A CONTENT ANALYSIS OF THE

COMPREHENSIVE TESTS OF BASIC SKILLS READING TEST,

FORM Q, LEVEL 4

A project submitted in partial satisfaction of the requirements for the degree of Master of Arts in

Education

by

Barbara Grundy Scheibel

June, 1973
The project of Barbara Grundy Scheibel is approved:

California State University, Northridge

June, 1973
ACKNOWLEDGMENT

The author is indebted to Mrs. Christine Smith for her cooperation as a member of the Reading Committee. Appreciation is particularly expressed to Dr. Louise Grindstaff, Graduate Adviser, for her patience, guidance and encouragement in the development of this study.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th></th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I.</td>
<td>THE PROBLEM</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Justification for the Study</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Choice of Test under Study</td>
<td></td>
</tr>
<tr>
<td></td>
<td>The Problems of the Study</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Classification of Categories</td>
<td></td>
</tr>
<tr>
<td>II.</td>
<td>A REVIEW OF RELATED LITERATURE</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Introduction</td>
<td></td>
</tr>
<tr>
<td></td>
<td>The Nature of Reading Comprehension</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Passage-related Tests</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A Facet Design</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Subject Matter and Reading Comprehension</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Content Analysis</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Summary</td>
<td></td>
</tr>
<tr>
<td>III.</td>
<td>PROCEDURE</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>Method</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Classifications for the Analysis of Relationship between Reading Selections and Correct Answers</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Classifications for the Analysis of Relationships between Distractor Statements and the Reading Selections</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Classifications of Subject Matter</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Classification of Previous Knowledge</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Classification of Value Judgments</td>
<td></td>
</tr>
<tr>
<td>IV.</td>
<td>RESULTS</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>Relationship between the Reading Selections and Correct Responses</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Relationship of Distractor Statements to the Reading Selection, Question and Correct Answers</td>
<td></td>
</tr>
<tr>
<td></td>
<td>The Nature of the Subject Matter</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Previous Knowledge Required</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Value Judgments</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Summary of Findings</td>
<td></td>
</tr>
<tr>
<td>CHAPTER</td>
<td>SUMMARY, CONCLUSIONS AND RECOMMENDATIONS</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>V.</td>
<td>SUMMARY</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>Conclusions</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Recommendations for Further Study</td>
<td></td>
</tr>
<tr>
<td>REFERENCES</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Relationship Between Reading Selections and Correct Answers</td>
<td>17</td>
</tr>
<tr>
<td>2. Relationship of Distractor Statements to the Reading Selections, Questions and Correct Answers</td>
<td>20</td>
</tr>
<tr>
<td>3. Subject Matter of Reading Selections</td>
<td>23</td>
</tr>
<tr>
<td>4. Previous Knowledge Required</td>
<td>23</td>
</tr>
<tr>
<td>5. Value Judgments</td>
<td>24</td>
</tr>
</tbody>
</table>
ABSTRACT

A CONTENT ANALYSIS OF THE
COMPREHENSIVE TESTS OF BASIC SKILLS READING TEST,
FORM Q, LEVEL 4

by

Barbara Grundy Scheibel

Master of Arts in Education

June, 1973

A study of the Comprehensive Tests of Basic Skills, Form Q, Level 4, was undertaken to analyze the following relationships: the relationship between the reading selections and the correct answers; and the relationship between the distractor statements and other facets comprising the test item, specifically the reading selection, question and correct answer. In addition, the study analyzed the nature of the subject matter and the extent to which previous knowledge and value judgments were necessary. The analysis of the correct answers indicated that most of them required concept recognition. The distractors generally had a textual-grammatical-associative relationship with the selections, questions and correct answers. The subject matter was divided between non-academic selections, literature, language usage, science and humanities. Previous knowledge was required in over one-quarter of the questions. Value judgments could be made in over ten per cent of the questions. Because of interference of previous knowledge
and value judgments, combined with low-interest subject matter for the average student, the researcher concluded that the test was inadequate for non-academic students and particularly so for students of bi-lingual and minority cultures.
CHAPTER I

THE PROBLEM

Justification for the Study

For over fifty years the question of reading comprehension and measurement has been the subject of debate and research. Many of the early reading tests primarily measured intelligence and general knowledge. Educators developing new tests became more sophisticated in their attempts to write passage-related questions, but still, the influence of previous knowledge pervaded the tests. However, an even more basic problem perplexed educators: the possibility that reading comprehension was not a single entity but a group of separate skills which might be separately tested. Researchers have spent many years attempting to identify these skills and write tests which would evaluate performance of them.

