LANGHOLZ, Armin Paul, 1929-
A STUDY OF THE RELATIONSHIP OF LISTENING TEST SCORES TO TEST ITEM DIFFICULTY.
The Ohio State University, Ph.D., 1965
Speech-Theater

University Microfilms, Inc., Ann Arbor, Michigan
A STUDY OF THE RELATIONSHIP OF LISTENING TEST SCORES TO TEST ITEM DIFFICULTY

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Armin Paul Langholz, B.S. in Ed., A.M.

* * * * * * *

The Ohio State University
1965

Approved by

Franklin H. Hower
Adviser
Department of Speech
ACKNOWLEDGMENTS

Appreciation is expressed to Dr. Franklin H. Knower for his wise counsel and patient encouragement during the entire course of the investigation.

Grateful acknowledgment is made to Dr. Keith Brooks and Dr. Richard Mall for reading the manuscript and for helpful suggestions during its revisions, to Dr. Wallace Fotheringham for his assistance in the statistical analysis of the data, and to my colleagues Dr. Harold Edmondson and Dr. Thomas Ludlum for their cooperation in the administration of test materials to the college freshmen who served as subjects for the study.
VITA

June 25, 1929  Born - St. Paul, Minnesota

1951 .......... B.S. in Ed., Capital University, Columbus, Ohio

1955 ......... M.A., The Ohio State University, Columbus, Ohio

1955-1965 .. Instructor, Assistant Professor of Speech
Capital University, Columbus, Ohio

FIELDS OF STUDY

Major Field: Speech

Studies in General Communication. Professors Franklin H. Knower
and Wallace C. Fotheringham

Studies in Radio and Television. Professors Harrison B. Summers
and Richard M. Mall

Studies in Public Address. Professors W. Hayes Yeager and Paul
Carmack
# CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>What is listening?</td>
<td></td>
</tr>
<tr>
<td>The importance of listening</td>
<td></td>
</tr>
<tr>
<td>The importance of listening measurement</td>
<td></td>
</tr>
<tr>
<td>The problem</td>
<td></td>
</tr>
<tr>
<td>Organization of the dissertation</td>
<td></td>
</tr>
<tr>
<td>II. REVIEW OF THE LITERATURE</td>
<td>14</td>
</tr>
<tr>
<td>Introduction</td>
<td></td>
</tr>
<tr>
<td>Listening research landmarks</td>
<td></td>
</tr>
<tr>
<td>The construction of listening tests</td>
<td></td>
</tr>
<tr>
<td>The development of the Brown-Carlsen Listening Comprehension Test</td>
<td></td>
</tr>
<tr>
<td>Question refinement techniques</td>
<td></td>
</tr>
<tr>
<td>III. PROCEDURE</td>
<td>29</td>
</tr>
<tr>
<td>The selection of a standardized test of listening comprehension</td>
<td></td>
</tr>
<tr>
<td>The analysis of the Brown-Carlsen test for the application of question refinement techniques</td>
<td></td>
</tr>
<tr>
<td>A summary of question changes for the experimental test form</td>
<td></td>
</tr>
<tr>
<td>The method of administering and scoring the experimental test form</td>
<td></td>
</tr>
<tr>
<td>IV. ANALYSIS OF THE DATA</td>
<td>61</td>
</tr>
<tr>
<td>Total test results</td>
<td></td>
</tr>
<tr>
<td>Revised test items results</td>
<td></td>
</tr>
<tr>
<td>V. SUMMARY AND CONCLUSIONS</td>
<td>79</td>
</tr>
<tr>
<td>Summary of data analysis</td>
<td></td>
</tr>
<tr>
<td>Conclusions</td>
<td></td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>85</td>
</tr>
</tbody>
</table>
TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Validity and Difficulty Data for Each of the 47 Revised Test Items in the Experimental Test Form</td>
<td>67</td>
</tr>
<tr>
<td>2. Validity and Difficulty Data for Revised Items in the Experimental Test Form</td>
<td>68</td>
</tr>
</tbody>
</table>
CHAPTER I

INTRODUCTION

About 1450, Johann Gutenberg of Mainz, Germany, originated a practical method of printing. By bringing together his movable types, the paper and ink of European block printers, and a wooden bed-and-platen press, he introduced a unique manner to produce books in quantity. His genius did away with laborious hand copying, increased the circulation of books and knowledge, and contributed to a new era in which primacy was given to the reading-writing process.

Now, after more than four centuries of increasing print-mindedness, of measuring literacy in terms of writing and reading, the twentieth century has witnessed the re-establishment of the importance of the speaking-listening process. With the advent of commercial radio broadcasting in 1920 and the rapid growth of television following World War II, an increasing awareness of the speaking-listening process developed. John Dolman's significant observation about radio can be applied to television and the moving pictures as well. He said, "The radio has taught us to be more critical and aware of our speech behavior and the speech behavior of others, and, more particularly, of the role of listeners in discriminating certain speech attitudes."\(^1\) To be literate in the twentieth century one must be able to speak as well as

\(^1\)John Dolman, "From the Listener's Point of View," Quarterly Journal of Speech, XX (1934), 203.
write and be able to listen efficiently as well as read efficiently.\(^2\)

The four skills of speaking, writing, listening, and reading are generally considered to be the basic elements involved in the process of communication. Speaking and writing are classified as expressive skills and listening and reading as receptive or assimilative skills. When the assimilative skills, the basic media of learning, are not adequately developed, the entire educational process suffers. Weaver and Strausbaugh refer to the listener or the receiver as "the pivotal point in the communicative process. It is what happens in him that determines the effect of the communicative act. The sender can only send, and he may never know the effect of his message."\(^3\)

**What is listening?**—The preceding statement suggests that listening is something more than hearing. Listening goes beyond the channeling of sounds, signals, words, vocalized ideas and information through nature's intricate and sensitive system known as the ear. Haberland calls listening "the conscious act of an attentive person hearing language symbols."\(^4\) Johnson says listening is "the ability to understand and respond effectively to oral communication."\(^5\) Baird and Knower refer to listening as "an active process of receiving and interpreting


messages ... (it) has to do with perception, comprehension, and other
mental activities." Barbe and Meyer include the relating of spoken
language symbols to past experiences and future courses of study in
their definition of listening. Don Brown, who coined the word auding
as a substitute for listening, defines this new term as "the process of
hearing, listening to, recognizing, and interpreting the spoken lan-
guage."®

Fessenden's theory on the levels of listening analyzes the se-
quential steps involved in the listening process and includes much of
what has been presented in the above mentioned definitions with the
addition of several new concepts. The levels of listening, which Fes-
senden states are in constant change and flux, include: (1) isolation
of sounds, ideas, facts, etc., with no evaluation or analysis, (2) iden-
tification, in which meaning is given to those aspects isolated,
(3) integration of what we hear with past experiences, (4) inspection of
the new as evaluation begins, (5) interpretation, where concern with the
subtle implications of the idea develops, (6) interpolation, in which
the listener supplies in part that which the speaker cannot supply, and
(7) introspection, where the listener notes the effect of that which he

6A. Craig Baird and Franklin H. Knower, General Speech (New York:

7Walter Barbe and Robert Meyer, "Developing Listening Ability in
Children," Elementary English, XXXI (February, 1954), 82.

8John Caffrey, "Auding," Review of Educational Research, XXV
(April, 1955), 121.
hears upon him and also the effect that this knowing how he is being affected affects him.\footnote{9}

Fessenden's theory underscores the point that listening is an amalgamation of physical and mental events in which meaning is attached to aural symbols.\footnote{10}

The importance of listening.--As early as 1929 Rankin pointed out that "listening is the most frequently used language activity, the average adult spending approximately three times as much time listening as reading." This statistic lends support to Wendell Johnson's declaration that listening plays the key role in human relations.\footnote{12}

Within the educational climate students spend most of their time listening. Wilt's elementary school investigation indicated that the median percentage of school time spent in listening by pupils in grades one through seven is 57.5 per cent.\footnote{13} A careful survey in a small women's college revealed that, on the average, students spend 42 per cent of their waking hours listening to other people and only 25 per cent of their time talking. Thirty-eight per cent of the women who observed and


\footnote{11}Paul T. Rankin, "Listening Ability: Its Importance, Measurement, and Development," Chicago Schools Journal, XII (June, 1930), 177.

\footnote{12}Wendell Johnson, "Do You Know How to Listen?" ETC, VII (Autumn, 1949), 4.

\footnote{13}Miriam Wilt, "A Study of Teacher Awareness of Listening as a Factor in Elementary Education," Journal of Educational Research, XLIII (1950), 626-636.
timed their own communicative behavior for this survey believed that listening is more important than reading, while only 18 per cent believed that reading is more important than listening. With listening occupying such a large percentage of a student's school time, it becomes apparent that learning and scholastic achievement are dependent upon efficient listening. The Commission on the English Curriculum concluded that pupils from pre-school through college learn more frequently by listening than by other means. McClendon, in his experimental investigation of listening, reports a positive correlation between listening comprehension and scholastic aptitude, and Blewett found that scholastic achievement depends almost equally on the ability to listen and the ability to read.

Many American teacher training institutions are giving due recognition to the importance of listening by providing listening units in special teaching methods courses taken by prospective teachers. Markgraf solicited information from educators in Speech, English, Education, and Elementary Education Departments in the 411 member institutions of the American Association of Colleges for Teacher Education. All but five of the institutions responded and it was learned that of the 670 special


16 Paul I. McClendon, "An Experimental Study of the Effects of Speaking Rate Upon Listenability," Speech Monographs, XXIV (June, 1957), 90.

teaching method courses offered by the member colleges, 44.5 per cent included a unit on the methods of teaching listening. In 38.7 per cent of the special teaching method courses, practice teachers were given the opportunity to observe the teaching of listening.18

On the graduate level of education, research done in the field of communication continues to give a constantly increasing degree of attention to the subject of listening. Duker discovered that a substantial proportion of such research is found in doctoral work. In the 1920's two doctoral dissertations were written on the subject of listening. The following decade witnessed the writing of five dissertations and eleven more doctoral studies were completed in the 1940's. A very impressive volume of listening research was compiled in the 1950's when 88 candidates for the doctoral degree presented the results of investigations on the subject. Research continues in the 1960's. Already more than 22 dissertations have been written.19

A review by Duker of the number of master's theses on the subject of listening indicates a similar pattern of increasing interest and investigation. Two master's theses were written in the 1930's, eleven in the 1940's and 74 in the 1950's. Thus far in the 1960's, more than twenty master's studies have been completed.20


The information resulting from the numerous listening research studies is being utilized by textbook authors in the field of oral communication to inform students of the importance of listening in the communicative act. An examination of representative college speech texts reveals that authors point to efficient listening as one of the basic skills sought in a study of oral communication. Entire chapters on listening are included in textbooks written by such authors as Anderson, Lewis, and Murray; Baird and Knowler; Brigance; Buehler and Linkugel; Capp; McBurney and Wrange; Smith; and Weaver and Ness, to list just a few. In each instance, the authors call attention to the importance of listening as they analyze the nature of the listening process and point out barriers to and prerequisites for effective listening. Irvin's conclusion that an individual "will be better or worse in


23 Brigance, op. cit., pp. 86-96.


proportion to his skill in listening" is the general thesis of the chapters on listening found in representative college texts.

The importance of listening measurement.--The measurement and evaluation of the communicative skill which occupies a major portion of an individual's waking hours and which plays a key role in the learning process is a vital concern of educators. Through suitable measuring techniques a student's listening comprehension skills may be diagnosed and the results employed in the application of a training program designed to improve listening comprehension efficiency. Also, measurement tools provide information about the listening process itself.

Informal teacher-made listening tests and evaluations designed to fit a specific classroom situation can be very useful. However, the more formal standardized tests have certain obvious advantages. They enable a teacher to compare her students with those making up a sizable representative norm group. They help the teacher interpret results by providing percentile ranks and standard scores. They have a known reliability, validity, and difficulty, and a predetermined relationship with other test instruments, established through research. Ease of administration and scoring are also matters of importance to the busy teacher. 30

The development of listening tests for the elementary, high school, and college levels of education has been an integral part of


communication research since 1945 (see Chapter II). From the development of these experimental tests several standardized tests have resulted which are used to good advantage in numerous schools and institutions across the United States. Keller, in his 1960 review of major findings in listening,\(^31\) refers to the standardization and continued use and refinement of the Brown-Carlsen Listening Comprehension Test\(^32\) and the Sequential Tests of Educational Progress--Listening\(^33\) as significant steps in the history of listening research. Duker lends his endorsement to the Brown-Carlsen test by noting that the wide use of and the general satisfaction with the test is considerable evidence of its acceptability.\(^34\)

**The problem.**--A factor of particular interest to this investigator as he reviewed the findings of listening comprehension measurement research was the report of the level of efficiency demonstrated by people when tested on their ability to listen to spoken communication in a setting in which both oral and visual cues were present. Cartier reported that people remember only 25 per cent of what they hear.\(^35\)


\(^{34}\)Sam Duker and Charles R. Petrie, "What Do We Know About Listening: Continuation of a Controversy," *Journal of Communication*, XIV (December, 1964), 246.

