DOCUMENT RESUME

ED 064 273 24 TE 002 953

AUTHOR Shilkret, Robert; Wiener, Morton
TITLE The Contribution of Syntactic and Para-Syntactic Cues in the Comprehension of Spoken and Written Language. Final Report.
INSTITUTION Clark Univ., Worcester, Mass.
SPONS AGENCY Office of Education (DHEW), Washington, D.C. Bureau of Research.
REPORT NO Proj-0-A-019
PUB DATE Feb 72
GRANT OEG-1-70-0009(509)
NOTE 89p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Child Language; Comparative Analysis; *Elementary School Students; Grade 4; Language Ability; *Language Rhythm; Memory; Numbers; Reading Ability; *Recall (Psychological); Sentences; Speech; *Syntax; *Verbal Stimuli

ABSTRACT

Two studies were conducted with English speakers to investigate (1) the facilitative effects of melodic features of speech, and (2) whether poor readers (without evidence of sensory defect) show a greater impairment than good readers when melodic features are made unavailable in the speech input. It was hypothesized that when melodic cues are not available, sentences of high syntactic complexity are harder to process than sentences of lower syntactic complexity. A modification of the Savin and Perchonock (1965) "overflow" procedure was employed. An auditory verbal stimulus was presented, followed by a digit list. The subject was asked on each trial to recall both sentence and digits. Half of the sentences had melodic features. Subjects were 40 fourth grade children randomly assigned to Sentence Type groups. There were 25 trials per subject. Results showed (1) melodic cues facilitated processing for Regular Sentences, and (2) these cues were relatively more important for sentences of greater complexity. In a second study, it was hypothesized that selected poor readers would have greater difficulty than normal readers in processing spoken language when melodic features were absent. A modification of the method used in the first study was used. Poor readers had greater difficulty in processing all stimulus types. Serendipitous findings suggest that standardized group-administered reading and intelligence tests are confounded measures. (Author/DB)

NCEP
TE
PRS C A-O4
PA 24





FINAL REPORT
Project No. O-A-019
Grant No. OEG-1-70-0009 (509)

# THE CONTRIBUTION OF SYNTACTIC AND PARA-SYNTACTIC CUES IN THE COMPREHENSION OF SPOKEN AND WRITTEN LANGUAGE

February 1972

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

FINAL REPORT

Project No. O-A-019
Grant No. OEG-1-70-0009 (509)

# The Contribution of Syntactic and Para-Syntactic Cues in the Comprehension of Spoken and Written Language

Robert Shilkret and Morton Wiener

Clark University
Worcester, Mass.

February 1972

The research reported herein was performed pursuant to a grant with the Office of Education, U. S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

U. S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

Acknowledgments

We are indebted to a great number of individuals, each of whom made it possible to undertake and complete this investigation. We acknowledge and thank the many helpful and cooperative Administrative Personnel, Principals, and Teachers in the Worcester Massachusetts School System for permitting us access to the schools and classrooms, and the children who so willingly participated. We thank the many parents who gave us permission to study their children during a part of the school day in the hope that our study might contribute some small measure towards trying to improve educational practice.

One other person warrants a special note of thanks, Jo Ann Bennett, who far beyond the "cull of duty" carried out the testing of the children in the second study, who helped in all aspects of collecting and analyzing these data, and who was a constant source of new insights and suggestions.

Table of Contents

Summary

**Although language behavior occurs as a temporal sequence of elements,** efficient language performance (comprehension) requires some imposition of organization on the elements--both by grouping and by cueing as to relationships among groups. In spoken English there are various "melodic" features (variations in tone and pitch, patterns of hesitations) which we hold make these kinds of organization explicit, in addition to specifying meaning not indicated by syntax alone (e.g., the difference between "I LOVE you" and "I love YOU," marked by stress). In contrast to the spoken form, in written English, syntax seems to be the primary cue for grouping, since there are relatively few other indicators of organization (e.g., punctuation) corresponding to melodic features of spoken English. The absence of graphic forms in written English corresponding to melodic cues could well account for some of the difficulties in reading comprehension. We hold that such reading (comprehension) difficulty does not necessarily derive from the reading activity (identification) per se; if the melodic cues are "removed" from spoken English, corresponding loss of efficiency should ensue for certain poor readers. This framework implies that the degree to which a language has explicit graphic forms (which remain invariant over contexts) which correspond to vocal organizational features of the language, the easier such a language will be to read. Two studies were conducted with English speakers to investigate (1) the facilitative effects of melodic features of speech, and (2) whether poor readers (without evidence of sensory defect) show a greater impairment than good readers when melodic features are made unavailable in the speech input.

The view that melodic features aid in speech processing implies (1) the "absence" of one or more of these features should result in increased difficulty, and (2) melodic features are more important the more complex the organization of the thought. Thus, it was hypothesized that when melodic cues are not available, sentences of high syntactic complexity are harder to process than sentences of lower syntactic complexity.

A modification of the Savin and Perchonock (1965) "overflow" procedure was employed. An auditory verbal stimulus (Regular Sentence, Anomalous Sentence, or Random String) was presented, followed by a digit list. The subject was asked on each trial to recall both sentence and digits; memory for sentence and digits was taken as an index of relative degree of difficulty in processing the sentence. Sentences consisted of 4 Sentence Types, assumed to vary in syntactic complexity; half had melodic features "present," half had these features "absent." Subjects were 40 4th grade children randomly assigned to independent groups according to the Sentence Types. There were 25 trials per subject.

Results showed (1) melodic cues facilitated processing for Regular Sentences, and (2) these cues were relatively more important for sentences of greater complexity, compared with less complex sentences. There was no difference in difficulty between the most complex sentence when melodic features were present and the least complex sentence when these features were absent. As expected, a "pure" syntactic model was

less accurate in predicting results for sentences with melodic features "present" than for the same sentences without these cues.

In a second study, it was hypothesized that selected poor readers would have relatively greater difficulty than normal readers in processing spoken language when the melodic features were "absent." When these features were present, or when the input was syntactically simple, poor readers were expected to perform the same as normal readers.

A modification of the method used in the first study was employed. Only Regular Sentences were used; additional Sentence Types were included to control more adequately for sentence length. Each subject was presented all stimuli (48). Subjects were 40 4th-grade children, half normal readers, half poor readers matched for IQ, sex, and socio-economic status.

Findings of the first study were generally replicated; an additional finding suggested sentence length also contributes to difficulty in sentence processing. Poor readers had greater difficulty than their normal reader controls in processing all stimulus types (melodic features "present" and "absent"). There was no significant evidence that melodic features (or syntactic simplicity) helped these poor readers, within the range of verbal stimuli employed here.

A set of serendipitous findings from this study suggested that standardized group-administered "reading" and "intelligence" tests are confounded measures, even within a "non-deprived" sample with equal educational opportunity. Implications of these findings were discussed and methods were proposed for systematically investigating the implications of these results.

## Introduction

This investigation explores the general view that efficient language performance (comprehension; memory) requires the presence and utilization of the variety of discriminable explicit features in the language "signal," features that constitute organizational patterns of a message. The locus of these features may vary from syntax to "melodic" features such as stress, pitch variations, or patterns of hesitations, and they may occur independently or in combination, or even in conflict with one another--e.g., when the sarcasm of an utterance is "carried" by the intonational contour. Sets of such organizational features may be more or less redundant, as, for example, in the instance "Is it raining today?"--in contrast to "It is raining today?" In this example, the question may be indicated by an intonational contour with a terminal rise in pitch, with or without the characteristic syntactic form of the question, in this case the subject-verb reversal. In sum, although language behavior occurs as a temporal sequence of elements (e.g., phonemes; words), the concomitant presence of the other characteristics of the message serves to group the elements and to indicate the particular relationship among groups. The "complexity" of any language instance can be characterized as a function of (1) the complexity of the event or thought to be represented linguistically--for example, as indicated by the degree of differentiation of components of the event, and (2) the complexity of the particular linguistic form of the representation of any given event or thought. It follows then, that presence of organizational features (e.g., melodic features of speech) should be relatively more important for efficient language performance (comprehension; memory) the more complex the event or the representation of the event in language (e.g., as syntactic complexity increases).

We further hypothesize that language performance of some individuals is hindered by failure to discriminate such organizational cues when they are available or by failure to learn to utilize such cues. Further, reliance by some individuals on any one set of such features (e.g., melodic features), while responding less to others (e.g., syntax), may result in difficulties under special conditions (e.g., written language) when particular sets of features may be minimal. For example, in written language, capitalization and punctuation may serve as analogues to melodic features of speech, but these features in writing seem to compose a more restricted set than the analogous set of such vocal features in speech. Thus, the relative absence, at least in written English, of graphic forms corresponding to melodic features of spoken English could well account for some of the difficulty in reading comprehension for some readers. We hold that such reading (comprehension) difficulty need not necessarily derive from a difficulty in word identification _per se_; rather, some reading difficulties may be attributable to the relative absence of organizational cues in written language, and the failure of those individuals to respond to those cues which _are_ available. If this view is tenable, then "removal" of the corresponding organizational features (e.g., melodic features) from spoken language should result in a corresponding decrement in comprehension, particularly for such readers.

The framework suggested above implies that the degree to which a particular language has explicit graphic forms which correspond to vocal organizational features of the language, the easier such a language will be to learn to read. Writing can be viewed as a notational system for representing the spoken form of language. As a graphic system which represents spoken language, different written languages may vary in terms of (1) explicitness, that is, the extent to which there is a different graphic form for each phonemic variation, and (2) articulateness, that is, the extent to which each grapheme has the same phonemic realization in a variety of contexts. Written English is neither fully explicit (e.g., commas are not used to indicate pauses within many grammatical constituents, although they are used to indicate pauses between certain constituents) nor fully articulate (e.g., a given grapheme may have several phonemic realizations depending on context of occurrence--as, for example, in the case where the initial vowel sound is dependent on the presence or absence of a subsequent "silent e," "cut" versus "cute"). Of great relevance for this view is Makita's (1968) exemplification of the high degree of correspondence in Japanese (as compared with Western languages) between graphic forms and phonemes, and the concomitant lesser difficulty of reading for Japanese children, compared to Western children at equivalent points during acquisition. While Japanese seems both more explicit and articulate than English, notational systems even more explicit and articulate might be possible (e.g., the notational system of formal logic).

The framework of the present research implies that part of learning to read English must involve learning to organize the input appropriately on the basis of the "part-cues" to the organization which are available; in this sense, learning to read English involves learning to impose organization on the visual input which is neither fully explicit nor articulate. The accomplished reader of English has mastered a complex set of such "rules" for organizing the visual input; for such an accomplished reader, it would not be expected that making the visual input more explicit and/or articulate (e.g., by providing additional visual cues) would facilitate comprehension. This expectation is supported by the results of Cromer (1970) and Carver (1968), which are discussed more fully below.

In the present investigation two studies were conducted with English speakers (4th-graders) to examine (1) the facilitative effects of melodic features of speech in a task involving immediate memory for speech, and (2) whether certain poor readers (4th-graders), without evidence of sensory defect, show a greater impairment than matched good readers in such a task when melodic features of speech are made unavailable in the speech input.

Starting with the hypothesis that melodic features facilitate language comprehension, it follows that "absence" of any one or more of these features should result in some degree of difficulty in a performance task (e.g., memory for language). Given this assertion, as the general methodology of this investigation, the degree of difficulty found without the availability of a particular component is used as an index of the relative importance of that particular component in comprehension. For example, in spoken language a sentence said in an atonal manner should be more difficult to remember and comprehend than the same sentence said with the usual tonal pattern, since under the atonal condition the total information available to the hearer (i.e., explicitly present in the input) does not include stress and pitch variations, which we hypothesize are helpful in discriminating the syntactic organization of the sentence and indicating which words are to be grouped, as well as indicating relationships among such groups of words.

The most direct measures of "comprehension" would seem to be paraphrasing the sentence, that is, repeating the meaning in some other form, or taking appropriate action in some task, that is, acting out the meaning of the sentence. However, each of these more direct measures of comprehension raises certain problems. In long utterances in spoken discourse an appropriate measure of comprehension might be the paraphrase; however, the short sentences (e.g., "The ball was hit by the boy") used in these studies would be quite difficult to paraphrase. Furthermore, paraphrasing as a measure of comprehension may be confounded by factors such as the verbal fluency of the subject. A second possible measure of comprehension, requiring appropriate action in terms of the material comprehended, raises difficulty in terms of constructing appropriate items. Furthermore, such a measure would establish an all-or-none criterion which may be too gross to measure subtle differences in degree of difficulty of comprehending the various kinds of stimuli used in these studies. Given the problems associated with direct measures of comprehension, some indirect measure, such as immediate memory, seems more appropriate. While we do not assume that verbatim immediate memory is a sufficient condition for comprehension, we do assume that some memory is necessary for the possibility of comprehension (e.g., if one is possibly to understand a spoken sentence, he must remember something about the beginning of the sentence when the end is reached).

Several recent investigations have been concerned with the role of syntactic organization in language performance, as measured by a memory task. The conceptualization of immediate memory shared by most of these studies derives from Miller's (1956) formulation, proposing that the input sequence is grouped into "chunks," and these chunks are the units which are processed. Immediate memory is conceived as relatively small and of fixed capacity, but the "size" of the chunks themselves (as measured by the physical parameters of the stimulus input) is variable. If the stimulus input is organized in some way, or if the chunks are indicated, it is presumed that memory will be facilitated. The general hypothesis common to these studies has been stated by O'Connell, Turner, and Onuska (1968): "Any stimulus characteristic which imposes organization on the stimulus or affords subjects potential groupings of the items in the stimulus list should facilitate learning. Syntax, by reason of its hierarchical structure, offers such potential groupings of stimulus items" (p. 111). Several studies (e.g., Epstein, 1961, 1962, 1967; Marks and Miller, 1964; O'Connell, Turner, and Onuska, 1968) have provided results which support this hypothesis.

A related set of studies has investigated an hypothesis derived from a transformational model of grammar, proposing that the degree of difficulty of processing (remembering) a sentence is directly related to the number of transformations (Chomsky, 1965) required to derive that sentence from the "simple" form (active-declarative-affirmative). These studies (e.g., Miller, 1962; Mehler, 1963; Savin and Perchonock, 1965; Clifton and Odom, 1966) have provided some evidence that memory for sentences is related to syntactic complexity, and that the degree of difficulty in recalling a sentence is related to the number of transformations, applied additively, required to derive the sentence from the simple form.

Several other studies have used measures other than immediate recall (e.g., errors and speed of recognition and reproduction; judged similarity; prompted recall; confusions in discrimination; perception over a masking stimulus) to test hypotheses derived from this model of sentence processing (see Clifton and Odom, 1966; Garrett and Fodor, 1968; Miller and McNeill, 1969; Lane and Schneider, 1963; Mehler and Carey, 1967).

These sets of studies have been questioned on several grounds: (1) the replicability and/or generalizability of results (e.g., Fodor and Garrett, 1966; Bregman and Strasberg, 1968; Matthews, 1968); (2) the results, at least in part, are believed to be attributable to possible artifacts of particular experimental tasks, and that the more general phenomenon might be due to "interference" and/or "rehearsal" rather than syntactic complexity (e.g., Epstein, 1969; Glucksberg and Danks, 1971); (3) some other model of grammar (e.g., Yngve, 1960) might better account for findings than a transformational model (e.g., Martin and Roberts, 1967; Martin, Roberts, and Collins, 1968; Wright, 1969; Wearing and Crowder, 1971); (4) the transformational complexity is often confounded with sentence length, which also contributes to performance difficulty (e.g., Martin and Roberts, 1967; Wearing, 1969; Perfetti, 1969; Orenstein and Schumsky, 1970; Foss and Cairns, 1970); (5) memory of syntactic forms occurs only when the task constrains the kinds of organization available to the subject; when possible, verbal material seems to be recalled in terms of "meaning," usually not specified further, and seems to minimize the particular syntactic form (e.g., Sachs, 1967; Fillenbaum, 1966, 1971).

While somewhat hesitant to walk again over this well-trodden ground, it is our contention that there are other characteristics of the language input (e.g., melodic features of speech), not fully redundant with the syntactic organization of the input, which may indicate organization and facilitate comprehension. If this view is tenable, then any "pure" syntactic model of language performance (e.g., memory) should be a relatively accurate predictor of experimental results only under some special conditions when other possible organizational cues are "removed" from the input (e.g., a sentence said atonally). From this point of view, a performance model which focusses exclusively on syntactic organization, to the exclusion of other possible means of organization available to the subject, is, at best, incomplete, and, at worst, reduces language behavior to a concern with "form" rather than "content." We are in agreement with those investigators (e.g., Sachs, 1967; Fillenbaum, 1966, 1971) who have noted that memory for the specific syntactic form of a sentence is more likely when other forms of organization are relatively unavailable to the subject. With other organizational cues "removed" (atonal sentences) sentences which are more syntactically complex should be more difficult to comprehend (remember) than sentences which are less syntactically complex. Once this is demonstrated, it is then possible to assess the facilitative effects of other organizational cues (melodic features) at various levels of syntactic complexity. In sum again, it is hypothesized, then, that the melodic features of speech serve to "compensate," in effect, for an increase in comprehension (memory) difficulty resulting from an increase in syntactic complexity.

In the present framework, syntactic complexity is assumed to be correlated with the "event complexity" which is represented in a particular linguistic form. The report of an event involving an actor and an object of action is usually given in the sequence of an actor acting on an object (e.g., "He hit the ball"). It is assumed that this "conceptual" sequence (i.e., an actor acting on an object) is the simplest form. The "experience" of "causality" in this sequence also corresponds with the temporal sequence of lexical elements in the simple sentence (i.e., subject-verb-object). Other linguistic forms (i.e., syntactic transformations) may report a similar event, but with variations in the temporal sequence of lexical elements. For example, in a passive sentence (i.e., object--"was"--(verb)--"by"--subject) the occurrence of the words ". . .was (verb) . . . by . . ." signals a linguistic inversion of this "conceptual" sequence (i.e., an object acted upon by an actor). "Complexity" is used in the sense that linguists have used the term, but it may be considered an operational definition of this kind of relationship, that is, the degree of correspondence between some ideal "conceptual" sequence and the linguistic sequence. In this sense, greater "complexity" implies less correspondence (e.g., temporally) between the two. We do not hold that any two different linguistic forms represent the "same" event; for example, the active and passive forms may distinguish differences in "theme" or emphasis (e.g., Halliday, 1967, 1968). However, we do assume that variations in syntactic complexity can be said to be, at least in part, related to variations in the complexity of the "thought" represented. While in the present studies syntactic complexity was used as a way of systematically varying performance (memory) difficulty, other possible means could be employed (e.g., variations in sentence length; variations in frequency of occurrence of words composing the stimulus sentences; variations in placement of "nested" elements of certain sentences). It is assumed that once a range of performance difficulty is demonstrated for any of these dimensions alone, that is, without the availability of melodic features, it is then possible to assess the "compensatory" effects of the presence of other organizational cues, that is, by comparison with the corresponding conditions with melodic features present at the various "levels" of difficulty.

While there has been much recent work concerning the role of syntactic organization in a variety of language tasks, there has been much less concern with the role of melodic contour in facilitating comprehension, or with the importance of intonational features during language development. However, beginning as early as 1894, G. E. Mueller, with Schumann and Pilzecker, conducted a series of studies which demonstrated a confounding in Ebbinghaus' earlier (1885) investigations of memory for nonsense syllables; the Mueller studies indicated the importance of such features as pause and stress in speech in "group formation." These features served to indicate groups of otherwise unrelated material, and thus facilitated memory (see Woodworth, 1938; Katona, 1940). Lipsky (1907) outlined a category system of the "rhythmic" characteristics of prose, using this to discriminate stylistic variations of a number of popular writers. While some earlier investigators mentioned the importance of intonational features of speech during early language development (e.g., Allport, 1924; and Huey, 1908, who discussed the use of intonational variation to indicate variation in meaning of the holophrase, "one-word sentence"), W. Leopold seems to have been the first specifically to point out the importance of intonational features in speech development (see Werner and Kaplan, 1963; Ingram, 1971). More recently, evidence has appeared that infants (8 mos.) discriminate certain intonational contours (Kaplan and Kaplan, 1971; Morse, 1971); that children imitate (e.g., Miller and Ervine, 1964) and produce spontaneously (e.g., Gruber, 1967; Kagan, 1964; Gleitman and Shipley, 1971; Weeks, 1971) the intonational features of utterances prior to the syntactic features which can serve to discriminate various utterances. Recent theories of language development as diverse as Werner and Kaplan (1963) and Braine (1963) have pointed to the importance of intonational cues to indicate appropriate word groupings and to specify meaning.

The few studies which have investigated the role of melodic features in facilitating language performance (recall) have dealt with nonsense materials, or they have been concerned with the effects of intonation only for a limited range of linguistic materials. O'Connell, Turner, and Onuska (1968) investigated the effects of intonation and grammatical structure for the recall of nonsense strings. It was found that both intonation and grammatical structure facilitated recall, and that the most facilitation occurred when both were present. The authors reasoned that the presence of intonation in their stimuli served as a cue to indicate the auditory stimulus was "sentence-like," and thus the grammatical tags were used to facilitate recall. First, the nature of the stimuli used (nonsense strings) makes it difficult to generalize these results to sentence recall. Second, and more important, from these results it is difficult to know the grounds for inferring that either syntax or intonation is the more "salient" cue to "sentencehood." O'Connell, Turner, and Onuska seem to imply that intonation is used as a "first" cue to indicate "sentencehood," and that intonational features are then no longer utilized by the hearer. In contrast, we would argue that both syntax and intonation can serve as such a cue, and that they function in concern, even after the stimulus is perceived to be "sentence-like."

A study of Scholes (1969), in which all stimuli were presented without intonational cues, led to the conclusion that with maturation the child comes to "ignore" intonational features and to "concentrate" on grammatical features. This conclusion seems unwarranted in that all the stimuli were non-intonated (i.e., there were not the appropriate control conditions--intonated stimuli), and, further, a direct comparison of the child data and adult data was not possible in that study.

Ford (1970) found that for children 5 to 11 years of age intonation facilitated performance in an imitation task, and further that only those pauses which occurred within phrase boundaries were found to be facilitative. It was concluded that such pauses enable the subject to "take advantage" sooner of the grammatical structure of the sentence, as compared with sentences with no pauses within phrases. While this conclusion may be tenable, again, as with the O'Connell, Turner, and Onuska study, the logic of this particular inference from the data seems unclear.

