R E P O R T R E S U M E S

ED 014 567 08 VT 003 257

THE DEVELOPMENT AND TESTING OF AN EVALUATION MODEL FOR VOCATIONAL PILOT PROGRAMS.

BY- TUCKMAN, BRUCE W.
RUTGERS, THE STATE UNIV., NEW BRUNSWICK, N.J.

REPORT NUMBER BR-6-8355-PR-2 PUB DATE APR 67
GRANT OEG-1-7-008355-2039
EDRS PRICE MF-$0.25 HC-$2.00 48P.

DESCRIPTORS- *TEST CONSTRUCTION, *CURRICULUM EVALUATION, *VOCATIONAL EDUCATION, *EVALUATION TECHNIQUES, DATA PROCESSING, *BEHAVIORAL OBJECTIVES, TASK PERFORMANCE, COURSE OBJECTIVES, TESTS, MODELS, PILOT PROJECTS, PROGRAM EVALUATION, ACHIEVEMENT TESTS, CHECK TECHNIQUE,

A PROGRESS REPORT ON THE DEVELOPMENT OF A "CURRICULUM HIERARCHY FOR THE EVALUATION OF COURSE KNOWLEDGE," THE "CHECK" TECHNIQUE, IS PRESENTED. THE TECHNIQUE IS BASED UPON A MODEL WHICH INITIALLY NECESSITATES TRANSLATING THE LEARNING PROCESSES AND OBJECTIVES INTO EASILY IDENTIFIABLE BEHAVIORAL RESPONSES. THE PROCESS OF TRANSLATING VAGUE COURSE GOALS INTO PRECISE BEHAVIORAL ONES IS DESCRIBED. AFTER BEHAVIORAL GOALS HAVE BEEN TRANSLATED, THEY CAN BE ANALYZED INTO A SEQUENCE OF PREREQUISITE BEHAVIORS BY TASK ANALYSIS. THIS SEQUENCE OF PREREQUISITE BEHAVIORS IS THEN USED AS A FRAME OF REFERENCE FOR THE DEVELOPMENT OF CONTENT VALID TEST ITEMS. IN THE FINAL PHASE, NOT DESCRIBED IN THIS REPORT, CONTENT VALID TEST DATA CAN BE ANALYZED, AND RESULTS CAN BE USED FOR PINPOINT LOCATION OF COURSE STRENGTHS AND WEAKNESSES. A SAMPLE CHECK TEST FOR DATA PROCESSING -- KEY PUNCHING IS INCLUDED. (FS)

PROGRESS REPORT #2

BR ~~Project No.~~ 6-8355
Grant No. OEG-1-7-008355-2039

PA 08

THE DEVELOPMENT AND TESTING OF AN EVALUATION MODEL FOR VOCATIONAL PILOT PROGRAMS

April 1967

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

Office of Education
Bureau of Research

THE DEVELOPMENT AND TESTING OF AN EVALUATION MODEL FOR VOCATIONAL PILOT PROGRAMS

Project No. 6-8355
Grant No. OEG-1-7-008355-2039

Bruce W. Tuckman

April 1967

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

Rutgers-The State University

New Brunswick, New Jersey

Progress Report #2

ACKNOWLEDGEMENTS

The principal investigator would like to acknowledge the contributions of his assistant, Mr. Lawrence Oliver, to this project.

The present method of evaluation has been entitled and will hereafter be referred to as the "CHECK" Technique. The letters of this title summarize the basic methodology of this technique: The development of a "Curriculum Hierarchy for the Evaluation of Course Knowledge."

In the first progress report, it was indicated that we had abstracted all the sub-tasks included in the Data Processing: Key Punch training course. Subsequent to this abstraction, these sub-tasks have been reworded to clarify their operational definitions. This rewording has utilized the list of operational terms included in Appendix C of the previous progress report.

Once these sub-tasks were satisfactorily operationally defined, a final course objective which encompassed these sub-tasks was determined. This enabled us then to determine the specific sub-task interrelationships. The most efficient method of determining these interrelationships is a task analysis which begins with the final course objective and ends with the most basic sub-tasks. That is, the hierarchy is developed from top-to-bottom However, since this method demands more knowledge of course content than the investigators possess, (and a limited budget makes the "buying" of experts impossible) a simpler bottom-to-top development of the hierarchy was utilized. Thus, we began with the most basic sub-tasks a student needs to perform in order to begin progressing toward the final course objective. These most basic sub-tasks form the base of the hierarchy. They are as follows:

The student can "Identify A Key Punch And Describe Its Function."
The student can "Identify A Punch Card."
The student can "Identify An Original Source Document."

Once the base of the hierarchy was identified, we constructed the remainder of it by continually asking the following question:

"What is the next higher-level sub-task a student needs to perform in order to progress toward the final course objective?"

Using this question in this way, the hierarchical interrelationships among the sub-tasks were soon apparent.

Our next step involved the schematic presentation of these sub-task interrelationships. Several rough drafts of these interrelationships were considered. Problems in the schematic presentation of these interrelationships centered around our initial inability to present a clear, and at the same time conceptually sound, picture of the sub-task interrelationships. The hierarchy appearing in the Appendix (see p. 22 of the appended manual) is the final outcome of these efforts.

Once the hierarchy was developed, our next step consisted of generating content valid test items. These items were considered to be "test situations" specifically designed to elicit the sub-task behaviors included within the hierarchy. Each test item measures a different sub-task behavior. An attempt was made, whenever possible, to develop paper and pencil items rather than performance items which would require special equipment. For example, consider the following hierarchical sub-task:

The student can "Identify The Names of Specific Parts of A Key Punch."

There are two alternative types of test items which appropriately measure this behavior. That is, a student might be asked to point to specific parts of an actual Key Punch, or he might be presented with pictures of a Key Punch in which each of the Key Punch parts are numbered. He might then be required to match the part numbers with the names of the parts. In all instances, these latter paper-and-pencil test items were chosen whenever they could validly be used to measure sub-task behaviors. Of course, some of the sub-tasks, by their very nature, prevented their measurement by any method other than the use of actual Key Punch equipment. Sub-tasks which required actual equipment for their measurement were more frequently encountered at higher hierarchical levels. Appropriate performance test items were designed to measure these sub-task behaviors.

Once all test items had been generated and refined, we found that several of the paper-and-pencil test items required such materials as pictures of Key Punch equipment, original source documents, and Punching and Verification Instructions. Thus, it was decided that a test kit which included these materials would be required for each student.

The construction of a model test kit containing all these necessary materials was a tedious process. Specific difficulties were encountered in our attempts to secure pictures of a Key Punch which were appropriate for our test items. After a series of contacts with International Business Machines (IBM), they generously sent us the proper pictures.

The sequence of test items within the test also required some consideration. That is, we wanted the sequence of items within the test to reflect accurately the increasing sub-task levels within the behavioral hierarchy. At the same time, however, the items could be grouped in other ways in order to facilitate smooth test administration. Our solution to this problem was a compromise. That is, some groupings were made to facilitate smooth administration, but we attempted to retain the item sequence as an accurate reflection of the sub-task levels within the behavioral hierarchy. A copy of the test is presented in the Appendix.

