DURING THIS PERIOD FROM SEPTEMBER 1 THROUGH NOVEMBER 30, 1966, 35 FIELD CONSULTANTS COMPLETED ANALYSIS OF THEIR INDIVIDUAL TRADE AND TECHNICAL CURRICULUMS. THESE ANALYSES WERE DEVELOPED INTO AN OUTLINE TO SERVE AS A GUIDE FOR DEVELOPING ACHIEVEMENT TESTS. THE FINAL OUTLINE WAS DIVIDED INTO AS MANY DIFFERENT AREAS AS THE CONSULTANTS FELT NECESSARY TO REPRESENT INDEPENDENT AREAS OF INSTRUCTION. EACH INDEPENDENT AREA WAS THEN SUBDIVIDED TO DETAIL THE VARIOUS ELEMENTS OF THE CURRICULUM. THE BREAKDOWN WILL BE USED TO INSURE A TEST ITEM POOL REPRESENTATIVE OF ALL AREAS OF THE CURRICULUM. REFERENCE TESTS WITH KNOWN RELIABILITY AND VALIDITY ARE BEING EVALUATED TO DETERMINE WHICH WOULD BE MOST APPROPRIATE IN THE INITIAL TEST BATTERY AS A BASIS FOR COMPARING PROJECT TEST RESULTS. TWELVE NORTH CAROLINA TECHNICAL INSTITUTES PARTICIPATED IN THE INITIAL PHASE OF THE PROJECT, AND 10 ADDITIONAL INSTITUTIONS HAVE AGREED TO PARTICIPATE IN ADMINISTERING THE PRELIMINARY FORMS OF THE TEST. METHODS AND TECHNIQUES OF EVALUATING THE PRELIMINARY FORMS AND THE PROBLEMS OF DATA STORAGE AND RETRIEVAL HAVE BEEN STUDIED. EXPERIMENTAL STUDIES OF KINESTHETIC SENSITIVITY WERE CONDUCTED AND APPARATUS BUILT TO INVESTIGATE THIS AREA IN MORE DETAIL. AUDITORY RESPONSE AND ITS RELATIONSHIP TO ACHIEVEMENT IN SEVERAL OF THE VOCATIONAL AREAS WERE INVESTIGATED, AND HIGH-FIDELITY RECORDING EQUIPMENT HAS BEEN REQUISITIONED TO RECORD NORMAL AND ABNORMAL SOUNDS FOR DIAGNOSTIC WORK IN AUTO MECHANICS, MACHINE SHOP, AND RADIO-TV REPAIR. TWO PAPERS ARE GIVEN IN THE APPENDIXES — "USE OF OBJECTIVES IN ITEM CONSTRUCTION" AND "AUDITORY DIAGNOSIS STUDY." (HC)
THE DEVELOPMENT OF ACHIEVEMENT MEASURES
FOR TRADE AND TECHNICAL EDUCATION

Thomas S. Baldwin

Office of Education Grant Number OEG-2-6-000517-0585
The Vocational Education Act of 1963, P.L. 88-210,
Section 4(c)

North Carolina State University at Raleigh
Raleigh, North Carolina

1 Sept. - 30 Nov. 1966
Since September 1, all additional instructors needed as field consultants have been contacted and placed under contract. Each of these instructors was visited and the objectives and operating procedures of the research project were explained in detail. At this time, the total number of instructors working as field consultants number approximately 35 for an average of five field consultants per curriculum.

Each instructor working on the project was asked to do a complete analysis of that portion of the curriculum considered to be the trade or technical specialty. This analysis was completed by mid-October. During the last three weekends in October, meetings were held with the field consultants in each of the seven technical or trade curricula included in the research project. The purpose of these meetings was first to arrive at a consensus concerning the curriculum analysis, secondly to plan ways and means of developing performance tests, and thirdly to evaluate items which had already been developed for the test item pool.

A consensus concerning the curriculum analysis was reached at the meetings, and the resulting outline is to serve as a guide in the development of the achievement tests. The final outline was broken into as many different areas as the field consultants felt necessary to represent what they considered to be independent areas of instruction. Each independent area was then subdivided to detail the various elements of the curriculum contained therein. This breakdown will be used to insure that items for the item pool will be representative of all areas of the curriculum.

At each of the meetings, the concept of performance measurement was discussed. In some of the curricula, particularly the two technical curricula, doubts were voiced as to the appropriateness of performance
measurement. Instructors in these areas felt that the objectives of these curricula were the assimilation, understanding and application of technical knowledge with minimum emphasis on the development of manual skills.