Classroom experiments conducted by Nichols at the University of Minnesota revealed a 50 per cent correct response in immediate test situations following the presentation of ten-minute lectures. The figure dropped to 25 per cent after two weeks had passed. H. E. Jones at Columbia University had the same experience and the use of guest lecturers did not alter the figures. Nichols concluded, "I think it is accurate and conservative to say that we operate at almost precisely a 25 per cent level of efficiency when listening to a ten-minute talk."  

Nichols's conclusion was based upon an analysis of test results. However, to fully understand the interpretation of such a conclusion one must be aware of basic principles of test construction which have a direct relationship with test results. An important principle of test construction is appropriate test difficulty. The basis of all objective approaches to test difficulty is performance on the test by the subjects for whom the test is intended. As a general rule, the average difficulty of the items in a test should correspond to the average ability of the subjects; i.e., the items should be such that, on the average, about half of the subjects will answer them correctly. The single item that best indicates the level of ability of an individual is one for which his probability of passing is 50 per cent. The most accurate test for such an individual is composed of items all of which are at the 50 per cent difficulty level.  

---

36 Ralph G. Nichols, "Do We Know How to Listen? Practical Helps in a Modern Age," The Speech Teacher, X (March, 1961, 118-124.  
It is a well-known fact that research experts who work on the development of diagnostic tools of measurement utilize the 50 per cent difficulty level in the design of their instruments. Brown,\textsuperscript{38} for instance, employed this principle in the early development of the Brown-Carlsen Listening Comprehension Test which is one of the most widely used standardized tools on the high school and college freshman levels.

With this background of information, the following questions arise: If a test is designed at the 50 per cent difficulty level, can results above the 50 per cent level of efficiency be expected? Do the published statements that people listen at a 25 to 50 per cent level of efficiency reflect accurately the ability of people to listen or the ability of test developers to construct instruments at a 50 per cent level of difficulty? Can a standardized measuring device with the 50 per cent difficulty level principle designed into it be improved through the application of basic question formation techniques so as to yield results with a higher efficiency level and still be recognized as a valid and reliable tool?

It thus became the purpose of this study to apply to a widely used and generally accepted standardized test of listening comprehension certain question refinement techniques intended to improve the clarity and comprehensibility of the test, and to note any changes in the efficiency of listener response that could be attributed to the revised test.

\textsuperscript{38} James I. Brown, "The Development of a Diagnostic Test of Listening Comprehension" (unpublished Ph.D dissertation, University of Colorado, 1949).
The standardized test chosen for the investigation was the Brown-Carlsen Listening Comprehension Test--Form Bm.\(^39\) The refinement techniques intended to improve test clarity and comprehensibility included the shortening of sentences, changes in the choice of words, the repositioning of facts within a question, changes in answer alternatives, and the replacement of certain questions with items designed to be more appropriate. These refinement techniques reflect research information in articles written by Black,\(^{40}\) Brown,\(^{41}\) Coleman,\(^{42}\) Ehrensberger,\(^{43}\) and Jersild,\(^{44}\) and in books written by Adkins\(^{45}\) and by Guilford.\(^{46}\)

**Organization of the dissertation.**--Chapter II of the dissertation will review applicable research in the area of listening comprehension measurement with particular attention given to the research by Brown

\(^{39}\)Brown and Carlsen, loc. cit.


\(^{41}\)James I. Brown, "Vocabulary--Key to Communication," *Education*, LXXX (October, 1959), 80-84.


\(^{45}\)Adkins, op. cit., p. 64.

which led to the development of the Brown-Carlsen Listening Comprehension Test. Chapter III will describe the procedure utilized to conduct the investigation, and Chapter IV will present the results of the study. Conclusions based upon an analysis of the data will be contained in Chapter V.
CHAPTER II

REVIEW OF THE LITERATURE

Introduction.--Research in the area of listening comprehension has been particularly voluminous within the past two decades. Although there is evidence of doctoral dissertation work as early as the 1920's, and although Rankin's 1929 pioneer investigation of time spent in communicative behavior is referred to frequently in listening literature, it was during and immediately following World War II that communication experts began intensive inquiry into the factors attendant to the skill of receiving spoken information.

Wartime research was undertaken at the request of Civilian Pilot Training and the United States Army to investigate possible training programs that would increase the effectiveness of interphone and radio communication concerned with aircraft. One such training program was established at the University of Wichita which became the regional head of seven states in the Civilian Pilot Training Program shortly before Pearl Harbor. Two years of experimentation by the university's Department of Speech produced results in the improvement of sending and receiving messages via microphone that made CPT ready to extend the


training to all regional units. When CPT gave way to Army Air Corps training, the work was continued.  

An extensive edited review by Black of some of the other wartime studies indicates that although speaker training resulting in improved speaking was the principal factor accounting for increased communicative efficiency, listening training was an integral part of the investigative process in the studies.

**Listening research landmarks.**—Several excellent reviews summarizing post-war experimental studies and major findings in listening research make it superfluous to engage in an item by item account of the completed works in this paper. Beighley, Duker, Keller, Toussaint, and Witty and Sizemore, all present detailed reports of pertinent listening inquiries and helpful bibliographies. Significant

---


listening research landmarks revealed by the cited reviews include the following:

1. Light has been shed on fundamental differences that exist between reading skills and listening skills.

2. An increasing number of factors thought to influence listening have been explored. The factors include the intelligence of the listener, the personality of the listener, listener anticipation, the theoretical interest of the listener, notetaking by the listener, the method of presentation, vocal skill in presentation, the rate of presentation, degrees of loudness in presentation, the prestige of the speaker, emotional vs. intellectual appeals in presentation, impairment of presentation by music and interruptions, the organization of the materials presented, and the difficulty of the materials presented.

3. Substantial evidence has been advanced to show that listening skill can be improved by instruction and can be measured objectively.

4. Standardized tests of listening comprehension have become a reality.

The development of objective tests of listening comprehension is a research landmark of particular significance for this review. Duker reports that no fewer than seven Master's theses and fourteen doctoral dissertations have been concerned with the construction and evaluation of experimental forms of listening test instruments for the various levels of education. In numerous other graduate degree studies, tests of various phases of listening were constructed as an incidental part of the research, or, existing tests were employed to seek out factors influencing the listening situation.56

doctoral research was accomplished by Blewett at the University of Missouri in 1945,\textsuperscript{57} Brown at the University of Colorado in 1949,\textsuperscript{58} Cartier at the University of Southern California in 1951,\textsuperscript{59} Heilman at the State University of Iowa in 1951,\textsuperscript{60} Dow at Michigan State College in 1952,\textsuperscript{61} Caffrey at the University of California in 1953,\textsuperscript{62} Biggs, at the University of Denver in 1954,\textsuperscript{63} Lewis at the Colorado State College of Education in 1954,\textsuperscript{64} Hollow at Fordham University in 1955,\textsuperscript{65}

\textsuperscript{57}Thomas T. Blewett, "An Experiment in the Measurement of Listening at the College Level" (unpublished Ph.D. dissertation, University of Missouri, 1945).

\textsuperscript{58}James I. Brown, "The Development of a Diagnostic Test of Listening Comprehension" (unpublished Ph.D. dissertation, University of Colorado, 1949).

\textsuperscript{59}Francis A. Cartier, Jr., "An Experimental Study of the Effect of Human Interest Factors on Listenability" (unpublished Ph.D. dissertation, University of Southern California, 1951).

\textsuperscript{60}Arthur W. Heilman, "An Investigation in Measuring Improving Listening Ability of College Freshmen" (unpublished Ph.D. dissertation, State University of Iowa, 1951).


Goodman-Malamuth at the University of Southern California in 1956,\textsuperscript{66} Gruszcynski at Fordham University in 1957,\textsuperscript{67} Hayes at Boston University in 1957,\textsuperscript{68} McDonald at Boston University in 1957,\textsuperscript{69} and Wright at Washington University in 1957.\textsuperscript{70}

A look at several of the experimental listening tests constructed for use on the college level of education in connection with doctoral research will reveal the general nature of the tests and the factors determined to be most important when measuring listening comprehension.

Blewett's experiment included the preparation of two subtests. The first subtest was labeled Content Retention. It measured the ability of students to recall the names of people, places, streets, buildings, etc., which had been contained in two hundred-word fictitious passages presented to the students by tape recording at a rate of 120-140 words per minute. The second subtest was labeled Drawing Conclusions. It measured the ability of students to draw inferences and

\textsuperscript{66}Leo Goodman-Malamuth, II, "An Experimental Study of the Effects of Rate of Speaking Upon Listenability" (unpublished Ph.D dissertation, University of Southern California, 1956).

\textsuperscript{67}Sister M. L. Gruszcynski, "An Experimental Study of Functional Reading and Listening Skills in the Fourth Grade" (unpublished Ph.D. dissertation, Fordham University, 1957).

\textsuperscript{68}Mary T. Hayes, "Construction and Evaluation of Comparable Measures of English Language Comprehension in Reading and Listening" (unpublished Ph.D. dissertation, Boston University, 1957).


identify speaker attitudes from a series of related ideas within orally presented information passages. Subtest I contained forty items with an average difficulty index of 46 per cent. The items were answerable in one to three words. Subtest II contained forty items with an average difficulty index of 55 per cent. The items were of the multiple response type. 71

Heilman's work involved the development of two comparable forms of a test of listening comprehension with each form having three subtests. Two of the subtests were based on sustained material in recorded 1600-1800 word information passages. Forty-one recorded multiple choice questions were prepared to measure comprehension. The third subtest had two parts with a total of forty-one recorded items. Part I involved the following of directions and the identification of figures. Part II contained stimulus statements which were followed by four test statements. The test statements were judged by the listener to be similar to or different from the stimulus statement. Test items for the three subtests on the two comparable test forms fell within difficulty limits of 20-80 per cent, with approximately one-half of the items within the 40-60 per cent range. 72

Dow's experimental research resulted in the instrument which is known as the Michigan State Listening Test. Implications from a speech


text and four reading manuals suggested to Dow seven "standardized" foils which were indicative of major aspects of listening comprehension. The seven foils became seven possible choices that a student could assign to a statement based on a passage of information. The three-minute passage of information was presented to the listening students by tape recording at a rate of delivery not to exceed 140 words per minute nor fall below a rate of one hundred words per minute. Following the presentation of information, 32 tape-recorded statements were heard by the students. To each statement the student was to assign one of the following seven possible choices: 1. This is an exact repetition of the central idea of the passage. 2. This statement contains the same central idea, but it is stated in different words from those used in the passage. 3. This is a false, twisted, or wrong statement of the central idea of the passage. 4. This is an exact repetition of a detail in the passage. 5. This statement contains a same, or similar detail, but it is stated in different words from those used in the passage. 6. This is a false, twisted, or wrong statement of a detail in the passage. 7. This statement is unrelated to either the central idea or a detail in the passage. It is not given nor implied.73

The diagnostic instrument designed by Biggs utilized a battery of seven tests to measure the listening comprehension of students. Each of the seven tests was to measure one of the following listening skills:

1. The ability to grasp the central idea of a passage, which was measured

---

by ranking the adequacy of one-sentence summaries of a speech. 2. The ability to retain pertinent content and make valid inferences, which was measured by objective-type items on speech material. 3. The ability to recognize main ideas and supporting ideas in a passage of information, which was measured by indicating whether a statement represented a main idea, a supporting idea, or an idea not discussed in a speech. 4. The ability to hear differences in meaning in similarly worded statements, which was measured by indicating whether paired statements were the same or different in meaning. 5. The ability to recognize the correct or incorrect usage of a word, which was measured by indicating whether or not all words in a sentence had been used correctly according to their meaning. 6. The ability to grasp the meaning of a word from contextual clues, which was measured by supplying a meaningful word or words for BLANK which appeared in each of a series of statements. 7. The ability to comprehend oral instructions, which was measured by following simple oral directions. The tests were administered in a natural classroom environment.74

The development of the Brown-Carlsen Listening Comprehension Test. Brown's experimental efforts in the development of a diagnostic test of listening comprehension75 laid the groundwork for the Brown-Carlsen Listening Comprehension Test, a widely used standardized tool which was chosen by this writer for his investigation of the application of

75 Brown, loc. cit.
question refinement techniques designed to increase the clarity and comprehensibility of a measuring instrument.