In the near future, both the completed test kit and copies of the behavioral hierarchy will be forwarded to Mr. William Bux, the teacher of the Key Punch training course at Princeton High School. His evaluation of the hierarchy and the model test will be discussed, and if necessary, appropriate revisions will be undertaken.

Concomitant with the development of the Key Punch hierarchy and its related test, we have written a majority of the chapters to be included in the manual which is the major goal of this project. This manual will convey both a conceptual model of the present technique of evaluation, and an actual description of the steps that one must perform when applying this technique to a specific course of study. It will include practical examples from the evaluation that we are presently undertaking. The purpose of these examples will be to clarify the practical application of this technique. We presently have one more chapter to revise before this manual is ready to be critiqued by our consultant.

As a second application of the present technique, we have chosen to evaluate a section of a course in Agri-Business which is taught at New Brunswick High School and at Oakcrest High School. This course section

focuses on the recording of sales for a wholesale business and is here labelled "Agri-Business: Recordkeeping." The sub-tasks included in this section have been abstracted and reworded in operational terms. The final course objective which encompasses these sub-tasks has also been determined and operationally defined. The sub-task interrelationships were then determined, again using the same technique that was utilized in the development of the Data Processing: Key Punching behavioral hierarchy. That is, we began with the most basic sub-tasks a student needs to know in order to begin progressing toward the final course objective. Again, we realized that a top-to-bottom approach is more efficient, but we felt more secure, due to our knowledge limitations, using a bottom-to-top development of the hierarchy. These sub-tasks were:

The student can "Identify A Purchase Order."
The student can "Identify A Sales Order."
The student can "Identify a Sales Invoice."
The student can "Identify a Sales Journal."
The student can "Identify a Cash Receipts Journal."
The student can "Identify a Sales Returns and Allowances Journal."
The student can "Identify a Credit Memo."

We then continually asked the question:

> "What is the next higher-level sub-task a student needs to perform in order to progress toward the final course objective?"

The answers to this question enabled us to determine the interrelationships among the sub-tasks. Once these interrelationships were determined, their schematic presentation was considered. After several rough drafts, we considered as satisfactory a behavioral hierarchy. (This hierarchy is being prepared for printing and will appear in the next progress report.)

We have already begun to generate test items designed to measure the sub-tasks in this behavioral hierarchy. As a next step, these test items will be refined and included in a complete test similar in form to the test developed for the Data Processing: Key Punching training course.

The total scope of this project includes the following:

(1) Develop a behavioral hierarchy and content valid test for Data Processing: Key Punching;
(2) Develop a behavioral hierarchy and content valid test for Agri-Business: Recordkeeping;
(3) Administer both tests to students in relevant programs for evaluation purposes;
(4) Develop a manual of the CHECK technique;
(5) Present the CHECK technique to a variety of program administrators and State Department of Education officials in a workshop.

Thus far we have:

(1) Developed a behavioral hierarchy and a content valid test for Data Processing: Key Punching.
(2) Developed a behavioral hierarchy for Agri-Business: Recordkeeping.
(3) Developed a first draft of a manual of the CHECK technique.

Our immediate plans include:

(1) Further refinement of our Date Processing: Key Punching training course test kit and hierarchy based on consultive advice.

(2) A further revision of our manual based on consultive advice.

(3) A development of a test kit for Agri-Business: Recordkeeping.

Our future plans include:

(1) Administration of both tests in relevant programs for evaluation purposes.

(2) The presentation of the CHECK technique during a workshop tentatively scheduled for June.

APPENDIX

NAME _____

CHECK TEST FOR DATA PROCESSING: KEY PUNCHING

1. Take Picture 1 out of your test kit. Circle the letter below which is the same as the letter above the Key Punch in the picture. Ia

a

b

c

2. Circle the letter in front of the correct answer. The major function of a Key Punch is to: Ia

   a. Sort original source documents into their proper order.

   b. Prepare final reports for management purposes.

   c. Convert data to IBM cards for future processing.

   d. Verify the correctness of original source documents.

3. Take a punch card out of your test kit. Print your name in the upper left-hand corner of the punch card. Then take your pencil and make "Xs" in the 2, 5, 8, 10, 11, and 12 punching positions in column five. Ib, IIb

4. Using the same punch card that you just marked, write a "9" and the nine edge and a "12" at the twelve edge. Write these numbers at the extreme right edge of the card. Be sure to put the punch card you have just used back into the test kit. IIb

5. Circle the letter next to the correct answer. A number of columns which contain a specific type of information (such as the date) is properly termed a: IIIb

   a. Punching position.

   b. Field.

   c. Area.

   d. Unit.

6. Take Pictures 2 and 3 out of your test kit. Notice that the parts in the Pictures are numbered. Write below the numbers (as shown in the pictures) of: IIa

   a. Any two Functional Control Keys _____ _____

   b. Any two Functional Control Switches _____ _____

   c. Any two Special Punching Keys _____ _____

   d. Any two Operative Parts _____ _____

7. Continue to use Pictures 2 and 3. Write the number of the part shown in the picture in the blank space below next to the name of the part. Ills

_____ NUM

_____ FEED

_____ CARD HOPPER

_____ READING BOARD

_____ DUP

_____ ALPH

_____ PROGRAM UNIT

_____ READING STATION

_____ SKIP (11 zone)

_____ REL

_____ CARD STACKER

_____ PRINT

_____ REG

_____ PUNCHING STATION

_____ BACK SPACE KEY

_____ 6

_____ P (12 zone)

_____ SPACE BAR

_____ AUTO FEED

_____ MAIN LINE SWITCH

_____ COLUMN INDICATOR

_____ SKIP

_____ 0

_____ / (10 zone)

_____ MULT PCH

_____ PROGRAM CONTROL LEVER

_____ AUTO SKIP

_____ AUTO DUP

8. Continue to use Pictures 2 and 3. Write the number of the part shown in the picture in the blank space below next to the function of the part. IVa

_____ Permits numeric punching in a field programmed for alphabetic punching.

_____ Causes the machine to back space.

_____ The area that holds original source documents and acts like a desk.

_____ Where the cards are read.

_____ Used to manually duplicate information from one card into another.

_____ Causes cards to skip a required number of columns.

_____ Where the cards are actually punched.

_____ The 10 zone punch.

_____ Permits alphabetic punching in a field programmed for numeric punching.

_____ The 12 zone or Y punch.

_____ Places cards in the punching and reading stations so they may be punched or read.

_____ Permits the printing of each column as it is punched.

_____ The lever used to place the machine under programmed control.

_____ The 11 zone or X punch.

_____ Feeds cards, registers cards, and stacks cards.

_____ Where the cards are stacked and finally removed from the machine.

_____ The switch used to turn on the machine.

_____ Unlocks the machine.

_____ Used to manually punch special characters.

_____ Where you would look to find out the next column number to be punched.

_____ Permits automatic skipping and duplicating when the machine is programmed.

_____ Enables cards to move past the punching and reading stations into the stacker.

_____ Permits automatic feeding of cards from the hopper.

_____ Where the cards are initially loaded into the machine.

9. Circle the letter next to the correct answer. As the size of a field in an original source document increases, the size of its corresponding punch card field: IVb

a. Increases.

b. Decreases.

c. Remains the same.

d. Decreases, then increases.