Much of the recent research on reading comprehension tests indicates that high verbal ability and academic knowledge in such subjects as social studies and science will enable students to perform at a higher level in the tests than students deficient in these abilities. This suggests that reading comprehension tests are not measuring what it is intended they measure.

A further problem exists where value judgments are required, not comprehension alone. Although in recent years educators have become aware that not all students share the same middle-class values, there appears to be no allowance made for this in standardized tests.
Choice of Test under Study

Because of the many faults prevalent in reading tests, the researcher has chosen to analyze a test which is used throughout the state to measure the reading comprehension of ninth-grade students: the Comprehensive Tests of Basic Skills (CTBS) Reading Test, Form Q, Level 4. It is frequently used to determine whether students should be placed in remedial reading programs. Readers desiring a copy of the test should write to CTB/McGraw-Hill, Del Monte Research Park, Monterey, California, 93940.

According to the Examiner's Manual of the CTBS (1968), the forty-five questions of the reading comprehension test, all of which are multiple-choice, "measure the student's ability to comprehend the meaning of ideas by paraphrasing; to interpret what is read by identifying the main idea, perceiving relationships, drawing conclusions, and making inferences; to extend interpretation beyond stated information and recognize the author's intention."

The Problems of the Study

The present study sought to explore and analyze relationships between the reading selections and corresponding correct responses in order to determine the range of difficulty and extent of balance in the test, i.e., whether the correct answer called for word recognition, semantic differences or probable concepts.

Secondly, the present study sought to explore and analyze the relationships between the distractor statements (incorrect multiple choices) and the reading selections, questions and correct answers. Theoretically, the categorization of distractor statements would indicate to some extent the level of misunderstanding in an incorrect answer, i.e., whether the error was due to a lack of word recognition or a more complex mistake.
A second thrust of the study sought to answer the following questions:

1. What was the nature of the subject matter used for reading selections?
2. To what extent was previous knowledge necessary?
3. To what extent were value judgments called for?

In general, the purpose of the analysis was to determine to whatever extent possible the level of comprehension being measured, the quality of the distractor statements, and what obvious drawbacks to the test could be observed so that improvements in future tests could be made.

**Classification of Categories**

In order to answer the questions posed by this study, the categories established by Auerbach (1971) were used, with minor modification. The categories are fully described in Chapter III.
**CHAPTER II**

A REVIEW OF RELATED LITERATURE

**Introduction**

It has been due primarily to recent research in reading comprehension tests that this investigation was designed. The work of researchers such as Auerbach (1971), Kretschmer (1972), Schlesinger and Weiser (1970) have been vital in bringing to the researcher an understanding of the complexities involved in test analysis.

**The Nature of Reading Comprehension**

Long before current researchers started their exacting studies of the problem, Thorndike (1917) was one of the first to be concerned with the measurement of reading comprehension. He was aware that many test questions required previous knowledge rather than an understanding of the selection. He also realized that the construction of the reading passages, as well as the phrasing of the questions, could influence the student's answers. The complexity of measuring reading comprehension was becoming apparent.

Research in reading comprehension continued. Davis (1944) concluded in a series of tests that there were at least eight unique skills involved in reading comprehension. He identified them as:

1. Recalling word meaning
2. Inferring word meaning from context
3. Understanding explicitly stated content
4. Ability to select main idea
5. Following the structure of a passage
6. Ability to answer semantically paraphrased questions
7. Making inferences
8. Identifying the author's literary techniques
9. Recognizing the author's mood, tone and purpose

However, there has been considerable debate whether reading comprehension may be measured as a variety of factors or only one. Following Davis' study, Thurstone (1946) re-examined Davis' work using a different method of computation. His results indicated generally that there was one common factor. Hunt (1957) chose to re-evaluate six out of the nine factors which Davis found most significant. Using the same method of analysis as Thurstone, Hunt concluded that, at the most, two factors could be identified in comprehension, but basically most of comprehension would come under one general skill. Nevertheless, Hunt believed that additional study of item construction and student response would reveal a variety of skills not apparent at that time.