For the purposes of preparing an initial form of the listening test, Brown determined through his survey of listening studies, existing tests in silent reading, and the opinions of experts in the field of communication, that listening is a composite of receptive and reflective skills. He decided to explore accuracy of reception in terms of getting lecture details, following oral directions, and keeping a sequence of details in mind. Reflective, or critical, listening was explored in terms of getting central ideas, drawing inferences, distinguishing relevant from irrelevant material, using contextual clues to determine word meanings, and identifying transitional elements.

The earliest version of the test consisted of two sections. Section I was planned to explore the possibility of measuring several of the identified listening skills by timesaving methods. This section contained 104 items which measured the ability of students to (1) follow oral directions, by having them write the correct answer in response to directions based on a group of numbers and letters to which the student referred while listening to the directions, (2) remember a sequence of details until questioned, by having them enter the correct detail from a choice of details, (3) use contextual clues in determining word meanings, by selecting the correct word meaning from several options, and (4) use transitional elements, by determining whether a statement was introductory, concluding, or transitional. In Section II, an attempt was made to get as close to a normal classroom lecture situation as was possible without losing the objectivity of a controlled script. After
listening to a fifteen-minute lecture, students were questioned about its details, central idea, inferences, and relevancy. All 86 test items in this section were of the multiple response type.

This experimental form of the test was administered to two small groups in order to eliminate non-functioning items and collect data on timing, clarity of directions, difficulty, and interest. This test, with resulting minor modifications, was then administered to 110 high school and 150 college students. The method of administration was personal rather than recorded because Brown thought of listening as being aural assimilation in a face to face speaker-audience situation. Other reasons advanced for the personal administration of the test were that the test could be used everywhere despite the lack of playback equipment, pronunciation and speech pattern variables would be reduced, a more natural school situation would result, and a more natural teacher-student or speaker-audience relationship would exist. Regular class periods were used for the presentation of test materials.

The test data were analyzed to shed light on the relationships between listening comprehension and intelligence, reading comprehension, and scholastic achievement; to determine subtest interrelationships; to check the validity of the test items; and to determine subtest and total test reliability. The average difficulty index of the test was approximately sixty per cent rather than the anticipated fifty per cent.

The data resulting from these early test tryouts, which were a part of Brown's experimental doctoral research, served as a basis for the construction of a second experimental form consisting of four
subtests: (1) Immediate Recall, (2) Following Directions, (3) Word Clues, which included recognizing transitions and recognizing word meanings, and (4) Lecture Comprehension. All test questions were changed to a multiple-choice type and otherwise adapted for use with a machine scoring answer sheet. The options for all answers were printed on the answer sheets. This test was administered in the spring of 1951 to 447 students in Grade 10, 327 in Grade 11, 348 in Grade 12, 611 college freshmen, and 192 college sophomores, juniors, and seniors. All examiners were asked to criticize the test and these criticisms were considered in developing the final form of the test.  

Form Am of the Brown-Carlsen Listening Comprehension Test was then constructed on the basis of data obtained during the tryout of the second experimental form. This form, comprising 76 items, measures what the experimental tryouts had indicated were five important listening skills. They are: (A) Immediate Recall, which measures the ability to keep a sequence of details in mind until a question is asked which requires thinking back over the sequence, (B) Following Directions, which measures the ability to follow oral directions, (C) Recognizing Transitions, which measures awareness of the function of transitional words and phrases within sentence contexts, (D) Recognizing Word Meanings, which measures the ability to recognize meanings of words from context, and (E) Lecture Comprehension, which measures the ability to listen for details, get the central idea, draw inferences, understand the organization, and note degree of relevancy in a brief lecture presentation read

---

by the examiner. Although the test is divided into parts, each measuring to some extent a different aspect of listening skill, a total listening score only is derived.77

An experimental edition of a second form of the test, Form Bm, closely paralleling Form Am in organization and type of content, was prepared and administered together with Form Am in a rotation-type experiment. A total of 1367 students was employed. A random half of the classes took Form Am first; the other half, Form Bm first. Both forms were then item analyzed following the same procedures used with the original experimental Form Am. Non-functioning and other items were successfully eliminated from Form Bm until the distributions of difficulty and validity indices for separate parts and total matched as closely as possible corresponding distributions of Form Am indices. The mean difficulty index for both forms was 63.5 per cent. Mean validity indices were for Form Am 28.4 per cent, and for Form Bm 34.0 per cent.78

The Brown-Carlsen Listening Comprehension Test was administered for standardization purposes in May, 1952, to approximately 8000 students in 25 high schools from 16 states, and in September, 1952, to more than three hundred college freshmen. Data concerning the Brown-Carlsen norm group indicated it to be fairly representative of the national population with respect to age and ability level.79

Before turning to a brief review of studies which point out

77 Ibid., p. 3.
79 Ibid., p. 3.
certain question refinement techniques that can be employed to increase the clarity and comprehensibility of measurement instruments, it should be noted that in the development of the standardized Brown-Carlsen Listening Comprehension Test a principle of test construction was utilized which bears particular significance in relation to the present investigation. The principle referred to is appropriate test difficulty discussed in Chapter I. In the early stages of construction and in the following stages of development, Brown used the 50 per cent difficulty level index as a guide for the inclusion or rejection of test items.

The analysis of data for the final test forms, which revealed that the mean difficulty indices for both forms of the standardized test were at the 60 per cent level, is an indication that Brown was fairly successful in achieving the aforementioned principle of test difficulty.

The question arises, however, about what the interpretation of these results says about the persons who participated as examinees in the measurement of listening skills. With the test designed at a particular difficulty level, would the general level of listening efficiency demonstrated by the examinees be above that level? Do the results of the test reflect accurately the listening efficiency of the participants or the ability of the test developers to construct a tool at a particular level of difficulty? If the application of basic question refinement techniques to a standardized measuring device with the 50 per cent difficulty level designed into it resulted in a higher level of listening efficiency, what interpretations could be made?

Question refinement techniques.--The question refinement techniques alluded to above include the shortening of sentences, changes in
choice of words, the repositioning of facts within a statement, changes in answer alternatives, and the replacement of certain questions with items designed to be more appropriate.

Adkins, in her book on test construction, points out that the statement of the problem may need to be remolded to rule out certain unintended interpretations leading to other than the best answer, that one too nearly correct alternative had better be replaced by another, that certain alternatives attract no competitors, that certain items are too easy or too difficult, or that although certain items are appropriate in difficulty, they could well be replaced by more valid items.\footnote{Dorothy C. Adkins, Construction and Analysis of Achievement Tests (Washington: U. S. Government Printing Office, 1947), p. 147.}

The work of Jersild and Ehrensberger has introduced an area of controversy in communication research which is worthy of investigation whenever the materials involved are appropriate for an application of their findings. Jersild was interested in discovering the effect of the position of materials within a speech upon the listener. He found the position of primacy definitely superior to recency as a means of emphasis in oral discourse. Later research by Ehrensberger showed a definite superiority of recency over primacy, which, of course, was in contradiction of Jersild's findings. It was determined by this investigator to apply the factor of information position within a test item where applicable to further inquire into this controversy.

Chapter III of the dissertation will describe the procedure utilized to investigate the application of some of the above mentioned question refinement techniques to the standardized Brown-Carlsen Listening Comprehension Test.


CHAPTER III

PROCEDURE

The present chapter contains a detailed report of the procedures involved in selecting a standardized measure of listening comprehension, analyzing the instrument for the application of question refinement techniques, and determining the method of administering and scoring the experimental test form.

The selection of a standardized test of listening comprehension.

The Brown-Carlsen Listening Comprehension Test--Form Bm was chosen by this writer as a suitable standardized tool for the experimental investigation of question refinement techniques designed to increase the clarity and comprehensibility of a measuring instrument. As was indicated earlier in the dissertation, the wide use of and the general satisfaction with the Brown-Carlsen test is considerable evidence of its acceptability. According to Duker, it has been used more frequently than any other existing listening test in experimental investigations conducted as a part of doctoral studies.\(^8\) In addition, Haberland discovered that of three different tests of listening comprehension, the Brown-Carlsen test showed the highest degree of correlation when compared with the results of the linguistic sections of such standardized tests as the ACE Psychological Examination, College Level; the Cooperative English

\(^8\)Sam Duker, "Doctoral Dissertations on Listening," *Journal of Communication*, XIII (June, 1963), 106-117.
Another factor dictating the selection of the Brown-Carlsen test was the principle of appropriate test difficulty. From its initial stages of development, the test had designed into it the 50 per cent difficulty level principle, a principle of test construction which this writer feels has a direct bearing upon the interpretation of listening efficiency results.

Finally, the Brown-Carlsen Listening Comprehension Test lends itself readily to the utilization of basic question refinement techniques. The five parts of the test—Immediate Recall, Following Directions, Recognizing Transitions, Recognizing Word Meanings, Lecture Comprehension—contain a variety of question forms to which several different question refinement techniques can be applied.

The analysis of the Brown-Carlsen test for the application of question refinement techniques.--Part A of the Brown-Carlsen Listening Comprehension Test—Form Bm is labeled Immediate Recall. It is designed to measure the ability of a person to keep a sequence of details in mind until a question is asked which requires thinking back over the sequence. There are seventeen questions in Part A. In questions one through six, a series of numbers is read aloud by the examiner. Following the reading of each of the series of numbers, the listener is

---

requested to recall a particular number in the series. Questions seven through twelve utilize the same approach with the substitution of two-letter words for numbers in each of the series making up a question. Questions thirteen through seventeen are more involved in that the person must listen to a set of directions and then reply to questions based upon the information contained in the directions. Questions thirteen and fourteen refer to one set of directions, question fifteen to a second set of directions, and questions sixteen and seventeen to a third set of directions.

For each question in Part A, the listener is provided with a choice of five answers which appear on a printed answer sheet. The respondent marks the answer he estimates to be correct by making a heavy black mark as long as the pair of lines on the answer sheet directly under the printed alternative he chooses. The fifth alternative for each question is labeled "N". If the correct choice in response to a question is none of those listed in the preceding four choices, the respondent fills in the answer space under "N".

Examination of the questions and answer alternatives in Part A, Immediate Recall, suggested several general changes that could be introduced for the purpose of improving the clarity of the test without disrupting the general form of the content.

The first general change involved the removal of an unimportant and incidental phrase in questions one, two, five, and nine, which, it seemed to this writer, serves as a distracting foil for the listener. The deletion of the phrase shortens the cited questions and seems to improve their clarity. Following are examples of the questions as
printed in their original form and as constructed in their revised form with the phrase omitted:

Question 1. (original form) "In the series of numbers 4-6-2-8-3-7, a series containing six numbers, the FIRST number is ______?"

Question 1. (revised form) "In the series of numbers 4-6-2-8-3-7, the FIRST number is ______?"

Question 2. (original form) "In the series of numbers 3-2-4-7-5, a series containing five numbers, the FOURTH number is ______?"

Question 2. (revised form) "In the series of numbers 3-2-4-7-5, the FOURTH number is ______?"

Question 5. (original form) "In the series of numbers 6-8-9-4-2-7, a series containing both odd and even numbers, the SECOND number is ______?"

Question 5. (revised form) "In the series of numbers 6-8-9-4-2-7, the SECOND number is ______?"

Question 9. (original form) "In the list of words by-of-in-an-at, a list containing five words, the SECOND word is ______?"

Question 9. (revised form) "In the list of words by-of-in-an-at, the SECOND word is ______?"

The second general change in Part A, Immediate Recall, was related to questions seven through twelve. As has been described, questions seven through twelve utilize a series of two-letter words which are read aloud by the examiner. Following the reading of each of the series of words, the listener is requested to recall a particular word in the series.

An examination of the two-letter words employed in questions seven through twelve suggested the possibility of confusion resulting for the listener if the enunciation of the examiner were inaccurate on some of the two-letter words which are similar in sound. The two-letter words used in the questions are by, an, of, in, on, at, and to.
To avoid the similarity in sound of some of the two-letter words and to improve the clarity of the questions as read by the examiner, the following two revisions were made: 1. The word *an* was discarded completely because of its similarity in sound to the word *in* (in questions seven, nine, and twelve of the Brown-Carlsen test, both *an* and *in* appear) and the word *be* was introduced as a substitute for the discarded word *an*. 2. In the positioning of words within each series, the words beginning with a vowel or ending in the same sound were not placed in immediate succession. Thus, the words were alternated in such a manner that if the first letter of the first word in a series were a vowel, the first letter of the second word in the series would be a consonant. This avoided the possibility, for instance, of the words *in* and *on* appearing in immediate succession.

By substituting the word *be* for the word *an*, and by alternating the two-letter words within each series so that the words beginning with a vowel or ending with the same sound were not in immediate succession, the following changes resulted:

**Question 7.** (original form) "In the list of words by-an-of-in-on-at, the LAST word is ______?"

**Question 7.** (revised form) "In the list of words by-of-be-in-to-at, the LAST word is ______?"

**Question 8.** (original form) "In the list of words in-on-at-by-of, the word beginning with *a* is ______?"

**Question 8.** (revised form) "In the list of words in-by-at-to-on, the word beginning with *a* is ______?"

**Question 9.** (original form) "In the list of words by-of-in-an-at, a list containing five words, the SECOND word is ______?"

