least as early as the second year of life.* If stress is used to make any grammatical distinctions in the two-word-utterance stage, it is most probably focus or contrastive stress that is used to signal these distinctions. Furthermore, possible distinctions made by stress in this two-word stage like the "Chrísty room--Christy róom" and "báby book"--"baby róom" examples cited by Miller and Ervin (1964) and Ervin-Trip (1966) are probably not distinctions marking possessive and locative, but the more general topic-comment relation (see Chapter 1 for further details). Children at this stage are probably using contrastive stress to mark the comment--i.e., the new information. Contrastive stress and focus plus memorized lexical stress patterns are thus probably the only kinds of stress the child has for at least the first two or three years of life, and possibly longer. (It is not clear when he acquires rules for producing sentence stress; possibly for a long time he uses no linguistic stress as in the adult grammar with sentences, but only emphasis and focus.)

---

*Some uses of contrastive stress may not be learned until much later, of course. For example, the use of contrastive stress with pronominal reference and focus of sentence negation tested by Chomsky (1971) and described in Chapter 1 may be later acquisitions.

At some later point, the child begins to use unemphatic stress to make meaning differences. He then begins to acquire ability with stress in compounds and phrases. This occurs after the age of six and coincides with the time when he turns his focus away from contrastive stress as the primary way of marking the topic-comment relationship and begins to depend more on alternative devices such as word order to signal this distinction. It is at this stage, when his attention is diverted from contrastive stress, that learning of the compound-phrasal stress distinction begins. As contrastive stress continues to become less predominant in marking the topic-comment relationship, it apparently plays less role in interfering with the learning of compound and phrasal stress and makes it possible for ability with this latter type of stress to be acquired and gradually perfected. According to a lexical-learning hypothesis, the stage at which the child begins to learn the compound and phrasal stress distinction is when he notices that words like blackboard and black board, which have different meanings, can have these lexical differences signaled by stress. This is when he learns the proper stress pattern for both members of such minimal pairs. This learning would take place gradually, item by item; for example, a child would now learn the appropriate stress patterns to associate with greenhouse and green house, having learned the two meanings and probably one of the stress patterns earlier. On the other hand, according to a rule-learning hypothesis, at this stage meaning differences could now be signaled by stress because the child has begun learning rules like the Main Stress Rule, which accounts for stress in noun and verb pairs, and the Compound Rule and Nuclear Stress Rule, which assign stress to compounds and phrases. As the rules are learned and their use is perfected, the stress patterns for each lexical item are no longer just memorized, but, as the rules take over, predictable. This process would be analogous in some ways to a child's having a lexical items like feet or went which had been learned and used as lexical items before the rules for inflection of plural and past were learned. Once these rules are learned, the child drops the learned lexical items, and instead *foots and *goed appear as the rules are applied, in these cases incorrectly. The child then must re-learn feet and went as exceptions to these plural and past rules which he has now learned. In the same way he must learn the exceptions to the Compound and Nuclear Stress Rules (e.g. Fifth Street but Fifth Avenue). This analogy is not perfect since not every word that a child learns has a plural, but every word has stress. Also, whereas children sometimes drop a learned irregular plural like feet when they first learn the plural rule, there is no evidence that they drop an irregular stress pattern when a stress rule is first learned. The few data available on children's use of stress at this stage suggest that children do not generally use stress patterns incorrectly in this way, and they apparently do not make major errors with lexical stress (c.f. Weir, 1962).*

* For example, in an informal experiment with a three-year old girl, we found that she made no errors in placing stress on multisyllabic nouns when naming pictures. Also, in an observational study with tapes collected by Dr. Margaret Bullowa, we found that generally speaking, children do not misplace stress. (See Bullowa, Jones, and Bever, 1964, for a description of these tapes).

