ABSTRACT

Designed to measure the syntactic maturity of oral speech, the K-Ratio Index was devised for use in an investigation of the relationships between certain measures of syntactic maturity of oral languages and silent reading comprehension scores. Preparation for computing the ratio was accomplished by transcribing oral speech samples, excluding syntactically irrelevant matter, dividing into T-units, placing each T-unit at the top of its own data sheet, and counting the number of kernel structures imbedded in each T-unit.

[This document is one of those reviewed in The Research Instruments Project (TRIP) monograph, "Measures for Research and Evaluation in the English Language Arts" to be published by the Committee of Research of the National Council of Teachers of English in cooperation with the ERIC Clearinghouse on Reading and Communication Skills. A TRIP review which precedes the document lists its category (Language Development), title, author, date, and age range (intermediate), and describes the instrument's purpose and physical characteristics.] (RB)
The attached document contains one of the measures reviewed in the TRIP committee monograph titled:

**Measures for Research and Evaluation in the English Language Arts**

TRIP is an acronym which signifies an effort to abstract and make readily available measures for research and evaluation in the English language arts. These measures relate to language development, listening, literature, reading, standard English as a second language or dialect, teacher competencies, or writing. In order to make these instruments more readily available, the ERIC Clearinghouse on Reading and Communication Skills has supported the TRIP committee sponsored by the Committee on Research of the National Council of Teachers of English and has processed the material into the ERIC system. The ERIC Clearinghouse accession numbers that encompass most of these documents are CS201320-CS201375.

**TRIP Committee:**

W.T. Fagan, Chairman  
University of Alberta, Edmonton

Charles R. Cooper  
State University of New York at Buffalo

Julie M. Jensen  
The University of Texas at Austin

Bernard O'Donnell  
Director, ERIC/RCS

Roy C. O'Donnell  
The University of Georgia  
Liaison to NCTE Committee on Research
Category: Language Development
Title: "K-Ratio"
Author: Kenneth H. Calvert
Age Range: intermediate

Description of Instrument:
Purpose - To measure the syntactic maturity of oral speech
Date - 1971

Physical Description - The K-Ratio Index was devised for use in an investigation of the relationships between certain measures of syntactic maturity of oral languages and silent reading comprehension scores. Level of reading achievement was compared with the following measures: mean T-unit length, subordination ratio, KERNEL STRUCTURE RATIO, frequency of multi-clause T-units, frequency of long T-units, frequency of surface structure patterns, and frequency of kernel structure patterns. "K-Ratio," the abbreviation for "kernel structure ratio," is computed by dividing the number of T-units in a speech sample by the number of kernel structures. Preparation for computing the ratio was accomplished by transcribing oral speech samples, excluding syntactically irrelevant matter, dividing into T-units, placing each T-unit at the top of its own data sheet, and counting the number of kernel structures imbedded in each T-unit.

[see attached]

---

A-2
(Student A, T-unit #2)       (T-unit)
(# of words in T-unit)

The old lady who owned the little dog
lived in a huge house.
(SV)
(Surface structure pattern)

(Kernel structures and their patterns)

1. The lady lived in a house.
   (SV)

2. The lady is old.
   (SVCa)

3. The lady owns a dog.
   (SVO)

4. The dog is little.
   (SVCa)

5. The house is huge.
   (SVCa)

K-Ratio = Total No. of T-units
         Total No. of Kernel Structures
Validity, Reliability, and Normative Data:

The researcher sought a measure that would account for subclausal structures resulting from deletion transformations. Hunt, in his study, questioned the insensitivity of the subordination ratio to reduced clauses (a reduced clause is a former clause which lost its clausal state by way of deletion transformations, e.g. "The boy who was large" became "the large boy."). Calvert, investigating the possibility that the reduced clauses may have also suppressed the sensitivity of the T-unit, revealed that the K-Ratio is a better predictor of reading achievement (Point Biserial Correlation = 0.406; p < .05) than is mean T-unit length (r_pb = 0.351; p = .05) or subordination ratio (r_pb = 0.320; p > .05).