Passage-related Tests

Attempting to exclude all sources of information from reading tests except the passages themselves, Tuinman (1971) developed a series of tests which he identified as Test Acquisitional Information. This included three kinds of passages: TAI-P (Phantasy); TAI-LS (Lexical Substitution) substituted an artificial word for a key word in the passage; TAI-SA (Semantic Anomaly) substituted a word which would not ordinarily be used in that context. Although testees had no particular difficulty with the TAI-P and
Questioning the validity of reading tests, an experiment was made by Preston (1964), who believed that because of their general knowledge and the manner in which test questions were composed, students as a group would score better than chance on a reading comprehension test when presented with questions but not the reading passage which usually accompanied them. Thirty questions from the Cooperative English Test, Reading Comprehension, were changed only enough to clarify the questions before they were presented to the students, who continued with the rest of the test in the usual fashion. Seventy-seven per cent achieved better than chance on the thirty questions. The observed mean for the group was 8.34, with 6 being chance.

A Facet Design

Schlesinger and Weiser (1970), concerned with the relationship between the text and test question, developed a facet design for reading tests which concentrated on this element rather than on supposed skills involved in comprehension.

They felt that one of the major facets involved is whether a statement is explicit or implicit in the text. For example, two statements may be fairly explicit in the text, but to miss one of them in a question might mean only having skipped a minor detail; to miss the other might involve missing the whole understanding of the text because the information might be further implicitly imbedded in the entire passage. In the same way, a distractor statement which contradicts information in the text may only contradict a simple statement; another distractor statement may run contrary to the entire meaning of the passage. Therefore, distractor statements also become an important part of the facet design.
Another major facet is the testee's source of information. Since it is not unusual for the testee to have information on a subject gained outside of the text, either by formal or informal education, this factor may be of great importance.

Detailed analysis of the distractor statements was an important part of the facet design. The researchers divided the distractor statements into two parts: the constant (the part of the distractor which is repeated in each of the distractor statements) and the variable (which is different in each of the distractor statements). The analysis becomes three-part when the constant and the variable are considered together. A distractor statement might be chosen as the right answer even when it contradicts the text because part of it, either the constant or variable, is explicit in the text. Both the constant and the variable may be explicit in the text but not in conjunction with each other, thereby providing another explanation for an incorrect answer.

Schlesinger and Weiser believed that the use of a facet design such as this would do much to make the construction of questions and distractor statements a more exact and systematic process. In addition, it would enable parallel tests to be just that because the construction of questions and distractor statements for each test could include the same facet variables.

**Subject Matter and Reading Comprehension**

Another area of research has been the importance of subject matter in comprehension tests. Many classroom teachers have long agreed that some of their underachievers in reading showed startling results when the material they were reading was of sufficient interest.

Bernstein (1955) tested a group of thirty homogeneous I.Q. and
reading-ability students with a thirty-question comprehensive test based on one of the duller passages from The House of Seven Gables and an exciting story about a runaway boy, both passages having been rewritten so that they were approximately at the sixth grade reading level. An interest reading scale was also included in the test. Those students who indicated high interest in the story about the runaway boy answered the questions at a significantly higher comprehension level, as well as reading it at a faster rate.

Considering that the subject matter of standardized tests could have a strong bearing on the results of such tests, Kretschmer (1972) classified sixteen tests under eighteen types of content headings. The tests ranged from primary to grade fourteen and included many of the well-known standardized reading tests such as the Nelson, Gates-MacGinitie, Iowa and California Reading Tests.

The categories were:

- Adventure Holidays
- Agriculture Humor
- Animals Money
- Circus - Theater Outdoors (description of landscapes)
- Cowboys Personal incidents
- Family Poems - Sayings
- Geography Science
- Health - Safety Sports
- History Vocational

The results of his classification indicated that 33.5 per cent of all passages were in Science and Animals; 25 per cent were in History, Geography and Agriculture. Thus more than fifty per cent of the items
represented only five categories, all of which reflected school studies. Therefore, not only might some students show lower scores because of lack of interest, but other students with high interest in academic studies might be reflecting general knowledge rather than reading ability. The subjects in which non-academic students might show considerable interest represented only a small number of passages: Family, 3 per cent; Sports, 1.5 per cent; Humor, .5 per cent.

In addition, Kretschmer felt that many of the passages were biased toward a middle-class culture, a serious complaint which has been made about standardized tests by many educators.

Content Analysis

Analyzing several standardized reading tests, Livingston (1972) concluded that they were more a test of academic subject matter than reading. One of the complaints about the Stanford Achievement Test was that it confused synonyms with inference; another complaint was that the test gave signals for inferences by the use of words like "probably," "apparently," and "most likely." Analyzing the Metropolitan Reading Test, Livingston concluded that in one passage with six questions, five of the questions called for synonyms or literal comprehension and one question required an inference to be made.