10. Circle the letter next to the correct answer. Which of the following parts controls the STAR WHEELS? Va

a. ALPH KEY.

b. REG KEY.

c. PROGRAM CONTROL LEVER.

d. PRINT SWITCH.

11. Circle the letter next to the correct answer. The major function of a program card is: Vb

a. To control automatically certain Key Punch operations.

b. To verify original source document information.

c. To place information in a specific order.

d. To prepare and print final reports.

12. From the following list of operations, select only the ones that you must perform when preparing a Key Punch for operation. Then put the operations in the order that you must perform them by writing a number before each operation to indicate when it should be performed. Write the number "1" before the first operation, the number "2" before the second, and so on. Operations not required should be left blank. V1a,V1b

_____ Depress the MULT PUNCH KEY.

_____ Depress the REL KEY.

_____ Check to see that a blank card is on the PROGRAM DRUM.

_____ Press the FEED KEY twice.

_____ Align the cards properly and square them.

_____ Place the pressure plate against the cards in the HOPPER.

_____ Place the cards in the HOPPER.

_____ Fan the cards.

_____ Check the control wiring of the ALPH KEY.

_____ Check to see that the STAR WHEELS are down.

_____ Turn on the MAIN LINE SWITCH.

_____ Check to see that the PROGRAM CONTROL LEVER is pushed in.

13. Write the missing number in the blank space. A date field composed of six digits which begins in Column 2 will end in Column _____. V1b

14. Circle the letters next to the operations that can be performed automatically by a Key Punch under programmed control. V1c

a. Report writing.

b. Verification.

c. Duplication.

d. Placing in alphabetic shift.

e. Sequencing.

f. Placing in numeric shift.

g. Skipping a field.

h. Correcting errors.

15. Print the names of the operations that can be automatically performed by a Key Punch under programmed control in the spaces below under "OPERATION". In the blank space under FIRST COLUMN, write the symbol you would punch into the first column of the programmed field. In the blank space under REMAINDER, write the symbol you would punch into the remainder of the programmed field. VIIb

OPERATION	FIRST COLUMN	REMAINDER

16. For this item you will need a Key Punch or Key Punch Typewriter.

Take an unused punch card out of your test kit. Write a "2" in the upper left-hand corner of this card. Then beginning at column one, punch the following numbers into the punch card without skipping any columns.

1, 3, 5, 7, 9, 2, 4, 6, 8, 1, 3, 5, 7, 9, 2, 4, 6, 8. VIIIa

Be sure to put the punch card you have just used back into the test kit.

17. Continue to use a Key Punch or Key Punch Typewriter. Take an unused punch card out of your test kit. Write a "3" in the upper right-hand corner of this card. Then beginning at Column 5, punch the following letters and numbers into the punch card without skipping any columns. IXa

A, B, C, D, 1, 3, 5, 7, E, F, G, H, 2, 4, 6, 8, I, J, K, L, 1, 5, 7, 9,

M, N, O, P, Q, 4, 6, 8, Q, R, S, T, 1, 3, 5, 7, U, V, W, X, Y, Z.

Be sure to put the punch card you have just used back into the test kit.

21. Take the PAYROLL SUMMARY and a blank PUNCHING AND VERIFICATION INSTRUCTIONS form out of your test kit. Using the information in the PAYROLL SUMMARY: Complete the PUNCHING AND VERIFICATION INSTRUCTIONS form. That is, fill in the PUNCHING AND VERIFICATION INSTRUCTIONS so that it becomes a complete plan for punching the PAYROLL SUMMARY. Allow 15 columns for the Employee name field. Begin at Column One, and skip a column between each field. Write the words "PAYROLL SUMMARY" at the top of the PUNCHING AND VERIFICATION INSTRUCTIONS form that you have just completed in the space marked "SOURCE." Be sure to put your completed PUNCHING AND VERIFICATION form back into the test kit. IXb

22. Place a check mark before each operation that must be used in removing a Program Card from a Program Drum. Then number these necessary operations to indicate the order in which they are done. Write the number "1" before the first operation, the number "2" before the second, and so on. Operations that are not necessary should be left blank. Xa

____Press the BACK SPACE KEY.

____Continue turning the CLAMPING STRIP HANDLE toward your left hand as far as it will go.

____Remove the PROGRAM CARD from the PROGRAM DRUM.

____Turn on the MAIN LINE SWITCH.

____Turn the PROGRAM CONTROL LEVER to the right.

____Turn the PROGRAM CONTROL LEVER to the right.

____Hold the PROGRAM DRUM in your left hand with the CLAMPING STRIP HANDLE on the right.

____Lift the PROGRAM DRUM from the SPINDLE.

____Turn the CLAMPING STRIP HANDLE toward your left hand ½ turn.

____Fan the cards and place them in the HOPPER.

23. Place a check mark before each operation that must be used in mounting a Program Card on a Program Drum. Then number those necessary operations to indicate the order in which they are done. Write the number "1" before the first operation, the number "2" before the second, and so on. Operations that are not necessary should be left blank. Xa

____Press the MULT PUNCH KEY.

____Turn the PROGRAM CONTROL LEVER to the left.

____Hold the PROGRAM DRUM in your left hand.

____Move the CLAMPING STRIP HANDLE ½ distance to the right.

____Insert the 80th Column under the smooth edge of the CLAMPING STRIP.

____Wrap the PROGRAM CARD tightly around the PROGRAM DRUM until Col.1 is aligned and slipped under the toothed edge of the CLAMPING STRIP.

____Turn on the AUTO SKIP / AUTO DUP SWITCH.

____Move the CLAMPING STRIP HANDLE to the extreme right.

____Remove the cards from the STACKER.

____Check the ALIGNMENT CHECK HOLES to make sure that the card is flush with the metal edge under the CLAMPING STRIP.

____Insert the PROGRAM DRUM into the machine.

24. Print in your own words the main reason for punching and proving the first card in a series before placing a Key Punch under programmed control. Xla

_____

_____

_____

25. Place a number before each of the following steps to indicate their correct order. Write the number "1" before the first operation, the number "2" before the second, and so on. Do not leave any blank. XIIa

____Punch a PROGRAM CARD.

____Punch and prove the first card in a series.

____Punch the second card in a series.

____Mount the PROGRAM CARD on the PROGRAM DRUM.

____Turn on the AUTO SKIP/AUTO DUP SWITCH.

Continue on Key Punch.

26. For this item and the two following you will need a Key Punch. (You may not use a Key Punch Typewriter - often called a Simulator.) Take three unused punch cards out of your test kit. Punch these cards using the information given in the following Original Source Document. Xb

QUANTITY	DESCRIPTION	ITEM NUMBER	UNIT PRICE
24	Coffee	20356	.71
56	Orange Soda	6538	.08
62	Soup Mix	473	.36

Begin the QUANTITY field in Column 5, the DESCRIPTION field in Column 10, the ITEM NUMBER field in Column 30, and the UNIT PRICE field in Column 40 Be sure to put the punch cards you have just used back into the test kit. Continue on the Key Punch.