Ordering Information:

ED

Related Documents:

AN INVESTIGATION OF RELATIONSHIPS BETWEEN
THE SYNTACTIC MATURITY OF ORAL LANGUAGE
AND READING COMPREHENSION SCORES

Kenneith H. Calvert

Dissertation Abstract and Summary

University of Alabama
1971

Committee:

Jeannine N. Webb, Chairman (now at University of Florida)
James B. McMillan, then chairman of English Department
Carlton Boyer
Robert K. Leigh
AN INVESTIGATION OF RELATIONSHIPS BETWEEN THE SYNTACTIC MATURITY OF ORAL LANGUAGE AND READING COMPREHENSION SCORES

Kenneith H. Calvert

Abstract

This study investigated the relationships of oral language—specifically, syntactic maturity—and reading achievement.

Audio-taped language samples from two groups, reading achievers and reading underachievers, were transcribed and compared for syntactic maturity. The major language maturity measure was a composite of three measures of complexity—T-unit length, subordination and kernel structures. The components of the complexity measure were also compared for predictability of reading achievement.

A higher degree of syntactic maturity was found among subjects classified as reading achievers, as measured by the composite complexity measure. No significant difference was found between the two groups when surface structures (subj-verb; subj-verb-obj, etc.) were compared. The findings suggest that achievers and underachievers put their language together in the same way, but the language of achievers is more content-loaded.

Kernel structures, in this study, were more predictive of reading achievement than T-unit length and subordination.
The Problem

Learning to read is a problem for many students in the schools. The number of poor readers has been estimated to be one-third of the elementary school enrollment (Thompson, 1966). According to reports on remedial and compensatory education programs of recent years (Austin & Morrison, 1963; Westinghouse Report, 1969), efforts to ameliorate the problem have been disappointing. Lasting gains in scholastic achievement and language development have, reportedly, not been made. Significance might be attached to the fact that scholastic achievement and language development were the two areas reported as not having such gains. Relationships between the two may be more significant than even earlier research (e.g., Loban, 1963; McCarthy, 1954) has suggested.

This study investigated the relationships of oral language production, specifically syntactic maturity, and reading achievement. The study was to identify any existing correlation between language and level of achievement by analyzing the language of reading achievers and underachievers for syntactic maturity.

Definition of Terms

Several terms which are important in the report of this study are defined as follows:

1. Reading achiever. A student who scores at or above grade level on a standardized reading achievement test.
2. **Reading underachiever.** A student who scores at least one year below grade level on a standardized reading achievement test.

3. **T-unit.** A short name for "minimal terminal unit" (Hunt, 1964). This syntactic unit consists of a single predication together with any subordinate clauses that may be related to it ("the big black dog on the porch bit the postman who was knocking at the door" but not "the big black dog bit the man and the man threw a shoe at him").

4. **Multi-clause T-unit.** A T-unit containing more than one clause, i.e., it has a main clause as well as one or more subordinate clauses.

5. **Long T-unit.** A T-unit having 20 or more words in it.

6. **Surface structure.** The form of a sentence, i.e., the sentence that we hear and read ("the black dog bit the man").

7. **Kernel structure.** An intermediary structure which lies between the deep structure, the structure giving meaning to the sentence, and the surface structure. It is an abstraction presumed to underlie the surface sentences and to grow out of the ultimate deep structure, according to the system of transformational-generative grammar as proposed by Chomsky (1965). An example is "the dog is black" which underlies the surface phrase, "the black dog."

8. **K-ratio.** The abbreviation for "kernel structure ratio." This ratio, computed by dividing the number of T-units by the number of kernel structures/ accounts for underlying structures embedded in surface structures.*

9. **S-ratio.** The abbreviation for "subordination ratio." This ratio is derived by dividing the number of subordinate clauses by the total number of all clauses.