The construction of test items is well underway for each area. As instructors develop items, the items are sent to the project staff. The items are then edited and evaluated in terms of grammatical correctness, format consistency and adherence to accepted principles of test construction. Finally, the items are being reproduced in booklet form for final evaluation by the committee of consultants.

A meeting of all field consultants has been scheduled for Friday night and Saturday, December 2 and 3. At this meeting, all items which have been processed as described above will be evaluated by the committee in terms of appropriateness of item, correctness of answer, and validity of item detractors. All items which pass this evaluation will be placed in the item pool from which tests will be compiled.

As stated in a previous progress report, twelve technical institutes and/or community colleges agreed to participate in the initial phase of the research project. Since that time, all other institutions within the North Carolina Department of Community Colleges which offer one or more of the curricula included in the research project have been contacted and asked to participate in the second phase of the research project. The second phase of the project will include the administration of the preliminary forms of the achievement tests. To this time, an additional ten institutions have agreed to participate in the initial testing program. One institution declined to participate on the grounds that this is its first year of operation and to this time has been unable
to complete the move into its permanent facilities. Twelve institutions consisting primarily of units of technical institutes or community colleges have not been heard from to this date.

In order to have some means of comparing student scores on preliminary forms of achievement tests developed by this project with an instrument of known reliability and validity, it was decided by the staff that each achievement battery should contain one or more reference tests, either in the area of achievement or aptitude, which possesses a high degree of reliability and a known validity. To determine the test or tests which would be most appropriate for each curriculum an extensive review of aptitude and achievement tests in print is underway. Tests in six different aptitude areas are being evaluated. These aptitude areas are: numerical ability, mechanical ability, verbal reasoning, linguistic ability, spatial reasoning, abstract reasoning. Standardized achievement tests in certain related areas such as mathematics, physical sciences and English are also being evaluated.

Problems concerning methods and techniques of evaluation of preliminary forms of the achievement tests have also been considered. At this time, it has been decided that various factor analyses will be performed between the various subtests in each curriculum. It was recognized that the weighting of the factor analyzed variables would be very important in the final forms of the achievement tests. Various weighting techniques are being studied and compared.

The problem of data storage and retrieval is being studied concurrently with methods and techniques of data evaluation. A study is currently underway to determine what types and quantities of equipment
such as card punches, sorters, and test scoring equipment will be needed to process for computer evaluation the large quantity of data anticipated from the first administration of the achievement tests.

As mentioned in the preceding progress report, the task of categorizing the various levels of learning and behavioral objectives as to number and kind presented a most difficult problem. The Taxonomy of Education Objectives, Cognitive Domain by Benjamin Bloom was examined. Due to the fact that there is much disagreement and confusion among psychologists and educators in applying this system to test items, it was decided by the staff to investigate a system of classification of educational objectives which would be more suitable to our needs. The system finally decided upon was a modification of Bloom's Taxonomy and consists of four levels of classification. These four levels of classification are: knowledge, comprehension, and two levels of application. Operational definitions of the four levels of classification have been developed and a paper for use by the staff and field consultants has been prepared. (Appendix A).

In the experimental area, much time was devoted to the tactile kinesthetic sensitivity experiment. As stated in the previous progress report this study involved the ability to discriminate small differences in weight. All attempts to replicate the study of Fleishman and Rich (1961) proved futile.

From our preliminary studies, it seems a difference in weight as small as two grams cannot be reliably discriminated. Our investigations indicated a six gram difference to be the smallest which could be reliably discriminated. Using 29 introductory psychology students as subjects yielded a 0.67 reliability coefficient on a 24 hour test-retest.
This result was obtained by correlating the total number of incorrect choices on the first day with those of the second day. The use of difference limens on the same data yielded reliabilities that were not significantly different from zero.

The above study was performed under what we consider to be crude circumstances in that the subjects were actually lifting the weights. The apparatus which we have designed for testing tactile kinesthetic response was described in the previous progress report; this apparatus will be ready for use in the near future and will, we feel, overcome some of the difficulties of the initial study. Other methods which have more face validity in the vocational setting are being designed to test tactile kinesthetic response. These will include the use of feeler gages, micrometers and rods and bushings.