**Question 9.** (revised form) "In the list of words by-of-to-in-be, the SECOND word is ______?" (NOTE: omission of incidental phrase plus realignment of words in the series)
Question 10. (original form) "In the list of words on-to-in-at-by, the THIRD word is _____?

Question 10. (revised form) "In the list of words on-to-in-by-at, the THIRD word is _____?

Question 11. (original form) "In the list of words at-of-in-by-to, the FOURTH word is _____?

Question 11. (revised form) "In the list of words at-be-in-by-of, the FOURTH word is _____?

Question 12. (original form) "In the list of words an-to-in-at-of-on, the FIFTH word is _____?

Question 12. (revised form) "In the list of words in-be-at-by-of-to, the FIFTH word is _____?

The third general change in Part A, Immediate Recall, concerned the possibility of rephrasing selected questions so that the information desired appeared in the question before the series of numbers or words were read rather than after the series of numbers or words were read. It was felt by the investigator that an objection to the proposed change might be that the listener no longer would be involved in thinking back over a series of numbers or words but would be required to think through numbers or words as they were read in sequence. However, it was determined that the occasional inclusion of this rephrasing would be acceptable for experimental inquiry. Thus, the following questions reflect this change:

Question 3. (original form) "In the series of numbers 8-9-6-4-7, the FOURTH number is _____?

Question 3. (revised form) "What is the FOURTH number in the series of numbers 8-9-6-4-7?"

Question 4. (original form) "In the series of numbers 5-8-3-5-7-1, the FIFTH number is _____?

Question 4. (revised form) "What is the FIFTH number in the series of numbers 5-8-3-5-7-1?"
Question 6. (original form) "In the series of numbers 7-1-8-4-2, the THIRD number is _________?"

Question 6. (revised form) "What is the THIRD number in the series of numbers 7-1-8-4-2?"

Question 8. (original form) "In the list of words in-on-at-by-of, the word beginning with a is _________?"

Question 8. (revised form) "What word begins with a in the list of words in-by-at-to-on?" (NOTE: changed sentence order plus the realignment of words within the series)

Question 11. (original form) "In the list of words at-of-in-by-to, the FOURTH word is _________?"

Question 11. (revised form) "What is the FOURTH word in the list of words at-be-in-by-of?" (NOTE: changed sentence order plus the realignment of words within the series)

The final changes in Part A, Immediate Recall, were related to questions thirteen through seventeen which are more involved in that the person must listen to a set of directions and then reply to questions based upon the information contained in the directions.

Questions thirteen and fourteen refer to a single set of directions. The information asked for in question thirteen appears early in the statement of directions read aloud by the examiner and the information requested in question fourteen occurs in the latter portion of the directions. It was determined that Ehrensberger's findings on the effect of the position of materials within oral discourse upon a listener\(^8\) could be applied to the set of directions for questions thirteen and fourteen. Thus, the statement of directions was rephrased so that the information requested in both questions would be in a

"recency" position, the position Ehrensberger found to be superior to the primacy position when arranging materials in spoken communication. As will be noted in the examples which follow, question thirteen remains unchanged even though the directions are rephrased in the revised form of the test. Question fourteen, however, asks for different information in the revised form of the test so that the "recency" principle can be carried out.

Question 13. (original form) "In the statement, 'Send your name and address with four soap wrappers and 30 cents to Minneapolis 14, Minnesota, Box 20, to receive the special seed offer,' the NUMBER OF WRAPPERS is _______?"

Question 14. (original form) "In the statement just read, the BOX NUMBER is _______?"

Question 13. (revised form) "In the statement, 'To receive the special seed offer, send your name and address to Box 20, Minneapolis 14, Minnesota, with four soap wrappers and 30 cents,' the NUMBER OF WRAPPERS is _______?"

Question 14. (revised form) "In the statement just read, the AMOUNT OF MONEY TO BE SENT is _______?"

No changes were made in question fifteen of Part A. Following is the wording of the question:

Question 15. (no change) "Listen to this statement 'Place the two braces, marked A on the diagram, in the slots marked D on the shelf. Insert them an inch and a half; then place the shelf against the wall and fix it in place with the four screws marked C on the diagram.' The NUMBER OF INCHES the braces are inserted into the shelf is _______?"

The final two questions in Part A, questions sixteen and seventeen, are based upon information giving directions to reach a particular destination. In question sixteen, the listener is required to think back over the directions read aloud by the examiner and state the exact number of street blocks which had to be covered to reach the
destination. Question seventeen requests the address of the house which serves as the destination point.

It was determined that question sixteen lends itself to the experimental inquiry of rephrasing the directions so that the information sought in the question appeared before the number of street blocks to be covered was enumerated rather than after the number of street blocks was read. This revision reflects the same change made in questions three, four, six, eight, and eleven. No change was made in question seventeen.

Question 16. (original form) "Listen to these directions: 'Seven of you should walk three blocks up this street, turn to your right and continue to Pine Street, two blocks away, then angle off Pine for three more blocks to the green house, 51 Pine.' The TOTAL NUMBER OF BLOCKS to the house on Pine is ______?"

Question 16. (revised form) "What is the TOTAL NUMBER OF BLOCKS to the house on Pine Street in the following directions: 'Seven of you should walk three blocks up this street, turn to your right and continue to Pine Street, two blocks away, then angle off on Pine for three more blocks to the green house, 51 Pine.'"

Question 17. (no change) "The ADDRESS of the house on Pine Street is ______?"

Part B of the Brown-Carlsen Listening Comprehension Test--Form Bm is labeled Following Directions. It is designed to measure the ability of a person to follow oral directions. There are twenty questions in Part B. As each question is read aloud by the examiner, the listener refers to a group of numbers and letters located in the upper lefthand corner of the printed answer sheet. The numbers are arranged in a straight line and in sequence with number one beginning the sequence and number eight concluding the sequence. Directly below the numbers is a line of letters arranged in an A through H sequence. The
letter A is directly below the number one and letter H is directly below the number eight, with the intermediate numbers and letters arranged accordingly. The odd numbers, one, three, five, and seven, and the vowels, A and E, are underlined; the numbers four and seven are encircled. A typical question may ask the listener, "What number is above the first vowel?" In recording the answer, the listener is provided with a choice of five answers on the printed answer sheet just as was the case in Part A.

In analyzing the wording of the twenty questions in Part B, it was determined that in some of the questions terminology could be employed which would be clearer and more easily followed than that which is used in the printed test form. Since all of the questions in Part B, with the exception of the first question, require the listener to utilize addition, subtraction, or multiplication in following the oral directions of the examiner, an attempt was made to introduce the most simple and clear language of mathematical figuring into those questions where such application would improve the comprehensibility of the questions.

Basically, the changes involved the replacement of such phrases as "the number one less than" and "the number one greater than" with the phrases "subtract one from" and "add one to." Following is a complete list of the questions in Part B, Following Directions, as they appear in their original form and as the changed questions appear in the experimental form of the test.

Question 18. (no change) "The number above the first vowel is _______?"
Question 19. (original form) "Two more than the second underlined number is _____?"

(revised form) "Add two to the second underlined number. The answer is _____?"

Question 20. (original form) "The number one less than the smallest number circled is _____?"

(revised form) "Subtract one from the smallest number circled. The answer is _____?"

Question 21. (no change) "Add the smallest number to one half of the largest number. The answer is _____?"

Question 22. (no change) "Add the smallest even number to the largest number and take half the sum. The answer is _____?"

Question 23. (original form) "Subtract the smallest number from the next to the largest number. The number that is four less than that is _____?"

(revised form) "Subtract the smallest number from the next to the largest number. From that figure subtract four. The answer is _____?"

Question 24. (no change) "Add the smallest number to the next to the largest number. Half of that sum is _____?"

Question 25. (no change) "Multiply the next to the largest number by the smallest number and add the smallest number to it. The answer is _____?"

Question 26. (original form) "Subtract two from the largest number. The next larger number is _____?"

(revised form) "Subtract two from the largest number. Add one to that figure. The answer is _____?"

Question 27. (original form) "Add the next to the smallest odd number to the next to the largest even number. The number one less than that is _____?"

(revised form) "Add the next to the smallest odd number to the next to the largest even number. Subtract one from the total. The answer is _____?"

Question 28. (no change) "Subtract the first underlined number from the second circled number. The letter directly below that number is _____?"
Question 29. (original form) "The number one greater than the number directly above the second underlined letter is ________?"

(revised form) "Add one to the number directly above the second underlined letter. The answer is ________?"

Question 30. (no change) "Add the second smallest number to the next to the largest number. Subtract the smallest number from the result. The answer is ________?"

Question 31. (no change) "Add the number above the third consonant to the next to the smallest number. Half that sum is ________?"

Question 32. (original form) "Add the number above the first consonant to the number above the second vowel. The number one greater than that number is ________?"

(revised form) "Add the number above the first consonant to the number above the second vowel. To that sum add one. The answer is ________?"

Question 33. (original form) "The number three greater than the number above the first letter of the word CANDY is _____?"

(revised form) "Add three to the number above the first letter of the word CANDY. The answer is ________?"

Question 34. (original form) "Subtract the number above the first vowel from the number above the first consonant. Then add the number above the second consonant to that sum. Two more than the resulting sum is ________?"

(revised form) "Subtract the number above the first vowel from the number above the first consonant. Then add the number above the second consonant to that sum. To the resulting sum add two. The answer is ________?"

Question 35. (original form) "Subtract the first underlined number from the second circled number. The letter directly below one greater than the answer is ________?"

(revised form) "Subtract the first underlined number from the second circled number. Add one to that figure. The letter directly below the resulting number is ________?"

Question 36. (original form) "From the sum of the two circled numbers, subtract the number above the second underlined letter. Two less than that answer is ________?"

(revised form) "From the sum of the two circled numbers, subtract the number above the second underlined letter. From the resulting total subtract two. The answer is ________?"
Question 37. (no change) "Add the number above the letter immediately following the first vowel to the next to the largest number. From this sum subtract the number above the letter which follows the second consonant. One less than the result is ________?"

Part C of the Brown-Carlsen Listening Comprehension Test is labeled Recognizing Transitions. It is designed to measure one's awareness of the function of transitional words and phrases within sentence contexts. As the examiner reads aloud each of the eight sentences which compose Part C, the listener is required to decide what function, if any, the sentence would serve within the context of oral discourse. If the listener thinks the sentence is introductory, he fills in the answer space under "I" on the printed answer sheet. If the listener thinks the sentence is transitional, he fills in the answer space under "T", and if he thinks the sentence is a concluding sentence, he fills in the answer space under "C". If the sentence read aloud by the examiner is none of the preceding types, the listener fills in the answer space under "N".

An examination of this part of the Brown-Carlsen test revealed no opportunities for the application of question refinement techniques. The sentences are clear and concise and seem to fulfill the purpose of measuring one's awareness of the function of transitional words and phrases within sentence contexts. The sentences are reproduced below.

"38. And thus, for people in Europe as well as in America, this move seems to indicate better future relationships.

39. Also, new aspects of the work are encouraging.

40. There are four countries whose actions should be discussed at some length."
41. In the same way, the forests of the Far West are being protected.

42. Because of these disastrous failures, it seems time that schools change their programs.

43. Today the lecture will be on 'Building a Vocabulary.'

44. For the facts in the case, let us examine these files.

45. Emergency teachers during the war have become permanent teachers in the post-war years."

Recognizing Word Meanings is the title of Part D of the Brown-Carlsen Listening Comprehension Test. The purpose of this section is to measure the ability of the listener to recognize meanings of words from context. Ten questions are included in Part D. The listener's responsibility is to decide which one of five correct dictionary definitions that appear on the printed answer sheet is the meaning intended in the particular sentence read aloud by the examiner. The test is so designed that for every two sentences there is one set of five correct dictionary definitions.

Part D suggested one major means for experimental inquiry involving the formation of test questions. The ten sentences which are included in Part D are so phrased that the word for which the listener must determine a meaning is stated before the sentence containing the word is read. An example of this phrasing is, "What does RUN mean in the sentence, 'The vine will run up the fence'?" In the experimental test form it was decided by the investigator to rephrase each of the ten sentences so that the word for which the listener had to determine a meaning would occur after the sentence containing the word had been read by the examiner. An example of the revised phrasing is, "In the
sentence, 'The vine will run up the fence,' what does the word RUN mean?"

Following is the complete list of ten questions which make up Part D of the Brown-Carlsen test. They appear in both original form and revised form.