When the child begins to master the ability with stress in compounds and phrases, he is able to perform certain experimental tasks before others, which again suggests gradual mastery of ability. Most types of comprehension skill precede those of production and follow the ability to imitate these contrasts, which is present very early. In general, the child is first able to perform a preference task, selecting the stress pattern which "sounds best" when given both the right and wrong patterns to compare. He later gains skill in identifying compounds and phrases in isolation on the basis of stress alone. Still later, he starts to develop production ability, following a similar sequence to that observed with comprehension skills. First he is able to produce these compounds and phrases correctly when he says both members of a minimal pair in succession and can emphasize the difference. Later he can produce each member correctly in isolation. At this point he can identify his own productions, since this comprehension skill was learned earlier. After he is able to perform all these comprehension and productions tasks almost perfectly, he is finally able to perform the Semantic/Acoustic Test task; that is, he can use the stress patterns to identify compounds or phrases when they are embedded in a contradictory sentential context. This final task is generally not mastered until the child is 12 years old, though some children reach this point earlier. Individuals vary with respect to the age at which they acquire these skills, but the order of acquisition appears quite consistent. (The best evidence to support this is seen in the results of the Production studies, discussed in Chapter 5.) The time gap between each stage was not determined; it may also vary to some extent. With some children there may be very little time between these steps, with others more; and in some cases the gap may be small enough to make the steps appear simultaneous. Whereas the general order of acquisition appears to be imitation, comprehension, and production, in the case of the Semantic/Acoustic Test, comprehension follows production. The difficulty and type of task should therefore be taken into account in any consideration of order of acquisition.

The findings of this study, particularly concerning order of acquisition, are specific to children learning English, since rules like the Compound and Nuclear Stress and Main Stress Rules account for stress in English. However, it is possible that certain general aspects of this order are universal. Perhaps in all languages with stress, the child begins with a phonetic awareness of stress; then he learns to use emphasis or contrastive stress; finally and considerably later, he acquires the ability to use stress to mark finer grammatical distinctions.

Suggestions for Future Research

The results and conclusions of this study suggest a number of hypotheses and unanswered questions which should be investigated further. Additional studies of children's acquisition of stress in compounds and phrases should be aimed at determining to what degree children acquire rules for stress placement as opposed to memorizing and storing stress patterns item by item, since this question was not clearly answered by the data provided by this

study. The best way to determine use of rules would be to use nonsense words; however, as discussed in Chapter 2, it is extremely difficult to find tests with nonsense words where contrastive stress could not interfere. Sorting out compound stress and contrastive stress is perhaps the greatest difficulty in studying these stress contrasts, as has been suggested throughout this study. If no good way can be found to test these stress distinctions with nonsense words, the method of examining lexical error patterns can be used, as was done here. With such a method, many tokens of the same type should be used with as many children as possible, in order to obtain a clear picture of existing trends. Perhaps using nonsense words in conjunction with lexical error pattern examination of real words would produce the best results, and would suggest the stage at which children can generalize stress to nonsense words.

Other kinds of task could also be employed in future studies with compound and phrasal stress. A paraphrasing task is a possibility for comprehension studies, though it may prove too complicated with very young children. It may be hard to separate children's errors in stress comprehension from their difficulty in accurately defining what they hear. Tasks like those of preference and identification may be more reliable as tests of stress comprehension. A type of comprehension task which was not used in this study, and which need not be used in future experiments, is a same/different one. The main thing it would show is that children have the ability to perceive these stress differences at a phonetic level and can tell that the patterns are not the same. It seems reasonable that children have this ability in view of the literature on infants' sensitivity to the acoustic parameters of stress, and this is verified by children's ability to imitate these patterns successfully. Thus, since questions concerning stress rules involve comprehension of stress at a phonological level rather than perception of stress at a phonetic level, a same/different task would not provide important data on stress rule acquisition.

In future production studies, an attempt should be made to examine children's productions of compounds and phrases in sentences. This could most easily be done with sentences elicited with pictures, in dialogues, etc. Children's spontaneous productions of these compounds, and particularly of phrases in a nontesting situation, would be more difficult to collect, as discussed earlier, but they would be useful. For example, samples of unsolicited phrases observed in naturalistic settings would help to determine whether the tendency to produce primary stress on the first syllable seen in this study resulted from contrastive stress influence due to the test tasks or from a preference for the compound stress pattern.

An acoustic analysis of children's productions of compounds and phrases would also be interesting. This is particularly true with those children who realized that a difference existed between members of pairs but were unsure of how to make it, and therefore experimented with ways to do so.