A more recent study by Weener (1971), using subjects from kindergarten through third grade, investigated the effects of intonation (present versus absent), syntax (simple active sentences versus these same sentences with words in reverse order), and associativity (meaningful versus anomalous sentences, as "Swift deer jump high fences" versus "Last foxes sail silver gardens") in a free recall task. It was found that intonation facilitated recall for all age groups, that the facilitative effects of associational patterns were present at all age levels and increased with age; that the effects of syntax on recall were evidenced only for the oldest subjects. There was also a significant interaction between intonation and associativity: adding intonation to messages without associativity resulted in small increases in recallability, while adding intonation to messages with associativity produced larger increases in recallability. The obtained interaction in the Weener study (1971) between intonation and associativity might be seen as evidence against the hypothesis of the present investigators that intonational features are relatively more important for efficient language performance (memory) for more difficult material, compared with less difficult material ("difficulty" scaled along some other dimension, such as degree of associativity of component words in the Weener study, or syntactic complexity in the present studies). However, it seems likely that Weener's low associativity stimuli (anomalous sentences) were of sufficient difficulty, particularly for the subjects in that study, that the addition of other organizational cues (i.e., intonational cues) did not facilitate recall more than found in that study. A test of this interpretation is possible, since, coincidentally, in the first of the studies reported here there are stimulus conditions which seem comparable to Weener's (i.e., Simple-active Regular Sentences versus Anomalous Sentences, with and without intonation), as well as other stimuli presumed to vary in difficulty in terms of another dimension (i.e., syntactic complexity).

In sum, there is evidence from these studies that melodic features facilitate memory for non-linguistic material (e.g., the Mueller studies) as well as linguistic material in a variety of tasks, for children and adults. There is further evidence that such features of speech are more facilitative to the extent that the linguistic stimuli resembles "normal" speech (e.g., O'Connell, Turner, and Onuska) and if the material is not exceedingly difficult or unusual (e.g., Weener). However, none of these studies has been concerned with the role of intonational features within a range of linguistic input varying in difficulty in terms of some other dimension such as syntactic complexity.

The present framework regards melodic components of speech as an aid in comprehending speech as for example by making the syntactic organization more discriminable or by indicating the "chunks" to be processed. It follows that the absence of any one or more of these components should result in some degree of difficulty in processing speech input; the degree of difficulty found without the availability of a particular component can be used as an index of the relative importance of that particular component. For example, in spoken language a sentence with the melodic features "removed" (e.g., said in an atonal manner, with equal pauses between words) should be more difficult to comprehend than the same sentence said with the usual tonal and hesitation patterns. Further, as the speech input becomes more "difficult" to process, along some other dimension (e.g., as syntactic complexity increases), the availability of melodic features should become more important for efficient performance. In the first of the present studies these expectations were tested, using immediate memory as an index of the relative difficulty of comprehending speech input. A modification of the Savin and Perchonock (1965) "overflow" procedure was used: an auditory verbal stimulus (Regular Sentence, Anomalous Sentence, or Random String, with or without intonational features) was presented, followed by a digit list. Subject was asked on each trial to recall both sentence and digits: for sentences recalled verbatim, memory for digits was taken as an index of relative degree of difficulty in processing the sentence.[1]

The general expectations for this study, then, can be outlined as follows:

1. Sentences of low syntactic complexity are less difficult to process (i.e., remember) than sentences of greater syntactic complexity, and this increased difficulty is not accountable simply as a function of sentence length.

2. Melodic components facilitate performance (memory), and the presence of these components is relatively more important for the recall of sentences which are syntactically more complex than for the recall of sentences which are syntactically less complex.

3. Consistent with earlier findings by other investigators, meaningful sentences will be easier to recall than anomalous sentences (i.e., words used in an unconventional manner), and that these in turn will be easier to recall than random strings of words.

From the framework of this investigation, reading can be viewed as a special condition where many of the organizational characteristics of language (the melodic features of spoken language) are not fully and explicitly represented. In written language, although the stimulus input is ordered sequentially (i.e., words ordered from left to right), several kinds of evidence suggest that good readers do not process written material solely in the sequential word-by-word as they appear in writing. Several studies (e.g., Neisser, 1967) have even shown that a good reader does not attain a full perceptual representation of each letter or phoneme he "reads." Rather, certain "part-cues" seem to be elaborated (Kempler and Wiener, 1963; Cromer and Wiener, 1966), and the input is "chunked" (Cromer, 1970) by some set of cues not apparently given explicitly. While in spoken language, the form of input for the hearer also consists of words in temporal sequence, we hold that the inclusion of melodic features facilitates comprehension. In that written English does not include all these melodic features represented graphically, some other cues must account for the organization imposed by a reader. As the linguist Charles Hockett (1967) has stated, ". . . [in writing] the successive words are not tied together into groups by variations of stress, by intonation, and by junctures, as are the successive words of ordinary talking. A written text can be regarded as a set of instructions telling a reader what to say (aloud or silently as the case may be). But the instructions are incomplete. There are no special marks to tell him how to distribute stresses, and only skimpy indications of intonation" (p. 919). Thus, it would seem that part of learning to read effectively involves learning to organize from the partial information available (that is, with melodic features absent) in written English.

Wiener and Cromer (1967) outlined a conceptual framework for dealing with issues and problems in reading and reading difficulties. They emphasized the requirements of definition, and suggested the necessity of the following distinctions: (1) Identification versus Comprehension (e.g., when one speaks of reading difficulties, one must specify whether he means difficulties in "saying" and/or "understanding"); (2) Absolute versus Relative criteria (e.g., whether the criteria for designating reading achievement are based on some arbitrary ideal, or whether they are based on a normative reference group); (3) Acquisition versus Accomplished reading (e.g., whether the reading difficulty occurs only during the acquisition of reading skills, or whether it appears even with an accomplished reader); (4) Reading versus Language skills. This last distinction is most germane to the present investigation. Typically, investigators who consider reading as identification have little concern with previously acquired auditory capabilities. When reading is considered in terms of comprehension, many investigators have not made explicit the relationship between reading skills and language skills. Further, what is taken to be "poor reading" is often not considered in terms of the language competence of the reader.[2]

Some recent investigations have begun to study the relationship of the organization of linguistic material to reading. As part of one study, Cromer (1970) identified, in a junior college population, one group of poor readers who were identified as reading in a word-by-word fashion. They were matched with a good reader group, and the effects for comprehension of "chunking," or "pre-organizing" the reading material in various ways investigated. The reading material was presented visually four different ways: (1) regular paragraph form; (2) one word at a time; (3) words grouped together arbitrarily, that is, with minimal regard for the syntactic organization; and,(4) words grouped by phrases, that is, in terms of syntactic organization. These word-by-word readers did worse that the good readers under the first three conditions, where the material was presented visually without organization consistent with grammatical constituents, but were indistinguishable from the good readers under the pre-grouped, syntactically appropriate condition, that is, the 4th condition.[3]

The efficacy of "chunking" reading materials into various kinds of units was also investigated by Carver (1968). He used only good readers, and found no differences between "chunked" and "unchunked" reading material, a result which would not be unexpected in terms of the present framework, and in terms of Cromer's results, who also found that good readers were not facilitated by the additional explicit visual organization. Since good readers utilize various principles of organization whenever they read, that is, they do not "read" every word, but rather, process units, it would not be expected that "chunking" of reading material would result in facilitation of comprehension for them. In a sense, good readers "chunk" the material whenever they read, whether or not the "chunks" are made explicit visually.

Whereas the above studies examined only reading (i.e., decoding of visual input), as part of a study of the relationship between identification and comprehension in good and poor readers, Oakan, Wiener, and Cromer (1971) investigated the effects of organization of input in the auditory as well as in the visual mode. An attempt was made to study the effects on comprehension of good readers and of poor readers who had both "good" and "poor" verbal inputs, auditorily and visually.

It was found that there was no significant increase in comprehension of the poor readers under "good" visual input. For these poor readers, it can be assumed that the problem in comprehension is not simply an inability to identify the words they read. Most important for the present investigation was the finding that, while good readers performed similarly under "good" and "poor" auditory input, the poor readers performed worse under "poor" auditory input than under "good" auditory input. Under good auditory input, the poor readers and the good readers performed similarly. It would seem, then, that while the poor readers had difficulty imposing organization on disrupted auditory input, the good readers had no such difficulty, even for auditory material.

In sum, it appears that for some poor readers, who read orally in a word-by-word manner,[4] the difficulty would not seem to be one of word identification alone; some poor readers do not seem to "chunk" the material unless it is explicitly "chunked" for them. Further, the findings of Oakan, Wiener, and Cromer suggest that certain poor readers are also impaired by poor auditory input as compared with matched good readers. If we hold that the difficulties of such poor readers can be said to be linguistic usage rather than a reading difficulty (in light of the parallel difficulty with auditory material), then further study of the linguistic response patterns of these kinds of poor readers is warranted.

It is hypothesized that when organizational cues (i.e., melodic features) are "removed" experimentally from auditory linguistic input, these poor readers will show a relatively greater decrement in performance (memory) than matched good readers. To this extent, such a "reading" problem might better be called a problem in utilization of organizational features. Further, if it is the case that melodic features have greater importance for the comprehension (memory) of high syntactic complexity than for sentences of low syntactic complexity (the first study here), then it is expected that these poor readers should do progressively less well, compared with good readers, as the syntactic complexity of the input increases and where melodic features are unavailable in the input.[5] To test these hypotheses, a second study, similar to the first, was conducted, comparing certain 4th-grade poor readers with matched 4th-grade good readers.

## Method

### I. Subjects

The subjects for this study were forty children[6] enrolled in the fourth grade of a public school, which was located in a upper-lower to lower-middle class community in a New England city. The subjects were divided randomly into five independent groups, with an equal number of boys and girls in each group. Thus, each group consisted of eight subjects, four boys and four girls. The subjects ranged in age from 9.0 to 10.5 years, with a mean of 9.6 years. See Table 1.

### II. General Procedure

All of the children who participated in the experiment had parental permission--a letter was sent home requesting approval. All subjects were tested individually during school hours in a small room within the school building. To keep external noise at a minimum, no testing was done when there was a class in an adjoining room. A testing session required approximately forty-five minutes.

#### A. Materials

The stimulus materials consisted of lists of sentences and lists of digits which were constructed in the following manner.

First, five simple active declarative sentences, each consisting of five words, were constructed. These sentences all conformed to the basic sentence frame:

The--Noun--Verb--the--noun.

Thus, of the five words in these sentences, there were two articles (i.e., "the") and three content words, two nouns and a verb in each sentence. One of these nouns was an agent, and the other an object. All agents in these sentences were members of the general meaning-class "people" (e.g., girl, woman, child, boy, man), and all objects were concrete nouns (e.g., dog, food, ball, letter, car). All verbs were transitive and were placed in the simple past tense throughout (e.g., walked, ate, caught, wrote, drove).

The selection of the particular words used as stimuli was constrained by several factors. First, there was no special semantic relationship between any of the five agents and any of the five objects. This constraint was imposed to allow interchangeability between these components and to prevent special associations which might serve as an aid for the memory of one or more of the sentences. The sentences were constructed so that no one sentence would be easier to remember in terms of its internal construction. For example, a sentence like, "The butcher cut the meat" was not included in the stimulus materials, since the semantic relationship between "butcher" and "cut" might serve as an aid in the memory of that sentence. However, relationships of this kind were allowed between verb and object. Second, all of the

Table 1

Age Characteristics of the Subject Groups
(in months)

| | | Boys | Girls | Combined |
|---|---|---|---|---|
| Group 1 | Mean | 115 | 114 | 114.4 |
| | Median | 115.5 | 113.0 | 114.0 |
| | Range | 112-118 | 111-116 | 111-118 |
| Group 2 | Mean | 113 | 116 | 114.4 |
| | Median | 113.0 | 117.5 | 117.0 |
| | Range | 108-119 | 110-118 | 108-118 |
| Group 3 | Mean | 112 | 119 | 115.6 |
| | Median | 112.5 | 120.0 | 114.5 |
| | Range | 109-116 | 110-125 | 109-125 |
| Group 4 | Mean | 112 | 112 | 111.5 |
| | Median | 112.0 | 110.5 | 111.0 |
| | Range | 108-114 | 109-116 | 108-116 |
| Group 5 | Mean | 116 | 116 | 115.9 |
| | Median | 118.0 | 113.5 | 117.0 |
| | Range | 109-119 | 110-126 | 109-126 |

words in the stimulus materials were selected from the 1000 most frequently occurring words in the English language, as determined by the Thorndike-Lorge word count (1944). This constraint was imposed to try to control for familiarity effects in the memory for particular words.

The five simple active declarative sentences were then "transformed" to their corresponding question, passive, and passive-question forms, yielding a total of twenty sentences. For example, the sentence "The girl walked the dog" yielded the sentences "Did the girl walk the dog?," "The dog was walked by the girl," and "Was the dog walked by the girl?" In generating these additional sentence types, the length of the sentences was increased in a systematic fashion. The questions were one word longer than the Simple Sentences (i.e., the addition of "did"), and the Passives and Passive-Questions were each two words longer than the Simple Sentences (i.e., the addition of the words "was" and "by"). Since immediate memory for verbal material is, in part, a function of the length of the stimulus input, it was decided to include an additional sentence type as an indirect control for sentence length, independent of complexity. The sentence type chosen for this group of sentences was the Who-Question, which is one word shorter than its corresponding simple active declarative form (i.e., the sentence "Who walked the dog?"), yet is syntactically more complex than the simple form (Harris, 1957). These four sentence types and the additional Who-Question sentence type (with five exemplars of each sentence type) comprised the "Regular Sentence" stimulus materials.

An equal number of "Anomalous Sentences" was constructed so that (1) the sentence frames of the Anomalous Sentences was the same as the corresponding sentence frames of the Regular Sentences for each sentence type, and (2) the same words appeared with the same frequency in the corresponding Anomalous Sentences and Regular Sentences for each sentence type. The five sentences in the Regular Sentence, Simple Sentence type group can be represented as follows:

1. The $A_1$ $V_1$ the $O_1$.
2. The $A_2$ $V_2$ the $O_2$.
3. The $A_3$ $V_3$ the $O_3$.
4. The $A_4$ $V_4$ the $O_4$.
5. The $A_5$ $V_5$ the $O_5$.

(A represents the agent, V represents the verb, and O represents the object.)

The Anomalous Sentences were constructed in the following manner. The agent from the first sentence was combined with the verb from the second sentence and the object from the third sentence. The agent from the second sentence was combined with the verb from the third sentence and the object from the fourth sentence, and so on. In this way, each content word appeared only once in any given group of sentences comprising each sentence type. Further, for the Anomalous Sentences, the agents and objects were reversed, to yield a more unusual combination of words (e.g., a sentence generated by the above procedure, "The woman caught the letter," became "The letter caught the woman"). The resulting Anomalous Sentences can be represented as follows:

1. The $O_3$ $V_2$ the $A_1$.

2. The $O_4$ $V_3$ the $A_2$.

3. The $O_5$ $V_4$ the $A_3$.

4. The $O_1$ $V_5$ the $A_4$.

5. The $O_2$ $V_1$ the $A_5$.

As before, the five simple active anomalous sentences were then "transformed" to their corresponding question, passive and passive-question forms, yielding a total of 20 sentences for the Anomalous Sentences stimulus materials.

For each of the Sentence Types, five random word lists (Random Strings) were generated. To obtain these Random Strings, the Anomalous Sentences for each sentence type were arranged in different random orders, each Anomalous Sentence "scrambled" to yield a random word list. The following constraints on constructing the different random orders were imposed: (1) No group of three or more words appeared in a random string which constituted a syntactically correct unit. For example, the group "walked the food" did not appear in a random word list. (2) The two words "the the" did not appear in sequence.

For each stimulus (that is, the Regular Sentences, Anamalous Sentences and Random Strings) a list of six digits, which constituted the second recall task on each trial, was drawn from a table of random numbers. In constructing these lists of digits, the following constraints were imposed: (1) The digit "O" was not used. (2) A given digit appeared only once in a given list of digits. (3) No two digits appeared in correct serial order in a list of digits (e.g., "..78.." was not allowed).

To summarize, five sentences were constructed for each of four different Sentence Types (i.e., Simple, Question, Passive, and Passive-Question). Five additional sentences of the Who-Question Sentence Type were generated to serve as a control for length (in total number of words) of the sentences. These constituted the Regular

Sentences. Five sentences were then constructed for each of the four Sentence Types (i.e., Simple, Question, Passive, and Passive-Question) in such a way to render them semantically anomalous (Anomalous Sentences). Finally, five random lists of words were constructed for each of the Sentence Types (Random Strings). For each one of the verbal stimuli a six-digit list of numbers was also generated.

Each of the verbal stimuli (i.e., the sentences and word lists) was recorded on a Bell and Howell Language Master card. Each of the Regular Sentences and the Anomalous Sentences was recorded under two conditions, Tonal and Atonal (i.e., conditions of melodic features "present" and "absent"). In general, for the Tonal condition the sentences were recorded as they would be spoken naturally; that is, under this condition the sentences included stress, pitch and tonal variations and juncture pauses as they occur in speech. However, none of these components was exaggerated. Each sentence was also recorded under an Atonal condition. For this condition stress, pitch, and tonal variations were kept at a minimum, and the words were recorded at a constant rate as if a list were being read. The total times for each sentence under the Tonal and Atonal conditions were approximately equal; the rate of recording for both conditions was approximately two words per second. The Random Strings were recorded in the same manner as the Atonal condition, with stress, pitch and tonal variations kept at a minimum, and words recorded at a constant rate.[7]

The six digits followed the verbal stimulus on each recorded card. The time interval between the end of the verbal stimulus and the beginning of the list of digits was approximately 1/2 second. The digits were recorded in a uniform manner (i.e., with constant stress and pitch for each digit) at a rate of approximately one digit per second. The final digit of each digit list was recorded with a slight drop in pitch to signal the end of that particular trial.

The recording controlled for volume was done by an adult male accomplished in public speaking. All stimuli were judged by an independent rater, and those judged not to meet criteria were remade.

The experimental stimuli were arranged in five groups, corresponding to the five sentence types employed in the study (i.e., active, question, passive, passive-question, and who-question). In each group of stimuli there were five Regular Sentence-Tonal stimuli, five Regular Sentence-Atonal stimuli, five Anomalous Sentence-Tonal stimuli, five Anomalous Sentence-Atonal stimuli, and five Random String stimuli. Thus, for each sentence type, there were a total of twenty-five experimental stimuli. (A set of the experimental stimuli can be found in Appendix A.)

B. Procedure

The subjects were assigned randomly to an experimental group, except that there was an equal number of boys and girls in each of the groups. After establishing initial rapport, the subject was presented the following instructions:

> I am interested in how school children remember things. That's why I came to your school today, and I will be working with each of the students in your class the same way we're going to work together now.
>
> This is not a test, like a school test, and what you do here does not have anything to do with your school work.
>
> I am going to ask you to listen very carefully to some things you will hear, and then after you hear each one, your job will be to say back what you've heard as best you can. Each time you will hear some words and then some numbers, and when the numbers stop, your job is to say the words and the numbers as best you can. Sometimes, what you hear may sound funny, other times it may sound O.K., but each time there will be some words and then some numbers, and each time your job will be to say back the words and the numbers you hear as best you can. Do you think you know what I would like you to do? (Repeat if necessary.)
>
> If you get tired and want to stop, just let me know, or, if you would like to stop and go back to your class, that will be O.K.
>
> The things you will hear will come from this machine, and from these cards which I have brought with me (demonstrate Language Master). The things you hear and the things you say back will be recorded on this machine (demonstrate tape recorder).
>
> Before we start, I want you to know that some of the things you will hear will be pretty hard to remember, but don't worry about that, because that's the way these cards were made. Even though it may be hard sometimes, I would like you to try really hard to remember as much as you can each time and say back as much as you can remember. Don't worry if there are things you can't remember, because these cards were made so nobody can remember everything he (she) hears. Again, your job each time is to say back as much as you can remember.
>
> If there are things you are not sure about it is O.K. to guess, but don't just make up words or numbers.
>
> Do you have any questions before we get started?

A trial consisted of the presentation of one of the stimuli on the Language Master (a sentence or string of words followed by a list of digits) and the subject's immediate recall of the stimulus. Each subject completed the twenty-five trials of one Sentence Type group. The trials were presented in randomized orders, and the intertrial interval was approximately ten seconds. No subject had difficulty understanding the instructions or completing the experimental task. All subjects seemed to enjoy participating in the study. The experimental sessions were tape recorded, and were later transcribed verbatim for coding and scoring.

## Results

### I. Scoring Criteria

Memory for the second task (i.e., digits) was taken as an index of relative degree of difficulty in processing the first task (i.e., sentence or word list), and constituted the data analyzed. Two scoring systems were devised to represent as accurately as possible the recall of the digits. The first of these systems scored the kinds of errors made in the recall of digits, while the second scored only for digits correctly recalled.

Errors in digit recall were classified according to four categories of errors.[8] Since examination of the data indicated that the distribution of the kinds of errors was approximately the same for the independent groups of subjects (i.e., Sentence Type groups), the numbers of errors for the four error categories were added to obtain a single Error Score for each trial. The reliability of this scoring system for 100 trials selected at random was .93. No disagreement between judges was more than one Error Score apart; for the few disagreements which occurred, the experimenter's scoring was taken as the value used in the analyses.[9]

### II. Analyses of the Data

The data in the analyses, then, were the mean Error Scores for each subject for the five trials of each of five different experimental conditions (i.e., sentences and word strings) administered. The data were analyzed by several "mixed" design analyses of variance with repeated measurements.

A 2 x 4 x 2 x 2 analysis of variance was computed on the Mean Error Scores to test whether (1) atonal sentences were more difficult to process than tonal sentences (i.e., yield greater Error Scores), (2) more complex sentences were more difficult to process than less complex sentences, (3) intonational features were relatively more important for more complex sentences, as compared with less complex sentences, and (4) whether the above three relationships were found for meaningless sentences (Anomalous Sentences) as well as meaningful sentences (Regular Sentences). The variables in this analysis were: (1) Sex; (2) Sentence Type, i.e., Simple, Question, Passive, or Passive-Question; (3) Intonation, i.e., Tonal versus Atonal verbal stimuli; and (4) Meaning, i.e., Regular versus Anomalous Sentences. The scores for the Random

Strings were not included in this analysis. A summary of this analysis is presented in Table 2.