Question 46. (original form) "What does RUN mean in the sentence, 'The vine will run up the fence'?

(revised form) "In the sentence, 'The vine will run up the fence', what does the word RUN mean?"

Question 47. (original form) "What does RUN mean in the sentence, 'The colors will not run'?"

(revised form) "In the sentence, 'The colors will not run', what does the word RUN mean?"

Question 48. (original form) "What does PART mean in the sentence, 'What part did John act in the play'?

(revised form) "In the sentence, 'What part did John act in the play', what does the word PART mean?"

Question 49. (original form) "What does PART mean in the sentence, 'Your part is to help the chairman'?

(revised form) "In the sentence, 'Your part is to help the chairman', what does the word PART mean?"

Question 50. (original form) "What does BRIGHT mean in the sentence, 'The Elizabethan period was a bright period of history'?

(revised form) "In the sentence, 'The Elizabethan period was a bright period of history', what does the word BRIGHT mean?"

Question 51. (original form) "What does BRIGHT mean in the sentence, 'His prospects for advancement were bright'?

(revised form) "In the sentence, 'His prospects for advancement were bright', what does the word BRIGHT mean?"

Question 52. (original form) "What does FALSE mean in the sentence, 'There were false supports used in constructing the bridge'?

(revised form) "In the sentence, 'There were false supports used in constructing the bridge', what does the word FALSE mean?"
The fifth and final section of the Brown-Carlsen Listening Comprehension Test is Part E, Lecture Comprehension. It is designed to measure the ability of a person to listen for details, get the central idea, draw inferences, understand the organization, and note the degree of relevancy in a brief lecture presentation read aloud by the examiner. The listener is directed to listen carefully, but to take no notes on that which is read by the examiner. Following the brief lecture presentation, 21 questions are asked of the listener. For each question, the listener must decide which of the five answer choices on the printed answer sheet is the correct response to the question.

The lecture presentation in Form Bm of the Brown-Carlsen test is titled "Improving Your Reading Ability." The lecture begins with the enumeration of four reasons why good reading is important. The reasons are, (1) reading gives power to learn, (2) through reading one may gain inspiration for life, (3) reading helps one understand his
own experience, and (4) reading often will mean the difference between an individual being a boring or an interesting person. The lecture uses examples and illustrations to support each of the four reasons stating the importance of good reading.

The lecture content turns next to five methods that one may use to improve his reading skill. The methods are, (1) the forcing method, in which one simply forces himself to read at a faster rate of speed, (2) the observation approach, in which one trains himself to see groups of words instead of looking at each word by itself, (3) the billboard or license plate game, which trains one to read things quickly and accurately through the process of attempting to grasp as much information as possible at one glance, (4) the process of learning to pick out the general idea in printed copy plus exactness in detail, and (5) the use of machines such as the Metronoscope which are designed to improve reading speed. Illustrations and examples are used to describe each of the five methods that may be used to improve reading.

The lecture concludes by stressing the importance of an individual using the methods that will give him help with his particular reading difficulty.

Careful and detailed analysis of the lecture content and the accompanying 21 questions revealed a number of possibilities for the application of question refinement techniques that would improve the clarity and comprehensibility of the materials for the listener being tested on his ability to listen efficiently. The following paragraphs will present a question-by-question review of the 21 items in Part E, Lecture Comprehension, and will indicate the nature of the changes
applied. It will be noted that ten of the 21 questions remain unchanged, while the remaining eleven items are subjected to some type of modification.

The first question in Part E is number 56, and it asks for a detail contained in the lecture. The question reads, "What relative of Lincoln was mentioned in the Lincoln story?" The answer alternatives include (a) mother, (b) father, (c) half brother, (d) stepmother, and (e) sister. Two factors interfere with the clarity of the question. First, both the father and the stepmother are mentioned in the Lincoln story as it is related in the lecture. Second, the stepmother is referred to as "stepmother" in one sentence, but as "mother" in a following sentence. Thus, the listener is confronted with a dilemma when consulting the answer alternatives.

Two changes were made to improve question 56. First, in the lecture content, the word "mother" was changed to "stepmother" so that the two references to the woman would be consistent. Second, the wording of the question was changed completely to offer the listener a fairer choice in the selection of the correct answer among the available alternatives which, incidentally, were left unchanged. The revised form of question 56 reads, "What relative of Lincoln guided his reading when Lincoln was a youth?" Answer alternative "d" is the correct answer.

Question 57 in the test also was subjected to major change. In its original form, question 57 asks, "How old was Edison when his teacher said he could not learn to read?" It was felt by the investigator that the question asks for an insignificant detail mentioned
only once in the lecture that is relatively unimportant to the lis-
tener. A detail felt to be more important for the listener to remember
was the name of the famous American about whom the illustration cen-
tered. Edison's name occurs once in the lecture illustration and the
term "inventor" is also used. Thus, question 57 was reworded to read,
"As a boy, what famous American was considered by his teacher as being
too stupid to learn to read?" The new answer alternatives offered
were, (f) Fulton, (g) Whitman, (h) Edison, (i) Lincoln, and (j) Unter-
meyer. The four latter names appeared at different times in the lec-
ture content and the name of Fulton was included because he too is
known as an American inventor.

Questions 58 and 59 in Part E were left unchanged in the re-
vised form of the test. Both queries involve details related to the
section of the lecture which stresses the importance of good reading.
The questions are clearly phrased and the answer options are appro-
priately challenging. The two questions with their answer choices read
as follows:

Question 58. "What author wrote about how to enjoy poetry?"
e. Untermeyer

Question 59. "What did the man at the dinner party sell?"
(Answers) f. gloves  g. books  h. shoes  i. silverware
j. dresses

Question sixty also was determined to be a good detail question,
but an analysis of the answer alternatives revealed the need for modi-
ification in the choice of words in three of the five alternatives so
that the terminology would be more specifically aligned with the lan-
guage used in the lecture. Question sixty reads, "Which subject did
the hostess of the dinner party NOT bring up?" The printed answer alternatives are, (a) baseball, (b) music, (c) art, (d) gardening, and (e) taxes. In the lecture read aloud by the examiner the term "Dodgers" is used in place of baseball, the word "paintings" is verbalized in reference to art, and the word "flowers" is used rather than the word gardening. Thus, in the revised answer alternatives for question sixty, the alternatives "baseball," "art," and "gardening" were changed to "Dodgers," "paintings," and "flowers," respectively. The other two answer alternatives were left in their original form.

The next four questions in Part E of Form Bm of the Brown-Carlsen Listening Comprehension Test, questions 61 through 64, underwent complete reconstruction for the experimental form of the test. Question 61, in the original form of the test, requests a seemingly unimportant detail which is mentioned only once in the lecture as a part of an illustration supporting the forcing method for the improvement of reading skill. The illustration refers to a Minnesota farmer named John Patterson who becomes exposed to the forcing method while in a barbershop reading a magazine article which gives the number of minutes it should take the average reader to complete a given article or story. He applies the method to his own reading skill and discovers that the training enables him to read his agricultural journals more rapidly.

The detail asked for in question 61 is the state in which John Patterson lives. A detail which seems more important to the illustration, because it is mentioned four times by the lecturer, is the fact that Patterson is a farmer. Thus, question 61 was changed from "In
which state did John Patterson live?" to "What was John Patterson's vocational work?" The answer options offered for the revised questions were, (f) gardener, (g) carpenter, (h) barber, (i) salesman, and (j) farmer.

Question 62 of the original test asks for another detail related to the same section of the lecture. It reads, "In what place did Patterson find an article that helped him?" The answer alternatives are, (a) barbershop, (b) restaurant, (c) bus station, (d) library, and (e) doctor's office. It should be noted at this point that while question 62 provides a second detail question about the story which illustrates the forcing method as a means of improving reading skill, the test provides no question to bring out the key idea of this part of the lecture. It was thus determined that a question focusing attention upon a key idea in the lecture would be more appropriate in a testing situation than a question asking for a once-mentioned detail that was a part of the illustration supporting the key idea. So, in the experimental form of the test, question 62 was changed by substituting the following question for the original query: "What method to improve reading did Patterson discover while reading a magazine article in the barbershop?" The answer alternatives made available to the listener were, (a) observation, (b) forcing, (c) billboard, (d) machine, and (e) eye regressions.

Questions 63 and 64 in the Brown-Carlsen test refer to an example used in the lecture presentation which develops the point that one can improve his reading by learning to recognize the general idea and the exactness of detail in a piece of written copy. The example
tells of an ex-school teacher, Miss Gavigan, who brought newspaper articles to the dinner table and drilled the children to recognize the most important idea and the significant details in each of the articles they read. Item 63 asks the question, "Where did Miss Gavigan give reading improvement drills?" The answer choices are, (f) on the lawn, (g) at the dinner table, (h) at the beach, (i) in the park, and (j) on the floor. Question 64 asks, "What is the first sentence of a typical news story called?" The answer options are, (a) data, (b) headline, (c) caption, (d) outline, and (e) lead.

Both of the aforementioned questions request a detail from the Miss Gavigan story. Each detail is mentioned once in the story, and, in question 64, the detail serves as a subordinate point to a clearly stated principle that the first sentence of a typical news story is supposed to summarize all the important information in the rest of the news story.

Analysis of the two detail questions suggested that one question should be retained in some form so that a detail from this portion of the lecture would be a part of the test questions. It was determined that question 64 concerned a more important detail than question 63, and that the question could be strengthened further by wording it to request the principle of what a first sentence in a typical news story should accomplish rather than what the first sentence of a typical news story is called. Following is the reworded question and the answer alternatives constructed to fit the question for the experimental form of the test:

Question 64. "The first sentence of a typical news story should do ... what?"
(Answers) a. summarize important information  
b. capture attention  
c. challenge the reader  
d. set the scene  
e. present the main idea.

The decision of what to do with question 63 was answered by the omission in the original test of a key idea question relating to one of the main reasons why good reading is important. The stated reasons occur in the first section of the lecture, and three of the four reasons are covered by key idea test questions. However, the reason which states that reading helps one to understand his own experience receives no coverage by a key idea question. This reason is supported in the lecture by a reference to the reading of poetry and Louis Untermeyer's contention that most young people, when confronted with a poem, make it an academic exercise rather than a means of understanding their own experience.

It was determined by the investigator that a question alluding to this important idea should be constructed for inclusion in the experimental test form and one of the less important questions in the original test be deleted in deference to the newly constructed item. Thus, question 63 in the Brown-Carlsen test was replaced by the following question:

Question 63. "What is most important in the reading of poetry?"
(Answers) f. to get information  
g. to notice how words are used  
h. to understand your own experience  
i. to recognize rhyme scheme  
j. to make it an academic exercise

Questions 65, 66, 67, 68, and 69 of Form Bm of the Brown-Carlsen test were left unchanged in the experimental version of the test with the exception of one minor modification in question 66. This particular question is so phrased that the listener is requested to note
something not given a separate place in the lecture presentation. In the experimental test form, the word "not" was underlined in question 66 so that the test reader would have a visual reminder to give proper interpretation to the reading of the statement.

An analysis of the five questions revealed that question 65 asks for a detail from the lecture and question 69 requires the listener to draw an inference. The other three queries within the group of five concern key ideas contained in the lecture content. The five questions with their respective answer alternatives are reproduced below. It will be noted that the word "not" is underlined in question 66.

Question 65. "The speed range of the reading machine is expressed in terms of the number of ... what per minute?"
(Answers) f. sentences  g. phrases  h. letters  i. words  j. lines

Question 66. "Which reading-improvement method was not given a separate place in the organization of the lecture?"
(Answers) a. newspaper clipping  b. billboard  c. forcing  d. eye regressions  e. machine

Question 67. "The Edison story was used in this lecture to show the importance of good reading for what purpose?"
(Answers) f. obtain information  g. become interesting  h. increase enjoyment  i. gain inspiration  j. gain power to learn

Question 68. "The dinner party story was meant to illustrate the importance of reading for what purpose?"
(Answers) a. obtain information  b. become interesting  c. increase enjoyment  d. gain inspiration  e. gain power to learn

Question 69. "Which would be the best title for the Miss Gavigan story?"
(Answers) f. An Interest Guest  g. Writing Newspaper Stories  h. The Three R's  i. Getting the Main Idea  j. Memory Work

An examination of question seventy in relation to the lecture presentation showed a clearly phrased and acceptable question, but
some doubt arose in connection with the correct answer for the query. The question reads, "What did the discussion of the use of machines to improve reading suggest?" The provided answer choices are, (a) It is the best method, (b) Americans are machine-minded, (c) It has faults as a method, (d) It increases comprehension, and (e) It works with serious cases. The third option, "It has faults as a method," is designated as the correct answer.