Analysis of children's imitations of stimuli might also be useful, since it would be clear which stimuli the child was producing, something that is not always obvious with other production tasks. (Experiments to determine the important acoustic correlates of stress have already been conducted, as noted above, and will be reported at a later time.)

It was suggested that one possible reason for children's late and gradual acquisition of compound and phrasal stress distinctions might be that adults sometimes fail to produce these differences in spontaneous speech because of performance factors; hence, it would be interesting to determine whether adults do sometimes fail to do this, and if so, how often. An experiment could be conducted in which adults are asked to participate in dialogues which contain these test stimuli. The compounds and phrases could be embedded in contexts where no doubt would exist as to which member of the pair was being discussed; at the same time, the dialogue could include enough other items and topics so that each subject would be unaware of the purpose of the experiment. The test pairs could then be excised from tape recordings of such dialogues. They could later be played in isolation to adult listeners (including the subjects who produced them) for identification. A number of these samples may indicate to what extent adults use these stress patterns correctly in a normal speech setting.

Since some individual variation was in evidence in this study, future work with acquisition of stress rules such as those in compounds and phrases should consider factors which could contribute to this variation. For example, it may be helpful to obtain as much background information as possible concerning children's education, I.Q. scores, socioeconomic status, etc.

Research should also be done on other constructions accounted for by the rules for compound and nuclear stress. For example, acquisition of compound verbs and adjectives could be examined and the results compared with data from acquisition of noun compounds. Studies should also be made to determine how early stress is consistently used on the nuclear syllable of sentences.

Another area which needs more consideration concerns the universal nature of the findings of this study, particularly of the hypothesis concerning order of acquisition. To test this, experiments of the type conducted here should be administered to children acquiring a first language other than English. Observational data should also be collected on structures involving stress.

Future study should also include examination of stress rules not dealt with in this study. It would be most informative to have a longitudinal study of children's acquisition, examining the emergence of various types of stress and paying particular attention to the interaction of syntax and phonology in stress acquisition. Comprehension and production of emphasis and stress should be observed in the first year, one- and two-word stages, and later as it occurs in monosyllabic and polysyllabic lexical items,

phrases, and sentences. In particular, the development of use of stress as predicted by rule and to contrast grammatical meaning should be observed longitudinally in order to provide data to support or negate the hypotheses suggested here. In these studies, an attempt should be made to separate emphasis or contrastive stress use from unemphatic stress, though this is an extremely difficult thing to do in some instances.

Studies with other stress rules would be helpful to determine when and how (and perhaps if) children learn use of the stress cycle. Many of these rules do not pose problems in using nonsense words, as did compound and phrasal stress. For example, a pilot study looking at questions of morphological relatedness, underlying forms, and vowel reduction using Ladefoged's and Fromkin's (1968) tasks and nonsense words (e.g. *sublanatory-*sublanation) indicated that children had no difficulty with these nonsense words or such tasks. Additional tasks like those described in Berko (1958), and a combination of the two where the stems were presented orally and the endings to be added were written, also were not difficult for children; in fact, they proved to have the most success with the latter task.*

Additional work is also needed with stress rules in nouns and verbs like those tested briefly in this study. Again, a primary line of investigation should be to establish whether the observed ability of children to select the correct noun or verb is based on use of a rule or on application of a memorized lexical item. As noted earlier, the main purpose of the experiment in this study was to obtain preliminary information which would make possible a tentative comparison of children's ability with regard to stress patterns distinguishing nouns from verbs and those in compounds and phrases. In order to have a more reliable comparison of performance with stress in the two types of structure, real words were used in the initial test with nouns and verbs, just as they were with compounds and phrases. When real words are used, a repetition of the test and a correlation of errors is necessary to determine whether stress rules or lexical items and their stress patterns are applied in the task. Use of nonsense words instead of real words would perhaps offer some clarification, since presumably if a rule is involved and learned, its application to a new situation should be possible.** It would be especially important to also test adults with nonsense words to determine whether they have this ability. If a rule is involved, it is not as simple and generally applicable as the rule for adding plural endings. In fact, it

---

*Though this work was only preliminary, it suggested that seven-to-twelve-year-olds may not have the same rules for vowel reduction with these types of words as adults, since four of the six children we tested failed to use reduced vowels in unstressed syllables.