The difference between the Error Scores for the Intonation variable (Tonal versus Atonal stimuli) was significant ($p < .01$, $F = 10.42$, $df = 1, 24$). The mean of the Error Scores for the Tonal stimuli ($m = 3.33$) was less than the mean of the Error Scores for the Atonal stimuli ($m = 3.71$). The difference between the Error Scores for the Meaning variable (Regular versus Anomalous Sentences) was also significant ($p < .01$, $F = 14.72$, $df = 1, 24$). The mean of the Error Scores for the Regular Sentences ($m = 3.32$) was less than the mean of the Error Scores for the Anomalous Sentences ($m = 3.73$). There were no differences between male and female subjects in Error Scores ($F = .07$, $df = 1, 24$), and there were no differences between Types of sentences ($F = 1.67$, $df = 3, 24$). However, Types of sentences for the two Meaning conditions showed a significant interaction ($p < .05$, $F = 3.50$, $df = 3, 24$). It can be seen from the means composing this interaction (Table 3), that for the Regular Sentences, there was an orderly increase in Error Scores corresponding with an increase in syntactic complexity of the Sentence Types. There are no such orderly increase in Error Scores for the Anomalous Sentences.

The effects of the experimental variables Intonation (Tonal versus Atonal Sentences) and Sentence Type (Simple, Question, Passive, or Passive-Question) were analyzed for the Regular Sentences only. A 4 x 2 analysis of variance was computed on the mean Error Scores for the Regular Sentences, Tonal and Atonal conditions (Table 4).

The main effect of sentence Types did not reach the arbitrary level of significance ($F = 2.39$, $df = 3, 28$, $p < .10$). The main effect of Intonation (Tonal versus Atonal stimuli) was found to be significant ($p < .001$, $F = 28.39$, $df = 3, 28$), with the mean Error Score for the Tonal stimuli ($m = 3.07$) less than the mean Error Score for the Atonal stimuli ($m = 3.57$). The interaction between Sentence Types and Intonation was also found to be significant ($p < .001$, $F = 12.65$, $df = 1, 28$). The Error Scores for this interaction are provided in Table 5, and these means are presented graphically in Figure 1. Using a Newman-Keuls procedure (Winer, 1962), the means in Table 5 were compared systematically with each other. The values of the differences between those pairs of means which differed significantly from each other are shown in Table 6. For the Tonal stimuli, the Simple Sentences score was significantly less than both the Passive Sentences score and the Passive-Questions score. However, the Simple Sentences score was not significantly less than the Questions score. The scores for the Questions, the Passives, and the Passive-Questions did not differ significantly from each other for the Tonal stimuli. For the Atonal stimuli, the Simple Sentences score was significantly less than both the Passives score and the Passive-Questions score. As with the Tonal stimuli, the Simple Sentences did not differ significantly from the Questions. However, for the

Table 2

Analysis of Variance of Error Scores for Four Sentence Types, Including Tonal and Atonal Sentences, and Regular and Anomalous Sentences

| Source of Variation | df | MS | F | p |
|---|---|---|---|---|
| Total | 127 | 0.720 | | |
| Between | 31 | 1.472 | | |
| Sex (S) | 1 | 0.096 | 0.07 | |
| Types (T) | 3 | 2.481 | 1.67 | |
| S x T | 3 | 0.795 | 0.53 | |
| Pooled I | 24 | 1.488 | | |
| Within | 96 | 0.477 | | |
| Intonation (I) | 1 | 4.535 | 10.42 | < .01 |
| S x I | 1 | 1.411 | 3.24 | |
| T x I | 3 | 0.304 | 0.70 | |
| S x T x I | 3 | 0.380 | 0.87 | |
| P I x I | 24 | 0.435 | | |
| Meaning (M) | 1 | 5.316 | 14.72 | < .01 |
| S x M | 1 | 0.027 | 0.07 | |
| T x M | 3 | 1.264 | 3.50 | < .05 |
| S x T x M | 3 | 0.348 | 0.96 | |
| P I x M | 24 | 0.361 | | |
| I x M | 1 | 0.466 | 1.94 | |
| S x I x M | 1 | 0.278 | 1.15 | |
| T x I x M | 3 | 0.087 | 0.36 | |
| S x T x I x M | 3 | 0.561 | 2.33 | |
| P I x I M | 24 | 0.241 | | |

Table 3

Mean Error Scores for Four Sentence Types, Regular and Anomalous Sentences

| | Sentence Type | | | |
|---|---|---|---|---|
| | Simple | Question | Passive | Passive-Question |
| Regular Sentences | 2.90 | 3.24 | 3.56 | 3.58 |
| Anomalous Sentences | 3.73 | 3.38 | 4.23 | 3.58 |

Table 4

Analysis of Variance of Error Scores for Four Sentence Types, Regular Sentences Only, Including Tonal and Atonal Conditions

| Source of Variation | df | MS | F | p |
|---|---|---|---|---|
| Total | 63 | | | |
| Between Groups | 31 | | | |
| Types (T) | 3 | 1.638 | 2.39 | < .10 |
| Error | 28 | .687 | | |
| Within Groups | 32 | | | |
| Intonation (I) | 1 | 4.000 | 28.92 | < .001 |
| I x T | 3 | 1.749 | 12.65 | < .001 |
| Error | 28 | .138 | | |

Table 5

Mean Error Scores, Intonation x Types Interaction

| | Sentence Type | | | |
|---|---|---|---|---|
| | Simple | Question | Passive | Passive-Question |
| Tonal | 2.70 | 3.03 | 3.35 | 3.20 |
| Atonal | 3.10 | 3.45 | 3.78 | 3.95 |

## Table 6

Significant Differences Between the Mean Error Scores for Four Sentence Types, Tonal and Atonal Conditions, Regular Sentences Only

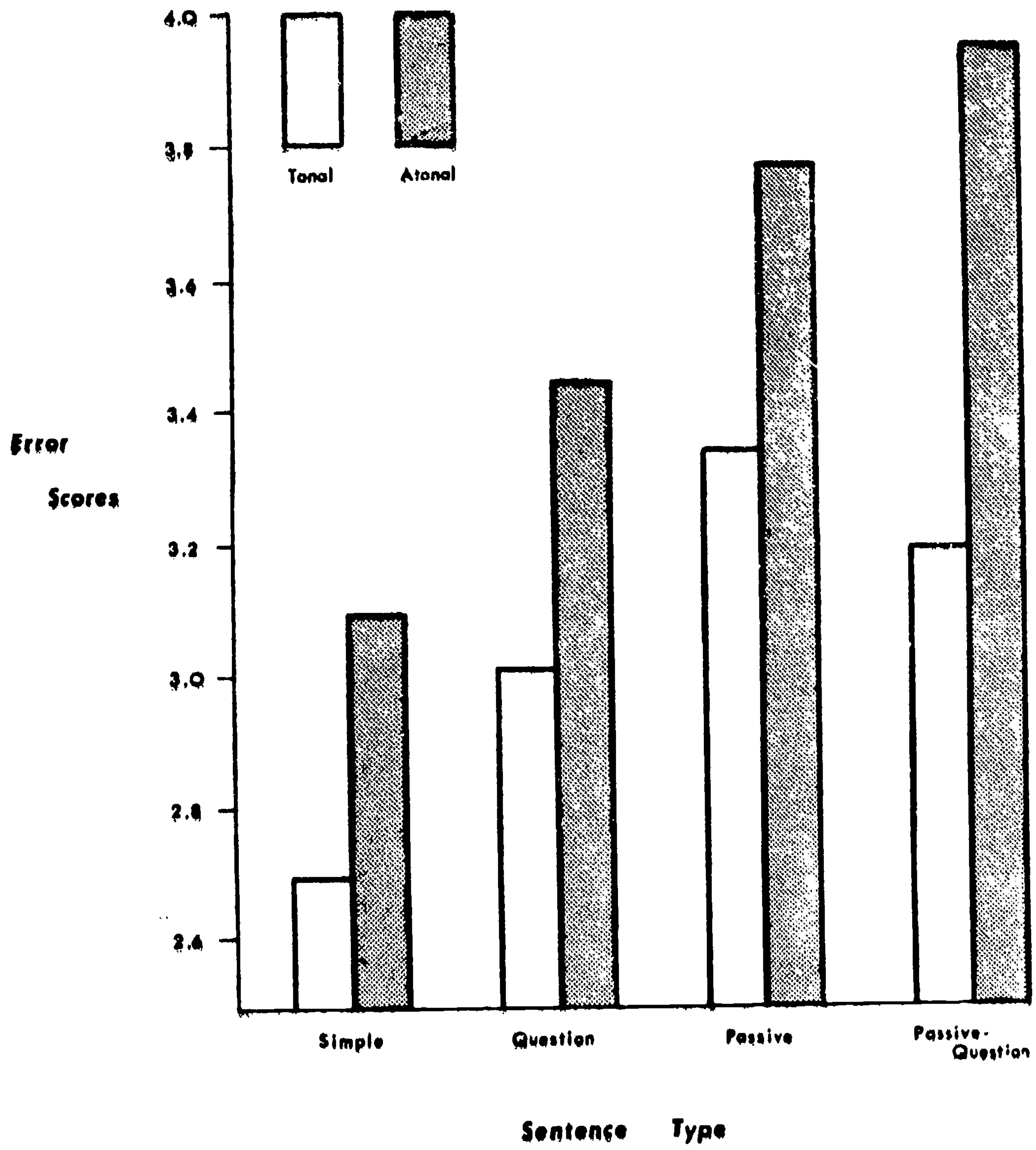
| | Tonal | | | | Atonal | | | |
|---|---|---|---|---|---|---|---|---|
| | S $M_1$ | Q $M_2$ | P $M_3$ | PQ $M_4$ | S $M_5$ | Q $M_6$ | P $M_7$ | PQ $M_8$ |
| $M_1$ | | | <u>.65</u> | .50 | .40 | <u>.75</u> | <u>1.08</u> | <u>1.25</u> |
| $M_2$ | | | | | | .43 | <u>.75</u> | <u>.93</u> |
| $M_3$ | | | | | | | .43 | <u>.60</u> |
| $M_4$ | | | | | | | <u>.58</u> | <u>.75</u> |
| $M_5$ | | | | | | | <u>.68</u> | <u>.85</u> |
| $M_6$ | | | | | | | | .50 |
| $M_7$ | | | | | | | | |

S = Simple
Q = Question
P = Passive
PQ = Passive-Question

Only significant differences between means are given in this table. All values given in this table are significant at the .05 level. The underlined values are significant at the .01 level.

FIGURE 1

Mean Error Scores for Four Sentence Types, Tonal and Atonal Conditions, Regular Sentences Only



Atonal stimuli, the score for the Questions was significantly less than the score for the Passive-Questions. The score for the Passives did not differ significantly from the score for the Passive-Questions, for the Atonal stimuli. Turning next to the comparisons of the Tonal and Atonal stimuli, it was found that for each of the Sentence Types taken in turn there was a significant difference between the Tonal and Atonal conditions, and that in each of these four comparisons, the score for the Tonal condition was less than for its associated Atonal pair. This set of results is represented graphically in Figure 1: the score for the Atonal stimuli is significantly greater than the score for the Tonal stimuli at each point (i.e., Sentence Type). It is of note that the syntactically simplest Sentence Type in the Atonal condition (Simple Sentence-Atonal) did not differ significantly from the syntactically most complex Sentence Type in the Tonal condition (Passive-Question-Tonal). The presence of Intonation made the most complex sentences in this study equivalent, in terms of the measures employed, to the simplest sentences with the Intonation "removed." [10]

The hypothesis relating to the increasing importance of Intonation with increasing syntactic complexity can be explored in the graphic representation in Figure 1. This hypothesis states that the difference between the two values (Tonal versus Atonal) for each Sentence Type should become greater as the syntactic complexity of the Sentence Type increases. To test this hypothesis, the Error Scores for the syntactically simplest Sentence Type and the syntactically most complex Sentence Type were converted to Error Difference Scores (i.e., Atonal Error Score - Tonal Error Score). One such Error Difference Score was derived for each subject in these two groups, using only the Regular Sentences trials, and a t-test was performed on these data. The difference between the Error Difference Scores of the two independent groups (Simple Sentences versus Passive-Questions) was significant ($p < .05$, $t = 2.19$, $df = 14$).

To test the possible interpretation that the above results might be a function of the length of the stimulus sentences, the fifth Sentence Type had been included as a control. This fifth Sentence Type was the Who-Question, and the stimulus sentences in this group were one word shorter than the corresponding stimulus sentences in the Simple Sentences group. If the results reported thus far were a function of the length of the sentences, then it would be expected that the Error Scores corresponding to the Who-Questions would be <u>less</u> than the Error Scores corresponding to the Simple Sentences. On the other hand, if the above results were a function of the syntactic complexity of the stimulus sentences, then it would be expected that the Error Scores for the Who-Question would be <u>greater</u> than the Error Scores for the Simple Sentences.

To examine the Who-Question in relation to the other Sentence Types, a $5 \times 2$ analysis of variance was performed for the mean Error Scores of the Regular Sentence trials, including the Tonal and Atonal stimulus conditions and Error Scores for the Who-

Question Sentence Type group as the fifth level of the first variable. A summary of this analysis is provided in Table 7. The results of this analysis were essentially the same as the analysis of variance utilizing four Sentence Types.

The mean Error Scores for the Who-Questions for tonal and atonal conditions were identical ($m = 3.23$). A Newman-Keuls procedure indicated that for tonal conditions, the mean Error Score for the Who-Questions was significantly greater than the mean Error Score of the Simple Sentences, but was not significantly different from the mean Error Scores of any of the other Sentence Types. For atonal conditions, the mean Error Score for Who-Questions was not significantly different from mean Error Scores of the Simple Sentences or Questions, but was significantly less than the mean Error Scores of both Passives and Passive-Questions (all at $p < .01$). Thus, it would seem that the finding that increasing complexity results in greater Error Scores for atonal stimuli is not due simply to the increasing length of the sentences.

To compare the Error Scores for the Random Strings with the Error Scores for the Regular Sentences and the Anomalous Sentences, a 2 x 4 x 3 analysis of variance was computed with the Error Scores for the Atonal conditions (Table 8). The variables included in this analysis were: (1) Sex; (2) Sentence Types, i.e., Simple, Question, Passive, or Passive-Question; and (3) Meaning, i.e., Regular Sentences, Anomalous Sentences, and Random Strings. The main effect of Meaning was significant ($p < .001$, $F = 29.55$, $df = 2, 48$). The means were ordered as expected; Regular Sentences ($m = 2.57$), Anomalous Sentences ($m = 3.85$), and Random Strings ($m = 4.76$), and each was significantly different from the other two.

With the repeated measurements design utilized in the present study, some confounding (i.e., across trials) might have occurred (e.g., practice effects, fatigue, proactive interference). Two questions can be asked: (1) Did such a systematic effect occur across trials within an experimental session?; and (2) If so, was this effect different for each of the experimental groups (Sentence Types)? It will be recalled that the order of trials was controlled by randomization in this study. This kind of control assured that any effects which occurred across trials were distributed randomly across subjects. This procedure controlled the order of presentation of the various experimental conditions to each subject, but there might have been a systematic effect across trials independent of the various experimental conditions presented to a given subject. To answer these questions, the data were examined in terms of the order of presentation of trials, that is, independently of the various experimental conditions as presented to each subject.

Table 7

Analysis of Variance of Error Scores for Five Sentence Types (Who-Question Included), Tonal and Atonal Conditions, Regular Sentences Only

| Source of Variation | df | MS | F | p |
|---|---|---|---|---|
| Total | 79 | | | |
| Between Groups | 39 | | | |
| Types (T) | 4 | 1.2565 | 1.29 | |
| Error | 35 | .97069 | | |
| Within Groups | 40 | | | |
| Intonation | 1 | 3.200 | 30.11 | < .001 |
| I x T | 4 | 2.340 | 22.02 | < .001 |
| Error | 35 | .1063 | | |

Table 8

Analysis of Variance of Error Scores for Regular Sentences, Anomalous Sentences, and Random Strings for Four Sentence Types

| Source of Variation | df | MS | F | p |
|---|---|---|---|---|
| Total | 95 | 1.001 | | |
| Between Subjects | 31 | 1.442 | | |
| Sex (S) | 1 | 1.516 | 1.11 | |
| Types (T) | 3 | 2.950 | 2.16 | |
| S x T | 3 | .530 | .39 | |
| Pooled I | 24 | 1.364 | | |
| Within Subjects | 64 | .788 | | |
| Meaning (M) | 2 | 12.331 | 29.55 | < .001 |
| S x M | 2 | .008 | .02 | |
| T x M | 6 | .643 | 1.54 | |
| S x T x M | 6 | .309 | .74 | |
| Pooled I x M | 48 | .417 | | |

A 2 x 4 x 5 analysis of variance, including trend components for the within-subjects variable, was performed on the Error Scores of the four Sentence Type groups. The variables in this analysis were: (1) Sex; (2) Sentence Type, i.e., Simple, Question, Passive, or Passive-Question; and (3) Blocks. Each subject had twenty-five trials. To test for trials effects, these trials were grouped into five Blocks of five trials each, according to the order of administration of the trials, independent of experimental conditions. The means of these five Blocks were used as scores in the analysis of variance, and the summary is presented in Table 9.

Blocks were found to be significant ($p < .01$, $F = 3.95$, $df = 4, 96$). This variable was also found to have a significant linear trend component ($p < .01$, $F = 10.13$, $df = 1, 96$), and a significant residual trend component ($p < .05$, $F = 4.10$, $df = 1, 96$). The means of the Error Scores for each of the five Blocks are presented in Figure 2. The results of the analysis of variance indicate that there was a systematic effect across trials within the experimental sessions. The Error Scores increased across trials within the experimental sessions, and this effect could be attributed to fatigue or proactive interference, rather than to a practice effect. While there was a systematic effect across trials within the experimental sessions such that the Error Scores increased non-monotonically with Blocks of trials, this effect was similar for the four Sentence Type groups (i.e., the differences between the experimental groups in terms of this trials effect were not statistically reliable). Thus, the results presented earlier for the Sentence Type groups may be interpreted in terms of the experimental conditions, rather than in terms of this trials effect.

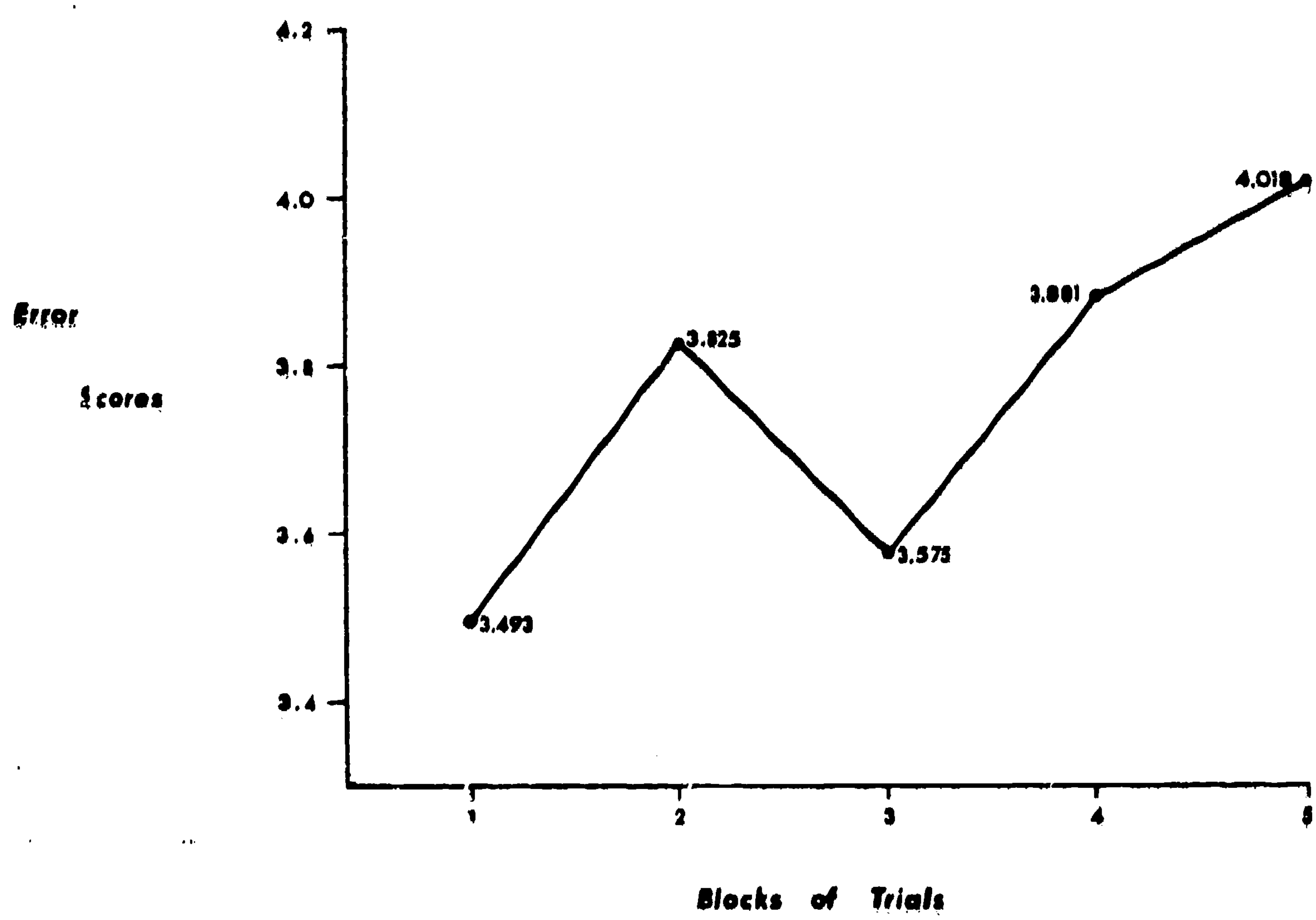
Table 9

Analysis of Variance of Error Scores for Four Sentence Types, Including Trend Components for Blocks of Trials

| Source of Variation | df | MS | F | p |
|---|---|---|---|---|
| Total | 159 | .667 | | |
| Between Subjects | 31 | 1.813 | | |
| Sex (S) | 1 | .272 | .15 | |
| Types (T) | 3 | 3.310 | 1.84 | |
| S x T | 3 | .910 | .50 | |
| Pooled I | 24 | 1.803 | | |
| Within Subjects | 128 | .389 | | |
| Blocks (B) | 4 | 1.527 | 3.95 | $< .01$ |
| Linear | 1 | 3.916 | 10.13 | $< .01$ |
| Quadratic | 1 | .065 | .17 | |
| Cubic | 1 | .544 | 1.41 | |
| Residual | 1 | 1.585 | 4.10 | $< .05$ |
| S x B | 4 | .097 | .25 | |
| Linear | 1 | .190 | .49 | |
| Quadratic | 1 | .032 | .08 | |
| Cubic | 1 | .128 | .33 | |
| Residual | 1 | .041 | .11 | |
| T x B | 12 | .324 | .84 | |
| Linear | 3 | .247 | .64 | |
| Quadratic | 3 | .188 | .49 | |
| Cubic | 3 | .473 | 1.22 | |
| Residual | 3 | .387 | 1.00 | |
| S x T x B | 12 | .195 | .51 | |
| Linear | 3 | .166 | .43 | |
| Quadratic | 3 | .360 | .93 | |
| Cubic | 3 | .158 | .41 | |
| Residual | 3 | .097 | .25 | |
| P I x B | 96 | .386 | | |

FIGURE 2

Mean Error Scores for Four Sentence Type Groups, for Each of Five Blocks of Trials Within Experimental Sessions



## Method

I. Subjects

The subjects were twenty boys and twenty girls enrolled in fourth grade classes in eleven schools in a New England Public School System. None of the subjects who participated in this study had participated in the earlier study. Good and poor readers were selected from a much larger sample on the basis of their scores on the Paragraph Meaning subtest of the Stanford Achievement Test (Primary II Battery, Form W), which had been administered the previous year by the school system.[11] Those subjects whose Paragraph Meaning scores were at or above the score expected for their grade level were designated good readers, and those subjects whose scores were at least one year below the score expected for their grade level were designated as poor readers. Each poor reader (10 boys, 10 girls) was matched with a good reader according to five criteria: (1) attendance in the same school, if possible in the same classroom in the same school, on the assumption that this would yield equivalent socioeconomic status and educational experience; (2) same sex; (3) same age; (4) same race (all subjects were Caucasian); and (5) equivalent intelligence test scores, as measured by the Otis-Lennon Test of Mental Ability (Elementary I Level, Form K), also administered the previous year. If the intelligence test scores of the two members of a good-poor reader pair differed, the good reader had the lower score (with one exception); all subjects had intelligence test scores within the range 85-115.