It was felt by the investigator that a more specific answer was needed to serve as the correct answer for question seventy. It is probable that all of the methods suggested in the lecture for the improvement of reading have faults. Thus, a more specific statement was constructed to serve as the correct answer alternative in the experimental version of the test. The alternative "It has faults as a method" was deleted and the statement "Machines are expensive and not readily available" was inserted in its place. The substitute alternative appeared in the lecture as the first sentence in a paragraph citing the problems connected with the use of machines for the improvement of reading.

Question 71 in the test refers to a portion of the lecture in which an interpretation is made of an experience that Walt Whitman may have had which motivated him to write the poem, "When I Heard the Learned Astronomer." The interpretation of the experience relates how Whitman may have attended a lecture given by one of the famous astronomers of his day. The sky searcher probably had speculated on the origin of the universe by using his charts and figures to express his knowledge on the topic. Whitman, unimpressed by the astronomer's
scientific analysis, wandered out into the clear night air and "looked up in perfect silence at the stars." The poem, the last line of which concludes the preceding sentence, appears in the test lecture presentation.

Question 71 asks the listener, "From the poem that Whitman wrote, what do you think his opinion of lecturers was?" The answers provided on the printed answer sheet include, (f) They give knowledge, (g) They miss the real truth, (h) They become disturbed by applause, (i) They are brilliant, and (j) They talk too long.

It was felt by the investigator that question 71 is inaccurate in that it implies a reaction by Whitman to all lecturers or to lecturers in general. To improve the question, it initially was decided to reword the question to solicit the listener's thinking on Whitman's opinion of this particular lecturer and to change the word "they" in each of the answer alternatives to "he." Further deliberation on a proposed revision resulted in the complete deletion of the original question which asks for an opinion by the listener, and the substitution of a new question requesting factual information contained in the Whitman story. Following is question 71 as it appeared in the experimental test form:

Question 71. (revised form) "Which of Whitman's poems was referred to in the lecture?"
(Answers) f. "I Hear America Singing" g. "Miracles" h. "Give Me the Splendid, Silent Sun" i. "When I Heard the Learned Astronomer" j. "For You O Democracy"

No change was made in question 72 of the Brown-Carlsen test. It requires the listener to draw an inference which points to the central theme of the lecture presentation. The question is clear and the
response alternatives are appropriate. Following is the question:

Question 72. "What did the lecture as a whole seem to suggest about learning to read?"

(Answers) a. Everyone can improve. b. Reading is easy. c. You should force yourself to read an hour a day. d. You shouldn't be ashamed if you don't read well. e. Everyone makes mistakes.

An attempt to improve the wording of the correct answer choice constituted the only change in question 73. The revised wording employs terminology more directly aligned with the language used in the lecture content. Answer alternative "j" is the correct answer for question 73. Following is the question and answer options as they appear in the original test form:

Question 73. (original form) "Which method did the author seem to feel was the best method for learning to read?"

(Answers) f. The machine method g. The one you like the best h. The billboard method i. No one, but all of them j. Depends upon how you read

The revised wording of the correct answer choice in the experimental test form is, (j) The one that will help you with your particular difficulty.

No change was made in question 74. It refers to a key idea in the lecture and is clearly presented. It reads as follows:

Question 74. "The reference to Lincoln shows how one can gain... what from reading?"

(Answers) a. understanding b. power c. polish d. inspiration e. wealth

Question 75 in the Brown-Carlsen test is, in the opinion of the investigator, a weak question in that the information requested is insignificant and calls for a subjective judgement by the listener. It refers to the ex-schoolteacher, Miss Gavigan, who is the central figure in the story supporting one of the methods to improve reading. The question follows.
Question 75. (original form) "What kind of a person do you think Miss Gavigan was?"
(Answers) f. old-maidish g. honest h. clever i. home-loving j. modest

In seeking a question to replace question 75, the investigator turned to a key idea in the lecture which had not been covered by any query in the original test form, but which was important enough to merit a question of some type. The key idea concerned the observation approach, which is one of the five methods suggested in the lecture for the improvement of reading.

It was determined to phrase the new question 75 so that the purpose of the observation approach would be solicited from the listener. Thus, question 75 became:

Question 75. (revised form) "What is the purpose of the observation approach as a method to improve reading?"
(Answers) f. Increase the numbers of words your eyes can see at a single stop. g. Recognize the main idea. h. Improve comprehension. i. Increase the number of stops your eyes make on a line. j. Increase the accuracy of the inaccurate reader.

The final question in Part E of Form Bm of the Brown-Carlsen Listening Comprehension Test was left unchanged. It is a relatively difficult but challenging question which requires the listener to analyze the structure of the lecture and draw a conclusion. The question is clearly phrased and the answer choices are appropriate.

Question 76. "Which part of the lecture was least directly related to the central idea?"

A summary of question changes for the experimental test form.
The preceding detailed account of the five parts of Form Bm of the Brown-Carlsen Listening Comprehension Test reveals the nature of the
changes made in 47 of the 76 test items. In Part A, Immediate Recall, fifteen of the seventeen questions were changed, and eleven of twenty items were revised in Part B, Following Directions. No modifications were made in the eight queries constituting Part C, Recognizing Transitions, but all ten questions in Part D, Recognizing Word Meanings, were subjected to revision. Of the remaining 21 questions which made up Part E, Lecture Comprehension, eleven items were changed.

The question refinement techniques utilized in the 47 items which were subjected to revision included the shortening of sentences, changes in choices of words, the repositioning of facts within a statement, changes in answer alternatives, and the replacement of certain questions with items designed to be more appropriate.

The method of administering and scoring the experimental test form.--Chapter II of the dissertation noted that the Brown-Carlsen Listening Comprehension Test was designed for personal rather than recorded presentation because Brown thought of listening as being aural assimilation in a face to face speaker-audience situation. It also was felt by Brown that the personal method of test administration would result in a more natural school situation and a more natural teacher-student or speaker-audience relationship. Regular class periods were used for the presentation of test materials in every phase of the development of the instrument.

Brown's thinking on the method of listening test administration has been supported in a number of experimental studies. While Westover found no group differences among college students in performance
on tests administered by listening and by reading, Beighley discovered the visual method of presentation superior to the oral method in his two studies. Gauger stresses the importance of the presence of the speaker in a listening situation, and both Kramer and O'Neill found a statistically significant advantage favoring the speaker who is present over the one who is not visible in their experimental investigations. Johnson and Frandsen tested the sound-on-film approach to administering the Brown-Carlsen Listening Comprehension Test, but learned that either "live" or audio tape administration yields significantly higher means and also lower variability than the sound-on-film method.

The present investigation utilized the personal method of


listening test administration in which the examiner and examinees were present in a normal face to face classroom setting. The subjects were college freshmen enrolled in the various sections of the Basic Speech course offered at Capital University in Columbus, Ohio during the 1963-64 school year. The subjects took the test as a part of the regular work constituting the listening unit in the schedule of Basic Speech assignments. Administrators of the test were the three members of the speech faculty whose course responsibilities included the teaching of the Basic Speech course. All subjects took the test during a regular class session under the administration of the instructor assigned to teach that particular section of Basic Speech. Although almost every section of the Basic Speech course taught during the two semesters of the 1963-64 school year contained a few students whose class rank was higher than that of freshman, the test data used for this investigation included only the results of those tests taken by the freshman members of the speech class sections.

The Brown-Carlsen Listening Comprehension Test--Form Bm was administered in its regular printed form to fifty college freshmen who composed three sections of the Basic Speech course. Each section was taught by a different member of the regular three-man teaching team. The experimental version of the test was administered to three hundred college freshmen who composed a number of sections of the Basic Speech course taught by the three members of the teaching team over a period of two semesters in the 1963-64 school year.

Through careful test design, the experimental test groups were able to use the printed test answer sheets which accompany the
published Brown-Carlsen Listening Comprehension Test--Form Bm for Parts A, B, C, and D of the test. However, the nature of the changes made in the questions and answer alternatives for Part E of the experimental test form necessitated the construction of a separate answer sheet. The separate answer sheet was so constructed that it was similar in design to the regular printed answer sheet. The fifty students making up the control group which took the test in its original form used, of course, the entire printed answer sheet for all parts of the listening test.

The chapter which follows, Chapter IV, will present an analysis of the data received through the administration of the Brown-Carlsen Listening Comprehension Test--Form Bm in its original form and in the experimental form designed for this investigation.
CHAPTER IV

ANALYSIS OF THE DATA

The preceding chapter included a question by question review of the 76 items which make up the five parts of the Brown-Carlsen Listening Comprehension Test--Form Bm. The review indicated that of the total number of items within the test, 47 were subjected to some type of revision for the experimental version of the test form. The purpose of the revisions was to increase the clarity and comprehensibility of the test materials.

In order to investigate the effectiveness of the revisions upon the measuring instrument, the test was administered to two groups of college freshmen. The first group, designated as the control group, consisted of fifty students who took the test in its original printed form. The second group, designated as the experimental group, included three hundred students who took the revised form of the test as it was constructed for this study.

As was pointed out in Chapter II of the dissertation, a total listening score only is derived from the administration of the Brown-Carlsen test. Although the test is divided into five parts, each part measuring to some extent a different aspect of listening skill, it is the total score achieved by the listener which is used for the determination of his listening efficiency. Thus, the analysis of the data
will give attention first to total test results and their significance in relation to the study. The analysis of total test results will be followed by a detailed examination of the 47 revised test items which were included in Parts A, B, D, and E of the experimental test.

Total test results.--Of initial interest to the total study was the general level of listening efficiency demonstrated by the students who composed the experimental group in comparison with the students who were a part of the control group. Through the application of a measure of central tendency to the total test scores achieved by the students in each group, it was learned that the mean or average raw score in the experimental group was 61.6 as compared with 57.6 in the control group.

A meaningful interpretation of the mean raw scores was made possible by a referral to the percentile norms for the Brown-Carlsen Listening Comprehension Test--Form Bm which are presented in Table 5b, page 16 of the printed test booklet. A mean raw score of 61 carries a 91 percentile ranking for a college freshman, while a mean raw score of 57 places a college freshman in the seventy-eighth percentile ranking.

Although the higher mean score with the resulting higher percentile ranking achieved by the experimental group was an encouraging sign of superior listening efficiency because of improved test items, the inadequacies of this measure of the difference between the means of the two groups prompted the use of measuring methods which would give some indication of the spread of the two sets of total test scores and the variation of performance among the members of each group.

It was first noted that the range of raw scores was the same in
both groups. The difference between the lowest and highest scores, plus one unit, was 28. However, the lowest score in the experimental group was 47 as compared with 42 in the control group. The corresponding high scores were 74 and 69, respectively. The maximum high raw score possible in each group was 76.

To further determine the individual differences or variations within the groups, the standard deviation was computed for each group. For the experimental group the standard deviation was 5.11. In the control group the standard deviation was 6.47.

Having obtained the standard deviation for each of the two groups, it became possible through the standard error concept to determine the reliability of the mean of the distribution in each group, and, more important, to determine the reliability or significance of the difference between the two means.

Following the computation of the standard error of the mean for both groups, the standard error of the difference between the mean of the experimental group and the mean of the control group was obtained. The difference was .961. The significance of this difference between the means of the two groups was computed to be 4.089. The difference can be considered a significant one in that a figure of 2.592 is needed for significance at the one per cent level of confidence.

Having learned that the difference between the mean of the experimental group and the mean of the control group was significant at the one per cent level of confidence, the investigator turned to a consideration of item difficulty. The preceding information would seem to indicate that there was a difference in the levels of difficulty of
the two tests which enabled the experimental group taking the revised test form to obtain a higher listening efficiency score than that obtained by the control group taking the test in its original printed form.

As a means of measuring the merit of the assumption concerning test difficulty, a t test for related measures was applied to the data. The 76 items in each of the two test forms were subjected to analysis which revealed the percentage of experimental group subjects and the percentage of control group subjects passing each item in the respective tests. The difference between the percentage passing each item in each test was noted. When the computation was completed, the t test for related measures revealed a value of 3.36, which indicated that the actual difference between the two tests in terms of difficulty was significantly greater than would occur by chance. A value of 2.65 is needed for significance at the one per cent level of confidence.

A final analysis of total test results was applied in the determination of test reliability for the experimental test form. The split-half method of estimating reliability, utilizing the deviation score method of computation to obtain the Pearson product-moment coefficient of correlation, was employed to determine the reliability coefficient of half the test. From this estimate, the reliability of the entire test was estimated by means of the Spearman-Brown prophecy formula.

One hundred cases representing the total population of three hundred experimental group subjects were used in the application of the split-half method. Careful selection of the one hundred cases
resulted in the proportional representation of all scores achieved by the experimental test group. The two halves of the test composed the odd-numbered items and the even-numbered items, respectively.

The Pearson product-moment coefficient of correlation for half the test was .625. The application of the Spearman-Brown prophecy formula indicated a reliability coefficient for the full test of .769, which is significant at the one per cent level of confidence.