An extensive analysis of reading comprehension tests was made by Auerbach (1971) on the basis of the subject matter, the readability, and the various relationships between the text, questions and choices. The tests under study were the California Achievement Test, 1963, Form W; the Gates-MacGinitie Reading Test, 1965, 1969, and the Comprehension subtest, GMRT; and the Stanford Achievement Test, 1964-65, Form X.
The division into subject matter was: (1) riddle (for elementary), (2) story (common occurrences, not academic), (3) language (usage or literature), (4) math, (5) social studies (history and geography), (6) social science (anthropology, psychology, sociology), (7) science, (8) humanities (philosophy, art and theology). Tests indicated a correlation between reading comprehension grades and social studies and science grades. Auerbach felt that a knowledge of science and social studies was required for success in the reading test.

A unique aspect of the analysis was Auerbach's categories for the relationship between the way the information was presented in the selection and in the correct answer:

recognition - answer requires recognizing the word that is the same as in the passage
contextual paraphrase - choosing the right word in a differently phrased statement
grammatical paraphrase - recognizing a word that is grammatically changed from passage to question; e.g., alter, alteration
semantic paraphrase - recognizing a word semantically changed; e.g., amazed, astonished

definite concept - recognizing common concepts; e.g., the children were working problems in addition, subtraction and multiplication; they were doing arithmetic
probable concepts - recognizing probable concepts; e.g., they saw tigers, lions and bears being led into the big tent; they were at the circus
language concepts - recognizing cliches, colloquialisms and relatives

previous knowledge - source of information other than test passage

word-picture matching (elementary)

Auerbach's analysis in the grade 9-14 range revealed 21 per cent of the correct answers to be contextual paraphrase; 24 per cent semantic paraphrase; 18 per cent probable concepts and 18 per cent previous knowledge (pp. 86-88).

The relationship of distractors to the passage, the question and to the correct answers were categorized as follows:

other (no association at all)
grammatical (fits grammatically)
associative (a slight semantic or conceptual relationship)
associative-grammatical
categorical (fits the same general category as correct choice)
categorical-grammatical
textual - word used in reading selection
textual-grammatical
textual-associative
textual-associative-grammatical
textual-categorical
textual-categorical-grammatical

Twenty-five per cent of the distractors had only a grammatical relationship; thirty-three per cent had an associative-grammatical relationship (p. 89).
There were various types of inappropriate distractors which could easily be detected, such as those with incorrect grammar, incorrect parts of speech, semantic anomalies (e.g., Pete is a house), and a lack of reality (e.g., the children were empty).

Summary

Although studies have not been able to determine conclusively what specific reading skills are being measured in reading comprehension tests, content analysis provides some clarification of what the test is measuring. Schlesinger and Weiser (1970) were particularly concerned with the relationship of the correct answers to the reading selection and the relationship of the distractor statements to the reading selection, question and correct answer. They believed that the source of information, such as previous knowledge, contributed considerably to an understanding of the testee's answer. Kretschmer (1972) believed that the subject matter had a strong influence on the testee's success or failure; therefore it was important to analyze the subject matter of a reading test. Auerbach's (1971) research encompassed all of these studies in her construction of categories for the analysis and comparison of a number of reading tests.
CHAPTER III

PROCEDURE

Method

The researcher sought to analyze the CTBS Reading Test in order to discover the following:

1. The relationships between the reading selections and the correct answers
2. The relationships between the distractor statements and the reading selections, questions, and correct answers
3. The nature of the subject matter
4. The extent to which previous knowledge was necessary
5. The extent to which value judgments were called for.

The following categories established by Auerbach (1971) were used to find the answers to the first three problems listed above, with only minor modifications. In addition, a separate tabulation was made to analyze problems 4 and 5. The results of the analysis were computed in percentages and reported in the tables found in Chapter IV.