As the subjects were being tested in the present study, the Public School System administered a reading and an intelligence test. The Reading Comprehension scores from the Comprehensive Tests of Basic Skills (Level 1, Form Q) provided the second school-administered measure of reading comprehension[12]; the second school-administered measure of intelligence was the Short Form Test of Academic Aptitude. The first set of tests used as part of the original subject selection criteria, was administered when the subjects were in the third grade; the second set of tests was administered 14 months later, when the subjects were in the fourth grade. The availability of the second set of school-administered tests (reading comprehension and intelligence), as well as our own assessment of the subject's oral reading, made it possible to examine the results in terms of several criteria of "good" versus "poor" readers. Subjects' scores on all four tests, ages, and scores on Oral Reading (described below) are presented in Table 10.

II. General Procedure

The general procedure was similar to that of the first study. Prospective subjects were contacted through their schools. Brief explanations of the study were sent home with each prospective subject (and usually with all his classmates), and a written reply was returned by his parents. Only children whose parents had given consent were tested. Whenever possible, the schools provided quiet rooms for the study, so that interruptions would be minimal. Each subject was tested individually during his time at school, with care being taken not to interrupt lunch, recess, or an especially important classroom activity. The experimental session lasted 35-45 minutes, and consisted of two tasks: (1) an immediate memory task, similar in procedure to that used in the first study; and (2) an oral reading task.

## Table 10

Pairs of Good and Poor Readers: Ages, Scores on Two School-Administered Tests of Reading Comprehension (RC) Scores on Two School-Administered Tests of Intelligence (I), and Oral Reading Scores (OR)

| | | Good Readers | | | | | | Poor Readers | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Age (mos.) | RC #1 | RC #2 | I #1 | I #2 | OR | Age (mos.) | RC #1 | RC #2 | I #1 | I #2 | OR |
| Male Pairs | 1. | 114 | 4.2 | 7.6 | 93 | 96 | 1.25 | 117 | 2.1 | 2.0 | 97 | -- | 3.50 |
| | 2. | 114 | 4.1 | 6.9 | 99 | 101 | 1.50 | 113 | 2.1 | 3.5 | 100 | 99 | 2.88 |
| | 3. | 113 | 3.6 | 4.2 | 104 | 113 | 2.25 | 117 | 1.9 | 5.7 | 104 | 104 | 1.75 |
| | 4. | 110 | 4.8 | 6.9 | 109 | 113 | 1.50 | 113 | 2.3 | 3.2 | 114 | 96 | 2.63 |
| | 5. | 116 | 4.0 | 5.0 | 96 | 106 | 2.13 | 118 | 2.0 | 2.9 | 98 | 92 | 3.25 |
| | 6. | 109 | 3.7 | 3.8 | 99 | 103 | 1.75 | 119 | 2.1 | 4.4 | 100 | 97 | 2.50 |
| | 7. | 115 | 3.4 | 5.3 | 112 | 115 | 1.88 | 116 | 2.0 | 3.6 | 112 | 99 | 2.13 |
| | 8. | 112 | 3.4 | 7.6 | 103 | 124 | 1.63 | 110 | 2.0 | 3.8 | 103 | 98 | 3.00 |
| | 9. | 114 | 3.6 | 5.7 | 89 | 107 | 1.50 | 121 | 1.7 | 3.4 | 92 | 102 | 3.50 |
| | 10. | 112 | 3.6 | 4.2 | 105 | 93 | 2.00 | 110 | 1.7 | 2.9 | 113 | 87 | 3.00 |
| Female Pairs | 1. | 114 | 4.1 | 7.6 | 106 | 110 | 1.00 | 117 | 2.2 | 5.7 | 107 | 102 | 2.38 |
| | 2. | 118 | 4.4 | 7.6 | 87 | 137 | 1.25 | 111 | 2.0 | 4.1 | 90 | 91 | 2.25 |
| | 3. | 112 | 3.3 | 5.0 | 100 | 111 | 1.13 | 109 | 2.2 | 3.8 | 103 | 100 | 2.00 |
| | 4. | 110 | 3.2 | 4.1 | 99 | 104 | 1.75 | 117 | 2.0 | -- | 98 | -- | 2.63 |
| | 5. | 120 | 4.0 | 5.7 | 104 | 107 | 1.67 | 120 | 2.1 | 3.9 | 106 | 93 | 1.63 |
| | 6. | 117 | 4.1 | 6.9 | 98 | 109 | 1.38 | 114 | 2.1 | 3.6 | 106 | 104 | 1.88 |
| | 7. | 118 | 3.3 | 5.7 | 94 | 103 | 1.00 | 109 | 2.2 | 4.4 | 100 | 100 | 2.38 |
| | 8. | 118 | 3.4 | 5.0 | 90 | 92 | 1.63 | 120 | 2.0 | 3.9 | 96 | 93 | 1.75 |
| | 9. | 115 | 4.2 | 5.7 | 105 | 103 | 1.50 | 110 | 2.3 | 3.4 | 107 | 91 | 2.38 |
| | 10. | 117 | 4.2 | 6.9 | 94 | 97 | 1.25 | 121 | 2.0 | 2.6 | 103 | 79 | 3.25 |

Each row in this table gives sets of scores for two subjects, a good-poor reader pair, as selected and matched by the original criteria (age and intelligence test #1). Reading comprehension test #1 was used as part of the original criteria to designate good versus poor readers. Those six Ss who changed reader designation as a function of reading comprehension test #2 were: male good readers number 6 and 10; male poor readers numbers 3 and 6; female good reader number 4; and, female poor reader number 1.

A. Materials

The stimuli for this study were 48 stimuli of the kind used in the earlier study; 40 as used previously and 8 new ones. The old stimuli were four of the Regular Sentences, each in the five sentence types (i.e., Simple active, Question, Passive, Passive-Questions and Who Questions). The 8 new stimuli were introduced to better control for the length of the verbal material (i.e., sentences which are grammatically more complex are often also longer, in terms of number of words, than sentences which are grammatically less complex). An indirect control for this possible confounding was the use, in this study as in the earlier one, of the Who-Question, which contains fewer words than the simple-active-declarative, but is, according to a transformational grammar, more complex and seems to be more difficult to recall (e.g., see Results of Study 1; Savin and Perchonock, 1965). In this study, an additional Sentence Type was included to more adequately control for possible effects of increased sentence length, independent of sentence complexity. This additional Sentence Type was the simple-active-declarative with two additional modifying words; thus, although grammatically simple, this Sentence Type has the same number of words (7) as the Passive and the Passive-Question, with the latter two assumed to be grammatically more complex. As for the other five Sentence Types, four exemplars of this additional Sentence Type (called the "Control" Sentence Type) were constructed: (1) "The girl walked the dog after school"; (2) "The woman ate some of the food"; (3) "The young boy wrote the long letter"; and (4) "The tall man drove the car away." These four sentences were recorded on Language Master cards by the same trained male adult speaker who recorded the stimuli used in Study 1; each of the four sentences was recorded in two ways, tonally and atonally, as described earlier, and for each of the eight new stimuli thus constructed, a string of random digits was recorded immediately following the sentence on the same Language Master card by the same speaker, in the same manner as was used in constructing all the other stimuli. The eight new stimuli were judged independently to be of similar length and intensity as the other stimuli used in this study, and to be of equivalent intonational quality (i.e., relative presence or absence of patterns of pauses, stress and pitch variations) as the other stimuli (i.e., sentences recorded in tonal versus atonal manner). In sum, there were a total of 48 stimuli employed in the recall task of this study (4 exemplars of each of 6 Sentence Types, each in a tonal and an atonal version). The particular Sentence Types included here allowed for analyses of possible effects of grammatical complexity independent of sentence length, as well as possible effects of sentence length independent of grammatical complexity.

In contrast to the earlier study, it was decided to use only 5 digits instead of 6 to decrease total errors for all subjects. This modification was accomplished by carefully masking the 6th digit of each of the stimuli, and thus there was now no downward inflection at the end of the digit list.

All subjects were presented all 48 stimuli.[13] The 48 stimuli were arranged in a fixed random order, satisfying the following conditions: each of four successive blocks of twelve trials contained (a) three instances of each of the four lexical variations; (b) two instances of each of the six Sentence Types; and (c) six instances of each of the two intonational variations (i.e., tonal-atonal). No lexical variation or Sentence Type appeared twice in succession, and no intonational type appeared more than twice in succession. This single fixed random (partially constrained) sequence of the 48 stimuli was presented in two Orders: Forward presentation to half the subjects, and Backward presentation of the same sequence to the other half. Order of stimulus presentation (i.e., Forward versus Backward) was completely counterbalanced for the two between-subjects variables (i.e., sex, and good versus poor readers).

For the oral reading task, five paragraphs from the Gray Oral Reading Test (Form A) were used. These passages (2 through 6) are graded in difficulty from the primer level through the fourth grade (Robinson, 1967). Each paragraph was accompanied by 4 questions concerning the information presented in that paragraph.

B. Procedure

The experiment was explained to the subjects as a study of children's memory for different types of things, such as words, numbers, and written stories. Instructions and procedure for the recall task (presented first to all subjects) were the same as Study 1. After all 48 stimuli of the recall task had been presented, and S had responded to each in turn, E said:

> That's all the cards I have, (name), but there's one more thing that I'd like for you to do before you leave. I brought some short paragraphs that I would like you to read out loud. You don't have to worry about reading quickly or about making mistakes, but you should try to remember what you read, because after you finish each one, I'll ask you a few questions about it.

As in the previous study, Ss were told their performance would have no bearing on their classroom work (e.g., that their performance would not be reported to their teacher), and that they were free to terminate the experiment whenever they wanted. All Ss completed the experimental task, and seemed to enjoy the experience. The entire session with each S was tape recorded.

## Results

As in the previous study, memory for the digit list on each trial was taken as an index of the relative degree of difficulty in processing the sentence which preceded it, and constituted one kind of data analyzed. Additionally, in this study, errors in recall of the sentences were analyzed independently.

1. Errors in Digit Recall

Recall of each digit list was scored. As before, for each trial, all types of errors were summed to obtain an Error Score for that trial. Interjudge reliability of kind of Error Score for 100 randomly selected trials was .97; one judge scored the remaining trials.

A. Differences Between Good and Poor Readers.

The primary purpose of this study was to test the hypotheses that (1) certain poor readers, compared with matched good readers, show greater difficulty in processing (remembering) spoken language when intonational features are "removed" from the stimulus, and (2) these poor readers also show relatively greater difficulty, compared with good readers, in processing (remembering) the "atonal" stimuli as syntactic complexity increases. There were several criteria possible for defining "good" versus "poor" readers; separate sets of analyses were performed on the Error Scores for reader groups as defined by each of the following criteria: (1) the original selection criteria (e.g., third grade school-administered test of reading comprehension); (2) a more recent (fourth grade) school-administered test of reading comprehension; (3) our own measure of subjects' oral reading, scaled according to the relative presence or absence of "melodic" features in subjects' oral reading; and (4) a combination of these three reading measures.

Although not planned in the original design of this study, comparisons between the two sets of school-administered tests (i.e., 3rd and 4th grades) were also possible. For all subjects pooled, good and poor readers combined, third and fourth grade scores showed some reasonable correspondence ($r = +.78$, $p < .005$). For good readers, considered separately, the correlation between thrid and fourth grade reading comprehension tests was significant ($r = +.59$, $p < .005$), although there was no apparent correlation between these two tests for the poor readers, considered separately ($r = +.12$, not significant). The correlation between tests for the good readers was significantly higher than the corresponding correlation for the poor readers ($z = 2.28$, $p < .05$, r to z transformation and approximation to the normal curve, Edwards, 1965), indicating that the lowered correlations for the two subgroups (i.e., good and poor readers) of the total sample are not simply due to attenuation of the correlation coefficients resulting from a decreased sample size. Thus, it would appear, for these Ss, the inter-test reliability of these two measures of reading comprehension was significant, but only for those Ss originally designated as good readers; for poor readers, there was no apparent correspondence between test scores.[14]

Surprisingly, there was no apparent relationship between intelligence test scores, for good and poor readers combined, as measured by the two sets of intelligence tests ($r = -.12$, not significant). The Ss had originally been selected and matched on the basis of their third grade scores so that intelligence test scores would not be significantly correlated with reading comprehension scores ($r = -.24$, not significant); that is, the procedure of matching for intelligence test scores yielded equivalent mean intelligence test scores for those Ss designated as good and poor readers, by the original selection criteria. According to the original matching procedures, when a good-poor reader pair differed in intelligence scores, whenever possible **the good reader had the lower intelligence score ; in all pairs the difference** between intelligence scores was less than 10 points. This procedure resulted in a mean third grade intelligence test score for good readers ($m = 99.3$) which was lower than the mean third grade intelligence test score for poor readers ($m = 102.5$), although this difference did not reach the arbitrary level of significance ($t = 1.50$, $df = 38$, $p > .10$).[15] Although Ss were matched for intelligence test scores, as measured by the third grade test, comparisons of third and fourth grade intelligence scores (Table 10) revealed that good readers' intelligence scores systematically increased from one year to the next, while poor readers' intelligence scores systematically decreased, as measured by two different tests of intelligence. Of the 20 Ss originally designated as good readers, 18 increased in intelligence score; of the 20 Ss originally designated as poor readers, 15 decreased in intelligence score (fourth grade intelligence test scores were unavailable for 2 poor readers). These observations were supported by t-tests for related measures: (1) for good readers, third grade intelligence score ($m = 99.3$) was significantly less than fourth grade intelligence score ($m = 107.2$) ($t = 2.92$, $df = 19$, $p < .01$); (2) for poor readers third grade intelligence score ($m = 102.5$) was significantly greater than fourth grade intelligence score ($m = 95.9$) ($t = 3.19$, $df = 17$, $p < .01$, two-tailed tests).

This apparent confounded validity of the school-administered tests was also suggested by correlations between the intelligence tests and the reading tests for the fourth grade. Subjects were selected and matched so that there would be no significant correlation between the first intelligence test and the first reading comprehension test ($r = -.24$, not significant). However, there was a significant positive correlation between the second intelligence test and the second reading comprehension test ($r = +.64$, $p < .005$).[16]

Given this apparent confounded validity for the standard school-administered tests of reading comprehension and intelligence, Ss in a second analysis were reassigned to good and poor reader groups on the basis of the more recent school-administered reading comprehension test, determined by a median split of subjects' scores on that test. For 6 Ss, this new classification reversed their original classification, as determined by the third grade tests (the Original Subject Grouping). For those Ss whose good-poor reader designation changed in the new grouping of Ss (the First Subject Regrouping), the original matching criteria were violated. Good and poor readers (First Subject Regrouping) had equal numbers of males and females, but the 6 Ss who changed classification were no longer matched for school attended or intelligence test scores. Thus, good readers in this analysis had significantly higher intelligence test scores than did poor readers ($t = 4.09$, $p < .001$), intelligence as measured by the fourth grade intelligence test. However, for this grouping of Ss, comparison of the intelligence scores as measured by the third grade test showed poor readers' intelligence scores were not significantly different from good readers' scores ($t = 1.105$, $p < .05$). In summary, for both the Original Subject Grouping and the First Subject Regrouping, the mean intelligence score of the good readers was not significantly different from the mean intelligence score of the poor readers; intelligence as measured by the third grade intelligence test (the original selection criterion). However, when the mean of the fourth grade intelligence test scores of the good and poor reader groups were compared, good readers had a significantly higher mean intelligence score than did poor readers.[17]

Subjects could also be assigned to good and poor reader groups on the basis of a sample of their oral reading taken as part of the experimental procedure. A rating system was devised which attempted to assess the degree that the oral reading included "melodic" features (i.e., the extent to which pauses, tonal variations, and other prosodic features corresponded with the sentence and phrase structure of the printed material). The attempt here was to assess the inclusion of such features, independent of the subject's possible occasional difficulty in word identification, by having each S read four passages of increasing difficulty. The assumption was that if a S does not organize reading material ( as indicated by the relative lack of "melodic" features in his oral reading), then this failure to organize should be apparent even on passages where he has no difficulty in word-identification (i.e., the easiest passages). A four-point rating scale was devised, ranging from (1) melodic features always present and appropriate to the material, to (4) melodic features absent--word-by-word reading. Each S's oral reading of each of the four passages was assessed independently by two judges who had been trained to use this scale, but who were unaware of the S's good-poor reader classification by any other criteria. Interjudge reliability was .89 ($df = 158$, $p < .01$). For each S, eight separate ratings (i.e., by two judges on four passages) were then averaged to obtain an Oral Reading Score for each S. By a median split, the 20 Ss having the better ratings for Oral Reading were designated as good readers and the 20 Ss having the poorer ratings on this scale were designated as poor readers.

Oral Reading Scores were compared with each of the four school-administered tests for all Ss pooled (good and poor readers combined). The Oral Reading Scores correlated relatively highly and consistently with the two school-administered measures of reading comprehension (for Oral Reading Scores and third grade Reading Comprehension scores, $r = +.73$, $p < .01$; for Oral Reading Scores and fourth grade Reading Comprehension scores, $r = +.78$, $p < .01$).[18] While both these correlations are significant, they are not as high as might be expected (i.e., they account for a little over one-half the variance), given that relative extremes on a scale of "reading" were selected (good versus poor readers). To the extent that these two kinds of reading measures are not more highly correlated (i.e., oral reading and reading comprehension, as measured by group-administered tests) for extremes of a "reading" scale, it is suggested that skills necessary for "good" oral reading (word identification and appropriate melodic organization) may be different from those skills necessary for "good" comprehension in silent reading, as measured by many group-administered standard tests. Indeed, even to call these both measures of "reading" (a single-class phenomenon: see Wiener and Cromer, 1967) may collapse distinctions which may be critical in distinguishing between various kinds of "reading" problems.

As might be expected, there was no apparent relationship between Oral Reading Scores and third grade intelligence test scores ($r = -.15$, not significant), however, the relationship between Oral Reading Scores and the fourth grade intelligence test scores was significant ($r = +.49$, $p < .01$). As reported earlier, the two intelligence tests themselves were apparently unrelated, at least for these Ss. The set of findings regarding the correspondence of the Oral Reading Scores to the two school-administered tests of intelligence would seem to lend some support to an interpretation that the first intelligence test was less "reading loaded" than the second intelligence test (to be discussed later).[19]

In addition to the three groupings of Ss described above (i.e., based on (1) the original selection criteria; (2) a more recent school-administered reading comprehension test; and (3) our own scale of samples of oral reading) an attempt was made to group Ss on all three measures of reading conjunctively. For this final grouping of Ss good-poor reader pairs were matched on the basis of (1) sex; (2) age; (3) race (all Caucasian); (4) school attended (not necessarily the same classroom within the school); and (5) intelligence test scores, based on the third grade test of intelligence. Only those Ss were included whose reading scores were consistent on all three measures of reading (upper or lower half of each respective distribution of scores). This requirement resulted in the elimination of 16 of the original 40 Ss, yielding 12 good-poor reader pairs.

To test for differences in Sentence Types and Intonation as a function of reader level, separate analyses of variance were computed for the four possible reader groupings: (1) third grade reading test; (2) fourth grade reading test; (3) Oral Reading Score; (4) combined scores. Each of these analyses was a 2 x 2 x 2 x 6 x 4 "mixed" design, with the following variables: (1) Readers (good versus poor); (2) Sex; (3) Intonation (tonal-atonal); (4) Sentence Types; (5) Trials. With the exception of readers grouped according to the fourth grade reading test, there was no evidence of any significant difference between the reader groups (good versus poor readers), nor was there evidence of a significant interaction between Readers and the other experimental variables (Intonation, Complexity). Of course, the results for the within-subjects variables (i.e. Intonation and Complexity for all Ss, good and poor readers combined) were the same in each analysis, and will be included in the analysis for Ss grouped by the fourth grade reading test, as reported below. The data for the fourth grade reading test grouping shows mean Error Score for poor readers ($m = 2.13$) was significantly greater than that obtained by good readers ($m = 1.59$) ($F = 5.81$, $df = 1.36$, $p < .05$). No interaction involving the Readers variable approached the arbitrary level of significance. Although the Sex main effect was not significant ($F = .73$, $df = 1, 36$), there was one interaction involving this variable which reached significance, Readers x Sex x Intonation x Complexity ($F = 3.37$, $df = 5, 180$, $p < .01$). (See Table 11)

The Trials main effect was not significant ($F = 1.24$, $df = 3, 108$), indicating that overall no one particular lexical variation was significantly harder or easier to process than the others. Although the Trials x Complexity interaction was significant ($F = 4.95$, $df = 15, 540$, $p < .01$), examination of the twenty-four means composing this interaction revealed no clearly interpretable pattern of results.[20]

The mean of the Error Scores for the atonal stimuli ($m = 1.92$) was significantly greater than the mean of the Error Scores for the tonal stimuli ($m = 1.80$), indicating that tonal stimuli were easier to process than atonal stimuli ($F = 9.51$, $df = 1, 36$, $p < .01$).