*Devised by the researcher especially for this study.
10. **SIS.** The abbreviation for "socioeconomic status."

11. **Comparison group.** A pair of groups to be compared in the study. For example, the group of achievers and the group of underachievers form one comparison group.

12. **Syntactic maturity.** That characteristic of language that is assumed to exist when complexity is present. For this study, complexity is considered the result of such properties as longer T-units, more subordination, and more kernel structures.

13. **Complexity measure.** A composite measure in which the mean T-unit length, S-ratio, and K-ratio combine to measure syntactic maturity.

Review of Related Literature and Research

Fries (1962) has said that reading depends first of all upon language control. Learning to read starts with and builds upon "whatever habits of language responses exist for the learner at the time (p. 187)." For the most part, findings from various studies agree with Fries' statement. Loban (1963), whose study has become a classic one, found that children who were advanced in general language ability were also advanced in reading.

McCarthy (1954) reported research which found that first graders who made unsatisfactory progress in reading also showed linguistic immaturity, particularly by the absence of elaborated sentences, the shortness of their average sentence, and lack of connectives. A variety of measures for identifying relationships of some aspect of oral language and reading were found by other researchers (Gibbons, 1941; Strickland, 1962; Ruddell, 1965, 1966; Saur, 1968; Skinner, 1969).
There were those who were not in complete agreement with the conclusion that definite relationships have been found between language and reading achievement (Winter, 1957; Weintraub, 1968; Bougere, 1969). This lack of agreement, however, may be a result of a lack of uniformity in measures of language maturity and reading comprehension (McCarthy, 1954). Obsolescence of measures as length of response, sentence length, and frequencies of parts of speech has also been noted (Loban, 1963; Hunt, 1964; O'Donnell, et al., 1967).

The assumption that language maturity can be effected by certain curricular and instructional considerations is supported in the research. Increased opportunities to use language within the school day is advocated, the rationale being that learning language comes by using it (Bromwich, 1968; Cazden, 1966; Chomsky, 1968; Hillerich, 1968; McCaffrey, 1970). Ruddell (1967) suggested that syntactic options can be extended by such oral enrichment activities as role playing, storytelling, group discussions of direct experiences, reading literature to children, and using experience charts. Cazden (1970) recommended giving children an opportunity to talk things over out loud with conversation focused on the development of ideas. These are ways the school curriculum becomes a part of a child's language development.

Purpose of the Study

The purpose of the study was to investigate the relationships between certain measures of syntactic maturity of oral language and silent reading comprehension scores. Specifically, can a statistically significant relationship be found between level of reading achievement and the following measures:
1. mean T-unit length?
2. subordination ratio?
3. kernel structure ratio?
4. frequency of multi-clause T-units?
5. frequency of long T-unit?
6. frequency of surface structure patterns?
7. frequency of kernel structure patterns?

When statistically significant relationships were found between achievers and underachievers, these subsequent questions were asked:

1. Does sex appear to significantly influence this relationship?
2. Does socioeconomic status appear to influence this relationship?

Design and Procedure

Subjects

The subjects in the study were 32 fifth and sixth grade students enrolled in the Harnett County, North Carolina schools during the 1970-1971 school year. These subjects, matched with respect to age, sex, SES, and IQ, were divided into two groups, 16 reading achievers and 16 reading underachievers. Subdivisions with respect to sex and socioeconomic status were made to form a total of nine comparison groups. These groups were:

1. Achievers and underachievers
2. Males and females
3. Male achievers and female achievers
4. Male underachievers and female underachievers
5. Low SES subjects and mid SES subjects
6. Low SES achievers and mid SES achievers
7. Low SES underachievers and mid SES underachievers
8. Low SES boys and mid SES boys
9. Low SES girls and mid SES girls

The level of reading achievement was established by administering the comprehension subtest of the Gates-MacGinitie Reading Test, Survey D, Form I. Grade level at the time of testing was 6.8 for sixth graders and 5.8 for fifth graders.