Classifications for the Analysis of Relationship between Reading Selections and Correct Answers

In order to analyze relationships between the reading selections and corresponding correct responses, the following classifications were established and a chart prepared showing the percentage of correct answers in
each category:
   a. Recognition - answer requires recognizing a word that
      is the same as in the selection
   b. Contextual paraphrase - choosing the right word in a
differently phrased statement
   c. Grammatical paraphrase - recognizing a word that is
gramatically changed from passage to
question; e.g., alter, alteration
   d. Semantic paraphrase - recognizing a synonym of a word
used in the reading selection
   e. Definite concept - recognizing common concepts
   f. Probable concepts - recognizing probable concepts by
drawing inferences
   g. Language concepts - recognizing cliches, colloquialisms
and idioms

A correct answer which called for previous knowledge or a value
judgment would be included in one of the above categories as well as being
recorded separately in the appropriate previous knowledge or value judgment
category.

Classifications for the Analysis of
Relationships between Distractor Statements and the Reading Selections

In order to analyze the relationships between the distractor state-
ments (incorrect multiple choices) and the reading selections, questions, and
correct answers, the following classifications were established and a table
prepared showing the relationships:
a. Other - no association to the selection, question or correct answer
b. Grammatical - fits the question grammatically
c. Associative - a slight semantic or conceptual relationship to the reading selection
d. Associative-grammatical
e. Categorical - fits the same general category as correct choice
f. Categorical-grammatical
g. Textual - word or phrase used in reading selection.
h. Textual-grammatical
i. Textual-associative-grammatical
j. Textual-associative
k. Textual-categorical
l. Textual-categorical-grammatical

Classifications of Subject Matter

In order to analyze the nature of the subject matter used in the reading selections, the following classifications were established and a chart prepared showing the percentage of passages in each category:

a. Non-academic (pertaining to every-day life)
b. Language Usage
c. Literature
d. Math
e. Social studies (history and geography)
f. Social science (anthropology, sociology, psychology)
g. Science
Classification of Previous Knowledge

In order to determine the extent to which previous knowledge was necessary, a separate record was kept. The basis for judgment was whether or not the answer requested was likely to be part of the curriculum or in normal usage if the information needed for the answer was not within the reading selection. In several cases the vocabulary in the multiple choice was inconsistent with the vocabulary level needed for comprehension of the reading selection. In such cases the question was classified as requiring previous knowledge.

Classification of Value Judgments

In order to determine the extent to which value judgments could be made, a separate record was kept. The basis for judgment was whether an alternate opinion could be an acceptable answer. The assumption was made that the researcher with a middle-class value system would be able to discern which questions called for value judgments which might not be common to all classes and cultures in our society.
CHAPTER IV

RESULTS

Relationship between the Reading Selections and Correct Responses

The purpose of this classification was to determine what depth of comprehension was being measured and how varied was the type of comprehension required (see Table 1).

TABLE 1

RELATIONSHIP BETWEEN READING SELECTIONS AND CORRECT ANSWERS

<table>
<thead>
<tr>
<th>Category</th>
<th>Question Numbers</th>
<th>Percentage of Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recognition</td>
<td>7, 36</td>
<td>4.4</td>
</tr>
<tr>
<td>Contextual Paraphrase</td>
<td></td>
<td>0.0</td>
</tr>
<tr>
<td>Grammatical Paraphrase</td>
<td></td>
<td>0.0</td>
</tr>
<tr>
<td>Semantic Paraphrase</td>
<td>12, 13, 19, 33, 37</td>
<td>11.1</td>
</tr>
<tr>
<td>Definite Concept</td>
<td>14, 43</td>
<td>4.4</td>
</tr>
<tr>
<td>Probable Concept</td>
<td>1, 2, 3, 4, 5, 6, 8, 9, 10, 11, 15, 16, 17, 18, 20, 28, 29, 30, 34, 35, 38, 39, 40, 41, 42, 44, 45</td>
<td>60.0</td>
</tr>
<tr>
<td>Language Concept</td>
<td>21, 22, 23, 24, 25, 26, 27, 31, 32</td>
<td>20.0</td>
</tr>
</tbody>
</table>
The findings indicate that sixty per cent of the answers required the
comprehension of a probable concept. Such comprehension requires greater
depth of understanding than the other types of comprehension measured in the
test, with the possible exception of language concepts. Many of these
questions were well-written for this purpose, but, according to this study,
thirty per cent of the probable concept answers required either value judg-
ments or previous knowledge. A more detailed explanation is made in this
chapter under the headings Value Judgments and Previous Knowledge.

Only one answer might have been considered inaccurate. The
answer to Question 30, "the twinkle of stars," might have been considered in-
correct by students knowledgeable in astronomy because stars do not twinkle
but only appear to because of atmospheric conditions; such students might
have chosen "the color of stars" as the most likely correct answer.