The complexity main effect was also significant ($F = 16.25$, $df = 5, 180$, $p < .001$). The means of the Error Scores of the six sentence types were ordered and subjected to systematic pair-wise comparisons by a Duncan Multiple-Range Test (Bruning and Kintz, 1968); differences between ordered mean Error Scores which were found to be significant are presented in Table 12. Of fifteen possible pair-wise comparisons (i.e., six means taken two at a time), ten such comparisons were found to be significant (all at $p < .01$). The mean Error Score of the Who-Question ($m = 1.47$) was less than the mean Error Score of the five other sentence types, the Active ($m = 1.72$), the Question ($m = 1.82$), the Passive ($m = 1.98$), the Passive-Question ($m = 2.04$), and the Control sentence type ($m = 2.13$). The mean Error Score of the Active sentence was less than the mean Error Scores of the Passive, the Passive-Question, and the Control sentence types. The mean Error Score of the Question was less than the mean Error Score of the Passive-Question and the Control sentence types. Thus, three sentence types, the Passive, the Passive-Question, and the Control, were significantly more difficult to process than each of the other three sentence types, but not significantly different from each other.

The Intonation x Complexity interaction was also found to be significant ($F = 2.24$, $df = 5, 180$, $p < .05$). The twelve mean Error Scores (i.e., tonal versus atonal for six sentence types) composing this interaction were compared systematically with each other using a Duncan Multiple-Range Test (Table 13). For the tonal stimuli, the Who-Questions score was significantly less than the Control mean Error Score. For the atonal stimuli, the Who-Question score was significantly less than the Question score, the Passive score, the Passive-Question score, and the Control score; the Active score was significantly less than the Control score. However, in contrast to the findings of Study 1, none of the tonal versus atonal comparisons within a given sentence type reached the arbitrary level of significance. (This finding was examined further by analyzing the data in terms of the length of the stimuli, independent of complexity, as reported below.)

## Table 11

Analysis of Variance of Error Scores for Good versus Poor Readers (4th-Grade Reading Test) for Six Sentence Types, Tonal and Atonal Conditions

| Source of Variation | df | MS | F | p |
|---|---|---|---|---|
| Total | 1919 | 1.712 | | |
| Between | 39 | 27.463 | | |
| Readers (R) | 1 | 144.272 | 5.80 | < .05 |
| Sex (S) | 1 | 18.191 | 0.73 | |
| R x S | 1 | 12.834 | 0.52 | |
| Error | 36 | 24.888 | | |
| Within | 1880 | 1.178 | | |
| Intonation (I) | 1 | 7.424 | 9.51 | < .01 |
| R x I | 1 | 0.750 | 0.96 | |
| S x I | 1 | 0.000* | 0.00 | |
| R x S x I | 1 | 0.337 | 0.43 | |
| Error | 36 | 0.781 | | |
| Complexity (C) | 5 | 18.667 | 16.25 | < .001 |
| R x C | 5 | 0.191 | 0.17 | |
| S x C | 5 | 0.588 | 0.51 | |
| R x S x C | 5 | 1.154 | 1.01 | |
| Error | 180 | 1.149 | | |
| Trials (T) | 3 | 1.693 | 1.24 | |
| R x T | 3 | 0.776 | 0.57 | |
| S x T | 3 | 1.675 | 1.22 | |
| R x S x T | 3 | 3.473 | 2.54 | (<.10) |
| Error | 108 | 1.370 | | |
| I x C | 5 | 1.894 | 2.24 | < .05 |
| R x I x C | 5 | 0.313 | 0.37 | |
| S x I x C | 5 | 2.852 | 3.37 | < .01 |
| R x S x I x C | 5 | 0.729 | 0.86 | |
| Error | 180 | 0.845 | | |
| I x T | 3 | 1.292 | 1.07 | |
| R x I x T | 3 | 1.041 | 0.86 | |
| S x I x T | 3 | 2.165 | 1.79 | |
| R x S x I x T | 3 | 0.832 | 0.69 | |
| Error | 108 | 1.209 | | |

*Computer rounding error

Table 11 - continued -

| Source of Variation | df | MS | F | p |
|---|---|---|---|---|
| C x T | 15 | 5.278 | 4.95 | < .001 |
| R x C x T | 15 | 1.271 | 1.19 | |
| S x C x T | 15 | 0.975 | 0.92 | |
| R x S x C x T | 15 | 1.143 | 1.07 | |
| Error | 540 | 1.066 | | |
| | | | | |
| I x C x T | 15 | 1.554 | 1.41 | |
| R x I x C x T | 15 | 0.763 | 0.69 | |
| S x I x C x T | 15 | 1.051 | 0.95 | |
| R x S x I x C x T | 15 | 1.254 | 1.14 | |
| Error | 540 | 1.103 | | |

Table 12

Significant Differences between mean Error Scores for Six Sentence Types

| | W? $M_1$ | A $M_2$ | Q $M_3$ | P $M_4$ | PQ $M_5$ | C $M_6$ |
|---|---|---|---|---|---|---|
| $M_1$ | | .25 | .34 | .50 | .57 | .66 |
| $M_2$ | | | | .25 | .32 | .41 |
| $M_3$ | | | | | .26 | .32 |
| $M_4$ | | | | | | |
| $M_5$ | | | | | | |

W? = Who-Question
A = Active
Q = Question
P = Passive
PQ = Passive - Question
C = Control

Only significant differences are given in this table (all $p < .01$).

## Table 13

Means and Significant Differences Between Mean Error Scores for Six Sentence Types, Tonal and Atonal Conditions

Means:

| | W? | A | Q | P | PQ | C |
|---|---|---|---|---|---|---|
| Tonal | 1.52 ($M_2$) | 1.70 ($M_4$) | 1.66 ($M_3$) | 1.96 ($M_6$) | 1.96 ($M_7$) | 2.00 ($M_{10}$) |
| Atonal | 1.43 ($M_1$) | 1.74 ($M_5$) | 1.98 ($M_8$) | 1.99 ($M_9$) | 2.13 ($M_{11}$) | 2.26 ($M_{12}$) |

Differences Between Means:

| | $M_6$ | $M_7$ | $M_8$ | $M_9$ | $M_{10}$ | $M_{11}$ | $M_{12}$ |
|---|---|---|---|---|---|---|---|
| $M_1$ | .53 | .53 | .55 | .57 | .58 | <u>.70</u> | <u>.84</u> |
| $M_2$ | | | | .48 | .48 | .61 | <u>.74</u> |
| $M_3$ | | | | | | .47 | .61 |
| $M_4$ | | | | | | | .56 |
| $M_5$ | | | | | | | .52 |

Only significant differences between means are given in this table. All values are significant at $P < .05$; underlined values are significant at $p < .01$.

W? = Who-Question
A = Active
Q = Question
P = Passive
PQ = Passive-Question
C = Control

In summary, there was no evidence from the analyses of Error Scores in digit recall supporting the hypothesis that these poor readers, however they were defined in this study, were facilitated relatively more than good readers by the presence of intonational features, or for the hypothesis that with increasing syntactic complexity poor readers were facilitated relatively more, when compared with good readers, by the presence of such features in the stimuli. Further, the analyses of Error Scores in digit recall indicated that only when good and poor reader groups were defined in terms of the more recent school-administered test of reading comprehension (i.e., the criteria for the First Subject Regrouping) was there a significant difference in the experimental task between these two groups. Since there were no apparent interactions between the Readers variable and the experimental variables (i.e., Complexity, Intonation, and Complexity x Intonation) for any of the subject groupings, the obtained significant difference between the good and poor reader groups (First Subject Regrouping) in Error Scores may possibly have been due to a difference between these groups in overall immediate memory. The original matching criteria included matching for "intelligence," as measured by a school-administered test, assuming matching on such a test would yield groups of good and poor readers equivalent in overall memory ability (as well as other skills such tests presumedly measure). Given the unexpected post hoc findings suggesting confounded validity for such tests of intelligence for these Ss, this assumption is questionable. Ideally, an independent test of immediate memory would allow equating groups of good and poor readers in terms of this skill.[21] Lacking such an independent assessment for these Ss, it was assumed that, of the three "levels" of syntactic complexity used in this study (i.e., (1) the Simple sentences, (2) the mean of the Questions and the Passives, and (3) the Passive-Questions), the Simple sentences in the tonal condition could be taken as a memory "baseline" (i.e., was the easiest to process). Difference scores were derived for each S for each of the six conditions (i.e., tonal and atonal conditions for each of three "levels" of syntactic complexity) by taking the mean Error Score for the Simple-tonal sentences and subtracting it from itself (to yield a zero difference score for each S for the "baseline" condition) and each of the other five conditions. An analysis of variance was performed on these derived difference scores, including the following variables: (1) Readers; (2) Sex; (3) Intonation; and (4) Complexity (three "levels"), with trend components.

Results of this analysis, indirectly controlling for overall differences in immediate memory, indicated no apparent difference between good and poor readers ($F = .55$, $df = 1, 36$, not significant). The mean difference score for girls ($m = .36$) was greater than the mean difference score for boys ($m = .01$, $F = 5.35$, $df = 1, 36$, $p < .05$). There was also a significant Sex x Intonation x Complexity interaction ($F = 8.10$, $df = 2, 72$, $p < .001$), with a significant linear trend component ($F = 13.28$, $df = 1, 72$, $p < 001$). Thus, it would appear that the difference between good and poor reader groups, based on fourth grade reading test, in mean Error Scores was due to a difference between these groups in overall memory ability. To the extent that the groups differed in terms of overall immediate memory ability, the experimental stimuli might have been of sufficient difficulty for these poor readers to obscure any possible greater facilitation of the presence of intonational features or syntactic simplicity, compared with the good readers.

B. Effects of Intonation and Complexity

In addition to the analyses relating to the primary purpose of this study (i.e., differences between good and poor readers), the data were also examined in terms of effects of the experimental variables (intonation; complexity) for combined good and poor readers.

The analysis, summarized in Table 11, indicated that three of the sentence types used in this study, the Passive, the Passive-Question, and the Control (i.e., a simple-active-declarative plus two modifiers) were each more difficult to process than the other three sentence types, yet not significantly different from each other. This last finding is somewhat unexpected from the point of view of a transformational grammar. Since the Passive, Passive-Question, and Control (active) differ from each other in terms of such a model of grammar, they might be expected to differ also in terms of difficulty of processing. It will be recalled that the Control sentence type was included in this study as a more adequate control for possible effects of sentence length; each of the Passives, Passive-Questions, and Control sentences contained seven words. These were the only stimulus sentences which were composed of seven words: the Who-Questions were composed of four words, the Actives were composed of five words, and the Questions were composed of six words.

An analysis was performed to examine the effects of sentence length (number of words) and Intonation for good readers versus poor readers. The variables in this analysis were (1) Readers (good versus poor, as defined by the fourth grade reading test); (2) Sex; (3) Intonation (tonal versus atonal); and, (4) Length of sentence (4-, 5-, 6-, and 7-word sentences). Ordering the sentence types used in this study in terms of sentence length results in a partial confounding of sentence length with sentence complexity; that is, the 4-word sentence (Who-Question) is assumed to be grammatically more complex than the 5-word sentence (Active), while the 5-word sentence is assumed to be grammatically less complex that the 6-word sentence (Question). Of the three sentence types composed of seven words, found not to differ from each other in terms of mean Error Scores, the Control was used in this analysis, assumed to be grammatically the simplest of the three possible 7-word sentence types. Since sentence length can be assumed to represent an interval scale, trend components were performed on the Length variable. A summary of this analysis is presented in Table 14.

The intonation main effect was found to be significant ($F = 7.46$, $df = 1,36$, $p < .01$), with the mean Error Score of the tonal stimuli ($m = 1.72$) less than the mean Error Score of the atonal stimuli ($m = 1.85$). Length was significant ($F = 20.67$, $df =$

Table 14

Analysis of Variance of Mean Error Scores of Good versus Poor Readers (4th-grade Reading Test) for Tonal versus Atonal Conditions and Four Sentence Lengths (4-words, 5-words, 6-words, 7-words), including Trend Components on Length

| Source of Variation | df | MS | F | p |
|---|---|---|---|---|
| Total | 319 | 0.806 | | |
| Between | 39 | 4.365 | | |
| Readers (R) | 1 | 23.249 | 5.94 | < .05 |
| Sex (S) | 1 | 3.252 | 0.83 | |
| R x S | 1 | 2.765 | 0.71 | |
| Error | 36 | 3.916 | | |
| Within | 280 | 0.311 | | |
| Intonation (I) | 1 | 1.413 | 7.46 | < .01 |
| R x I | 1 | 0.087 | 0.46 | |
| S x I | 1 | 0.329 | 1.74 | |
| R x S x I | 1 | 0.003 | 0.02 | |
| Error | 36 | 0.189 | | |
| Length of Sentence (L) | 3 | 5.944 | 20.67 | < .001 |
| Linear | 1 | 17.171 | 59.71 | < .001 |
| Quadratic | 1 | 0.086 | 0.30 | |
| Residual | 1 | 0.575 | 2.00 | |
| R x L | 3 | 0.087 | 0.30 | |
| Linear | 1 | 0.102 | 0.35 | |
| Quadratic | 1 | 0.005 | 0.02 | |
| Residual | 1 | 0.154 | 0.54 | |
| S x L | 3 | 0.238 | 0.83 | |
| Linear | 1 | 0.001 | 0.00 | |
| Quadratic | 1 | 0.549 | 1.91 | |
| Residual | 1 | 0.164 | 0.57 | |
| R x S x L | 3 | 0.142 | 0.50 | |
| Linear | 1 | 0.383 | 1.33 | |
| Quadratic | 1 | 0.016 | 0.06 | |
| Residual | 1 | 0.028 | 0.10 | |
| Error | 108 | 0.288 | | |
| I x L | 3 | 0.737 | 3.64 | < .05 |
| Linear | 1 | 1.806 | 8.91 | < .01 |
| Quadratic | 1 | 0.188 | 0.93 | |
| Residual | 1 | 0.218 | 1.08 | |

Table 14 - continued -

| Source of Variation | df | MS | F | p |
|---|---|---|---|---|
| R x I x L | 3 | 0.049 | 0.24 | |
| Linear | 1 | 0.118 | 0.58 | |
| Quadratic | 1 | 0.005 | 0.02 | |
| Residual | 1 | 0.023 | 0.12 | |
| S x I x L | 3 | 0.981 | 4.84 | < .01 |
| Linear | 1 | 0.145 | 0.72 | |
| Quadratic | 1 | 0.297 | 1.47 | |
| Residual | 1 | 2.499 | 12.33 | < .001 |
| R x S x I x L | 3 | 0.274 | 1.35 | |
| Linear | 1 | 0.165 | 0.81 | |
| Quadratic | 1 | 0.044 | 0.22 | |
| Residual | 1 | 0.613 | 3.02 | |
| Error | 108 | 0.203 | | |

3, 108, $p < .001$), with a significant linear trend component ($F = 59.71$, $df = 1, 108$, $p < .001$). The means were ordered linearly in terms of sentence length: four-word sentences ($m = 1.47$, five-word sentences ($m = 1.72$), six-word sentences ($m = 1.82$), and seven-word sentences ($m = 2.13$).[22]

The interaction of Intonation and Length was found to be significant ($F = 3.64$, $df = 3, 108$, $p < .05$), with a significant linear trend component ($F = 8.91$, $df = 1, 108$, $p < .01$). The eight mean Error Scores composing this interaction (i.e., tonal versus atonal stimuli for each of four sentence lengths) were ordered and compared systematically with each other by a Duncan Multiple-Range Test. The pair-wise comparisons which reached the arbitrary level of significance are presented in Table 15. For the tonal stimuli four- five- and six-word sentences are not different from each other, but all show less error than for the seven-word sentence. For the atonal stimuli, four-word sentences have less errors than five-word sentences, which are in turn less than six-word sentences, and less in turn than seven-word sentences. Thus, for atonal stimuli, all four sentence lengths were significantly different from each other, and the obtained means were ordered linearly with increasing sentence length. For comparisons of tonal versus atonal stimuli within a given sentence length, for the two shorter sentences (four-and five-word sentences) tonal and atonal conditions did not differ significantly. However, for the two longer sentences (six-and seven- word sentences) the atonal conditions yielded a significantly higher mean Error Score than the respective tonal conditions.

The interaction of Readers, Intonation, and Length was not found to be significant. Thus, while the presence of intonational features seemed to facilitate processing and seemed to be relatively more important for the longer sentences (six- and seven-word sentences, as compared with four- and five-word sentences), there was no apparent difference in this effect for poor readers, as compared with good readers.[23]

In summary, there was little evidence from these analyses of Error Scores in digit recall to support the hypothesis that the presence of intonational features is relatively more important for processing syntactically complex sentences, as compared with syntactically simple sentences. However, there was evidence suggesting that the presence of intonational features is relatively more important for efficient processing of longer sentences, as compared with shorter sentences, when length of the sentence is considered independently of syntactic complexity. The most straightforward test of a "pure" complexity hypothesis for the stimuli used in this study would examine only those sentence types which can be reasonably ordered in terms of their transformational histories (i.e., the "number" of transformations required to derive one sentence type from another). Starting with the Simple active sentence, a simple transformational grammar would characterize both the Passive and the Question as being derived by roughly the same "number" of transformations, although the transformations themselves are different. Since it is difficult to predict from such a grammar differences in

Table 15

Means and Significant Differences Between Mean Error Scores For Four Sentence Lengths, Tonal and Atonal Conditions

Means:

| | Sentence Length (number of words) | | | |
|---|---|---|---|---|
| | 4 | 5 | 6 | 7 |
| Tonal | 1.519 ($M_2$) | 1.700 ($M_4$) | 1.656 ($M_3$) | 2.00 ($M_7$) |
| Atonal | 1.425 ($M_1$) | 1.744 ($M_5$) | 1.975 ($M_6$) | 2.26 ($M_8$) |

Differences Between Means:

| | $M_3$ | $M_4$ | $M_5$ | $M_6$ | $M_7$ | $M_8$ |
|---|---|---|---|---|---|---|
| $M_1$ | .23 | .28 | .32 | .55 | .58 | .84 |
| $M_2$ | | | .23 | .46 | .48 | .74 |
| $M_3$ | | | | .32 | .34 | .61 |
| $M_4$ | | | | .28 | .30 | .56 |
| $M_5$ | | | | .23 | .26 | .52 |
| $M_6$ | | | | | | .29 |
| $M_7$ | | | | | | .26 |

Only significant differences between means are given in this table. All values are significant at $p < .05$; underlined values are significant at $p < .01$.

in complexity between different transformations, these two sentence types may be characterized as roughly equivalent in terms of complexity. Similarly, the Passive-Question may be characterized as more complex than the Simple, the Question, and the Passive, since it requires the application of those transformations necessary for both the Passive and the Question. An analysis was performed on the mean Error Scores, utilizing the stimuli in this manner. The variables in this analysis were (1) Readers (good versus poor, as defined by the fourth grade tests; (2) Sex; (3) Intonation (tonal versus atonal stimuli); and (4) Complexity (with three levels). Simple active sentences were used for the first level of complexity; the mean of the Questions and the Passives were used for the second level of complexity; and the Passive-Questions were used as the third level of the complexity variable. Since these three levels of complexity can be assumed to constitute an interval scale (i.e., the "distances" between levels one and two, and levels two and three are equivalent), trend components were performed on Complexity. A summary of this analysis is presented in Table 16.

The mean Error Score of the tonal stimuli (m = 1.82) was less than the mean Error Score of the atonal stimuli ($m = 1.96$, $F = 4.92$, $df = 1, 39$, $p < .05$). The Complexity main effect was significant ($F = 8.00$, $df = 2, 72$, $p < .001$), with a significant linear trend component ($F = 15.98$, $df = 1, 72$, $p < .001$), indicating that the three levels of Complexity were each significantly different from the others and ordered linearly (level 1, $m = 1.72$; level 2, $m = 1.89$; level 3, $m = 2.04$). However, the Intonation x Complexity interaction did not reach the arbitrary level of significance ($F = .57$, $df = 2, 72$, not significant).

Finally, as in the previous study, the data were analyzed in terms of possible systematic effects within the experimental session due to learning or practice (a systematic improvement within the experimental session) versus fatigue (a systematic decrement in performance over trials). This analysis of variance revealed no such systematic effects (F, for Blocks of twelve trials = 1.67, df = 3.96, not significant); further, the two orders of presentation of the stimuli were apparently equivalent ($F = .54$, $df = 1, 32$, not significant).

II. Errors in Sentence Recall

In general, errors which occurred in the recall of the first part of the stimulus trials (i.e., the sentences) were examined in two ways. First, those sentences recalled incorrectly in any way were examined independently of the number of errors or the kinds of errors within a given non-verbatim repetition of the sentence. Second, error categories were derived for these non-verbatim sentences and analyzed as described below.