Socioeconomic status was identified by parental occupation. The "Revised Occupational Rating Scale" from Warner, Meeker, and Bell's "Index of Status Characteristics" (Miller, 1970), established status on seven levels, Classes I through VII. Each class was represented among the subjects; however, over half of them clustered around the center, Classes IV and V.

An attempt to control intelligence was made by including only those subjects having IQ scores within the average range (90 through 110). However, within this range, a statistically significant difference was discovered between the achievers and underachievers, the achievers having the higher mean. A simple regression statistical technique was then applied to parcel out the variance caused by IQ.

Data Collection

The data were measures of language samples obtained from the subjects. These samples were oral responses to a picture shown the subjects by an interviewer. After an opportunity to become acquainted with the interviewer, the
subject was asked to tell a story about the picture. Responses were taped and later transcribed, without punctuation, into typewritten form. The transcripts were then segmented into T-units, the structures which were to be subjected to syntactic analysis. Only the first 20 T-units of each subject's response were used in order to keep the corpus manageable.

Analysis

Data. The analysis consisted first of removing all syntactically irrelevant matter such as audible pauses, false starts, redundant subjects, repetitions, and "fillers" (well, you know, and so). The responses were then divided into T-units, each T-unit being entered at the top of a 5 by 8 inch slip, on which all analytic data for the T-unit were to be written. The code for subject identification was entered in the upper left corner; for example, "A-3" indicated "the third T-unit of Subject A." Immediately below the T-unit entry, the structure pattern (subj-verb; subj-verb-obj; etc.) for that unit was entered in parentheses. The number of words in the T-unit was written in the upper right corner of the data slip; the number of kernel structures was entered in the lower left corner (see Appendix).

Data on the T-unit slips were used to tabulate and compute the various maturity index scores. The mean T-unit length for each of the two large comparison groups, achievers and underachievers, was computed by dividing the total number of words in all the T-units produced by the group by the number of T-units. The subordination ratio was derived by dividing the number of subordinate clauses by the total number of clauses. The kernel structure ratio was found by dividing the number of T-units by the total number of kernel structures. Frequencies of T-unit types (multi-clause and long), as well as structure patterns, were also tabulated from entries on
Below is a summary of the information obtained from the data slips:

1. Mean T-unit length
2. Subordination ratio (S-ratio)
3. Kernel structure ratio (K-ratio)
4. Frequency of multi-clause T-units
5. Frequency of long T-units
6. Frequencies of surface structure patterns
   a. THERE + sentence
   b. NEGATIVE + sentence
   c. Subject-verb (SV)
   d. Subject-verb-object (SVO)
   e. Subject-verb-indirect object-object (SVIO)
   f. Subject-verb-adjective complement (SVC\textsubscript{a})
   g. Subject-verb-noun complement (SVC\textsubscript{n})
7. Frequencies of kernel structure patterns
   a. Subject-verb (SV)
   b. Subject-verb-object (SVO)
   c. Subject-verb-adjective complement (SVC\textsubscript{a})
   d. Subject-verb-noun complement (SVC\textsubscript{n})

**Statistical**

The primary statistical procedure used to evaluate the complexity measure was Cattell's Coefficient of Profile Similarity. This statistical technique plotted profiles for each group comparison from the z-score means of the three elements in the complexity measure. Cattell's Index emphasizes the level of the different elements and their composite profile. By applying equal
weights to each variable (element), groups can be compared on as many variables as desired. Another advantage of this index is that it can tolerate profile elements that have a degree of correlation.

The coefficient of similarity yielded by the Cattell Index indicates the degree to which groups are alike or different. As the groups become more alike, the coefficient approaches +1.00. As the groups become more unlike, the coefficient approaches -1.00.

In order to determine significant separation between each of the three single elements in the profile, the t test was applied. Three t tests were required, and as a result, the degree to which achievers and underachievers differed on mean T-unit length, S-ratio, and K-ratio as they functioned independently was determined.