Much planning and investigation has been done in the area of auditory response (Appendix B), and its relationship to achievement in several of the vocational areas. Test recordings have been made of several normal and abnormal sounds produced by an operating automobile engine. The equipment used was not of the quality needed for recording for test purposes, but we have gained insight and experience. This appears to be a fruitful area for further investigation. Suitable quality recording and playback equipment has been requisitioned, and plans are underway for further recording in the areas of automotive mechanics, machine shop, and radio and TV repair.
A meeting of the advisory committee was held on October 25, 1966. A progress report was given and problems which we have encountered were discussed. The possibility of testing students in states other than North Carolina was discussed. The idea was enthusiastically endorsed by the advisory committee. We are currently studying the problems to ascertain how much additional testing can be done in other states within the confines of our existing budget and with the personnel currently assigned to the project.
APPENDIX A
USE OF OBJECTIVES IN ITEM CONSTRUCTION

Once concepts and skills (subject matter content) to be learned have been identified, a further breakdown in terms of intended behaviors must be accomplished in order to complete the specification of instructional objectives. It is suggested here that the intended behaviors might be classified as knowledge, understanding and application. There are two kinds of application (explained below), making, in all, four broad classifications of intended behaviors. Ideally, achievement tests require all of the intended behaviors, therefore, operational definitions of the four classifications, as they pertain to testing, are given in this paper.

Items of a test which measure knowledge require the repetition of responses that have been or should have been practiced in learning experiences prior to the time of taking the test. After learning, memory is the major requisite to correct performance.

Examples:

1. An alphabetic character in the Hollerith card code consists of:
   a. One numeric punch
   b. Two numeric punches
   c. One zone and one control punch
   * d. One numeric and one zone punch

2. A control punch in a card is used:
   a. To define fields
   * b. To identify different types of cards
   c. To control key punching
   d. To prevent punching errors

3. Identify the following schematic symbols:
   a. Capacitor, resistor, coil
   b. Condensor, inductor, rheostat
   c. Capacitor, coil, potentiometer
   * d. All of the above

4. Ohm’s Law is expressed by the following equation:
   a. \( W = E / R \)
   b. \( E = (W/R)^{1/2} \)
   c. \( EI = R \)
   * d. \( E = IR \)
Items of a test which measure understanding require responses in addition to those previously practiced and learned. The additional responses are likely to be interpretations, translations, summarizations, analyses, detection of similarities, detection of differences, etc. Items at this level do not set tasks which require solutions other than explorations in meaning.

Examples:

1. In storage, instruction and data:
   a. Must be separated by 100 positions of storage
   b. Should be grouped by length of field
   c. Can occupy the same locations simultaneously
   *d. Should be assigned separate areas

2. A computer can make logical decisions by:
   a. Comparing
   b. Arithmetic results
   c. Testing signs or characters
   *d. All of the above

Items of a test which measure application require the use of previously learned responses in the solution of problems set by the items. At this level, which we shall call application I, the problems are not new having been experienced by the testee before to the extent that responses that are necessary to find solutions are more or less routine.

Examples:
1. Card Type

2. Read A Card

3. Test Switch

Off

4. Turn Switch On

5. Add 5 to Quantity Accumulator

6. Test Card Type

7. Add 5 to Quantity Accumulator

8. Subtract 5 From Quantity Accumulator

9. Test Counter

10. Add 1 to Counter

11. Subtract 1 from Counter

12. Print Contents of Quantity Accumulator

13. Punch Card

14. Halt
Answer the following questions based on the above flow-chart. Assumptions are:
(1) switch is off initially, (2) counter is zero initially, (3) quantity is zero initially, (4) file consists of 10 cards. Each card represents a quantity of 5.

1. How many cards will be processed?
2. What is the total quantity printed on the first line?
3. What is the total quantity printed on the second line?
4. What is the final quantity amount?
5. How many cards will be punched?
6. What is the status of the switch when program halt is reached?
2. A dual trace oscilloscope is connected to points "P" and "Q" in the grounded base amplifier circuit shown below. Select the pattern you would see on the oscilloscope.

3. Calculate the time constant of the following circuits

\[ T_1 = \frac{1 \text{ microsecond}}{10 \text{ meg ohms}} \]

\[ T_2 = \frac{100 \text{ millihenries}}{10 \text{ kilohms}} \]

* A. \[ T_1 = 1 \text{ second}, \ T_2 = 0.0001 \text{ seconds} \]
B. \[ T_1 = 1 \text{ second}, \ T_2 = 10 \text{ seconds} \]
C. \[ T_1 = 1 \text{ microsecond}, \ T_2 = 1 \text{ microsecond} \]
D. \[ T_1 = 1 \text{ millisecond}, \ T_2 = 1 \text{ millisecond} \]
Given the following resonant, High "Q", circuit

\[ L = 9 \text{ mh} \]
\[ C = 0.01 \text{ uf} \]
\[ \text{Res. freq.} = 1.68 \text{ MHz} \]

1. At Resonance the meters, M1 and M2, will read as follows
   *A. M1 less than M2
   B. M2 less than M1
   C. M2 same as M1
   D. Can't tell

2. The inductor is adjusted in such a manner that "L" is reduced to 1 mh. The resulting frequency will be
   A. 15.1 MHz
   B. 0.18 MHz
   C. 5.04 MHz
   * D. 0.56 MHz

Application II is unlike application I in that solution requires responses of the understanding level. At least one element of the problem is new to the testee. The newness might appear in either the conditions of the problem or in the solution required, that is, a new, novel solution might be required.