27. Take PUNCHING AND VERIFICATION INSTRUCTIONS NO.2 and an unused punch card out of your test kit. Punch a Program Card from these PUNCHING AND VERIFICATION INSTRUCTIONS. Be sure to put the Program Card back into the test kit. Xc

28. Take the WILSON COMPANY form and several unused punch cards out of your test kit. Use the information in this form to complete the following steps:

    1. Plan a PROGRAM CARD by filling in a blank PUNCHING AND VERIFICATION INSTRUCTIONS form.

    2. Punch a PROGRAM CARD.

    3. Mount this PROGRAM CARD on the PROGRAM DRUM.

    4. Punch the information from the WILSON COMPANY form into punch cards while the Key Punch is under programmed control.

Be sure to put your completed PUNCHING AND VERIFICATION INSTRUCTIONS, PROGRAM CARD, AND PUNCH CARDS back into the test kit.

A Manual

for Evaluating Educational Programs

THE CHECK* TECHNIQUE

Bruce W. Tuckman

(with the assistance of Lawrence Oliver)

Rutgers-The State University

* Curriculum Hierarchies for the Evaluation of Course Knowledge

FORWARD

The purpose of a pilot program is to determine whether a particular course of study will lead to increased knowledge and/or skills for those exposed to the program. If those exposed show significant gains as a result of the program, then it can be incorporated into the overall curriculum on a permanent basis. The pilot program serves the function of enabling one to test the efficacy of an approach without making a full-scale committemnt to a specific program. Thus, administrative changes and overall costs can be saved if the program, when administered on a pilot basis, appears to be of limited usefulness.

The purpose of a pilot program is defeated if it is not adequately evaluated. If a program is established on a test basis, it does not serve as a test unless some sort of evaluation procedure is applied to the program. That is, the results or outcome of a pilot program must be evaluated in terms of specifiable criteria before one can confiently recommend its permanent inclusion in the curriculum.

Thus, evaluation must be an integral part of a pilot program. In designing the program, evaluation needs to be designed in. It should not come as an afterthought to an already existing program but as a basic part of the program itself. Individuals and groups who design and run pilot programs must be further sensitized to the necessity of evaluation. In addition, it would be useful if a format or model for evaluation were to be available to program developers so that they could refer to it while designing their programs. Such a model would need to be general enough so that it would apply to a wide diversity of programs. At the same time, it must be specific enough so that it applies to individual programs of particular content. This is an attempt at such a model.

The basic problem, of course, is what should such a general, all purpose Evaluation Model be like. Obviously, it must be more than just a credo which says: Evaluate. It must describe the form the program must take so that it lends itself to adequate evaluation procedures and it must describe the steps to be taken in such evaluation. It must outline both the reasons for evaluation and the means of evaluation. Finally, the Evaluation Model must itself be evaluated.

There are many ongoing pilot programs in vocational-technical education in most of the states, primarily stimulated (and of course funded) by the Vocational Education Act of 1963. New Jersey, as an example has ongoing programs costing over five million dollars (half federal and half state funds). Many of these programs will not be adequately evaluated as a result of limited resources in this area. Consequently, the pilot programs may not serve their purpose unless adequate techniques of evaluation are forthcoming.

Existing techniques of evaluation are lacking in precision. Many of them are highly subjective, consisting of raters armed with rating scales, check lists, and/or anecdotal records who descend upon a school to determine the adequacy of its offerings. Obviously, the accuracy of these techniques if weakened by unknown interactions among the personalities and frames of reference of the raters, the dimensions included in the rating scales and check lists, the selective perceptions of the raters in completing anecdotal records, and their limited observations of the ongoing programs, the dynamics of which they experience in only a very limited sampling. One judge might give a rating of "Excellent," a second judge "Satisfactory," and a third judge "Poor." To average these ratings is to overlook certain assumptions of measurement, while to consider them separately gives little information.

Other techniques of evaluation are more quantitative and thus objective, but they frequently quantify variables which are only superficially or perhaps completely unrelated to the efficacy of pilot programs. Such variables might include the physical facilities of the school, the number of students enrolled in a course, or the years of education and experience of an instructor.

Achievement tests also are frequently encountered in evaluation. While the potential of these tests, if constructed and used properly, is promising, they are frequently globally interpreted to discriminate among students for purposes of assigning grades. A better interpretation would include an item analysis to point up areas of subject difficulty. This would be for the purpose of course improvement rather than for student grading.

By far, the most rational method of course evaluation is to measure the effect it has upon a student's behavior at some future date. Frequently, years after students have completed their formal schooling and have entered the occupational world, they are measured on the following types of variables:

Ease of obtaining employment
Length of time on the job
Promotions
Potential for advancement
Job security
Job satisfaction
Job success

Once these variables are adequately defined and individuals can be measured along their dimensions, attempts are made to determine the degree of relationship existing between these criteria and previous formal training. For example, if "success" in a vocational area can be agreed upon and objectively measured, and if it can be demonstrated by experimental methods that a specific course produces this desired "success,"

then this provides strong support for the permanent inclusion of this course in the school's curriculum.

Follow-up evaluational techniques of this type encounter a number of procedural difficulties. First, the criterion needs to be adequately defined and quantified which is no small undertaking. Second, experimental methods requiring random placements of students into school programs must be used, which is not feasible within public school systems. And finally, several years need to pass beyond graduation before these techniques may be applied. When the decision about whether to continue a pilot program must be made immediately upon completion of the school year, a follow-up approach of this type is unrealistic.

Thus, it would seem that an ideal technique of evaluation would be one which is quantifiable, able to pinpoint the strenghts and weaknesses of a course, able to compare one course with another, and which can be appropriately applied immediately at the completion of a course rather than at some future date. The technique of evaluation presented herein incorporates these four positive characteristics. It is based upon a model which initially necessitates the translation of learning processes and objectives into easily identifiable behavioral responses.

CHAPTER ONE

OVERVIEW

Achievement tests for course evaluation

Achievement tests are typically used to differentiate among students for grading purposes. Students who perform well on these tests are assigned "A's" or some other symbol which is supposed to represent their excellent achievement of course content. Those who perform poorly are assigned "F's" to represent their failure to master course content.

Less frequently, achievement test are used to evaluate a course of study. They are used in the following manner:

> If every student in a class were to fail the final exam, this poor exam performance could be interpreted in either of two ways:
>
> 1. The exam was unfair because it did not measure what was taught in the course.
>
> 2. The exam was fair, that is, it did measure what was taught, but the students all failed to learn the course content.

If the exam was fair (Interpretation 2), and all students failed the exam, it is obvious that the course in some way needs modification. That is, there is something wrong with the way the students are being taught. Or, stated differently, the course is a poor one since none of the students are able to grasp its content. This is course evaluation.

Notice the underlined sentence in the previous paragraph. It begins with "If the exam was fair . . ." This is to point up the fact that the effectiveness of this type of evaluation depends upon the fairness of the final exam. The fairness of the final exam must be guaranteed before this type of evaluation can be considered valid. In attempting to guarantee the fairness of the final exam, one enters the realm of the test constructor.