The t test was used in the between group comparisons on all the remaining language measurement variables.

Finally, the relative ability of each of the three variables in the complexity measure (mean T-unit length, S-ratio, K-ratio) to predict level of reading achievement was determined by a Point Biserial Correlation. Hunt (1964) performed a similar evaluation to determine the validity of four language measures as predictors of grade level. Two of his measures were mean T-unit length and subordination ratio (S-ratio, in this study). He found T-unit length to be the best predictor of the four, and subordination ratio as third best.

Discussion of Findings

Complexity Measure

When achievement was the only variable, the language of achievers was
significantly more complex at the .01 level (Similarity Coefficient of -0.474) as measured by the complexity measure. The application of the t test found no statistical significant differences at each point on the profile (mean τ-unit length, S-ratio, K-ratio). This was not an unexpected occurrence since the Cattell Index and the t test are independent of each other; i.e., the scores on one do not depend on the scores of the other.

When sex was the only variable, the two groups, males and females, were more alike than what could be expected by chance (.05 level). This suggests some interaction between sex and some other variable(s). When both sex and achievement were variables, the language of female achievers was significantly more complex than male achievers (.01), while that of male underachievers was significantly more complex than female underachievers (.01). According to these findings, underachieving boys had higher language scores than underachieving girls even though they were reading at almost three grade levels below those girls.

Socioeconomic status was important in all comparisons in which it was a variable. When SES was the only variable, the two groups—low SES and middle SES—differed to a great extent (-0.725; p < .01); the language of middle SES was more complex. When both SES and achievement were variables, middle SES subjects exhibited more syntactic complexity (p < .01) in both achievement groups. When SES and sex were variables, the SES groups were significantly different in favor of the middle SES within their own sex (.01). (Table 1 summarizes all comparisons on the complexity measure.)
TABLE 1

Coefficients ($r_p$ values) and Probabilities for Comparison Groups on the Complexity Measure
(mean T-unit length + S-ratio + K-ratio)

<table>
<thead>
<tr>
<th>Comparison Group</th>
<th>Coefficients ($r_p$)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Achievers and Underachievers</td>
<td>-0.474</td>
<td>0.01</td>
</tr>
<tr>
<td>Males and Females</td>
<td>0.553</td>
<td>0.05</td>
</tr>
<tr>
<td>Low SES Subjects and Mid SES Subjects</td>
<td>-0.725</td>
<td>0.01</td>
</tr>
<tr>
<td>Male Achievers and Female Achievers</td>
<td>-0.467</td>
<td>0.01</td>
</tr>
<tr>
<td>Male Underachievers and Female Underachievers</td>
<td>-0.557</td>
<td>0.01</td>
</tr>
<tr>
<td>Low SES Achievers and Mid SES Achievers</td>
<td>-0.513</td>
<td>0.01</td>
</tr>
<tr>
<td>Low SES Underachievers and Mid SES Underachievers</td>
<td>-0.590</td>
<td>0.01</td>
</tr>
<tr>
<td>Low SES Females and Mid SES Females</td>
<td>-0.488</td>
<td>0.01</td>
</tr>
<tr>
<td>Low SES Males and Mid SES Males</td>
<td>-0.560</td>
<td>0.01</td>
</tr>
</tbody>
</table>

To summarize, the complexity measure indicated that the oral language samples of achievers contained longer T-units, more subordinate clauses, and more kernel structures than did the language samples of the underachievers. On the other hand, these language variables, when examined separately, were not statistically more frequent in the language of achievers than in that of the underachievers.

Sex, as a single variable, did not make a difference in the length of the T-unit, amount of subordination, and number of kernel structures. However,
sex and level of achievement as co-variables appeared to interact. Among the achievers, girls had longer T-units, more subordination, and more kernel, or underlying structures; while among the underachievers, boys had more of these qualifiers in their language samples.