**Revised test items results.**—Of particular interest to the investigator were the 47 revised test items which were used in the experimental test form. How effective were the revised items in discriminating between those subjects of high and low listening ability? How difficult were the revised items for the experimental group subjects? What were the differences in difficulty between the revised items and their counterparts in the original printed test? What question refinement techniques used to formulate the revised items were most effective in improving the clarity and comprehensibility of the test questions thus enabling the experimental group subjects to be more successful than the control group subjects in correctly answering the questions? The answers to these questions were sought in a detailed examination of the 47 revised test items.

In describing the effectiveness of test items to discriminate between those subjects of high and low ability, Kelley has shown that the most accurate determination of item validities or internal consistencies can be obtained by comparing approximately the upper and lower 27 per cent of the total population tested. Elimination of the middle 46 per cent leads to results that are more consistent from
sample to sample than those obtained from using all available cases. On the basis of this demonstration, Flanagan has prepared a chart, and also a table, for estimating product-moment correlations from data on percentage succeeding with the task in the upper and lower 27 per cent of the group. The table was reproduced by Thorndike and was used in this investigation to estimate the validity coefficients of the revised test items.

It was decided by the investigator to also use the upper and lower 27 per cent of the group to compute difficulty value for each of the 47 revised items. The per cent passing each item in the upper and lower 27 per cent of the sample were averaged to compute the difficulty values. Eighty-one cases in each category made up the top and bottom 27 per cent of the three hundred experimental group subjects.

Table 1 represents the validity and difficulty data for each of the 47 revised items in the experimental test form. Table 2 presents a summation of the validity and difficulty data for the revised items.

The results appearing in Table 1 indicate that 38 of the 47 revised items carried a correlation coefficient of 21 or higher and, thus, can be considered effective in discriminating between those

---


96 J. C. Flanagan, "General Considerations in the Selection of Test Items and a Short Method of Estimating the Product-Moment Coefficient from the Tails of the Distribution," *Journal of Educational Psychology*, XXX (1939), 674-680.

TABLE 1

VALIDITY AND DIFFICULTY DATA FOR EACH OF THE 47 REVISED TEST ITEMS
IN THE EXPERIMENTAL TEST FORM

<table>
<thead>
<tr>
<th>Item</th>
<th>Upper 27%</th>
<th>Lower 27%</th>
<th>Validity</th>
<th>Difficulty</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PART A</td>
<td>PART A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>100</td>
<td>96</td>
<td>23</td>
<td>98</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
<td>85</td>
<td>47</td>
<td>92</td>
</tr>
<tr>
<td>3</td>
<td>100</td>
<td>93</td>
<td>33</td>
<td>96</td>
</tr>
<tr>
<td>4</td>
<td>99</td>
<td>88</td>
<td>43</td>
<td>93</td>
</tr>
<tr>
<td>5</td>
<td>95</td>
<td>83</td>
<td>24</td>
<td>89</td>
</tr>
<tr>
<td>6</td>
<td>100</td>
<td>95</td>
<td>25</td>
<td>97</td>
</tr>
<tr>
<td>7</td>
<td>100</td>
<td>99</td>
<td>00</td>
<td>99</td>
</tr>
<tr>
<td>8</td>
<td>99</td>
<td>99</td>
<td>00</td>
<td>99</td>
</tr>
<tr>
<td>9</td>
<td>91</td>
<td>71</td>
<td>33</td>
<td>81</td>
</tr>
<tr>
<td>10</td>
<td>80</td>
<td>67</td>
<td>16</td>
<td>73</td>
</tr>
<tr>
<td>11</td>
<td>100</td>
<td>96</td>
<td>23</td>
<td>98</td>
</tr>
<tr>
<td>12</td>
<td>53</td>
<td>41</td>
<td>12</td>
<td>47</td>
</tr>
<tr>
<td>13</td>
<td>100</td>
<td>89</td>
<td>42</td>
<td>94</td>
</tr>
<tr>
<td>14</td>
<td>84</td>
<td>62</td>
<td>28</td>
<td>73</td>
</tr>
<tr>
<td>16</td>
<td>81</td>
<td>68</td>
<td>17</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>PART B</td>
<td>PART B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>99</td>
<td>95</td>
<td>26</td>
<td>97</td>
</tr>
<tr>
<td>20</td>
<td>100</td>
<td>99</td>
<td>00</td>
<td>99</td>
</tr>
<tr>
<td>23</td>
<td>97</td>
<td>81</td>
<td>43</td>
<td>89</td>
</tr>
<tr>
<td>26</td>
<td>100</td>
<td>97</td>
<td>17</td>
<td>98</td>
</tr>
<tr>
<td>27</td>
<td>99</td>
<td>75</td>
<td>58</td>
<td>87</td>
</tr>
<tr>
<td>29</td>
<td>96</td>
<td>89</td>
<td>21</td>
<td>93</td>
</tr>
<tr>
<td>32</td>
<td>96</td>
<td>70</td>
<td>46</td>
<td>83</td>
</tr>
<tr>
<td>33</td>
<td>97</td>
<td>90</td>
<td>24</td>
<td>94</td>
</tr>
<tr>
<td>34</td>
<td>85</td>
<td>60</td>
<td>31</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>PART C</td>
<td>PART C</td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>67</td>
<td>55</td>
<td>12</td>
<td>61</td>
</tr>
<tr>
<td>36</td>
<td>94</td>
<td>78</td>
<td>31</td>
<td>86</td>
</tr>
<tr>
<td>46</td>
<td>96</td>
<td>81</td>
<td>35</td>
<td>89</td>
</tr>
<tr>
<td>47</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>90</td>
</tr>
<tr>
<td>48</td>
<td>97</td>
<td>88</td>
<td>28</td>
<td>92</td>
</tr>
<tr>
<td>49</td>
<td>96</td>
<td>94</td>
<td>08</td>
<td>95</td>
</tr>
<tr>
<td>50</td>
<td>70</td>
<td>50</td>
<td>21</td>
<td>60</td>
</tr>
<tr>
<td>51</td>
<td>99</td>
<td>91</td>
<td>38</td>
<td>95</td>
</tr>
<tr>
<td>52</td>
<td>62</td>
<td>41</td>
<td>21</td>
<td>51</td>
</tr>
<tr>
<td>53</td>
<td>78</td>
<td>53</td>
<td>28</td>
<td>65</td>
</tr>
<tr>
<td>54</td>
<td>89</td>
<td>64</td>
<td>34</td>
<td>76</td>
</tr>
<tr>
<td>55</td>
<td>99</td>
<td>93</td>
<td>33</td>
<td>96</td>
</tr>
<tr>
<td></td>
<td>PART D</td>
<td>PART D</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PART E</td>
<td>PART E</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

67
### TABLE 2
VALIDITY AND DIFFICULTY DATA FOR REVISED ITEMS IN THE EXPERIMENTAL TEST FORM

<table>
<thead>
<tr>
<th>Part</th>
<th>Number of Revised Items</th>
<th>Validity Range</th>
<th>Validity Mean</th>
<th>Difficulty Range</th>
<th>Difficulty Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>15</td>
<td>00-47</td>
<td>24.4</td>
<td>47-99</td>
<td>86.9</td>
</tr>
<tr>
<td>B</td>
<td>11</td>
<td>00-58</td>
<td>28.1</td>
<td>61-99</td>
<td>87.2</td>
</tr>
<tr>
<td>C</td>
<td>No revisions made</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>10</td>
<td>08-40</td>
<td>28.6</td>
<td>51-96</td>
<td>81.4</td>
</tr>
<tr>
<td>E</td>
<td>11</td>
<td>23-62</td>
<td>38.9</td>
<td>50-98</td>
<td>79.1</td>
</tr>
<tr>
<td>Total</td>
<td>47</td>
<td>00-62</td>
<td>30.0</td>
<td>47.99</td>
<td>83.6</td>
</tr>
</tbody>
</table>

Validity indices are approximations of the item-total score correlation obtained from the upper-lower 27 per cent of the group by means of the Flanagan table. Difficulty values for each item were computed by averaging the per cent passing each item in the upper and lower 27 per cent of the group.

subject of high and low listening ability. Of the remaining nine items, the correlation coefficients were as follows: two items with 17, one item with 16, two items with 12, one item with 08, and three items with 00.

In Table 2, the mean difficulty values for the revised items in Parts A, B, D, and E of the experimental test form indicate an unusually high degree of success by the subjects in answering correctly the revised items.

For a more meaningful look at item difficulty, it was decided to subject each part of the listening test that contained revised items to a t for related measures. Although a total test computation
utilizing the \( t \) test for related measures had already revealed a significant difference between the experimental test and the original test in terms of difficulty, an analysis of only the revised items and their original test counterparts as they appeared in Parts A, B, D, and E of the two tests would reveal information more specific in nature.

The same procedure was employed in the \( t \) test computation for revised items only as was utilized in the \( t \) test computation for the total test referred to earlier in the chapter. Total test populations of the two groups were used in both computations.

In Part A, Immediate Recall, the actual difference between the items of the two tests in terms of difficulty was significantly high. A value of 4.257 was revealed when a value of only 2.947 was needed for significance at the one per cent level of confidence.

The greater degree of success experienced by the experimental group subjects in correctly answering the revised items than that experienced by the control group subjects in correctly answering the original test items can be attributed to the particular question refinement techniques used in Part A which were described in detail in Chapter III. A brief review will indicate the impact of the particular changes upon test results.

In questions one, two, five and nine of Part A, an unimportant and incidental phrase was deleted from the questions so as to shorten the question statements and improve their clarity. In three of the four cases, a higher percentage of the experimental group subjects than the control group subjects answered the questions correctly.

Questions seven through twelve in Part A utilized a series of
two-letter words which were read aloud by the examiner. Some of the two-letter words were very similar in sound, and it was felt by the investigator that confusion could result for the listener if the examiner were inaccurate in his enunciation of the words. To avoid similarity in sound on some of the two-letter words and to improve the clarity of the questions read by the examiner, one of the two-letter words was replaced by a more distinctive sounding two-letter word, and the words in each of the series were alternated in such a manner that the words beginning with a vowel or ending in the same sound were not placed in immediate succession.

In five of the six questions subjected to this modification, a higher percentage of experimental group subjects than control group subjects were successful in correctly answering the items. In the case of question nine, which was changed by both the above described modification and the omission of an incidental phrase, 83 per cent of the experimental group as opposed to 56 per cent of the control group correctly answered the item.

The third general change in Part A involved the rephrasing of questions three, four, six, eight, eleven, and sixteen, so that the information desired appeared in the questions before rather than after a series of numbers or words were read. The result of this change in question wording was that the difference in the percentage of experimental group subjects passing the items and control group subjects passing the items ranged from five per cent to 33 per cent. Obviously, the listeners in the experimental group benefitted from the revision.

The final change in Part A involved items in which the listeners
were required to listen to a set of directions and then reply to questions based upon the information contained in the directions. In the revised form of the test, the statement of directions was rephrased so that the information requested in the questions was in a "recency" position as opposed to a "primacy" position used in the original test.

Questions thirteen and fourteen of Part A reflected this particular form of refinement, and the test results indicated a definite superiority for information in the recency position. In item thirteen, 96 per cent of the experimental group subjects answered the question correctly while only 76 per cent of the control group subjects were successful. In question fourteen, the passing percentage figures were 76 for the experimental group and 40 for the control group.

Questions fifteen and seventeen in Part A were not revised in any way.

In Part B, Following Directions, eleven of the twenty items were revised by changing the terminology within the questions so that the most simple and clear language of mathematical figuring would be used. Phrases such as "the number one less than" and "the number one greater than" were replaced with the phrases "subtract one from" and "add one to."

The t test for related measures showed a difference in terms of difficulty between the experimental and control group test forms, but the value derived for Part B was 2.025 with a value of 2.201 needed for significance at the five per cent level. Although the derived value was not quite high enough to be significant, a greater percentage of the experimental group subjects were successful in correctly
answering eight of the eleven revised items than were the control group subjects in correctly answering the questions in their test form which were the counterparts of the revised items. Thus, the changes in terminology introduced into eleven of the twenty Part B questions were helpful in making eight of the eleven items clearer to the listener, but the increase in the percentage of subjects correctly answering the items as a result of the changes was not sufficient enough to be significant.

As was indicated in Chapter III, no revisions were made in the eight questions composing Part C, Recognizing Transitions.

The basic change in the wording of the questions which comprise Part D of the Brown-Carlsen Listening Comprehension Test--Form Bm, was introduced for experimental inquiry purposes only. Part D calls for the recognition of word meanings from context. The ten sentences in Part D are so phrased that the word for which the listener must determine a meaning is stated before the sentence containing the word is read. An example of this phrasing is, "What does PART mean in the sentence, 'What part did John act in the play?'" In the experimental test form it was decided to rephrase each of the ten sentences so that the word for which the listener had to determine a meaning would occur after the sentence containing the word had been read by the examiner. An example of the revised phrasing is, "In the sentence, 'What part did John act in the play?' what does the word PART mean?"