Twenty per cent of the answers in the test were classified as lan-
guage concepts; however over seventy-five per cent of them required previous
knowledge or value judgments, as is explained under those headings else-
where in this chapter.

Eleven per cent of the questions were semantic paraphrases, not
requiring an extensive vocabulary. However, one of the questions was classi-
fied as requiring previous knowledge because it required knowing the meaning
of the word "aeons."

Approximately four per cent of the questions required only recog-
nition of a word used in the reading selection as the correct answer.

Approximately four per cent of the questions required only the com-
prehension of a definite concept; that is, there was no inference to be drawn
because the information was set forth in the reading passage.
Relationship of Distractor Statements
to the Reading Selection, Question
and Correct Answers

The chief result of analyzing the distractor statements was to deter-
mine the level of comprehension of the reading selections required by the
distractor statements. As is apparent in Table 2, there were not too many
obvious, easy distractors. All of the distractors had something to do with
the reading selections and were in grammatical agreement. Only seven per
cent were considered to have no relationship other than grammatical.
Another seven per cent were textual-grammatical, which indicated the
repetition in the distractor of a word in the reading selection, but no other
relationship.

Although the textual-associative-grammatical classification included
almost fifty per cent of the distractors, not too much significance can be
placed on that fact because it was necessary to include many textual relation-
ships which were not key words and which might have made little impression
on the testee because of the frequency of their usage in the reading selection.
If more of these distractors had been in the associative-grammatical classi-
fication, where they might better have belonged for the purpose of analysis,
there would have been a more balanced picture of the results. In any event,
most of the distractors were associative.

It would appear that most of the questions had distractors which could
be seriously considered and which required reasonable comprehension of the
selection before they could be eliminated as the correct answers.

The Nature of the Subject Matter

The subject matter of the reading passages was fairly well-balanced
(see Table 3) among five categories: non-academic, literature, language
<table>
<thead>
<tr>
<th>Category</th>
<th>Distractor Numbers</th>
<th>Percentage of Distractors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other</td>
<td>None</td>
<td>0</td>
</tr>
<tr>
<td>Associative</td>
<td>None</td>
<td>0</td>
</tr>
<tr>
<td>Categorical</td>
<td>None</td>
<td>0</td>
</tr>
<tr>
<td>Textual</td>
<td>None</td>
<td>0</td>
</tr>
<tr>
<td>Textual-associative</td>
<td>None</td>
<td>0</td>
</tr>
<tr>
<td>Textual-categorical</td>
<td>None</td>
<td>0</td>
</tr>
</tbody>
</table>
usage, science and humanities (philosophy), with no reading selections in mathematics, social studies or social sciences. However, a closer analysis reveals the following:

One of the non-academic reading selections was a business letter concerning the qualities of a vacuum cleaner. Only one reading selection related to students' daily lives.

There were two reading selections in literature. One was a fairly complex poem by Sara Teasdale; the other a story based on ancient Greek history.

There were seven questions classified under language usage. Three of these required the definition of idiomatic expressions, possibly considered cliches, which are rarely used today by anybody, much less the modern teenager. One reading selection in this classification required the classification by derivation of four neologisms.

There was one reading selection in science, which concerned the eating habits of the shrew.

There was one reading selection in humanities (philosophy) which concerned the presence of soul in all living things.

Previous Knowledge Required

Because no specific research was made regarding the vocabulary used in the test, the researcher has accepted most of the words used in the test as conceivably within the range of average students. However, because comprehension of the reading selection should be the main purpose of the test, the vocabulary used in the multiple choices should not be of such difficulty that it prevents the expression of such comprehension.

For this reason the words used in the multiple choice answers for
Questions 15, 16, and 20 are classified as requiring previous knowledge. Question 15 required the testee to describe a letter as being: a) "a letter of complaint," b) "a personal letter," c) "an invitation," or d) "an adjustment letter." Question 16 required the testee to identify the tone of the letter as being: a) "defensive," b) "conciliatory," c) "neutral," or d) "belligerent." Question 20 required the testee to identify the recipient's reaction to the letter as a) "vengeance," b) "appreciation," c) "indignation," or d) "satisfaction."

Questions 30 and 33 are also classified as requiring previous knowledge because of the use of "myriads" and "aeons," neither of which could be defined by context clues.