Socioeconomic status was a strong variable when acting as a single variable or pairing with level of achievement and sex. Middle SES subjects in each comparison group had more complex language than low SES in the same group.

**Selected Language Measures**

There were no statistically significant differences between achievers and underachievers, when compared with the t test, on any of the selected language structures compared in this section. The language samples of achievers did, however, contain more multi-clause T-units, long T-units, and kernel structures than did the samples of the underachievers. Long T-units occurred very few times in both groups. The number of subject-verb and subject-verb-object patterns exceeded that of other patterns in both the surface and kernel structures for all subjects. The difference between the frequencies of the subject-verb-adjective complement and the subject-verb patterns in the kernel structures was less than in the surface structures.

**Predictors of Level of Reading Achievement**

To determine the relative ability to predict level of reading achievement, the Point Biserial Correlation was applied to the mean T-unit length, S-ratio, and K-ratio. Correlation between the K-ratio and level of achievement was highest ($r_{pb} = 0.406; p < .05$). Mean T-unit length was next with a coefficient of $0.351 (p = .05)$. S-ratio correlated least with reading achievement with a coefficient of $0.320 (p > .05)$. 
The indication that the K-ratio is most predictive of reading achievement may substantiate the question Hunt (1964) posed regarding the insensitivity of the subordination ratio to reduced clauses. These reduced clauses (e.g., adjective clauses in the underlying structures) may have also suppressed the sensitivity of the T-unit length in measuring syntactic complexity.

Conclusions and Implications

Certain measures of syntactic maturity in oral language do appear to be positively correlated with reading comprehension achievement. This apparent correlation warrants more consideration by curriculum designers. Educators need to experiment with programs in which students are allowed to use language as much or more than they hear the language to test their influence on language maturity and reading comprehension.

Reports by research that many present-day reading programs are not adequate for boys seem to be supported by this study. The fact that underachieving boys had higher language scores even though their reading scores were far below underachieving girls merits attention. With these findings, one hesitates to conclude that lack of oral language maturity is the major factor contributing to unsuccessful reading experiences for boys.

Socioeconomic status appears to make a significant difference in both syntactic maturity of oral language and reading achievement. One might conclude, then, that educational programs must be aimed at preschool children in the lower socioeconomic segment of our population. This is already being done;

---

1 A "reduced clause" in this study refers to a former clause which lost its clausal state by way of deletion transformations (e.g., "the boy who was large" became "the large boy").
but in view of reports which question the effectiveness of these programs, perhaps curriculum and instruction should be reconsidered.

The embedding of ideas in language structures has been found more prevalent among achievers in reading comprehension. Achievers were not different from underachievers in the kinds of sentence structures used, i.e., the two groups put their language together in the same way (subj-verb, subj-verb-obj; etc.) but the language of achievers is more content-loaded. Will a curriculum that emphasizes such oral enrichment activities as role playing, storytelling, discussions of direct experiences, and making experience charts effect more substance in language structures?
APPENDIX

Data Analysis Slip

A-Z
(Student A, T-unit #2)

The old lady who owned the little dog
lived in a huge house.

(SV)

(Surface structure pattern)

(Kernel structures and their patterns)

1. The lady lived in a house.
   (SV)

2. The lady is old.
   (SVC)

3. The lady owns a dog.
   (SVO)

4. The dog is little.
   (SVC)

5. The house is huge.
   (SVC)

(T-unit)

Total # of kernel structures = 5

K-ratio = \frac{\text{Total # of T-units}}{\text{Total # of kernel structures}}

Note. The size of the slip was 5-by-8 inches, each having one T-unit on it.


Bromwich, R. M. *Developing the Language of Young Disadvantaged Children.* *Education Digest,* 1968, 34, 19-22.


Ruddell, R. B. Variation in Syntactical Language Development and Reading Comprehension Achievement of Selected First Grade Children. International Reading Association (Vistas in Reading), 1966, 11, 420-425.