**It is possible, of course, that some of the younger children may not have been familiar with all the real words used in testing in this study and thus these words might have been equivalent to nonsense words to them. Use of nonsense words would therefore be good as a control of the degree of vocabulary item familiarity.

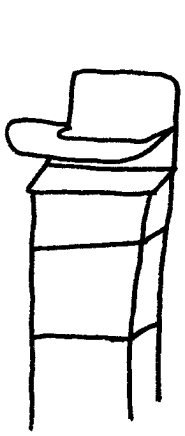
may be that the pattern will not be general enough for even adults to have success with nonsense words. Creating a test situation with nonsense words should not prove as difficult a task with nouns and verbs as it would with compounds and phrases, because contrastive stress would not create difficulties in testing noun and verb pairs. It should be possible to construct a test of perception with nonsense words similar to the one given here with real words and a test of production with nonsense words along lines described in Berko (1958).

A second type of investigation which could be pursued in future studies with stress in noun and verb pairs involves the role played by syntactic and vowel reduction factors. Among other explanations, it was suggested that the earlier acquisition of ability with stress in nouns and verbs could be due in part to the fact that nouns and verbs are syntactically more basic and more different than are compounds and phrases. On the other hand, it could be attributed to the vowel reduction cue, since it is possible that in this experiment children may have ignored the stress patterns and attended instead to vowel-reduction cues. Two types of words could be tested to shed light on this question: (1) noun and verb pairs which have differences of stress but no differences of vowel quality, such as impact, survey, insert, torment, permit, etc., and (2) noun and verb pairs which have differences of vowel quality but no difference of stress, such as delegate, advocate, document, etc. (However, it may be difficult to find items of these types which young children would know, as it was with the type of noun and verb used in this study.) Results of tests with these two types of words could then be compared both with each other and with the results from this study. Children's performance on words with only stress differences, words with only vowel reduction differences, and words with both stress and vowel reduction differences would give an indication of the relative importance and interaction of syntactic and vowel reduction cues.

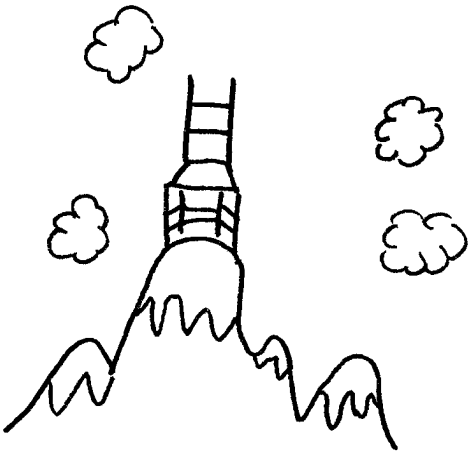
It is hoped that additional research such as that suggested with stress in nouns and verbs and compounds and phrases, as well as with stress rules in other constructions, will provide information to supplement the findings of this study concerning the acquisition of stress rules and the relationship between syntax, semantics, and phonology in the acquisition of language.

APPENDIX

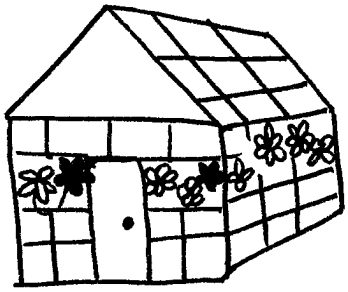
Pictures Used for Compound and Phrase Minimal Pair Test Items