Questions 21, 22 and 23 required the definition of the following idiomatic expressions: "He sets a good table," "Good words relieve worries and soften the heart," and "He is on the horns of a dilemma." These required previous knowledge, but few of today's students would find these expressions in use at home or in the classroom.

Although the concept of neologisms is explained in the reading selection, Questions 24, 25, 26 and 27 require previous knowledge because the student must describe the derivation of words which are not mentioned elsewhere in the reading selection; thus, the student would have to know that "mackintosh" is a proper name, while "frigidaire" is a trade name (see Table 4).

Value Judgments

Questions 3, 4, and 5 required value judgments. For example: A new teacher has been assigned to study hall. He is able to keep the group silent with his cold joyless manner. The passage states, "Never before had we been so--so good. The silence was noisier than all of yesterday's
### TABLE 3

**SUBJECT MATTER OF READING SELECTIONS**

<table>
<thead>
<tr>
<th>Category</th>
<th>Question Numbers</th>
<th>Question Percentage</th>
<th>No. of Selections</th>
<th>Selection Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-academic</td>
<td>1-6</td>
<td>27</td>
<td>2</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>15-20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Language Usage</td>
<td>21-23</td>
<td>15</td>
<td>2</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>24-27</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Literature</td>
<td>28-33</td>
<td>25</td>
<td>2</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>41-45</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Math</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Social Studies</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Social Science</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Science</td>
<td>7-14</td>
<td>18</td>
<td>1</td>
<td>12.5</td>
</tr>
<tr>
<td>Humanities (Philosophy)</td>
<td>34-40</td>
<td>15</td>
<td>1</td>
<td>12.5</td>
</tr>
</tbody>
</table>

### TABLE 4

**PREVIOUS KNOWLEDGE REQUIRED**

<table>
<thead>
<tr>
<th>Question No.</th>
<th>Percentage of Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>15, 16, 20, 21, 22, 23, 24, 25, 26, 27, 30, 33</td>
<td>27</td>
</tr>
</tbody>
</table>
A non-conforming testee might easily have chosen the word "disgust" or "disappointment" to indicate his feelings for the use of "good." On the other hand, a highly authoritarian testee might have chosen the word "pleasure." Based on this same incident, Question 4 requires the testee to judge the previous teacher as either "ineffective," "forceful," "Unfair," or "humorous."

Question 39 requires the testee to indicate in which expression "the author has introduced a touch of humor." Since the correct answer is rather simple humor, it would be possible for a sophisticated student to be more amused by "an animal is any one of our fellow beings," than by "in its crickety little way." The other choices were equally subjective (see Table 5).

### TABLE 5

<table>
<thead>
<tr>
<th>Question No.</th>
<th>Percentage of Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>3, 4, 5, 16, 39</td>
<td>11</td>
</tr>
</tbody>
</table>

Summary of Findings

The results of the study indicated the following:

1. Most of the correct answers required recognition of concepts by drawing inferences, but the value of the questions was often negated because the answer required previous knowledge or value judgment.

2. The distractor statements in all questions were related grammatically to the questions and generally had an associative relationship to the
reading selection. More than half of the choices included words used in the selection. The distractors generally required reading skill beyond mere word recognition.

3. The subject matter was balanced between literature, English usage, science, humanities and non-academic subjects.

4. Previous knowledge was required in over one-quarter of the questions, primarily in the usage section. Several questions which might have been understood by the average student used vocabulary choices for the answers which might have prevented the testee from displaying his comprehension of the selection.

5. Over ten per cent of the questions required answers which were deemed value judgments for which alternative answers could have been made. This occurred primarily in the selection which probably would have been the most interesting to non-academic students because of its subject matter.
CHAPTER V

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

Summary

Because of the uncertainties as to what reading comprehension tests actually measure, this study was undertaken to get a clearer view of the contents of a specific reading test, the CTBS Reading Test, Form Q, Level 4. The study sought to explore the following:

1. Relationships between the reading selections and corresponding correct answers.
2. The relationships between the distractor statements (incorrect multiple choices) and the reading selections, questions and correct answers.
3. The nature of the subject matter.
4. The extent to which previous knowledge was necessary.
5. The extent to which value judgments were called for.

Previous research concerning reading selection, question, correct answer and distractor statement relationships, as well as source of information, had been done by Schlesinger and Weiser (1970). A study on subject matter of reading tests was made by Kretschmer (1972). Auerbach’s (1971) study set up categories to analyze these relationships in a number of standardized reading tests. Categories established in this study were based primarily on those of Auerbach.