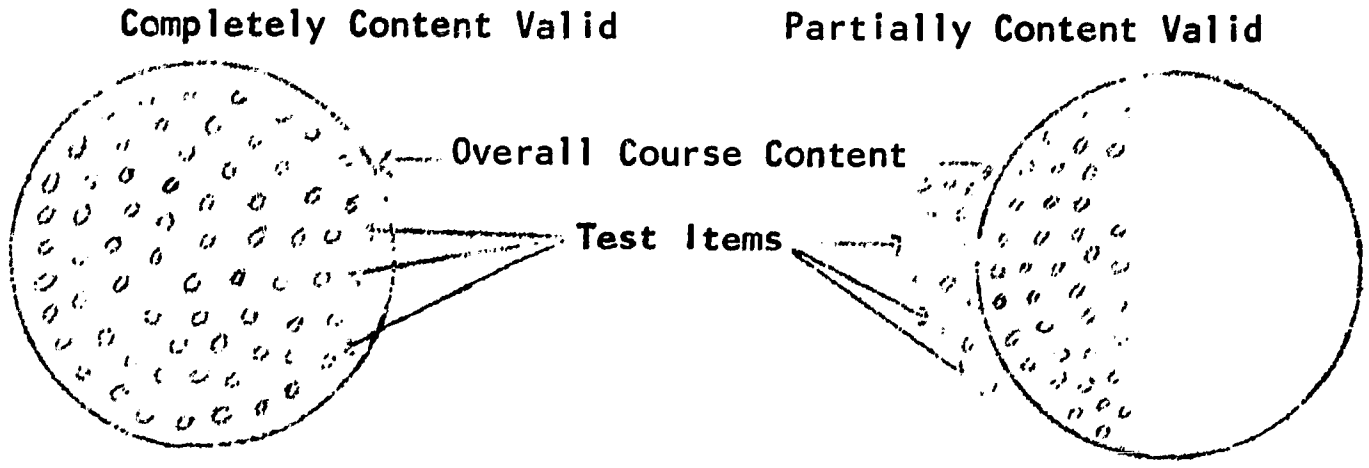
Test constructors do not use the terms "fair" and "unfair." Instead, they prefer to use the terms "valid" and "invalid" which means pretty much the same thing. Specifically, a test item is valid if it measures what it is supposed to measure and invalid if it fails to measure what it is supposed to measure.

There are many types of validity. One of them, content validity, is based on the extent to which the test is truly representative of the subject matter it is testing. As an example of content validity, if one wants to measure Johnny's arithmetic computation skills, an appropriate question might be, "How much is two times two?" This item clearly has content validity. One would not ask him how to spell "elephant," for this would measure his spelling achievement rather than his arithmetic computation skills. This item would have content validity for spelling, but would lack content validity for arithmetic computation skills.

These examples are clear-cut. Their content validity or lack of it is immediately obvious. However, when one attempts to construct a content valid test which covers an entire course of study such as wood-working or electronics, whether or not a specific test item is content valid becomes a difficult decision to make. Thus, the final test may be made up of some items which are content valid, and others which are not. These tests are presented schematically in Figure One.

FIGURE ONE

A Schematic Representation of a Completely Content Valid and a Partially Content Valid Final Examination



The two large circles in Figure One represent the content of a course. The small circles represent test items. The fact that all the small circles of the Completely Content Valid test fall within the large circle and are evenly distributed indicates that:

1. All the test items are measuring within the content of the course.
2. All aspects of the course are represented.
3. No aspects of the course are overrepresented.

That is, no test items measure things not related to the course. No aspect of what has been taught is overlooked, and no area of the course is more heavily measured than another (unless it has occupied a greater proportion of course time).

The fact that some of the small circles of the Partially Content Valid test fall outside the large circle indicates that some of the test items are measuring something other than the content of the course. Furthermore, their grouping within the large circle indicates that some areas of course content are being overlooked. That is, some of what has been taught is being overlooked in the final exam. Or, in other words, only half of what has been taught is being measured on the final exam. Moreover, some areas are being overtested. Thus, this final exam would not be an accurate measure of the effectiveness of the course since it

does not measure what has been learned as a result of the course. Half of the test is measuring something else.

In summary, course evaluation is related to content validity. With a content valid test, one can measure student achievement of course content. If no student achieves well, one can consider the course to be deficient in some areas, and appropriate modifications can then be made.

The problem in evaluation is thus directly related to the problem of content validity. To evaluate, one must have a content valid instrument. Thus, there is the need to develop a test with content validity.

The most efficient way of developing a test with content validity is to consider first the final objectives of a course. These final objectives give insight into course content, since it is reasonable to assume that what is taught is directly related to what is expected as a final result of teaching.

However, final objectives are typically vaguely stated in such terms as, "The student understands..." or "The student appreciates..." As they stand, such objectives are of little value for the test developer. With a little effort in translation, however, they can become the key factor of the entire process of evaluation. This translation requires the substitution of vaguely stated terms with behavioral ones. For example, terms like "demonstrate" require the student to perform a behavior which is observable by others. Being observable, this behavior can be measured and thus indicates the student's achievement of the final course objective. The process of translation of final objectives into behavioral terms is the topic of Chapter Two.

After final course objectives have been translated into behavioral terms, they can be analyzed into a sequence of prerequisite behaviors. This technique is explained in Chapter Three.

This sequence of prerequisite behaviors is then used as a frame of reference for the development of content valid test items. How these items are developed is the topic of Chapter Four.

Finally, Chapter Five presents methods of analyzing test data. Chapter Five is significant since it indicates the way in which the results of a content valid test can be used for pin-point location of course strengths and weaknesses. The entire technique is summarized in Chapter Six.

CHAPTER TWO

IDENTIFYING BEHAVIORAL OBJECTIVES

The significance of behavioral objectives

The field of education is presently far removed from the field of the physical sciences. Instead of dealing with physical units of length and weight dealt with by the physical scientist, the educator frequently deals with such abstract concepts as "understanding," "ability" and "learning." Too often, these abstract concepts appear when educational goals are expressed. As an example:

> The student, as a result of this course, will develop an "understanding" and "appreciation" of business practices.

The difficulty with this type of goal-setting becomes obvious whenever subsequent attempts are made to measure the success of an educational venture. The terms "understanding" and "appreciation" simply do not lend themselves to any known procedure of scientific measurement. There is no way of determining whether or not, or to what extent, the student "understands" or "appreciates" the course content.

One way of overcoming this difficulty is to translate these abstract concepts into easily observable student behaviors. As an example of what a behaviorally translated goal might be like, the goal of a data processing course might be stated as follows:

> The student can demonstrate a procedure for punching data from original source documents into the proper punch card columns after he has planned and placed machine operations under programmed control.

Stated in this way, one can determine whether or not a student has achieved a final course objectibe by observing his performance. That is, he can be observed while performing a task which requires him to

demonstrate the procedures associated with the operations stated in the final course objective. His success on this task can then be expressed in an objective statement of the degree to which he "understands" the course content. Thus, a measurement procedure of this type is both highly objective and easily communicated to others. The student's next teacher will not need to dispair over the vague statement that the student's understanding and appreciation are "adequate." Instead, he will know exactly what tasks the student has and has not mastered. He will then be in a position to gear his teaching approach to begin at the level at which the student is presently functioning and subsequently to facilitate the student's movement up to the next higher level of performance. In this way, a waste of time and effort are avoided, and the educational process becomes one of continuity of training which is highly individualized.