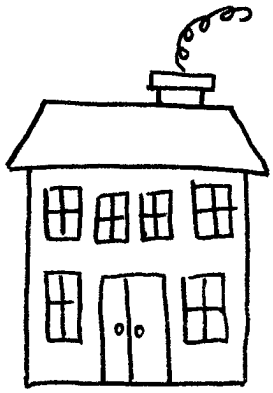
híghchair



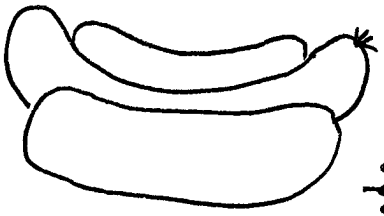
high cháir



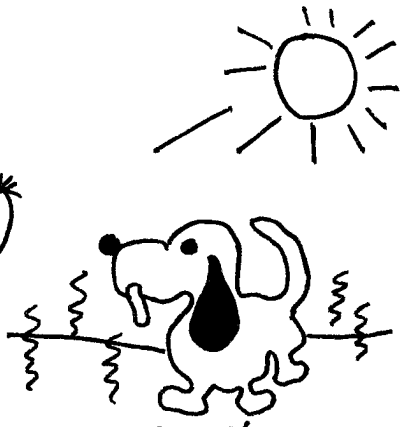
gréenhouse



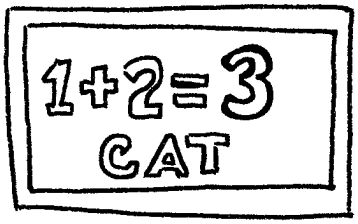
green hóuse



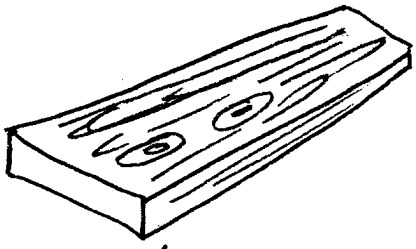
hót dog



hot dóg

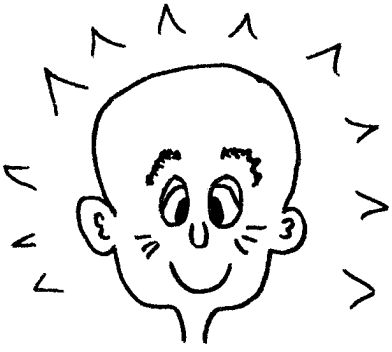


bláckboard

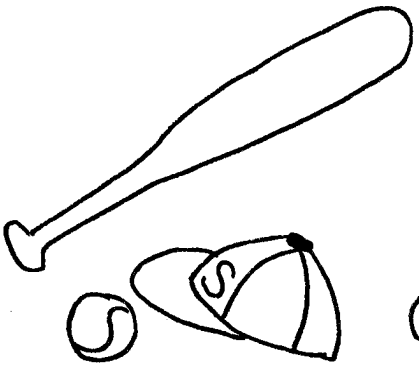


black bóard

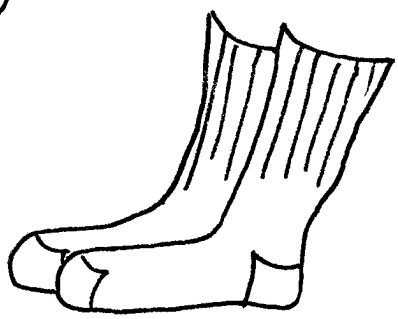
rédhead



red héad



Réd Sox



red sócks

References

Akmajian, A. and R. Jackendoff (1970) "Coreferentiality and Stress." Linguistic Inquiry, 1:1, 124-26.

Atkinson, K. (1968) "Language Identification from Nonsegmental Cues." Working Papers in Phonetics, 10, UCLA Phonetics Lab. 85-89. (Abstract in Journ. of the Acoust. Soc. of America, 44:1, 378.)

Atkinson, K., B. MacWhinney, and C. Stoel (1969) "An Experiment on the Recognition of Babbling." Working Paper, 15, Language Behavior Res. Lab.,U. of Calif.,Berkeley. (Another version of this paper is published in Papers and Reports on Child Lang. Development, 1, Committee on Linguistics, Stanford U. 1970, 71-76.)

Atkinson-King, K. (1970a) "Children's Perception of Stress Contrasts." Journ. of the Acoust. Soc. of America, 48:1, 83.

Atkinson-King, K. (1970b) "Children's Perception and Identification of Stress Contrasts." Quart. Prog. Re.,98, M.I.T. Res. Lab. of Electronics. 113-19.

Bates, R. (1969) "A Study of Acquisition of Language." Unpublished Doctoral Dissertation, U. of Texas.