The results of the study indicated:
1. Most of the correct answers required recognition of probable concepts.

2. The distractor statements were all related grammatically; they generally had an associative relationship to the reading selection. Although more than half of the choices included words used in the reading selection, they generally required reading skills beyond mere word recognition.

3. The subject matter was balanced between literature, usage, science, humanities and non-academic subjects.

4. Previous knowledge was required in over one-quarter of the questions, especially in the English usage category. In some cases the vocabulary in the multiple choice selections were more difficult than the reading selection, thereby preventing the testee from displaying his comprehension of the selection.

5. Over ten per cent of the questions were classified as requiring value judgments. This occurred primarily in the selection which probably would have been the most interesting to non-academic students because of its subject matter.

Conclusions

It was the researcher's conclusion that almost every reading selection in the test was deficient, in one way or another, as a means of measuring comprehension for many students:

Those not particularly interested in academic subjects would probably be very disinterested in many of the reading selections.

Bi-lingual and bi-dialectical students might be especially handicapped by the English usage section, which called for previous knowledge not likely to be in their possession.
The literature category, consisting of a difficult poem and a tale from ancient Greek history, would not be a fair test for students who have been reading primarily about young people like themselves and contemporary life, a popular theme used in working with average or below-average readers. While poetry is certainly an important part of literature, there are many teachers who do not introduce other than the simplest poems unless they have academic classes. Many teachers themselves are not comfortable with poetry and therefore avoid it. While there is a place for poetry in a reading test, the one selection carried too much weight for a test with only forty-five questions.

The non-academic selections, which comprised over a quarter of the total score, were marred by the need for value judgments or previous knowledge.

By analyzing the various facets of the test, the researcher was able to make a more discerning judgment on the value of the test as a measurement of reading comprehension. While it was apparent that many of the questions did measure the student's ability to draw conclusions and make inferences, many others were not able to do this accurately because of the previous knowledge needed and value judgments involved. In addition, if we may rely on other studies which have indicated that there is a correlation between the subject matter and test success, it is very possible that the test under study will not give a fair evaluation of all kinds of students. It therefore would seem to the researcher that there is value in making such a content analysis in order to learn a great deal more about a test than one can learn from the test manual furnished by the publisher.
Recommendation for Further Study

One reading skill which was not measured by the study but which became apparent to the researcher as having an influence on comprehension was the complexity of sentence structure. This was particularly noticeable in the science selection (CTBS, p. 13) and the literature selection about ancient Greece (CTBS, p. 14). The latter selection had a Biblical rhythm which might complicate comprehension for anyone not familiar with this style. While there is nothing wrong in testing understanding of complex sentence structure, the test should be so structured that administrators of the test are aware that this skill requirement is present and may have a strong influence on the level of comprehension.

An additional deterrent to the measurement of comprehension is the fact that the CTBS Reading Test is a timed test. The instructions for the test specify thirty minutes for completion of the test, with an additional five minutes allowed for giving instructions. The slow student who does not finish the test in the prescribed time will show a poorer comprehension score when he may actually be lacking in speed. Perhaps further study could investigate the effect of this factor on test scores.

The result of this study indicates to the researcher that comprehension measurement is still an uncertain process primarily because of the vast difference in experience which each individual brings to a test. Previous knowledge, value judgments and interest contribute so heavily to the results of the test that it becomes questionable whether the test provides any valid measurement.

As reading becomes less of a necessity for entertainment and excitement, due to other communication media, it becomes apparent that only those who read well enough to enjoy reading a great deal will develop enough
background and interest in words to be successful using the kind of test represented by the GTBS Reading Test. Not only will non-academic, average students be discouraged by their inability to perform well, many students may develop a sense of alienation from the institution which tests him with materials so foreign to his daily activities.

The major recommendation of this study is that greater effort should be made to develop tests which are a practical reflection of what is being taught in the non-academic, average classroom. Students who must constantly struggle for improvement may be psychologically damaged by a test that is to some degree alien to them. The harm is done to the student not only by feelings of frustration, but also through the resultant inaccurate score which, written into his records, encourages teachers to make misguided judgments.
REFERENCES


Davis, F. B. "Fundamental Factors of Comprehension in Reading," Psychometrika, IX, No. 3 (September, 1944), 185-197.


Thurstone, L. L. "Note on a Reanalysis of Davis Reading Tests," Psychometrika, XI (September, 1946), 185-188.