The results of the application of the t test for related measures to the ten questions in Part D indicated an insignificant difference in terms of difficulty between the experimental test form and
the original test form. However, the difference revealed by the statistical measure indicated that the subjects taking the test in its original form were more successful in correctly answering the questions than were the subjects using the experimental test form with the revised wording of questions. Thus, the question which is phrased so that the word for which the listener must determine a meaning is stated before the sentence containing the word is read elicits a correct response more frequently than when the word for which the listener must determine a meaning is stated after the sentence containing the word is read.

The fifth and final section of the test, Part E, Lecture Comprehension, contains 21 questions, eleven of which were modified for the experimental test form. The \( t \) test for related measures indicated some difference in difficulty between the eleven revised items and their counterparts in the original test form, but the difference was not significant. The value derived was .961.

Because of the variety of the changes made in eleven of the 21 questions in Part E, a detailed examination of the eleven questions in relation to test results seemed important to the analysis of data.

The first question in Part E, Lecture Comprehension, is number 56, and it asks for a detail contained in the lecture. The question reads, "What relative of Lincoln was mentioned in the Lincoln story?" The answer alternatives include mother, father, half brother, stepmother, and sister. The dilemma which confronts the listener when attempting to choose the correct answer alternative is that both the father and the stepmother are mentioned in the Lincoln story. To
further complicate the question, the stepmother is referred to as "stepmother" in one sentence of the lecture, but as "mother" in a following sentence.

To improve question 56, two changes were made in the experimental test form. First, in the lecture content, the word "mother" was changed to "stepmother" so that the two references to the woman would be consistent. Second, the question was reworded to read "What relative of Lincoln guided his reading when Lincoln was a youth?"

The results of the changes are reflected in the fact that 70 per cent of the experimental group subjects correctly answered the question while only 58 per cent of the control group subjects passed the item.

The second question to be revised in Part E was question 57. It asks for an unimportant detail mentioned only once in the lecture. In the experimental test form the question was revised to ask for a more important detail from the same lecture illustration. The result of the change was that 97 per cent of the experimental group subjects answered question 57 correctly while only 52 per cent of the control group subjects passed the item.

A modification in the choice of words in three of the five answer alternatives, so that the terminology would be more specifically aligned with the language used in the lecture, constituted the change for question sixty in Part E. The effect of the revision was very minimal. Eighty-seven per cent of the experimental group and 86 per cent of the control group correctly answered the item as it appeared in the respective test forms.
The next four questions in Part E, questions 61 through 64, underwent complete reconstruction for the experimental form of the test. The results of the changes were varied.

In question 61, the item was rephrased to ask for a detail mentioned four times in a lecture illustration. The original question asks for a detail from the same illustration which is mentioned only once. The positive effect of repetition of information upon a listener was evident in the test results which revealed that 92 per cent of the experimental group subjects correctly answered the revised question while only 56 per cent of the control group subjects passed the item in its original form.

Question 62 in the original test asks for another detail from the same section of the lecture covered by question 61. It was noted by the investigator, however, that the original test provides no question to bring out the key idea of this part of the lecture. It was thus determined that a question focusing attention upon a key idea in the lecture would be more appropriate in a testing situation than a question asking for a once-mentioned detail, particularly since one detail question had already been formed for the portion of the lecture under consideration. So, in the experimental form of the test, question 62 was changed to place emphasis upon a key idea rather than a detail. Test results indicated that the key idea question was correctly answered by 87 per cent of the experimental group while 88 per cent of the control group passed the detail question.

The changes made in items 63 and 64 of the original test resulted in more difficult questions for the experimental group. Both
of the aforementioned questions in the original test request a detail from one of the lecture illustrations. Each detail is mentioned once in the illustration. Analysis of the two questions suggested that one question should be retained in some form so that a detail from this portion of the lecture would be a part of the test questions. It was determined that question 64 concerned a more important detail than question 63, and that the question could be strengthened further by rewording it to request a clearly stated principle contained in the illustration rather than a detail which served as a subordinate point to the principle. The result of the change was that it was more difficult for the experimental group to remember the principle than it was for the control group to remember the detail. Seventy-three per cent of the experimental group and 86 per cent of the control group were successful with question 64.

The decision of what to do with question 63 was answered by the omission in the original test of a key idea question relating to one of the major points covered early in the lecture. It was felt that a question alluding to the major point should be constructed for inclusion in the experimental test form and the less important question 63 be deleted in deference to the newly constructed item. The result of this change was a more difficult question for the experimental group. While 94 per cent of the control group answered the original question 63 correctly, only 70 per cent of the experimental group was successful with the newly constructed item.

In question seventy, the next item subjected to revision in Part E, it was noted that the correct answer alternative in the original
test is worded in such general terms that the answer could apply to several principal points contained in the lecture. For the experimental test form, a more specific wording of the alternative was constructed which made the answer applicable only to the information covered by the particular test question under consideration. The result of the change was a slightly higher percentage of experimental group subjects passing the item than control group subjects. For the experimental group, 52 per cent of the subjects were successful in correctly answering the question. For the control group, the figure was 48 per cent.

The revision made in question 71 resulted in one of the largest differences in percentage of subjects between the two groups correctly answering the question. The original question was deemed by the investigator to be inaccurate in what it implies about the lecture illustration to which it refers. A new question was constructed for the experimental test form which called for factual information contained in the illustration. As a result of the revised question, 98 per cent of the experimental group subjects were successful in answering the query. Only 62 per cent of the control group subjects correctly answered the original question 71.

An attempt to improve the wording of the correct answer choice constituted the only change in question 73. The revised wording utilized the language of the lecture which expressed the information being tested in item 73. The revised wording had some effect, as 83 per cent of the experimental group responded correctly in comparison with 76 per cent of the control group.
Question 75 was the last question to be changed in Part E, Lecture Comprehension. The original question seems to be a weak question in that the information requested is insignificant and calls for a subjective judgement by the listener. However, the key idea question utilized in the experimental test form proved to be more difficult than the original question 75 which it replaced. Only 62 per cent of the experimental group passed the item, but 86 per cent of the control group correctly answered the question.

The preceding analysis concerning the fifteen revised items in Part A, Immediate Recall, the eleven revised items in Part B, Following Directions, the ten revised items in Part D, Recognizing Word Meanings, the eleven revised items in Part E, Lecture Comprehension, and their question counterparts in the original test form, has presented detailed information with interesting results. Before turning to Chapter V and a summary of the data analysis with conclusions, a final application of the $t$ test for related measures should be reported.

Whereas the previous applications of the $t$ test for related measures have concerned separately the four parts of the listening test counterparts, the application of the measure now reported involved all 47 revised items and their original test counterparts in a total grouping.

The results of the computation revealed a value of 3.158, which indicated that the actual difference between the 47 revised items and their original test counterparts in terms of difficulty was significantly greater than would occur by chance. A value of 2.683 is needed for significance at the one per cent level of confidence.
CHAPTER V

SUMMARY AND CONCLUSIONS

The purpose of this experimental study was to apply to a standardized test of listening comprehension certain question refinement techniques designed to improve the clarity and comprehensibility of the test materials, and to note any changes in the efficiency of listener response that could be attributed to the revised form of the standardized tool. Motivating the investigation were the reports of the level of efficiency demonstrated by people who were tested on their ability to listen to spoken communication in a setting in which both oral and visual cues were present. The reported listening efficiency level was approximately 50 per cent.

It was felt by the investigator that a principle of test construction known as appropriate test difficulty was related to the reported test results indicating a particular level of listening efficiency. The principle states that, as a general rule, the average difficulty of the items in a test should be such that about half of the subjects will answer them correctly. The most accurate test, according to this principle, is one composed of items all of which are at the 50 per cent level of difficulty.

The question arose as to whether the reports of listening efficiency reflected accurately the ability of people to listen or the
ability of test developers to construct instruments at a 50 per cent difficulty level. Also, could a standardized measuring device with the 50 per cent difficulty level principle designed into it be improved through the application of basic question refinement techniques so as to yield results with a higher efficiency level and still be recognized as a reliable tool?

Summary of data analysis.—Chapter IV reported the results of the statistical analysis of the data used in the present investigation. Considered first were total test results and their significance in relation to the study. The major findings of the analysis of total test results are reported below.

1. With a maximum raw score of 76 possible in each test form, a mean raw score of 61.6 was achieved by the experimental group subjects taking the revised test form. The mean raw score for the control group subjects taking the original form of the test was 57.6. The significance of the difference between the means of the two groups was 4.089, which is significant at the one per cent level of confidence.

2. The actual difference between the two test forms in terms of item difficulty was significantly greater than would occur by chance. A t test for related measures revealed a value of 3.36, which is significant at the one per cent level of confidence.

3. The reliability coefficient for the experimental test form was .769, which is significant at the one per cent level of confidence.

A detailed examination of the 47 revised test items which were included in Parts A, B, D, and E of the experimental test form resulted in the findings reported below.
1. Thirty-eight of the 47 revised items had a correlation coefficient of 21 or higher as computed by means of the Flanagan table, and thus, can be considered effective in discriminating between subjects of high and low listening ability.

2. The mean validity index for the 47 revised items was 30.0.

3. The mean difficulty index for the 47 revised items was 83.6.

4. The actual difference in terms of difficulty between the 47 revised items and their original test counterparts was significantly greater than would occur by chance. A t test for related measures revealed a value of 3.158, which is significant at the one per cent level of confidence.

5. In Part A, Immediate Recall, the question refinement techniques were very effective in making the actual difficulty difference between the revised items and their original test counterparts significant. The shortening of question statements, the avoidance of similarity of sound in a series of orally presented two-letter words, the rephrasing of question sentences, and the use of the "recency" principle for the repositioning of facts to be recalled from a set of directions, contributed to the significantly high difficulty difference.

6. In Part B, Following Directions, the introduction of the most simple and clear language of mathematical figuring into eleven of the twenty items resulted in a greater percentage of experimental group subjects than control group subjects correctly answering the
items, but the value denoting the difference between the percentages fell just below the level of significance.

7. The revisions made in Part D, Recognizing Word Meanings, were not effective in improving the clarity of the questions. The original test questions which were so phrased that the word for which the listener must determine a meaning is stated before the sentence containing the word is read elicited a correct response more frequently than when the word was stated after the reading of the sentence, as was done in the experimental test form.

8. The variety of the revisions made in eleven of the 21 questions in Part E, Lecture Comprehension, created varied results in terms of item difficulty, but all eleven revised items were effective in discriminating between those subjects of high and low listening ability. The questions which benefitted most in terms of difficulty from refinement techniques were three detail questions in which requests for unimportant detail information were replaced by requests for important details mentioned several times in an illustration, and one inaccurate question requiring the drawing of an inference by the listener which was replaced by a question requesting factual information. Moderate improvement resulted from an attempt to increase the clarity of answer alternatives in three questions. Several revisions involving the replacement of detail questions with key idea queries caused the items to be more difficult in their revised form than they had been in their original form.
Conclusions.--On the basis of the above reported findings of this investigation, the following can be concluded:

1. Listening efficiency scores resulting from the administration of a listening comprehension test reflect, in part, the difficulty of the items comprising the measuring instrument.

2. The interpretation of listening efficiency scores should be based upon the difficulty index of the particular tool employed to measure listening skill.

3. A reliable measure of listening comprehension can be developed from a standardized test of listening comprehension which will yield a higher level of listening efficiency than that yielded by the standardized measure through the application of basic question refinement techniques which decrease item difficulty by increasing the clarity and comprehensibility of the items.

4. The question refinement techniques which were most effective in improving item clarity and reducing item difficulty in this investigation are listed below in the order of their effectiveness:

   a. the repositioning of facts within a statement of directions so that the information requested in the questions was in a "recency" position as opposed to a "primacy" position,

   b. the replacement of questions requesting a once-mentioned lecture detail with questions requesting a detail repeated several times in the lecture,

   c. the replacement of questions requiring the drawing of an inference by the listener with questions seeking factual information,
d. the rephrasing of questions so that the information desired appeared in the question before rather than after a series of numbers or words were read,

e. the utilization of the most simple and clear language in the wording of questions,

f. the shortening of question stems through the removal of unimportant and incidental phrases.
BIBLIOGRAPHY

Books


**Articles**


Brown, James I. "Vocabulary--Key to Communication," Education, LXXX (October, 1959), 80-84.


Nichols, Ralph G. "Do We Know How to Listen? Practical Helps In a Modern Age," The Speech Teacher, X (March, 1961), 118-124.


Unpublished Material


Other Sources