Berko, J. (1958) "The Child's Learning of English Morphology." Word, 14, 150-77.

Bloom, L. (1970) Language Development: Form and Function in Emerging Grammars. Cambridge, Mass.:M.I.T. Press.

Blasdell, R.C. and P. Jensen (1970) "Stress and Word Position as Determinants of Imitation in First-Language Learners." Journ. of Speech and Hear. Re., 13:1, 193-202.

Blount, B.G. (1969) Acquisition of Language by Luo Children. Working Paper 19, Language-Behavior Re. Lab.,U. of Calif. Berkeley.

Bolinger, D.L. (1965) Forms of English: Accent, Morpheme, Order. Edited by I. Abe and T. Kanekiyo, Cambridge, Mass.:Harvard U. Press.

Bolinger, D.L. (1958) "A Theory of Pitch Accent in English." Word, 14, 109-49.

Braine, M.D.S. (1963) "The Ontogeny of English Phrase Structure." Language, 39, 1-13.

Brown, R. (1969) "A First Language: Stage I." Unpublished paper, Harvard U. (Revised version is to be published.)

Brown, R. (1970) "A First Language: Stage II." Unpublished paper, Harvard U. (Revised version is to be published.)

Brown, R. and C. Fraser (1963) "The Acquisition of Syntax," in C.N. Cofer and Barbara Musgrave (eds.) Verbal Behavior and Learning. N.Y.: McGraw-Hill. 158-97.

Bullowa, M.,L.G. Jones, and T. Bever (1964) "The Development from Vocal to Verbal Behavior in Children." Monographs of the Society for Research in Child Development, 29:1.

Chao, Y.R. (1951) "The Cantian Idiolect: An Analysis of the Chinese Spoken by a Twenty-Eight-Months-Old Child." Univ. Calif. P bl. Semitic Philology, 11, 27-44.

Chomsky, C. (1969) The Acquisition of Syntax in Children from 5 to 10. Cambridge, Mass.: M.I.T. Press.

Chomsky, C. (1971) Linguistic Development in Children from Six to Ten. Final Office of Education Report.

Chomsky, N. and M. Halle (1968) The Sound Pattern of English. N.Y.:Harper and Row.

Eisenberg, R. (1967) "Stimulus Significance as a Determinant of Newborn Responses to Sound." Paper presented at the Biennial Meet. of the Society for Research in Child Development, New York, N.Y., March 31, 1967.

Ervin-Tripp, S. M. (1966) "Language Development," in L.W. Hoffmann and M.L. Hoffman (eds.) Review of Child Development Research, Vol. 2. N.Y. Russell Sage Foundation.

Fischer, S.D. (1971) "The Acquisition of Verb-Particle and Dative Constructions." Unpublished Doctoral Dissertation, M.I.T.

Fraser, C., U. Bellugi, and R. Brown (1963) "Control of Grammar in Imitation, Comprehension, and Production." Journ. of Verbal Learn. and Verbal Behavior, 2, 121-35.

Fry, D.B. (1955) "Duration and Intensity as Physical Correlates of Linguistic Stress." Journ. of the Acoust. Society of America, 35, 765-69.

Fry, D.B. (1958) "Experiments in the Perception of Stress." Language and Speech, 1, 126-52.

Garrett, M., J.Fodor, and T. Bever (1966) "The Active Use of Grammar in Speech Perception." Perception and Psychophysics, 1:1, 30-32.

Gleason,J. Berko, K. Atkinson-King, and B. Mandelkorn (1970) "Improving Communication Skills in Preschool Children." Final re. submitted to Arthur McCaffrey, Director of Office of Economic Opportunity Project "Communicative Competence and the Disadvantaged Child," Harvard U.

Gleitman, L.R. and H. Gleitman (1970) Phrase and Paraphrase: Some Innovative Uses of Language. New York: W.W. Norton and Co.

Hayes, J.R. (1970) (ed.) Cognition and the Development of Language. New York: Wiley.

Hornby, P.A. (1970) "The Differential Effect of Word Order, Syntactic Structure, and Contrastive Stress on One Aspect of Sentence Meaning as a Function of Age." Unpublished Doctoral Dissertation, U. of Chicago.