At the same time that the student's performance is being evaluated, the course of instruction can be analyzed to facilitate its modification for improvement. For example, if the majority of students in a data processing course were unable to demonstrate a procedure for preparing a program card from the information on an original source document, then a closer look at the method of instruction in this particular course content area might disclose that there is a specific difficulty that students encounter in this area which might be avoided if an alternate instructional approach is considered.

In summary, an approach in the setting of course goals which stresses easily observable student behaviors has significant advantages over an approach which utilizes such abstract concepts as "learning" or "understanding." The former approach enables the instructor to measure accurately the degree to which a student is able to perform the behaviors

included in the final course objectives. This technique of measurement has the added advantages of being both objective and thus easily communicable to others. These advantages enable the process of education to become continuous and individualized for the student. At the same time, curriculum evaluation and enlightened modification for improvement is facilitated.

Writing behavioral objectives

It was seen in the previous chapter that the results of a content valid achievement test can be used to evaluate the effectiveness of a course. For example, if a final exam is content valid, and if all students perform poorly on it, there is a strong indication that the course itself needs modification.

The first step in developing a content valid test involves a process of translation. The final objectives of a course must be translated from the vague statements in which they are usually presented into behavioral terms. Thus, terms like "appreciates" or "understands" are translated into such behavioral terms as "demonstrates" or "constructs."

These behavioral terms enable an observer to measure accurately the performance of a student. In this way, measurement becomes objective and thus communicative to others. Furthermore, behavioral terms facilitate task analysis, a procedure which is required in evaluation, and which is explained in Chapter Three.

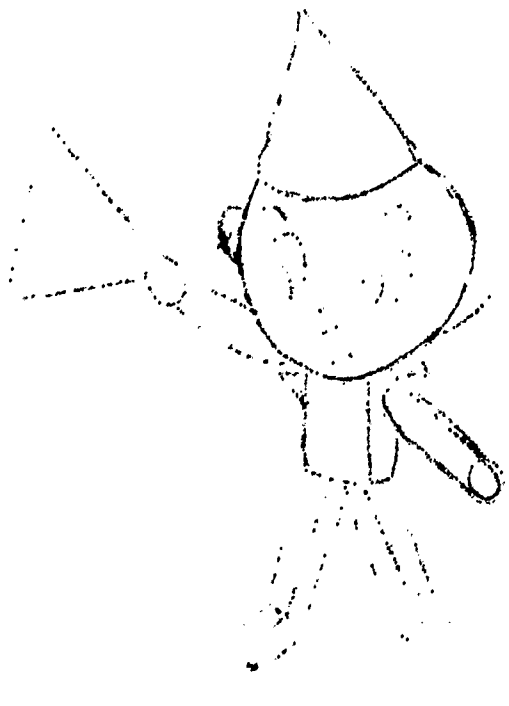
The term "behavioral" is an important one. An objective is behavioral if it specifies the behavior that one must observe to determine if the objective has been achieved. Because of the importance of behavioral objectives, a list of behavioral terms which may be used to develor behavioral objectives is presented below in Table One. Next to each term is its precise behavioral definition.

TABLE ONE

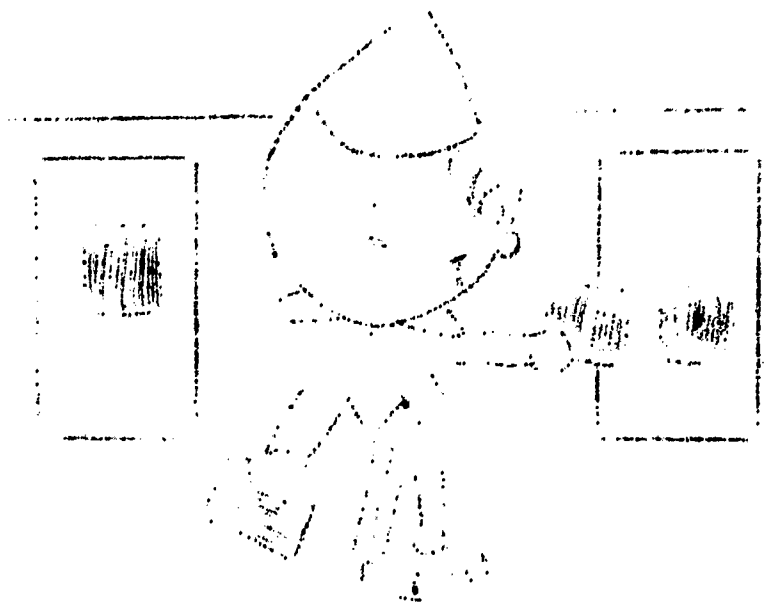
Definition of Action Words

The action words which are used as operational guides in the construction of the instructional objectives are:

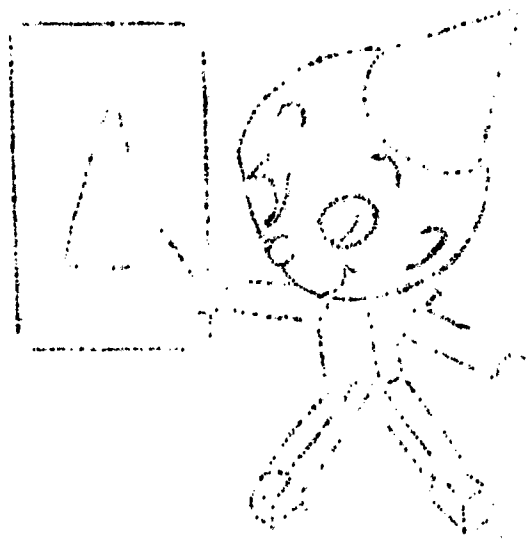
1. IDENTIFYING. The individual selects (by pointing to, touching, or picking up) the correct object of a class name. For example: Upon being asked, "Which key is a functional control key," the student is expected to respond by pointing to a functional control key; if the student is asked to point to the ALPH key, he is expected to point to the ALPH key. This class of performances also includes identifying object properties (such as rough, smooth, straight, curved) and, in addition, kinds of changes such as an increase or decrease in size.



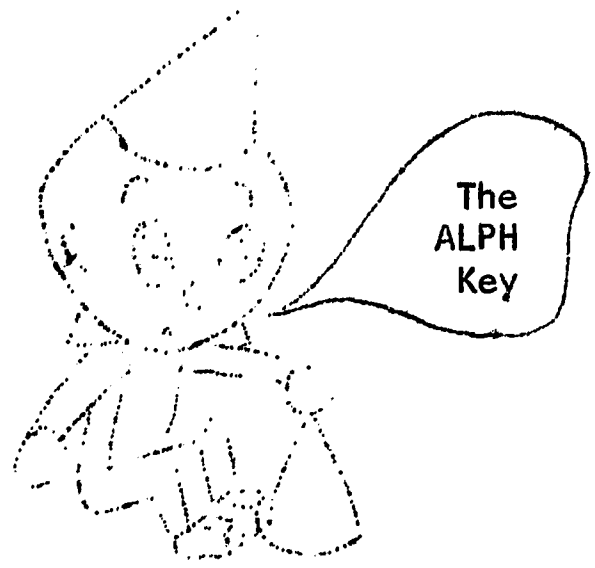
2. DISTINGUISHING. Identifying objects or events which are potentially confusable (square, rectangle), or when two contrasting identifications (such as right, left) are involved. For example: Upon being asked to point to the 11-zone key, the student is expected to respond by pointing to the 11-zone key.