A. Non-Verbatim Sentences.

There were a total of 164 sentences which were recalled less than perfectly; of a total of 1920 stimulus trials (40 Ss, 48 trials per S), this represents 8.5% of the stimulus trials. Inspection of the distribution of Non-Verbatim sentences

Table 16

Analysis of Variance of Mean Error Scores of Good vs. Poor Readers (4th-Grade Reading Test) for Tonal versus Atonal Conditions and Three Levels of Syntactic Complexity (Simple Sentences, Mean of Passives and Questions, Passive-Questions), Including Trend Components on Complexity

| Source of Variation | df | MS | F | p |
|---|---|---|---|---|
| Total | 239 | 0.801 | | |
| Between | 39 | 3.640 | | |
| Readers (R) | 1 | 17.402 | 5.19 | < .05 |
| Sex (S) | 1 | 1.445 | 0.43 | |
| R x S | 1 | 2.326 | 0.69 | |
| Error | 36 | 3.355 | | |
| Within | 200 | 0.248 | | |
| Intonation (I) | 1 | 0.985 | 4.92 | < .05 |
| R x I | 1 | 0.212 | 1.06 | |
| S x I | 1 | 0.242 | 1.21 | |
| R x S x I | 1 | 0.258 | 1.29 | |
| Error | 36 | 0.200 | | |
| Complexity (C) | 2 | 2.035 | 8.00 | < .001 |
| Linear | 1 | 4.064 | 15.98 | < .001 |
| Residual | 1 | 0.006 | 0.03 | |
| R x C | 2 | 0.104 | 0.41 | |
| Linear | 1 | 0.127 | 0.50 | |
| Residual | 1 | 0.081 | 0.32 | |
| S x C | 2 | 0.010 | 0.04 | |
| Linear | 1 | 0.014 | 0.06 | |
| Residual | 1 | 0.006 | 0.03 | |
| R x S x C | 2 | 0.260 | 1.02 | |
| Linear | 1 | 0.002 | 0.01 | |
| Residual | 1 | 0.517 | 2.03 | |
| Error | 72 | 0.254 | | |
| I x C | 2 | 0.107 | 0.57 | |
| Linear | 1 | 0.156 | 0.83 | |
| Residual | 1 | 0.057 | 0.31 | |
| R x I x C | 2 | 0.079 | 0.42 | |
| Linear | 1 | 0.100 | 0.53 | |
| Residual | 1 | 0.057 | 0.31 | |

Table 16 - continued -

| Source of Variation | df | MS | F | p |
|---|---|---|---|---|
| S x I x C | 2 | 1.525 | 8.10 | $< .01$ |
| Linear | 1 | 2.500 | 13.28 | $< .001$ |
| Residual | 1 | 0.550 | 2.92 | $(< .10)$ |
| R x S x I x C | 2 | 0.287 | 1.52 | |
| Linear | 1 | 0.056 | 0.30 | |
| Residual | 1 | 0.517 | 2.75 | |
| Error | 72 | 0.188 | | |

indicated they were not distributed randomly across subjects or across types of stimuli. Several non-parametric analyses were conducted to assess differences between subjects (Readers; Sex) and within experimental conditions (Intonation; Complexity).

To assess differences between good and poor readers (fourth grade test) and between male and female Ss in inaccuracy of sentence recall, a 2 x 2 randomization test was performed on the non-verbatim sentence trials for each S. This analysis included independent tests on the contrasts for Readers (good versus poor), for Sex, and for the Readers x Sex interaction.[24] The number of non-verbatim sentence trials of the good readers (63) was less than the number of non-verbatim sentence trials of the poor readers (101) at the $p = .06$ confidence level. The number of non-verbatim sentence trials by male Ss (64) differed from the number of non-verbatim sentence trials by female Ss (100) at the $p = .11$ confidence level. The interaction between Readers and Sex reached only the $p = .25$ confidence level.

Inspection of the distribution of non-verbatim sentences within the experimental conditions (Intonation; Complexity) suggested that more sentences were recalled non-verbatim for the Control (longer) sentences than for any other sentence type, and that fewer sentences were recalled non-verbatim for the Who-Question (shorter) sentences than for any other sentence type. These two sentence types were compared with the sentence types most similar to them in syntactic form (i.e., the Simple compared with the Control; the Question compared with the Who-Question). A score was derived for each S for each of these contrasts (non-verbatim sentences in Simple sentences versus in Control sentences for each S; in Questions versus in Who-Questions for each S); these scores were then ranked and analyzed by Wilcoxon Matched-Pairs Signed-Ranks Tests (Siegel, 1956). A greater number of non-verbatim sentences occurred for Control sentences than for Simple sentences ($T = 59.5$, $N = 33$, $p < .001$, two-tailed test, with z computed for N larger than 25), and a fewer number of such occurrences were evident for Who-Questions than for Questions ($T = 15$, $N = 14$, $p < .02$, two-tailed test). A third comparison indicated that the occurrence of non-verbatim sentences did not significantly differ for the tonal sentences, as compared with the atonal sentences ($T = 146$, $N = 24$, not significant).

Given that a sentence is recalled non-verbatim, it can be asked whether this error (or errors) hinders or facilitates processing the sentence. There seems to be no a priori rationale for expecting either facilitation (i.e., "simplification" of the stimulus sentence in recall) or non-facilitation (e.g., due to interference by S's possible recognition of the error) to occur in all cases of non-verbatim recall. The experimental task employed in this study (i.e., sentence plus digit list on each trial, thus allowing two recall measures) permitted a post hoc statistical test relating to this question. For each S the Error Score in digit recall for each non-verbatim sentence was compared with the Error Scores (digit recall) for the remaining sentences of the same stimulus type which were recalled verbatim. A Sign Test (Siegel, 1956) suggested that when a sentence was recalled less than perfectly, the Error **Score was greater than verbatim recall trials for sentences of the same stimulus** type ($F = 3.54$, $p < .05$, two-tailed test). Whether this apparent greater difficulty was the case for all types of sentence errors was also examined, as discussed below.

B. Sentence Errors

Each of the sentence errors in the non-verbatim sentences was classified according to the following derived error categories: (1) Omission (e.g., "The girl walked the dog after school" recalled as "The girl walked the dog"); (2) Substitution (e.g., "The woman ate the food" recalled as "The lady ate the food"); (3) Transformation (e.g., "Did the man drive the car?" recalled as "The car was driven by the man"); (4) Morphophonemic Irregularity (e.g., "Who drove the car?" recalled as "Who druve the car?"); and (5) Miscellaneous, including contaminated responses (e.g., "Was the car driven by the man?" recalled as "Was the car dritten by the man?"), verb tense changes (e.g., "The woman ate the food" recalled as "The woman was eating the food"), and subject-object reversals (e.g., "The dog was walked by the girl" recalled as "The girl was walked by the dog"). Sentence errors were scored by two judges, with all differences in scoring resolved between the two.

Since 27 of the 164 non-verbatim sentences contained more than one Sentence Error, the number of Sentence Errors (202) exceeded the number of non-verbatim sentences. Of the 202 Sentence Errors, 35.1% were Omissions; 26.7 % were Substitutions; 13.9% were Transformations; 19.8 were Morphophonemic Irregularities; and 4.5% were Miscellaneous errors.

Appropriate non-parametric analyses on Summed Sentence Errors, paralleling those performed on non-verbatim sentences, yielded similar results (i.e., good readers $<$ poor readers; boys $<$ girls; tonal stimuli = atonal stimuli; control [longer] sentences simple sentences; who-questions [shorter] $<$ questions).

Of the categories of Sentence Errors, it appeared by inspection that the frequencies of the various kinds of Sentence Errors were distributed randomly acorss the six sentence types, with the following exception. Of the 71 Omission errors, 61 (or 85.9%) occurred with the Control sentences; this represents 73.5% of all errors which occurred with this sentence type. This is not surprising, since an Omission error occurring with any of the other sentence types violates the syntactic integrity of the sentence (e.g., "The girl walked the dog" recalled as "The walked the dog"), whereas one of the two modifiers (or both) in the Control sentences could be omitted without violating the syntactic integrity of the sentence in recall.

Sign Tests (as described above) were performed comparing the Error Scores in digit recall of the trials with Sentence Errors with the trials of **verbatim recall of the same stimulus type, for each of the categories of** Sentence Errors, and for trials with only one versus multiple Sentence Errors. For those trials with single Omission errors, there was no apparent difference in the Error Scores for digit recall, compared with the verbatim recall trials of the same stimulus class ($z = .29$, not significant). Also, for trials with single Substitutions, this Sign Test failed to reach the arbitrary level required for significance ($z = 1.20$, not significant). However, a Sign Test for trials with single non-Omission errors (i.e., the four remaining categories of Sentence Errors, including Substitutions) was statistically reliable ($z = 3.91$, $p < .05$), with the occurrence of an error resulting in a greater Error Score in digit recall than the corresponding verbatim recall trials. Sign tests were significant on the categories of Transformations ($z = 2.75$, $p < .05$) and Morphophonemic Irregularities ($z = 2.30$, $p < .05$), and both tests indicated the same direction of the relationship; the occurrence of these types of errors yielded higher Error Scores in digit recall than corresponding verbatim recall trials. In summary, the results of these analyses indicate that although, overall, the failure to recall a sentence verbatim seems to "interfere" with sentence processing (i.e., results in a higher Error Score in digit recall), this "interference" did not occur uniformly for all categories of Sentence Errors. Specifically, this effect was found only for non-omission errors. Omission errors were particularly interesting, since, as discussed above, they occurred primarily in the Control sentences. For those Omissions which occurred in the Control sentences, all involved omission of one or both of the modifiers. Thus, it seems that Ss in this task recalled what was necessary to maintain the syntactic integrity of the stimulus sentence; when "additional information" was added in the form of modifiers (i.e., the Control sentences), Ss often simply omitted these modifiers in recall; further, such omissions seemed to facilitate sentence processing, as compared with other possible kinds of errors in recall which did not appear to facilitate processing (memory).

III. Total Errors

A third measure of errors in recall was derived by taking the sum, for each S on each trial, of the Error Score in digit recall and the number of Sentence Errors. Since on each trial, the Ss were presented a sentence and a digit list and asked to recall them both, and since Sentence Errors were not distributed randomly across groups of subjects (i.e., good versus poor readers; males versus females), there may have been a confounding of either of the two parts of each trial (i.e., Error Scores in digit recall and/or Sentence Errors when considered separately). Indeed, as shown above, certain Sentence Errors seemed to result in lower Error Scores, and these various Sentence Errors did not appear to be randomly distributed across sentence types. Further, if "short term memory" can be conceptualized as a limited and relatively fixed "capacity," when a S is required to "divide" that "capacity" between two distinct kinds of stimuli (i.e., sentence and digits), he may "focus" differentially on one kind of stimulus or the other as a function of either the stimulus type (i.e., the experimental conditions) or as a function of the individual (i.e., readers, sex). In a sense, the derived Total Error Score (i.e., Error Score in digit recall plus Sentence Errors) is a measure of the recall of the total trial, given that not all sentences were recalled verbatim.

An analysis of variance was performed on the mean Total Error Scores, including the following variables: (1) Readers (good versus poor, as defined by the fourth grade testing; (2) Sex; (3) Intonation (tonal versus atonal stimuli); and, (4) Complexity (six sentence types). A summary of this analysis is presented in Table 17.

The mean Total Error Score for good readers ($m = 1.66$) was significantly less than the mean Total Error Score of the poor readers ($m = 2.27$) ($F = 6.26$, $df = 1, 36$, $p < .05$). There was no apparent difference between the mean Total Error Scores of male versus female subjects ($F = 1.05$, $df = 1, 36$, not significant).

The mean Total Error Score for tonal stimuli was significantly lower than the mean total Error Score for atonal stimuli ($m = 2.02$) ($F = 6.59$, $df = 1, 36$, $p < .05$). The Complexity main effect was significant ($F = 19.32$, $df = 5, 180$, $p < .001$); a Duncan Multiple-Range Test revealed the following pattern of means (at $p < .05$ or better): Who-Question < Simple = Question < Passive < Passive-Question < Control. The interaction of Intonation x Complexity was also significant ($F = 2.22$, $df = 5, 180$, $p < .05$); a Duncan Multiple-Range Test performed on the 12 means composing this interaction (i.e., tonal and atonal versions for 6 sentence types) yielded a pattern of significant contrasts similar to that obtained for the analysis of Error Scores (Table 15). There was also a significant Readers x Sex x Intonation x Complexity interaction ($F = 3.99$, $df = 5, 180$, $p < .01$).

Table 17

**Analysis of Variance of Mean Total Error Scores of Good versus Poor Readers (4th-Grade Reading Test) for Tonal versus Atonal Conditions and Six Sentence Types**

| Source of Variation | df | MS | F | p |
|---|---|---|---|---|
| Total | 479 | 0.999 | | |
| Between | 39 | 7.891 | | |
| Readers (R) | 1 | 44.105 | 6.26 | < ,05 |
| Sex (S) | 1 | 7.375 | 1.05 | |
| R x S | 1 | 2.411 | 0.34 | |
| Error | 36 | 7.051 | | |
| Within | 440 | 0.388 | | |
| Intonation (I) | 1 | 1.575 | 6.59 | < ,05 |
| R x I | 1 | 0.411 | 1.72 | |
| S x I | 1 | 0.023 | 0.10 | |
| R x S x I | 1 | 0.003 | 0.00 | |
| Error | 36 | 0.239 | | |
| Complexity (C) | 5 | 7.154 | 19.32 | < ,001 |
| R x C | 5 | 0.034 | 0.09 | |
| S x C | 5 | 0.237 | 0.64 | |
| R x S x C | 5 | 0.264 | 0.71 | |
| Error | 180 | 0.370 | | |
| I x C | 5 | 0.556 | 2.22 | < ,05 |
| R x I x C | 5 | 0.142 | 0.57 | |
| S x I x C | 5 | 0.998 | 3.99 | < .01 |
| R x S x I x C | 5 | 0.294 | 1.17 | |
| Error | 180 | 0.250 | | |

## Discussion

### A. Effects of syntactic structure and melodic features on sentence processing.

Within the framework of the present investigation, it was posited that syntactic organization is a patterning of a linguistic input which facilitates performance. Results of the first study, consistent with results reported in other studies using similar kinds of materials (e.g., Epstein, 1969; Weener, 1971), indicated that random strings of words were more difficult to process than the same words ordered syntactically, both when the component words were used unconventionally (Anomalous Sentences) and conventionally (Regular Sentences). Further, results of the first study also indicated that the presence of melodic features can, in part, "compensate" for the added difficulty introduced when the sentence is syntactically more complex. Results of the second study indicated that the presence of such features can also "compensate" for the added difficulty of longer sentences, as compared with shorter sentences.

From the findings that melodic features are facilitative, it would be expected that any model of language performance which attempts to predict performance difficulty by using only one set of linguistic features, such as syntactic complexity or sentence length, would be relatively accurate only under the special conditions which other possible organizational features, such as the melodic features of speech, are relatively unavailable to the subject. For example, a "pure" syntactic model should be a more accurate performance model for the atonal stimuli in these studies, or with reading material, where not all organizational features are graphically explicit.

A model which has attempted to account for linguistic performance in terms of syntactic organization can be stated in two forms. A "general" syntactic model of performance difficulty would posit that any sentence which is syntactically more complex is more difficult to process than those which are less complex. A more explicit form of this kind of model would posit an equal-interval scale of difficulty, with difficulty determined by the number of transformations required to arrive at a particular syntactic form from the simple form. Said in another way, the more explicit form would posit that the increase in difficulty resulting, for example, from transforming the simple active sentence to a passive form is equal to the increase in difficulty resulting from transforming a question, for example, to a passive-question. If the Sentence Types in the present studies are compared in the first way, five specific predictions derive from the general form of a syntactic model of performance, that is, Simple $<$ Question; Simple $<$ Passive; Simple $<$ Passive-Question; Passive $<$ Passive-Question; Question $<$ Passive-Question. The more explicit form of the model would result in one additional prediction, that is, Passive minus Simple equals Passive-Question minus Question. Results from the first study indicated that for the atonal stimuli three of the five predictions derived from the general form of such a model were confirmed, as was the prediction of the more explicit form of the model. In contrast, for the tonal stimuli, only two of the predictions of the general form of the model was confirmed, but there was no evidence to support the prediction of the model in its more explicit form. Thus, it would seem, at least from the results of the first study, that a syntactic model of performance is relatively more accurate in predicting results under the atonal conditions, where the subject has to rely more on the syntactic organization of the stimuli, as compared with the tonal conditions, where other organizational cues are available.

Two questions posed by these findings are (1) the less than complete prediction of results of the syntactic model under the atonal conditions, and (2) the two confirmed predictions of this model in the tonal conditions--in the latter case, intonational features were expected to "compensate" for the added complexity. Syntactic complexity, which in principle may be scalable in terms of performance difficulty, may for particular testing instances not show this scaling. For example, lack of experience with all of the syntactic forms, as is the case with beginning speakers, may "dissolve" the scale since all forms might be of such great difficulty that any differences between them are so small as to be indistinguishable. Concommitantly, for individuals with a great deal of familiarity with all the forms of the scale, experienced language users for example, differences between items of the scale may not be manifest, since all forms may be extremely "easy" as a function of familiarity. It would not be expected that "removing" intonational features would make the input significantly more difficult than it already is for inexperienced speakers, nor would highly experienced speakers find the loss of this additional information, for them redundant, any hindrance. Early during acquisition of language, the added information of intonational features may be critical for comprehension and for the learning of new grammatical forms. With experience with particular forms, and mastery of them, intonational features may, in fact, become "redundant" rather than "compensatory," and, therefore, not as critical for comprehension. In a sense, then, having found a scaled series for a given level of language usage in our population, the compensation for added difficulty of some part of that scale by the tonal components is consistent with the original hypotheses. From the above argument, it should be possible to construct scales of syntactic complexity, which in terms of a performance measure are "too easy" for one population, "too difficult" for another population, but yet "scale out" for a third intermediate group. It is only when such scaling is possible, as in the intermediate group, that we would expect to be able to test for the posited compensatory effects of intonational features. Thus, for example, in the first study the finding that the passive-question was more difficult than the simple sentence, even in the tonal conditions, may have been due to the unfamiliarity of the passive-question for this population. Concomitantly, the finding that the question and simple sentence were equal, even in the atonal conditions, may have been due to the relatively great familiarity with these two forms for this population. It follows that presenting this same set of stimuli to two other populations, one greatly more experienced with the forms, and the other much less experienced, should result in failure to confirm both the hypothesis of greater difficulty for greater syntactic complexity, and the hypothesis of the compensatory effects of melodic features.

The discussion thus far should not be taken to imply that the only function of the intonational contour is to make the syntactic organization more discriminable. We hold, as do others, that intonational features serve other functions as well. It would be possible, for example, to construct sentences where particular intonational features are independent of the syntact organization, by including pauses within phrases or stresses which indicate emphasis or special meaning. In such cases, where intonational features "carry" meaning independent of the syntactic form of the utterance, it follows that "removal" of such features should result in difficulty, or even failure to comprehend the specific meaning of the message, even were the syntax "known."

Further, there are other kinds of factors in addition to the intonational features and syntax which may interact in a complex way and affect performance. Such possible variables might be (1) the frequency of occurrence of component words of the sentence, (2) degree of abstractness versus concreteness of the stimuli, that is, their "imagability" (e.g., Paivio, 1971), (3) degree of conventionality of the combinations of words of the sentence, and (4) the length of the sentence. In the present studies, attempts were made to control the first two of these variables, that is, the frequency of occurrence of words and their "imagability." It seems obvious, perhaps even trivial to note that a syntactically simple sentence with low frequency words might well be more difficult to comprehend than a syntactically complex sentence with high frequency words.

As for the third variable, it was shown in the first study that syntactically "correct" sentences with unconventional word combinations, that is, the Anomalous Sentences, were more difficult to process than similar sentences with conventional word combinations, that is, the Regular Sentences. In this study, there was no evidence with Anomalous Sentences to support any of the predictions derived from a syntactic model of performance, and correspondingly, for these Anomalous Sentences, "adding" intonational features did not significantly reduce performance difficulty.

The fourth variable which was found to affect performance in these studies is the length of the sentence. The longer sentences were more difficult to process than the shorter ones. This finding is consistent with findings of other investigators such as Martin and Roberts (1967);Orenstein and Schumsky (1970). It is not argued that length alone could account for all of the results, and a direct test of this view was not possible in these studies in that length was partially confounded with syntactic complexity. It does not seem evident how sentences of different lengths can be constructed which control directly for syntactic complexity, since implicitly, at least, those holding a syntactic model could argue that the addition of any word changes the "deep" structure. However, to the extent that length has been systematically varied in these studies, it accounts for more of the difficulty than syntactic complexity. This effect is more notable in the atonal conditions than in the tonal conditions. Results from the second study indicated that when melodic features were "absent," the difficulty of each of the four different sentence lengths was significantly different from the other three and all were ordered linearly. In contrast, when melodic features were "present," three of the four sentence lengths were not significantly different from each other; only the longest sentence was significantly more difficult than the others. Further, the tonal versus atonal comparisons for each sentence length indicated that for the two shorter lengths the presence of melodic features did not significantly reduce difficulty, while for the two longer sentences, presence of such melodic features did reduce difficulty.

The finding that the Control sentence in the second study was more difficult, even in the tonal conditions, than three other sentence types of shorter lengths,may have been due to the additional "semantic" content of the Control sentences, not represented in the other three sentence types. It has been demonstrated by Perfitti (1969a) that adding such "content" words as adjectives, or, in his terms increasing the "lexical density" of a sentence, while holding the complexity constant (for Perfitti, the syntactic model investigated was that of Yngve, 1960) increases performance difficulty. It must be noted again that for the stimuli used in these studies, the melodic features were related to syntactic organization and only incidentally to sentence length. It would be possible to construct items which would directly test the compensatory effects of melodic features for variations in length, for example, in compound noun phrases (e.g., "the large green house") with and without pauses. Such an investigation would have to take into account Perfetti's findings--it may well be that for a given level of language usage, the addition of pauses, or even decreasing presentation rate, may, within limits, "compensate" for an increase in difficulty of greater "lexical density."

In sum, the results of the two studies suggest that syntactic complexity, sentence length, and conventionality of usage of word combinations affect difficulty of performance, but that no one factor, considered in isolation, constitutes an adequate basis for predicting difficulty noted. More important, it is suggested that melodic features can, within limits, "compensate" for the increased difficulty introduced by a variation in one or more of these other variables. It is also suggested that there are many language variables which appear to interact in a complex way, and that a "pure" syntactic model may be relatively accurate only under restricred conditions--"all other things being equal"--and when other possible organizational cues, such as melodic features, are unavailable in the input.

One further finding from the second study can be discussed here, although not related directly to the hypotheses regarding melodic features. At least one recent study of memory for sentences (Mehler, 1963) with adult subjects and a prompted recall technique reported a "regression to the kernel" in recalling complex sentences. That is, in that study simple active sentences were the easiest to recall of the sentence types used, and when an error in recall of a syntactically complex sentence occurred, the recalled sentence was syntactically simpler than the stimulus sentence. Hayhurst (1967) reported similar findings that children generally made errors of simplification, that is, the "kernel," or simple active form was recalled for a stimulus sentence in the passive form. Such evidence has been taken to support Miller's earlier (1963) "schema-plus-correction," or "kernel-plus-code" hypothesis, namely, that complex sentences are "understood" as their underlying "kernels" plus independently processed (recalled) "transformation tags" specifying the syntactic form of the stimulus. While questions can be raised about this hypothesis on conceptual grounds and from evidence in this and other studies, the focus here will be on the evidence derived in the second study which bears most directly on this hypothesis.