Hornby, P.A. (1971) "Structure and Function in Child Language." Unpublished Paper, State U. of New York, Plattsburgh.

Hornby, P.A. and W.A. Hass (1970) "Use of Contrastive Stress by Preschool Children." J. of Speech and Hear. Re.,19:13, 395- 99.

Kaplan, E.L. (1969) "The Role of Intonation in the Acquisition of Language." Unpublished doctoral dissertation, Cornell U.

Ladefoged, P. (1967) Three Areas of Experimental Phonetics. London: Oxford U. Press.

Ladefoged, P. and V.A. Fromkin (1968) "Experiments on Competence and Performance." IEEE Transactions on Audio and Electroacoustics, AU-16:1, 130- 36.

Lees, R.B. (1960) The Grammar of English Nominalizations. The Hague: Mouton.

Lees, R.B. (1970) "Problems in the Grammatical Analysis of English Nominal Compounds," in M. Bierwisch and K.E. Heidolph (eds.), Progress in Linguistics. The Hague: Mouton. 174-86.

Lenneberg, E.H. (1967) Biological Foundations of Language. New York: Wiley.

Leopold, W.F. (1939-1949) Speech Development of a Bilingual Child. Evanston, Ill.: Northwestern U. Press. 4 Vols.

Lewis, M.M. (1936) Infant Speech: A Study of the Beginnings of Language. London: Kegan Paul, Trench, Trubner and Co., Ltd.

Li, C.N. (1970)"Topics in Semantics."Draft of doctoral dissertation, U. of Calif.,Berkeley.

Lieberman, P. (1960) "Some Acoustic Correlates of Word Stress in American English." Journ. of the Acoust. Soc. of Amer. 32, 451-54.

Lieberman, P. (1965) "On the Acoustic Basis of the Perception of Intonation by Linguists." Word, 21, 40-54.

Lieberman, P. (1967) Intonation, Perception, and Language. Cambridge, Mass.: M.I.T. Press.

Livant, W.P. (1962) "Productive Grammatical Operations. I: The Noun Compounding of 5-year-olds." Language Learning, 12:1, 15-26.

Marchand, H. (1966) The Categories and Types of Present-Day English Word Formation. U. of Alabama, U. of Alabama Press.

Marchand, H. (1967) "On the Description of Compounds." Word, 23, 379-87.

McCaffrey, A. (1969, 1970) "Communicative Competence and the Disadvantaged Child." Office of Economic Opportunity Reports. Harvard U.

McCarthy, D. (1954) "Language Development in Children," in L. Carmichael (ed.) A Manual of Child Psychology. 2nd ed. New York: Wiley. 492-630.

McNeill, D. (1966) "Developmental Psycholinguistics," in F. Smith and G. A. Miller (eds.) The Genesis of Language: A Psycholinguistic Approach. Cambridge, Mass.: M.I.T. Press, 15-84.

Menyuk, P. (1971) The Acquisition and Development of Language. Englewood Cliffs, New Jersey: Prentice-Hall.

Menyuk, P. (1972) The Development of Speech. The Bobbs-Merrill Studies in Communicative Disorders. New York and Indianapolis: Bobbs-Merrill.

Menyuk, P. and S. Anderson (1969) "Children's Identification and Reproduction of /w/, /r/, and /l/." Journ. of Speech and Hear. Res., 12:1, 39-52.

Miller, W.R. and S. Ervin (1964) "The Development of Grammar in Child Language," Monographs of the Society for Research in Child Development, 29:1, 9-33.

Morton, J. and W. Jassem (1965) "Acoustic Correlates of Stress." Language and Speech, 8, 159-81.

Moskowitz, A.I. (1970) "The Two-year-old Stage in the Acquisition of English Phonology." Language, 46:2, 426-41.

Moskowitz, A.I. (forthcoming) "On the Status of Vowel Shift in English." (To be published in a volume on developmental psycholinguistics, edited by T.E. Moore.)

Murai, J-I. (1960) "Speech Development of Infants." Psychologia, 3, 27-35.

Murai, J-I. (1963-64) "The Sounds of Infants." Studia Phonologica, 3, 21-24.