Examples:
Answer the following questions based on the above programs.
Assumptions: (1) No other information is stored in memory,
(2) instructions are stored consecutively starting with position
06000, (3) instructions are 1620 machine language.

1. After instruction #2 is executed, what data is stored in
locations 01001 and 01002?

2. After instruction #3 is executed, what data is stored in
location 01001?

3. What is the purpose of instruction #5?

4. What memory positions contain the digit 6?

5. How many cards are punched? What is the code?

6. What character is printed?

7. What is the purpose of the last instruction?
Given the following circuit -
A Free-running multivibrator

1. Which of the following statements describe the circuit's operation?
   A. The operation is non-symmetrical.
   B. The time constants for \( C_1R_7 \) and \( C_2R_5 \) are not equal
   C. The period of operation is 110 milliseconds
   D. All of the above
2. Two meters, \( M_1 \) and \( M_2 \), are placed in an RF amplifier circuit as shown. The tuned circuit in the plate circuit is adjusted to resonant frequency. Each meter will

\[\text{Diagram of RF amplifier circuit with meters} \]

A. \( M_1 \) will increase, \( M_2 \) will decrease
B. \( M_1 \) will increase, \( M_2 \) will increase
C. \( M_1 \) will decrease, \( M_2 \) will decrease
* D. \( M_1 \) will decrease, \( M_2 \) will increase
One of the most apparent findings to come out of our initial investigations into the components of skill development is the emphasis placed upon diagnostic ability by the various trades and technologies. This ability, rather than simple manipulative ability, is viewed by many as being the differentiating factor between "good" and "average" students. At first glance diagnostic ability might seem to depend entirely on intellectual capability. Yet, while this is recognized as being of considerable importance, it seems that much more is necessary to make a good diagnostician. It is obvious that the senses, such as auditory, visual, kinesthetic, olfactory, etc., play a major role in the diagnostic process. A good diagnostician may spend as much time listening to and touching a malfunctioning engine as actually using the numerous mechanical tools available to him. At present we are conducting studies in the determination of kinesthetic sensitivity and we now feel our next step is to investigate the auditory sense. This proposed step has been reinforced by the enthusiasm shown by both members of vocational education and industry, including General Motors, Chrysler, and Ford.

Our choice of the auditory sense is made on several accounts; (1) in our discussions with instructors in certain areas the consensus of opinion seems to be that the auditory sense is invaluable for diagnostic ability, (2) it is a sense which we feel we can adequately investigate given our resources, (3) it is a sense, unlike smell, taste, temperature, etc., which can feasibly be tested outside of the laboratory situation.

The primary question to be answered here is whether certain diagnostic sounds do discriminate among students, and if so, whether this discrimination is independent of that made by paper and pencil tests.

At present indications are that auditory diagnosis is important in the following areas being studied under this grant: auto mechanics, machine shop,
radio-TV repair, and air conditioning-refrigeration. The area of auto mechanics is used below as being illustrative of the work which will be undertaken in this part of the study.

**Procedure**

The initial phase of this study will be to have auto mechanic instructors in the various institutions identify those sounds which are useful in doing diagnostic work. This will be supplemented where feasible by obtaining similar information from master mechanics in industry. This initial phase corresponds fairly closely to the delineation of critical areas of cognitive behavior that should be included in a paper and pencil test. Instructors will be contacted individually and later brought together as a committee to determine which sounds should be included in such a test.

A contract will be written with one of the institutions to "build in" the malfunctions that have been specified by the committee. This might involve inserting a bad connecting rod bearing into an engine, installing a burned valve, etc. (It should be mentioned that "normal" engine sounds will also be included in the test.)

The malfunctioning engine will then be recorded using a high fidelity binaural recording system. This phase of the project will be performed under the direction of a consulting sound engineer.

The tape produced in the above step will then consist of a series of perhaps 50 to 100 sounds of malfunctioning parts of an automobile. In consultation with instructors four alternatives will be constructed for each item in the test. That is, four possible causes of such a sound will be presented to the student. The student's task will be to select from the several alternatives the one which correctly identifies the malfunction. The entire test will be recorded on tape so that the test can be administered without benefit of specially trained psychometricians. The sound will
be played back using a high fidelity tape recorder and binaural headphones for each student.