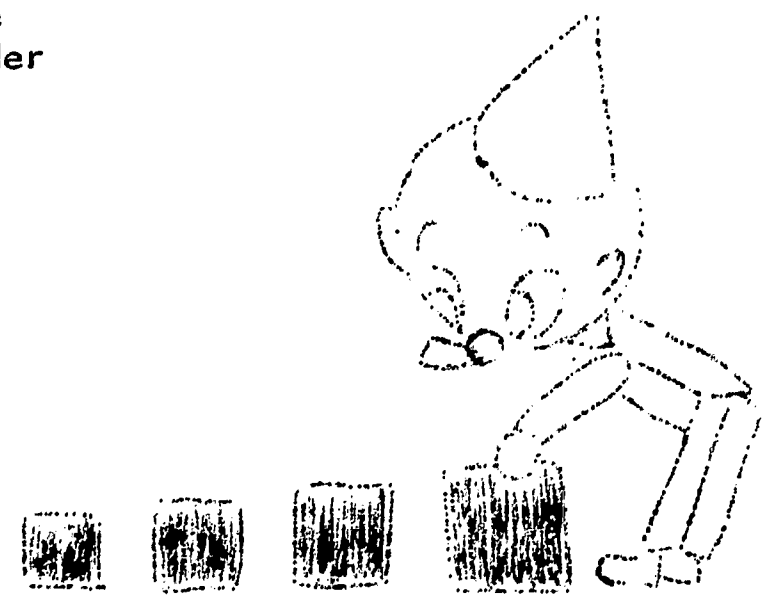
3. CONSTRUCTING. Generating a construction or drawing which identifies a designated object or set of conditions. Example: Beginning with an original source document, the request is made, "Plan and punch a program card for this data."



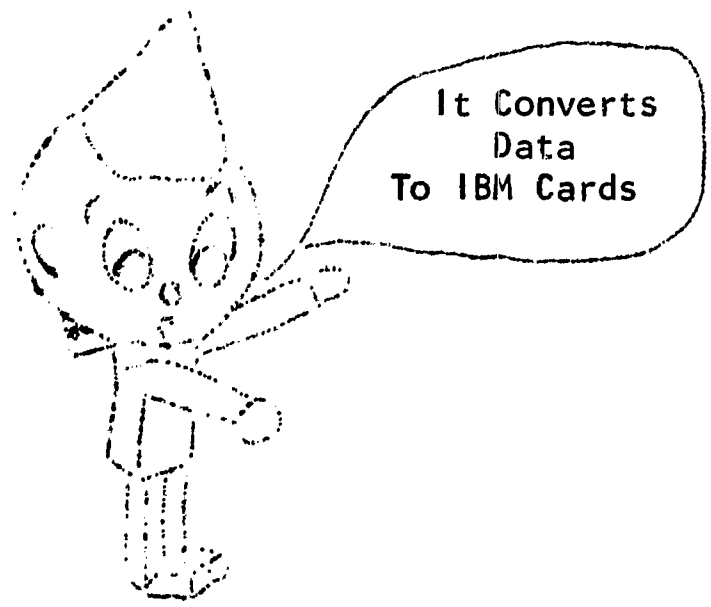
4. NAMING. Supplying the correct name (orally or in written form) for a class of objects or events. Example: "Which key permits alphabetic punching in a field programmed for numeric punching?"



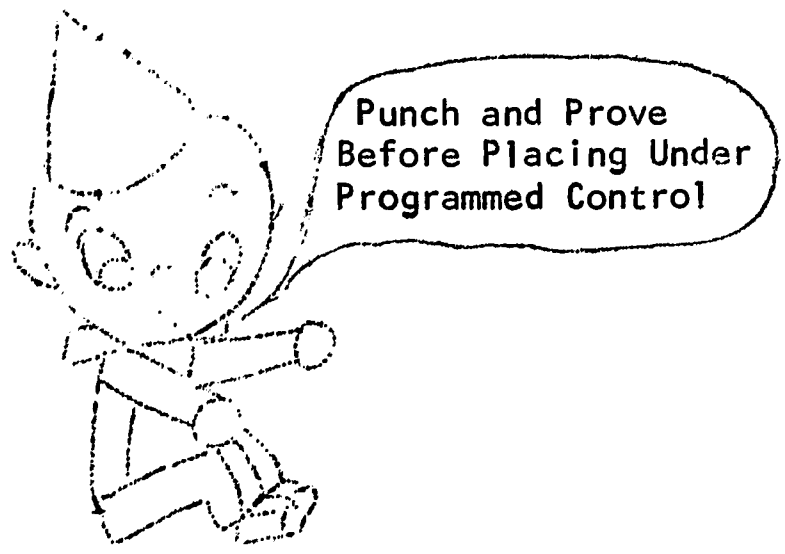
5. ORDERING. Arranging two or more objects or events in proper order in accordance with a stated category. Example: "Arrange these machine operations in their proper sequence."



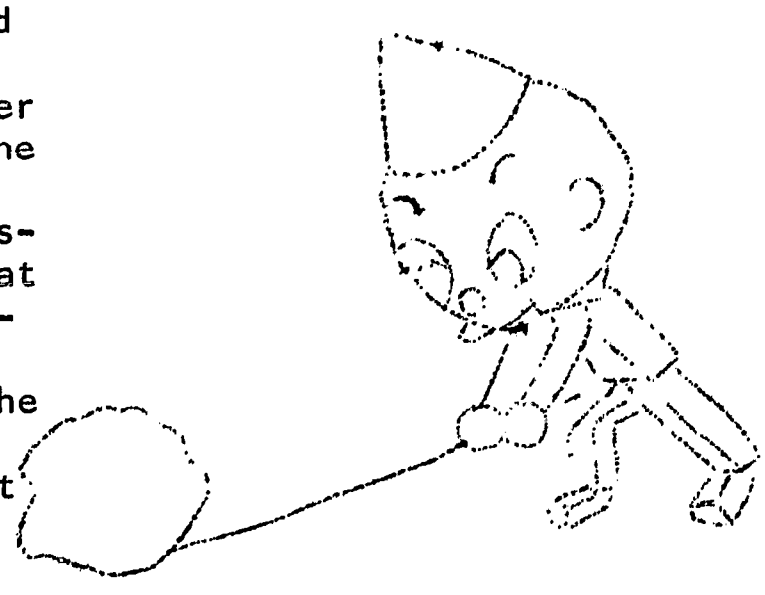
6. DESCRIBING. Generating and naming all of the necessary categories of objects, object properties, or event properties, that are relevant to the description of a designated situation. Example: 'Describe a Key Punch." The student's description is considered sufficiently complete when there is a probability or approximately on that any other individual is able to use the description to identify the object or event.