Nakazima, S. (1962) "A Comparative Study of Speech Developments of Japanese and American English in Childhood." Studia Phonologica, 2, 27-39.

Ohala, J. (1970) "Aspects of the Control and Production of Speech." Working Papers in Phonetics, 15, UCLA Phonetics Lab.

Pike, E. (1949) "Controlled Infant Intonation." Language Learning, 2:1, 21-24.

Read, C. (1971) "Pre-school Children's Knowledge of English Phonology." Harvard Educational Review, 41, 1-34.

Reed, C. (1970), (ed.) The Learning of Language. New York: Scribners.

Ross, J. (1969) "A Reanalysis of English Word Stress." Unpublished paper, M.I.T.

Scholes, R. (1970) "On Functors and Contentives in Children's Imitations of Word Strings." Journ. of Verbal Learning and Verbal Behavior, 9, 167-70.

Shipley, E., C. Smith,and L. Gleitman (1969) "A Study in the Acquisition of Language: Free Response to Commands." Language, 45:2, 322-43.

Slobin, D.I. (1966) "Soviet Psycholinguistics," in N. O'Connor (ed.), Present-Day Russian Psychology. Oxford: Pergamon Press.

Slobin, D.I. (1968) "Early Grammatical Development in Several Languages, with Special Attention to Soviet Research." Working Paper, 11, Language-Behavior Research Lab. U. of Calif.,Berkeley.

Slobin, D.I. (1970a) "Suggested Universals in the Ontogenesis of Grammar." Working Paper, 32, Language-Behavior Research Lab.,U. of Calif.,Berkeley.

Slobin, D.I. (1970b) "Universals of Grammatical Development in Children," in B.G. Flores d'Arcais and W.J.M. Levett (eds.), Advances in Psycholinguistics. Amsterdam: North-Holland, 174-86.

Slobin, D.I. (1971a) (ed.) The Ontogenesis of Grammar: A Theoretical Symposium. New York: Academic Press.

Slobin, D.I. (1971b) "Developmental Psycholinguistics," in W.O. Dingwall (ed.). A Survey of Linguistic Science. College Park, Md.: U. of Maryland Linguistics Program. 299-400.

Slobin, D.I. and C.A. Welsh (1967) "Elicited Imitation as a Research Tool in Developmental Psycholinguistics." Unpublished paper which will appear in C.A. Ferguson and D.I. Slobin (eds.), Studies of Child Language Development. New York: Holt,Rinehart, and Winston.

Smith, F. and G.A. Miller (1966), (eds.) The Genesis of Language: A Psycholinguistics Approach. Cambridge, Mass.: M.I.T. Press.

Templin, M.C. (1957) Certain Language Skills in Children. Minneapolis: U. of Minnesota Press.

Templin, M.C. (1966) "The Study of Articulation and Language Development During the Early School Years," in F. Smith and G.A. Miller (eds.), The Genesis of Language: A Psycholinguistic Approach. Cambridge, Mass.: M.I.T. Press.

Todd, R. and N.J. Reynolds (in press) "Stress as a Prompt for Verbal Imitation." Journ. of Applied Behavior Analysis.

Trager, G.L. and H.L. Smith Jr. (1951) "An Outline of English Structure." Studies in Linguistics. Occasional Papers, 3. Norman, Oklahoma.

Vanderslice, R. (1968) "Synthetic Elocution." Working Papers in Phonetics, 8, UCLA Phonetics Lab.

Vanderslice, R. and P. Ladefoged (1971) "Binary Suprasegmental Features." Working Papers in Phonetics, 17, UCLA Phonetics Lab. 6-24. (Also to be published in Language.)

Velten, H.V. (1943) "The Growth of Phonemic and Lexical Patterns in Infant Language." Language, 19:4, 281-92.

Weir, R. (1962) Language in the Crib. The Hague: Mouton.

Weir, R. (1966) "Some Questions on the Child's Learning of Phonology," in F. Smith and G.A. Miller (eds.), The Genesis of Language: A Psycholinguistic Approach. Cambridge, Mass.: M.I.T. Press, 153-68.

Zimmer, K. (1970) "Some General Observations About Nominal Compounds." Working Papers for Language Universals Project, Stanford U.