Tests of hearing ability will be administered to each student in order to rule out any errors which might be due to a hearing defect. One of the questions yet to be answered is the amount and type of additional information which the subject must have in order to make a reasonable judgment. This additional information is necessitated by the fact that probably at no time does the person's diagnostic judgment depend solely upon such meager information as a characteristic sound alone would supply. Furthermore, this information would have to be of such a nature as not to allow the student to make a correct judgment from it alone.

The data produced by this test will be quite similar to that produced by conventional paper and pencil test. An item analysis, very similar to that performed on more conventional tests, will be conducted. This analysis will yield item difficulties, biserial correlations between the item and the total test score, proportion responding to each incorrect alternative, etc.

In an attempt to obtain data which will support the validity of the test, it will be administered to several groups. First, the test will be administered to a group of college students who presumably have little or no knowledge of auto mechanics. This group is felt to be necessary since there may be slight variations in the recording of the several alternatives which would give a clue as to which alternative is correct. To the extent that contaminating factors such as this are present, the college group should perform at a better than chance level. To the extent that such contaminating factors are not present and to the extent that college students are untrained in auditory diagnosis, performance should be at the chance level.

The second group will be a group of students entering auto mechanics in a selected sample of technical institutions. It seems reasonable to
hypothesize that this group might perform at a slightly better level than the college student since these students will very likely have shown an avocational interest in auto mechanics during earlier years and will have learned some of the diagnostic sounds with which we are concerned.

The third group will be a group of students who are ready to graduate from the auto mechanic program in North Carolina. Presumably these students will have acquired considerable knowledge about auditory diagnosis during their formal training. It is this group on which the majority of the item analysis will be performed.

The fourth group will consist of master mechanics who have worked for some years in industry who would be expected to perform at the highest level for all groups.

The data provided by these several comparison groups will yield information regarding the validity of this instrument. If the instrument is in fact measuring those behaviors which are relevant to auto mechanics it is reasonable to expect the mean performance to increase from groups one through four.

Reliability data will be obtained using internal consistency and test-retest techniques. Considerable item rewriting may be needed to provide the range of item difficulty required for this test, since it is hard to estimate the item difficulty appropriate for the student population with which we are concerned.

As a secondary criterion in this study we plan to gather peer nomination data as to the student's diagnostic ability. It would seem that this is an excellent instance in which peer nominations might be used since one would expect that students would be quite familiar with the abilities that their peers exhibit in diagnostic work. This would seem to be far more true in this case than in peer evaluations.
of intellectual ability, personality characteristics, etc.

Correlations between this test and the several dimensions of achievement being measured by the paper and pencil tests under development will be obtained to determine the extent to which this test measures independent dimensions of achievement.
SUMMARY OF PROGRESS REPORT NO. 3

Thirty-five field consultants were placed under contract and each completed an analysis of his curriculum. Meetings were held with the committees of consultants during October to arrive at a consensus concerning the curriculum analysis, to discuss possible performance tests, and to review test items.

A study of appropriate reference tests to include the initial battery was begun. Arrangements for participation of twelve North Carolina technical institutes and/or community colleges were completed. Plans are being made to test in institutions in other states.

A study of problems of data storage and retrieval and of appropriate statistical analyses to be conducted was begun.

Experimental studies of kinesthetic sensitivity were conducted and apparatus built to investigate this area in more detail.

A study on the use of the auditory modality in diagnostic work in auto mechanics, machine shop, and radio-TV repair was designed and preliminary recordings made. This appears to be a fruitful area for investigation and the necessary audio equipment has been requisitioned.
Thirty-five field consultants were placed under contract and each completed an analysis of his curriculum. Meetings were held with the committees of consultants during October to arrive at a consensus concerning the curriculum analysis, to discuss possible performance tests, and to review test items.

A study of appropriate reference tests to include the initial battery was begun. Arrangements for participation of twelve North Carolina technical institutes and/or community colleges were completed. Plans are being made to test in institutions in other states.

A study of problems of data storage and retrieval and of appropriate statistical analyses to be conducted was begun.

Experimental studies of kinesthetic sensitivity were conducted and apparatus built to investigate this area in more detail.

A study on the use of the auditory modality in diagnostic work in auto mechanics, machine shop, and radio-TV repair was designed and preliminary recordings made. This appears to be a fruitful area for investigation and the necessary audio equipment has been requisitioned.