In examining the Sentence Errors in the second study, it was possible to test the "kernel-plus-code" hypothesis in two ways, although it must be recognized that these were <u>post hoc</u> tests. First, if a Transformation error results in a "simplification" for the subject, then the errors in digit recall for such simplifications should be less than the errors for the corresponding verbatim recall of similar sentences. Results indicated a statistically reliable difference, but in a direction opposite that predicted by the "kernel-plus-code" hypothesis. When a Transformation error occurred, there was a <u>greater</u> Error Score in the digit recall than for the equivalent verbatim recall instances. <u>Second</u>, the 28 instances of transformation errors were examined in terms of their complexity. A "regression toward the kernel" hypothesis would predict errors in which the recalled sentences are transformed to simpler sentences. Of the 28 instances of transformation errors, errors in sentences, only 10 were "toward" the kernel, while 18 were "away from" the kernel, that is, the recalled sentence was <u>more</u> complex than the stimulus. However, such a comparison is not an adequate test of the "regression toward the kernel" hypothesis since it may be biased by the inclusion of the two extremes of high and low complexity of the sentence types used in this study. If a Transformation error occurred for a Simple active sentence (low complexity, or, "close to" the kernel), it was impossible in principle for the recalled sentence to be less complex than its corresponding stimulus. Similarly, if a Transformation error occurred for the most complex sentence type used in this study, the Passive-Question, the recalled sentence is more likely to be less complex than more complex. Thus, a more appropriate test of the "regression to the kernel" hypothesis could utilize only those sentence types of intermediate complexity (Passive, Question, and Who-Question). Of these sentence types, six of the sentence errors were "toward the kernel" and six were "away from the kernel." Thus, there was no significant evidence from this study to support a derivation from a "pure" syntactic model of performance, positing such a "regression to the kernel."

B. Good Readers versus Poor Readers.

**Results from the second study indicated that poor readers, as defined by** the most recent school-administered reading test, performed worse than matched good readers for all conditions of the experimental task. There was no significant evidence from this study that poor readers perform relatively more poorly than good readers when melodic features are "absent," or that this difference increases as syntactic complexity increases. There was a suggestion that these poor readers, compared with the good readers, show a greater performance decrement when melodic features are "absent" as the sentences become longer, but this finding was reliable only at the $p < .10$ confidence level.

The absence of evidence suggesting the hypotheses that poor readers would find it more difficult to process sentences which are more complex, particularly with tonal features absent, presents some difficulty. Cromer (1970) found that those poor readers, assumed to read in a word-by-word fashion, have difficulty, compared with good readers, imposing organization on visual linguistic material, but improve in comprehension when the appropriate organization is made visually explicit. Further, Oakan, Wiener, and Cromer (1971) showed that some poor readers, selected in much the same way as done in the study reported here, have what appears to be an analogous difficulty imposing organization on "disrupted" auditory linguistic material, but are indistinguishable from matched good readers when the organization is made explicit, that is, for auditory material which is not "disrupted." The discrepancy in the findings of this study with the previous studies can be accounted for in one or more ways. First, it may be the case, as suggested by Cromer (1970), that some poor readers can be said to have more of a problem of "organization" of material, while other poor readers can be said to have more of a problem in "identification." If so, then a more careful selection of poor readers, that is, to select specifically only word-by-word readers with no problem in word identification, might be required before the effects can be shown. A more stringent selection criterion was not employed here since the subjects in the Oakan, Wiener, and Cromer study showed that poor readers, defined by school tests, did show difficulty in imposing organization on "disrupted" auditory input. However, the concept of "organization," as used by Oakan, Wiener, and Cromer may not be strictly comparable to the use of that concept here. Oakan, Wiener, and Cromer used "disrupted" auditory input which included more than a disruption of the melodic contour, that is, words were also misidentified, substituted, repeated, and so on. It may be that Oakan, Wiener, and Cromer's good readers were better able than their poor readers to infer from the words which were correctly identified in the condition of "disrupted" auditory input to determine the correct sentence. In the second study here, no stimuli were "disrupted" in the sense that Oakan, Wiener, and Cromer constructed their "disrupted" stimuli; rather, in some conditions in the present study, intonational features were "removed" from the stimuli. In contrast to Oakan, Wiener, and Cromer's stimuli, it may well be that the stimuli used in the present study were not of sufficient difficulty that removing intonational features discriminated these particular good and poor readers. If this view is tenable, then stimulus material constructed to be of greater difficulty, such as sentences which are longer than those employed here, could well yield the hypothesized interactions between good versus poor readers and the presence versus absence of intonational features

There was also evidence that good and poor readers differed in overall memory, and it is not clear whether this difference could also account for the lack of evidence supporting the hypothesized interactions between the reader groups, intonation, and syntactic complexity. Given the results of this study, it would seem reasonable to reject the original hypotheses of the second study only after (1) the good and poor readers are equated in some other non-verbal memory task, and (2) after such equating, showing that making the material more "difficult" in some way other than removing organizational cues from the input, such as using uncommon words, would not discriminate between the good and poor readers. If these two controls were instituted, and if there were evidence confirming the original hypotheses for some poor readers, it would then seem reasonable to characterize the difficulty of those poor readers as one of "language organization"--that is, a failure to impose organization when not explicitly available.

The lack of apparent confirmation for the hypotheses regarding differences between good and poor readers in the second study, and the finding that overall differences between these groups were apparent only when "good" and "poor" readers were defined by the most recent school-administered test of reading, and not when defined by the original selection criteria, must also be considered in terms of the selection criteria themselves. In a real sense, the most important suggestions from the second study derive from the set of serendipitous results which show the apparent confounded validity of standard group-administered tests of "reading" and "intelligence." To the extent that such tests are confounded (e.g., a reading test measures "more than" reading), they may misclassify good and/or poor readers, as defined by some other, more stringent set of criteria. Needless to say, such tests could also be expected to fail to discriminate between kinds of reading difficulties, a discrimination which seems important in a study like the second one here, as discussed above, and which would be essential in suggesting appropriate remediation.

The use of group-administered "reading comprehension" and "intelligence" tests by many school systems is commonplace. While many investigators and educators have recognized and accepted the view that the validity of the estimates given by these tests does not apply to educationally and economically deprived children (as a function of the cultural bias of the tests), many seem to continue to hold the view that the validity estimates do indeed apply equally for other subgroups within a culturally homogeneous population, that is, for children with equal educational opportunity. The findings from the second study suggest that there may be systematic differences in the validity within a culturally homogeneous population. More specifically, the present findings suggest that the tests of "reading" and "intelligence" appear to be confounded measures for poor readers.

There are at least three possible interpretations of the results obtained regarding the relationships between the four school-administered tests (two "intelligence" tests, two "reading" tests) for good and poor readers:

1. The first view is that the second "intelligence" test was more heavily "reading loaded" than the first intelligence test, and thus the intelligence scores of the 4th-grade test reflect more the subjects' reading skills than an intelligence estimate alone.[25] Examination of the two particular intelligence tests used indicated that this interpretation is possible; the first intelligence test contained fewer seemingly "reading dependent" items. For example, it had relatively more items involving judgment of spatial relations and matching-to-sample of geometric figures, while the second had relatively more items requiring reading skills, such as definitions involving reading multiple-choice answers to a story which had been read to the examinee. Furthermore, examination of the two "intelligence" tests over the entire age range for which they are employed indicated that the level of reading skills required to take the test as usually administered--a subject must read the instructions and the items--seems to increase systematically with age level. If this interpretation is tenable, then a poor reader whose "true" intelligence test score remains constant over age, as measured by an intelligence test which does not require reading at all, will systematically show a decrease in apparent intelligence score over age, as measured by a "reading loaded" intelligence test. Correspondingly, a good reader whose "true" intelligence remains constant over age will increase in apparent intelligence from one year to the next, as measured by a "reading loaded" intelligence test, simply as a function of the successively higher level of reading skill.

One method for examining the degree of "reading loading" of an intelligence test is to administer the test in two forms to groups of matched good and poor readers: (1) visually, standard administration, and (2) auditorially. Only if these two forms of administration result in equivalent scores for poor readers does it seem reasonable to infer that the intelligence test is not "reading loaded." For good readers, it would be expected that visual and auditory administrations would be relatively equivalent, even if the intelligence test is reading-dependent.

2. A second possible explanation of the findings of the second study regarding the relationships between "reading" and "intelligence" tests is that the second school-administered reading test was more "intelligence loaded" than the first school-administered reading test. Both the first interpretation and this one are supported by the finding that while there was no apparent relation between the first intelligence test and the first reading test, there was, for all subjects, good and poor readers combined, a significant positive correlation between the second intelligence test scores and the

second reading comprehension test scores. However, to accept this second interpretation that the second reading test was more "intelligence loaded" than the first reading test, it would seem necessary to assume also that both intelligence tests were equally valid measures, independent of reading skills, Of course, given the finding of no significant correlation between the two intelligence tests, such an assumption seems questionable, at least for the subjects and the tests used in this study. Examination of a number of standard reading tests suggests that they do vary in terms of "intelligence loading," that is, answering comprehension questions requires making inferences not given directly in the reading passage or differences in memory requirements.

Incidentally, an unpublished study by Cromer and Anderson (personal communication, 1970) suggests that good readers perform better than poor readers in answering the multiple-choice questions of standard reading tests, even when the reading passages to which the questions refer had not been presented to be read. From this finding it would seem that the standard school-administered tests measure "more than" what most investigators take to be "reading," that is, the tests may measure differences between good and poor readers in terms of the general fund of information they bring to the reading task, or differences in test-taking ability, etc.

As noted earlier for intelligence tests, one way to measure the "intelligence loading" of a reading test is to administer the test in two forms, visually and auditorially, to good and poor readers over the age range such tests are used. It would seem reasonable to infer the reading test is a "pure" test of reading only if all the following conditions are found: (a) for auditory administration, there is no difference between good and poor readers, as selected by independent criteria; (b) for good readers, there is no difference between auditory and visual administrations of the same test; and (c) for poor readers, auditory administration results in significantly better comprehension than does visual administration.

3. The third possible interpretation of the findings regarding the test comparisons in this study involves a combination of the two other explanations offered above. Systematic exploration of the implications derived from the inconsistencies obtained here, and from earlier studies, will be a part of a program of future research. While the apparent confounded validity of these tests may account in part for the failure to verify some of the hypotheses of the second study here, certainly a more important implication concerns the possible misuse of such tests by many school systems.

Footnotes

1 **This methodology, as used by Savin and Perchonock, was complicated by** the nature of the stimuli to be recalled. In their study, each stimulus consisted of a sentence followed by a list of eight words; the subject was asked to recall both the sentence and the word list on each trial. These eight words on each trial came from eight different categories of meaning (e.g., "nature," "animal"), and the order of presentation of these categories was the same for each trial. In recall, the subject was shown a list of the category names, in the same order the words were given in the stimulus. The rationale for this procedure is unclear. The present studies used lists of digits for the second part of each trial, to prevent intra-stimulus associations and to simplify the method. Further, in the present studies, stimulus sentences were constructed equating for frequency of occurrence of component words (all high frequency). This procedure was used to make it more likely that all sentences would be recalled correctly. The rationale of this Archimedian "overflow" method makes it important to ensure verbatim recall of sentences. If all sentences are recalled, then memory for the second part of the trial (i.e., digits) can be taken as a index of difficulty of processing the preceding sentences.

2 There are any number of possibilities for the etiology of individual differences in language behavior. For example, children from subcultures which speak particular dialects of the language may have difficulty because their rules of organization (as well as the lexicon) are different from the rules of organization of the particular dialect of language they are faced with in learning to read. If their own rules of organization were to be used, the learning of reading should be facilitated. Or, a child may use a language form which is relatively non-complex (e.g., Bernstein's "restricted code," 1962), and which does not require the use or learning of certain organizational patterns for efficient communication. It may be inappropriate to call such difficulties "reading problems."

3 Cromer also identified and studied a second group of poor readers (who lacked word fluency or vocabulary skills). This group of poor readers showed no changes in comprehension as a function of grouping, that is, Conditions 3 and 4, and language organization was assumed not to be a problem for these "poor" readers.

4 Oral and silent reading are not seen as identical forms. For example, good readers obviously do not read the same way orally and silently; in the former case, they read (identify) all words, while in the latter, apparently they do not. However, for certain poor readers and beginning readers, the forms of oral and silent reading are assumed to be more similar than for accomplished readers.

5 We would expect that word-by-word readers are relatively less disrupted when the material read is of low complexity (compared with the disruption when the material is more complex). It follows that this kind of poor reader is probably not detected until intermediate grades, since the material to be read is minimally complex in the lower grades.

6 Children were used in the present study, in part, with the goal of developing a procedure which could be used in a later study comparing good readers with certain poor readers (see Discussion). The experimental task for the present study was constructed to be of moderate difficulty for the children used in this study. The present conceptual formulation would hypothesize that melodic features facilitate processing of sentences for adults as well as for children. There is a hypothesis regarding language acquisition (see Discussion) that melodic components are relatively more important earlier during the acquisition of language than later during acquisition, but this hypothesis was not tested in the present investigation.

7 An additional set of stimuli was recorded which was not used in the present study. These included the Random Strings recorded with para-syntactic components included. A set of such strings was constructed for each of the four sentence types (e.g., a random list of words recorded with the intonational pattern of a question). A study by O'Connell, Turner, and Onuska (1968) indicated that such intonational patterns facilitate the recall of lists of nonsense syllables, presumably by "chunking" the material. However, using the techniques of the present investigation, these stimuli were quite difficult to construct so that the tonal pattern could be reliably judged to be equivalent to a corresponding Regular or Anomalous sentence of the same sentence type.

8 For each trial, the following errors were scored:

1. Addition. When a number was added in recall which did not appear in the stimulus list, this was scored as an Addition error (e.g., if the stimulus list included "7 9 6 . . ." and the corresponding response was " 8 7 9 6 . . .").
2. Omission. When a failure to include a number of the stimulus list in the corresponding response occurred, this error was scored as an Omission error (e.g., if the above stimulus sequence was recalled as " 7 6 . . .").
3. Substitution. When a number was substituted in place of the corresponding number in the sequence of the stimulus list, this was scored as a Substitution error (e.g., if the above stimulus sequence was recalled as "7 8 6 . . .").
4. Reversal. If two numbers were recalled correctly, but reversed in sequence, this was scored as a Reversal error (e.g., if the above stimulus sequence was recalled as "7 6 9. . .").

[9] The digit recall data was also scored in terms of the digits recalled <u>correctly in sequence</u> for each trial. Thus, if the stimulus sequence "7 9 6 . . ." was recalled as "9 6 . . . ," it was assigned a Number Correct Score of 2. If this sequence was recalled as "8 7 9 6. . .," the Number Correct Score was 3, since this scoring system did not take into account the Addition error. The correlation between the Error Scores and the Number Correct Scores for 100 trials selected at random was -.78 ($p < .005$). Analyses performed on the Number Correct Scores yielded essentially the same results as analyses performed on the Error Scores, and are thus not reported here.

[10] It had been expected that the Passive-Questions would result in greater Error Scores than all the other Sentence Types for both the Tonal and Atonal conditions, since the Passive-Question was the most complex of the stimuli used in this study. This expectation was only partially confirmed (viz., the Error Scores for the Passive-Questions and some of the other Sentence Types did not differ significantly). One possible explanation for this result was examined further, as part of the test of the homogeneity of variance assumption underlying the analysis of variance. The Passive-Question was the only Sentence Type in the present study that could be decoded as either of two distinct syntactic forms (i.e., either a passive or a question). If such variation in decoding had occurred (i.e., the Passive-Questions sometimes decoded as passives or as questions, and at other times as passive-questions, then it would be expected that the variance for this Sentence Type would be greater than the variances of the other Sentence Types. The $F_{max}$ statistic (Winer, 1962) was used to test this possibility. For the Tonal conditions the $F_{max}$ was 3.40 ($df = 4, 7$), and for the Atonal conditions, the $F_{max}$ was 3.89 ($df = 4, 7$). Neither of these two values is statistically reliable.

[11] This test required the examinee to read short paragraphs from which words or phrases have been deleted. Using lists of words or phrases provided below each paragraph, the examinee is to select the appropriate completion (e.g., "John wanted to buy a cake. He went to the __1__ . He also bought some __2__ ." Choices: (1) country, baker, builder, airport; (2) butter, meat, fish, bread). The correct choice is dependent upon knowing the meanings of most of the words involved, knowledge of the grammatical functions of the deleted words and of the possible substitutions, general comprehension of the situation presented in the paragraph, and previous learning about similar situations. Insofar as reading comprehension includes any of these behaviors, the Paragraph Meaning subtest is a measure of reading comprehension

[12] This test requires the examinee to read sentences or paragraphs, and to answer questions concerning the information presented in those sentences or paragraphs (e.g., "I have a parrot named Perky Pete." "My parrot's name is: (1) Peter Perk; (2) Pecky Pete; (3) Peter Peck; (4) Perky Pete").

[13] The design of this study thus differed from that of Study 1, where independent groups of Ss were presented with a single Sentence Type. The design was changed for two reasons: (1) for purposes of simplification in a study with an additional between-Ss variable (good versus poor readers); and (2) to eliminate possible "set" effects for a particular grammatical type. To demonstrate differences in memory for different grammatical structures, it is necessary that Ss not expect any one particular grammatical structure, which they well may if only one such type is presented throughout the experimental task. Mehler and Carey (1967) have demonstrated that it is possible to induce such a set with adult Ss in an immediate memory task when sentences all of the same grammatical type are presented in a repeated-measurements design. While an independent groups for Sentence Types design was used in Study 1, it is unlikely that such a set was induced since Anomalous Sentences and Random Strings were also presented to each S as repeated measurements (as well as tonal and atonal stimuli). One might still argue that, in that study, when Ss heard meaningful sentences, they "knew" they were of a particular grammatical type (i.e., a set was induced for the meaningful sentences). Despite such a possibility, increased difficulty in sentence processing for increased grammatical complexity was demonstrated (for atonal sentences). In Study 2, however, where only meaningful sentences (i.e., Regular Sentences) were presented, the possibility of set-inducement is greater. The presentation of all Ss with all stimuli (heterogeneous with respect to grammatical type) in random order minimizes this possibility.

[14] A $2 \times 2$ analysis of variance performed on these two sets of reading comprehension test scores indicated that the improvement from one year to the next, as measured by these two tests, was not significantly greater for good readers, as compared with the improvement (increase in test scores) of the poor readers (for the Readers x Tests interaction, $F = 1.283$, $df = 1, 37$, $p > .10$).

[15] This matching procedure controls in two ways for an interpretation of observed experimental effects that might be attributed to differences in intelligence test scores alone. First, assuming the intelligence test measures abilities independent of abilities measured by the reading comprehension test, Ss designated as good readers have intelligence scores lower than Ss designated as poor readers. Second, to the extent that the intelligence test requires a certain level of reading skill, the intelligence test systematically underestimates the poor readers' "intelligence," compared with the test's estimation of the good readers' "intelligence." Thus, the "true" intelligence of the poor readers would be even higher than the good readers' intelligence, as measured by such a test. This procedure ensures that if there is any difference between intelligence scores of good and poor readers, the direction of difference is opposite that required by an interpretation of experimental differences due to intelligence alone.

[16] For good readers, considered separately, the correlation between the 3rd-grade reading test and the 3rd grade intelligence test was $r = +.03$ (not significant); **the correlation between 4th-grade reading and 4th-grade intelligence** was $r = +.36$ ($.10 < p < .05$). For poor readers, considered separately, the correlation between the 3rd-grade reading test and the 3rd-grade intelligence test was $r = +.25$ (not significant); the correlation between 4th-grade reading and 4th-grade intelligence was +.61 ($p < .01$). The correlation between the 4th-grade reading test and the 4th-grade intelligence test was not significantly higher for the poor readers than for the good readers ($F = .937$, $p > .10$, all two-tailed tests).

[17] One attempt to equate for good and poor reader groups' intelligence scores, as measured by the second intelligence test, was to omit those Ss who scored outside the 90-110 range on the second intelligence test. This resulted in the exclusion of 8 of the original Ss. For this grouping of Ss ($n = 32$), for all Ss pooled, again, there was no apparent relation between 3rd and 4th grade intelligence tests ($r = +.05$, not significant). For good readers, considered separately, there was a significant positive correlation between these two intelligence tests ($r = +.41$, $p < .02$); however, for poor readers, considered separately, no such relation was found ($r = .04$, not significant). For this grouping of Ss, poor readers' intelligence test scores (as measured by the first test) were equivalent to good readers' scores (first test) ($t = 1.58$, not significant); however, as measured by the second intelligence test, good readers' scores were significantly higher than poor readers' scores ($t = 2.96$, $p < .01$). Although separate sets of analyses were conducted on this grouping of Ss, they are not reported, since results obtained were equivalent to those obtained for the First Subject Regrouping.

[18] Correlations of Oral Reading Scores with scores on the school-administered tests of intelligence and reading comprehension are reported here reversing the algebraic sign of the correlation coefficient. Since the way Oral Reading was scaled, a numerically low Oral Reading Score denotes good oral reading and a numerically high reading comprehension score (or intelligence test score) denotes good reading comprehension (or high intelligence, as measured by that test), an algebraically negative correlation coefficient indicates a "positive" relationship. To avoid this possible confusion, correlations are reported with reversed signs.

[19] A complete set of correlations between Oral Reading Scores and the two sets of two school-administered tests, for good and poor reader groups considered separately (reader groups as defined by (a) the school reading tests; and (b) Oral Reading Scores) is given in Appendix B.

20 To examine a possible violation of the assumption of homogeniety of error variances, required for this analysis of variance, this analysis was also performed using a square-root transformation of the Error Scores ($x' = \sqrt{x} + \sqrt{x+1}$). Since analysis of the square-root transformation of scores yeilded the same results as analysis of the untransformed scores, the square-root transformation was not employed in subsequent analyses.

21 As measured, for example, by the Wechsler Memory Scale or the digit span subtest of the WISC. Analysis of covariance could then "partial out" effects due to overall differences between the groups in immediate memory.

22 A Duncan Multiple-Range Test comparing these means indicated that each of the four means was significantly different from the other three (all at $p < .01$), with the exception of the comparison of the five-word sentences with the six-word sentences, which did not reach this level of reliability.

23 An analysis of variance computed on derived difference scores (atonal-tonal) for good versus poor readers (4th-grade tests) and sentence length yielded a pattern of means consistent with the second hypothesis of this study (i.e., relatively greater facilitation of the "presence" of intonational features for poor readers, in comparison with good readers, as sentence length increases), but this Readers x Length interaction was reliable only at the $.10 < p < .20$ confidence level. Differences between mean Error Scores (atonal-tonal): (1) for 5-word sentences, .013 (good readers) versus .075 (poor readers); (2) for 6-word sentences, .300 (good readers) versus .337 (poor readers); and (3) for 7-word sentences, .163 (good readers) versus .363 (poor readers).

24 The multiple-factor randomization test was programmed by Dr. Neil Rankin, and performed by him on a Digital PDP-12 computer. Exact probability values for the randomization tests reported here are given for 250 randomizations of the distribution. This analysis is an extension for more than one variable of the Fisher randomization test (Kempthorne and Doerfler, 1969; Lohnes and Cooley (1968), which, in turn, is based on an extension of the logic of the Fisher exact test. The test uses a Monte Carlo procedure to obtain successive randomizations of the obtained distribution of scores; it is as efficient as the analysis of variance, but does not require parametric assumptions.

25 It may well be that any measure of "aptitude," as that concept is commonly understood, is correlated with reading skill to some degree. However, both administration manuals for the two tests used for these subjects provide "IQ equivalents," and these tests are often used as if they measured some ability independent of reading skill. Furthermore, while we discuss only the two particular tests used for these subjects, our argument should hold for most, if not all group-administered "intelligence" tests.

## REFERENCES


Allport, F. H. *Social psychology*. Boston: Houghton Mifflin, 1924.

Bernstein, B. Social class, linguistic codes, and grammatical elements: Linguistic codes, hesitation phenomena, and intelligence. *Language and Speech*, 1962, *5*.

Braine, M. D. S. On learning the grammatical order of words. *Psychological Review*, 1963, *70*, 323-348.

Bregman, A. S., & Strasberg. Memory for the syntactic form of sentences. *Journal of Verbal Learning and Verbal Behavior*, 1968, *7*, 396-403.

Bruning, J., & Kintz, B. L. *Computational handbook of statistics*. Glenview, Illinois: Scott-Foresman, 1968.

Carver, R. P. The efficacy of "chunking" reading materials. U. S. Department of Health, Education, and Welfare Document, Project No. 8-C-051, 1968.

Chomsky, N. *Aspects of the theory of syntax*. Cambridge, Mass.: M.I.T., 1965.

Clifton, C., & Odom, P. Similarity relations among certain English sentence constructions. *Psychological Monographs*, 1966, *80*, Whole No. 6B. (613).

Cromer, W. The difference model: A new explanation for some reading difficulties. *Journal of Educational Psychology*, 1970, *61*, 471-483.

Cromer, W., & Wiener, M. Idiosyncratic response patterns among good and poor readers. *Journal of Consulting Psychology*, 1966, *30*, 1-10.

Cromer, W., Wiener, M., & Shilkret, R. Do reading tests measure "reading"? Office of Education, Research Grant Proposal, January 1972.

Edwards, A. L. *Experimental design in psychological research*. New York: Holt, Rinehart & Winston, 1965.

Epstein, W. The influence of syntactical structure on learning. *American Journal of Psychology*, 1961, *74*, 80-85.

Epstein, W. A further study of the influence of syntactical structure on learning. *American Journal of Psychology*, 1962, *75*, 121-126.

Epstein, W. Some conditions of the influence of syntactical structure on learning: Grammatical transformation, learning instructions, and "chunking." Journal of Verbal Learning and Verbal Behavior, 1967, 6, 415-419.

Epstein, W. Recall of word lists following learning of sentences and of anomalous and random strings. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 20-25.

Fillenbaum, S. Memory for gist: some relevant variables. Language and Speech, 1966, 9, 217-227.

Fillenbaum, S. On the use of memorial techniques to assess syntactic structures. Psychological Bulletin, 1970, 73, 231-237.

Fodor, J., & Garrett, M. Some reflections on competence and performance. In Lyons and Wales (Eds.), Psycholinguistic papers. Edinburgh: University Press, 1966.

Ford, B. L. Children's imitation of sentences which vary in pause and intonational pattern. Unpublished Ph.D. Dissertation, Cornell University, 1970.

Foss, D. J., & Cairns, H. S. Some effects of memory limitation upon sentence comprehension and recall. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 541-547.

Garrett, M., & Fodor, J. Psychological theories and linguistic constructs. In Dixon and Horton (Eds.), Verbal behavior and general behavior theory. Englewood Cliffs, New Jersey: Prentice-Hall, 1968.

Gleitman, L., & Shipley, E. The acquisition of linguistic structure: The emergence of the child as grammarian. Technical Report, Appendix D, Grant MH-20041, July, 1971.

Glucksberg, S., & Danks, J. H. Grammatical structure and recall: A function of the space in immediate memory or recall delay? Perception and Psychophysics, 1969, 6, 113-117.

Gruber, J. S. Topicalization in child language. Foundations of Language, 1967, 3, 37-65.

Halliday, M. A. K. Notes on transitivity and theme in English. Journal of Linguistics. Part I: 1967, 3, 37-81; Part II: 1967, 3, 199-224; Part III: 1968, 4, 179-215.

Harris, Z. S. Co-occurrence and transformations in linguistic structure. Language, 1957, 33, 283-340.

Hayhurst, H. Some errors of young children in producing passive sentences. Journal of Verbal Learning and Verbal Behavior, 1967, 6, 654-660.

Hockett, C. F. Where the tongue slips, there slip I, in To honor Roman Jakobson. The Hague: Mouton, 1967.

Huey, E. B. The psychology and pedogogy of reading. New York: Macmillan, 1908.

Ingram, D. Transitivity in child language. Language, 1971, 47, 888-910.

Kaplan, Elenor, & Kaplan, G. The prelinguistic child. In Eliot, J. (Ed.) Human development and cognitive processes. New York: Holt, Rinehart, & Winston, 1971.

Kagan, J. Formal discussion. In Bellugi and Brown (Eds.), The acquisition of language. Society for Research in Child Development Monographs, 1964, 29, No. 1, 169-172.

Katona, G. Organizing and memorizing. New York: Columbia University Press, 1940.

Kempler, B., & Wiener, M. Personality and perception in the recognition threshold paradigm. Psychological Review, 1963, 4, .19-356.

Kempthorne, O., & Doerfler, T. E. The behaviour of some significance tests under experimental randomization. Biometrika, 1969, 56, 231-248.

Lane, H., & Schneider, B. Some discriminative properties of syntactic structures. Journal of Verbal Learning and Verbal Behavior, 1963, 2, 457-461.

Lipsky, A. Rhythm as a distinguishing characteristic of prose style. Archives of Psychology, 1907, No. 4.

Lohnes, P. R., & Cooley, W. W. Introduction to statistical procedures. New York: Wiley, 1968.

Makita, K. The rarity of reading disability in Japanese children. American Journal of Orthopsychiatry, 1968, 38, 599-614.

Marks, L. A., & Miller, G. A. The role of semantic and syntactic constraints in the memorization of English sentences. Journal of Verbal Learning and Verbal Behavior, 1964, 3, 1-5.

Martin, E., & Roberts, K. H. Grammatical factors in sentence retention. Journal of Verbal Learning and Verbal Behavior, 1966, 5, 211-218.

Martin, E., & Roberts, K. H. Sentence length and sentence retention in the free-learning situation. Psychonomic Science, 1967, 8, 535-536.

Martin, E., Roberts, K. H., and Collins, A. M. Short-term memory for sentences. Journal of Verbal Learning and Verbal Behavior, 1968, 7, 560-566.

Matthews, W. A. Transformational complexity and short term recall. Language and Speech, 1968, 11, 120-128.

Mehler, J. Some effects of grammatical transformations on the recall of English sentences. Journal of Verbal Learning and Verbal Behavior, 1963, 2, 346-351.

Mehler, J., & Carey, P. Role of surface and base structure in the perception of sentences. Journal of Verbal Learning and Verbal Behavior, 1967, 6, 335-338.

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

Miller, G. A. Some psychological studies of grammar. American Psychologist, 1962, 17, 748-762.

Miller, G. A., & McNeill, D. Psycholinguistics. In Lindzey and Aronson (Eds.) The handbook of social psychology, Vol. III, Reading, Mass.: Addison-Wesley, 1969.

Miller, W., & Ervin, Susan. The development of grammar in child language. In Bellugi and Brown (Eds.), The acquisition of Language. Society for Research in Child Development Monographs, 1964, 29, No. 1, 9-34.

Morse, P. A. Speech perception in six-week-old infants. Paper presented at Society for Research in Child Development Meetings, Minneapolis, April, 1971.

Neisser, U. Cognitive psychology. New York: Appleton-Century-Crofts, 1967.

Oakan, R., Wiener, M., & Cromer, W. Identification, organization, and reading comprehension for good and poor readers. Journal of Educational Psychology, 1971, 62, 71-78.

O'Connell, D., Turner, E., and Onuska, L. Intonation, grammatical structure, and contextual association in immediate recall. Journal of Verbal Learning and Verbal Behavior, 1968, 7, 110-116.

Orenstein, H. B., & Schumsky, D. A. The effects of sentence length and grammatical structure in a serial learning task. Psychonomic Science, 1970, 19, 243-244.

Paivio, A. Imagery and language. In Segal, S. J. (Ed.), Imagery. New York: Academic, 1971.

Perfetti, C. A. Sentence retention and the depth hypothesis. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 101-104.

Perfetti, C. A. Lexical density and phrase structure depth as variables in sentence retention. Journal of Verbal Learning and Verbal Behavior, 1969a, 8, 719-724.

Robinson, Helen M. (Ed.) Administrator's Manual, Gray Tests of oral reading. New York: Bobbs-Merrill, 1967.

Sachs, J. S. Recognition memory for syntactic and semantic aspects of connected discourse. Perception and Psychophysics, 1967, 2, 437-442.

Savin, H., & Perchonock, E. Grammatical structure and the immediate recall of English sentences. Journal of Verbal Learning and Verbal Behavior, 1965, 4, 348-353.

Scholes, R. J. The role of grammaticality in the imitation of word strings by children and adults. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 225-228.

Siegel, S. Nonparametric statistics for the behavioral sciences. New York: McGraw-Hill, 1956.

Steiner, R., Wiener, M., & Cromer, W. Comprehension training and identification for poor and good readers. Journal of Educational Psychology, 1971, 62, 506-513.

Thorndike, E., & Lorge, I. The teacher's word book of 30,000 words. New York: Teachers College, Columbia University, 1944.

Wearing, A. J. Recognition memory for sentences of varying length. Psychonomic Science, 1969, 15, 221-222.

Wearing, A. J., & Crowder, R. G. Dividing attention to study sentence acquisition. Journal of Verbal Learning and Verbal Behavior, 1971, 10, 254-261.

Weeks, T. E. Speech registers in young children. Child Development, 1971, 42, 1119-1131.

Weener, P. Language structure and the free recall of verbal messages by children. Developmental Psychology, 1971, 5, 237-243.

Werner, H., & Kaplan, B. Symbol Formation. New York: Wiley, 1963.

Wiener, M., & Cromer, W. Reading and reading difficulty: a conceptual analysis. Harvard Educational Review, 1967, 37, 620-643.

Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill, 1962.

Woodworth, R. S. Experimental Psychology. New York: Henry Holt, 1938.

Wright, P. Two studies of the depth hypothesis. British Journal of Psychology, 1969, 60, 63-69.

Yngve, V. H. A model and an hypothesis for language structure. Proceedings of American Philosophical Society, 1960, 104, 444-466.

## Appendix A

### Stimulus Materials - Study 1

**Simple--Tonal**

**Regular**

| | | |
|---|---|---|
| 1. | The girl walked the dog. | 1 4 3 6 9 7 |
| 2. | The woman ate the food. | 9 7 3 6 8 4 |
| 3. | The child caught the ball. | 8 5 9 1 7 6 |
| 4. | The boy wrote the letter | 8 2 9 4 1 3 |
| 5. | The man drove the car. | 7 3 8 4 6 2 |

**Anomalous**

| | | |
|---|---|---|
| 1. | The ball ate the girl. | 5 7 3 6 4 1 |
| 2. | The letter caught the woman. | 4 2 8 7 9 1 |
| 3. | The car wrote the child. | 8 5 1 6 4 9 |
| 4. | The dog drove the boy. | 7 3 1 4 6 9 |
| 5. | The food walked the man. | 2 9 5 7 3 8 |

**Simple--Atonal**

**Regular**

| | | |
|---|---|---|
| 1. | The girl walked the dog. | 7 9 6 4 3 1 |
| 2. | The woman ate the food. | 4 8 6 3 7 9 |
| 3. | The child caught the ball. | 6 3 1 9 5 8 |
| 4. | The boy wrote the letter | 3 1 4 9 2 8 |
| 5. | The man drove the car. | 2 6 4 8 3 7 |

**Anomalous**

| | | |
|---|---|---|
| 1. | The ball ate the girl. | 1 4 6 3 7 5 |
| 2. | The letter caught the woman. | 1 9 6 8 2 4 |
| 3. | The car wrote the child. | 9 4 6 1 5 8 |
| 4. | The dog drove the boy. | 9 6 4 1 3 7 |
| 5. | The food walked the man. | 8 3 7 5 9 2 |

Random

| | | |
|---|---|---|
| 1. | Caught car the girl the. | 9 2 8 6 3 1 |
| 2. | **Woman the dog the wrote.** | 5 7 3 2 9 1 |
| 3. | Food the child drove the | 8 1 4 6 9 5 |
| 4. | The walked boy ball the. | 2 5 8 6 3 7 |
| 5. | Letter the ate man the. | 3 7 1 5 4 2 |

Question--Tonal

Regular

| | | |
|---|---|---|
| 1. | Did the girl walk the dog? | 1 6 2 5 7 9 |
| 2. | Did the woman eat the food? | 8 6 3 9 2 5 |
| 3. | Did the child catch the ball? | 1 3 9 5 8 6 |
| 4. | Did the boy write the letter? | 3 6 4 7 1 9 |
| 5. | Did the man drive the car? | 6 1 7 2 8 3 |

Anomalous

| | | |
|---|---|---|
| 1. | Did the ball eat the girl? | 8 1 4 6 3 7 |
| 2. | Did the letter catch the woman? | 2 4 9 6 3 8 |
| 3. | Did the car write the child? | 6 2 5 7 1 4 |
| 4. | Did the dog drive the boy? | 7 1 4 9 6 3 |
| 5. | Did the food walk the man? | 7 9 5 3 1 4 |

Question--Atonal

Regular

| | | |
|---|---|---|
| 1. | Did the girl walk the dog. | 9 7 5 2 6 1 |
| 2. | Did the woman eat the food. | 5 2 9 3 6 8 |
| 3. | Did the child catch the ball. | 6 8 5 9 3 1 |
| 4. | Did the boy write the letter. | 9 1 7 4 6 3 |
| 5. | Did the man drive the car. | 3 8 2 7 1 6 |

Anomalous

| | | |
|---|---|---|
| 1. | Did the ball eat the girl. | 7 3 6 4 1 8 |
| 2. | Did the letter catch the woman. | 8 3 6 9 4 2 |
| 3. | Did the car write the child | 4 1 7 5 2 6 |
| 4. | Did the dog drive the boy. | 3 6 9 4 1 7 |
| 5. | Did the food walk the man. | 4 1 3 5 9 7 |

### Random

| | | |
|---|---|---|
| 1. | Did catch car the girl the. | 9 5 8 2 4 6 |
| 2. | **Woman did the dog the write.** | **1 4 2 8 6 3** |
| 3. | Food the did child drive the. | 2 5 9 1 4 7 |
| 4. | The walk boy ball did the. | 1 6 2 7 5 9 |
| 5. | Letter the eat man the did. | 4 2 5 9 6 8 |

## Passive--Tonal

### Regular

| | | |
|---|---|---|
| 1. | The dog was walked by the girl. | 2 4 9 3 8 1 |
| 2. | The food was eaten by the woman. | 8 4 2 6 3 9 |
| 3. | The ball was caught by the child. | 1 9 2 8 6 3 |
| 4. | The letter was written by the boy. | 9 7 4 8 5 3 |
| 5. | The car was driven by the man. | 2 1 6 8 5 3 |

### Anomalous

| | | |
|---|---|---|
| 1. | The girl was eaten by the ball. | 8 3 9 7 4 6 |
| 2. | The woman was caught by the letter. | 4 8 3 7 5 1 |
| 3. | The child was written by the car. | 7 5 8 4 1 3 |
| 4. | The boy was driven by the dog. | 5 9 7 3 6 1 |
| 5. | The man was walked by the food. | 6 8 3 9 7 1 |

## Passive--Atonal

### Regular

| | | |
|---|---|---|
| 1. | The dog was walked by the girl. | 1 8 3 9 4 2 |
| 2. | The food was eaten by the woman. | 9 3 6 2 4 8 |
| 3. | The ball was caught by the child. | 3 6 8 2 9 1 |
| 4. | The letter was written by the boy. | 3 5 8 4 7 9 |
| 5. | The car was driven by the man. | 3 5 8 6 1 7 |

### Anomalous

| | | |
|---|---|---|
| 1. | The girl was eaten by the ball. | 6 4 7 9 3 8 |
| 2. | The woman was caught by the letter. | 1 5 7 3 8 4 |
| 3. | The child was written by the car. | 3 1 4 8 5 7 |
| 4. | The boy was driven by the dog. | 1 6 3 7 9 5 |
| 5. | The man was walked by the food. | 1 7 9 3 8 6 |

### Random

| | | |
|---|---|---|
| 1. | By caught car the was girl the. | 1 9 6 8 5 2 |
| 2. | **Woman the was dog the by wrote.** | 7 1 3 6 4 8 |
| 3. | Food the was child drove the by. | 8 5 2 6 1 9 |
| 4. | The by walked was boy ball the. | 6 4 8 5 3 1 |
| 5. | Was letter the ate man by the. | 5 3 7 1 4 2 |

## Passive-Question--Tonal

### Regular

| | | |
|---|---|---|
| 1. | Was the dog walked by the girl? | 7 1 8 4 6 3 |
| 2. | Was the food eaten by the woman? | 7 9 5 2 3 6 |
| 3. | Was the ball caught by the child? | 4 7 1 9 5 8 |
| 4. | Was the letter written by the boy? | 8 3 5 9 6 1 |
| 5. | Was the car driven by the man? | 6 3 7 9 2 8 |

### Anomalous

| | | |
|---|---|---|
| 1. | Was the girl eaten by the ball? | 4 3 5 7 1 9 |
| 2. | Was the woman caught by the letter? | 2 9 1 6 3 8 |
| 3. | Was the child written by the car? | 5 4 6 8 1 9 |
| 4. | Was the boy driven by the dog? | 4 3 1 7 2 8 |
| 5. | Was the man walked by the food? | 3 8 2 1 5 7 |

## Passive-Question--Atonal

### Regular

| | | |
|---|---|---|
| 1. | Was the dog walked by the girl. | 3 5 4 8 1 7 |
| 2. | Was the food eaten by the woman. | 6 3 2 5 9 7 |
| 3. | Was the ball caught by the child. | 8 5 9 1 7 4 |
| 4. | Was the letter written by the boy. | 1 6 9 5 3 8 |
| 5. | Was the car driven by the man. | 8 2 9 7 3 6 |

### Anomalous

| | | |
|---|---|---|
| 1. | Was the girl eaten by the ball. | 9 1 7 5 4 3 |
| 2. | Was the woman caught by the letter. | 8 3 6 1 9 2 |
| 3. | Was the child written by the car. | 9 1 8 6 4 7 |
| 4. | Was the boy driven by the dog. | 8 2 7 1 3 5 |
| 5. | Was the man walked by the food. | 7 5 1 4 8 3 |

Random

| | | |
|---|---|---|
| 1. | Was by caught car the girl the. | 3 7 2 5 9 6 |
| 2. | **Woman the was dog the by wrote.** | 9 4 1 3 8 5 |
| 3. | Food the drove was child the by . | 5 8 3 6 2 9 |
| 4. | The by walked boy was ball the. | 9 4 1 6 8 5 |
| 5. | Letter the ate man by the was. | 3 1 6 9 5 8 |

Who-Question--Regular

Tonal

| | | |
|---|---|---|
| 1. | Who walked the dog? | 1 4 2 6 3 9 |
| 2. | Who ate the food? | 2 6 1 8 4 7 |
| 3. | Who caught the ball? | 7 2 4 8 1 9 |
| 4. | Who wrote the letter? | 4 9 2 6 3 8 |
| 5. | Who drove the car? | 6 1 4 8 2 9 |

Atonal

| | | |
|---|---|---|
| 1. | Who walked the dog. | 9 3 6 2 4 1 |
| 2. | Who ate the food. | 7 4 8 1 6 2 |
| 3. | Who caught the ball. | 9 1 8 4 2 7 |
| 4. | Who wrote the letter. | 8 3 6 2 9 4 |
| 5. | Who drove the car. | 9 2 8 4 1 6 |

Appendix B

**Correlations Between Oral Reading Scores and School-Administered**
Tests of Intelligence and Reading Comprehension for Good and Poor Readers Considered Separately

I. For good and poor reader groups, as defined by the First Subject Regrouping (i.e., school-administered reading test).

| | Good Readers (n=20) | Poor Readers (n=20) |
|---|---|---|
| Oral Reading--3rd-grade reading | +.22 | +.31 |
| Oral Reading--4th-grade reading | +.63 | +.61 |
| Oral Reading--3rd-grade intelligence | -.31 | +.23 |
| Oral Reading--4th-grade intelligence | +.08 | +.33 |

II. For Good and Poor Reader Groups, as determined by a median split of Oral Reading Scores

| | Good Readers (n=20) | Poor Readers (n=20) |
|---|---|---|
| Oral Reading--3rd-grade reading | +.43 | +.49 |
| Oral Reading--4th-grade reading | +.54 | +.64 (n=19) |
| Oral Reading--3rd-grade intelligence | -.33 | +.26 |
| Oral Reading--4th-grade intelligence | +.15 | +.42 (n=18) |

III. For Good and Poor Reader Groups, as determined by the upper and lower extremes of the Oral Reading Scores (first and fourth quartiles).

| | Good Readers (n=10) | Poor Readers (n=10) |
|---|---|---|
| Oral Reading--3rd-grade reading | -.46 | +.46 |
| Oral Reading--4th-grade reading | +.05 | +.69 |
| Oral Reaaing--3rd-grade intelligence | -.11 | +.44 |
| Oral Reading--4th-grade intelligence | +.07 | +.18 |

* $p < .05$
** $p < .02$
*** $p < .01$, all two-tailed tests. As before, all correlation coefficients are reported with algebraic signs reversed to indicate the "true" relationship (e.g., a positive r denotes (a) good oral reading--good reading comprehension (a high intelligence test score) and, (b) poor oral reading--poor reading comprehension (or low intelligence test score).