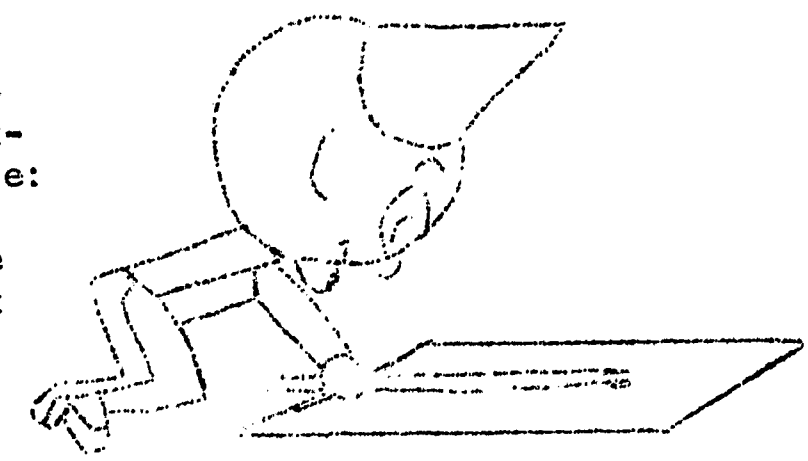
7. STATE A RULE. Makes a verbal statement (not necessarily in technical terms) which conveys a rule or a principle, including the names of the proper classes of objects or events in their correct order. Example: "What is the rule concerning the first card in a series when a key punch is to be placed under programmed control?" The acceptable response is that the first card is punched and proofed before the key punch is placed under programmed control.



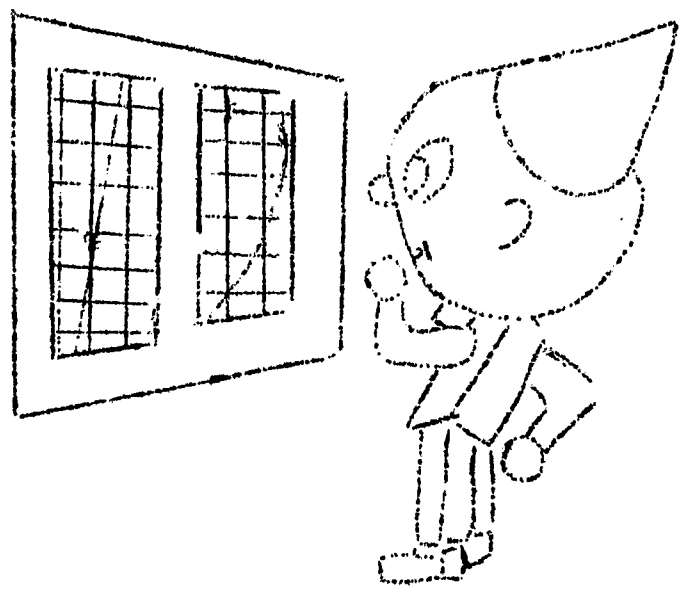
8. APPLYING A RULE. Using a learned principle or rule to derive an answer to a question. The answer may becorrect identification, the supplying of a name, or some other kind of response. The question is stated in such a way that the individual must employ a rational process to arrive at the answer. Example: "When should the DUP switch be turned on?" The student must know that the first card in a series is punched and proofed before the key punch is placed under programmed control to answer this question correctly.



9. DEMONSTRATING. Performing the operations necessary to the application of a rule or principle. Example: "First place the key punch under programmed control. Then punch the data properly." This requires that the student know the rule about punching and proofing the first card in the series.



10. INTERPRETING. The student should be able to identify objects and/or events in terms of their consequences. There will be a set of rules or principles always connected with this behavior. Example: "The same mistake is repeated throughout an entire series of punch cards. What has the operator forgotten to do?"



To clarify the proper use of these terms, consider the following translation of a vague course goal into a precise behavioral one.

Vague Goal: As a result of this course, the student will understand and appreciate the methods and significance of data processing.

To translate this goal, it is first necessary to confer with the instructor of the course to determine what is meant by the terms "understands" and "appreciates." Textbooks and instructional materials might also help to clarify what is included. As a result of conferences with the instructor and examination of the materials involved, it might be found that the student is really expected to perform the following operations:

1. Operate a key punch.
2. Prepare program cards from original source documents.
3. Place a key punch under programmed control.

Thus, the final goal might better be stated in the following behavioral terms which encompass the three delimited operations:

> As a result of this course, the student can <u>demonstrate a procedure</u> for punching proper data into proper columns after he has planned and prepared a key punch for programmed control.

Once the final goal is stated in these terms, an appropriate test to measure these <u>behaviors</u> can be developed. This test can use either actual equipment such as a key punch, or it can use simulated conditions which require the student to demonstrate the procedures under consideration. The important point to note in test development is that student <u>behaviors</u> are being measured.

Consider a second example of a vague final goal, its behavioral analysis, and its restatement in behavioral terms.

Vague Goal: As a result of this learning unit, the child will understand how to operate a lemonade stand.

Behavioral Analysis: The child must be able to:

1. Build a stand.
2. Gather supplies.
3. Make lemonade.
4. Give change.
5. Maintain utensils.

Restatement of the Final Goal:

As a result of this learning unit, the child can construct a lemonade stand, and demonstrate the procedures required for purchasing and storing supplies, mixing the drink, dealing with customers, and maintaining clean utensils.

An appropriate test can then be developed from this behavioral goal. Perhaps the test might require the child to operate a lemonade stand in a role-playing situation while the teacher observes his performance. It can readily be seen that if no child were able to operate the stand, the learning experience would have been inadequate in teaching this skill. Thus, a content valid test of this type could indicate the adequacy or inadequacy of the program, and at the same time, point up its pattern of strengths and weaknesses.

Using the list of behavioral terms presented above, translate the following vague goal into behavioral terms. Then compare your answer with the one presented immediately below.

Vague Goal: As a result of this lesson, the student will have knowledge of several European countries.

Translated Behavioral Goal:

As a result of this lesson, the student can identify four European countries on a map, name their major imports and exports, and describe their economic interrelationships.

Obviously, behavioral goals frequently differ, even when applied to the same courses. For example, in the above goal, one teacher might emphasize the history of European countries, another their social developments, and so on. And whatever is emphasized will be reflected

in the final goal. However, regardless of what aspects are considered important and thus reflected in the final goal, they must be stated in behavioral terms.

The next step in the development of a content valid test is a task analysis of the behaviorally stated final goal. This task analysis enables the test developer to construct a picture of all the subordinate behaviors which enter into the final task. How to undertake this task analysis is the main topic of the next chapter.

CHAPTER THREE

A TECHNIQUE OF TASK ANALYSIS

As mentioned in the previous chapter, conferences with an instructor and an analysis of the course materials he uses can facilitate the translation of a vague final goal into a precise behavioral one. Once this goal is stated in behavioral terms, it can be subjected to a task analysis. This task analysis which is the topic of this chapter, then results in a picture of the behavioral prerequisites needed by a student before he can complete the final course goal.

This picture directly facilitates the development of content valid test items which are useful in pinpointing the strenghts and weaknesses of a course. For example, if in a data processing test, all students were to fail on a test item which represents an operation that they need to know in order to perform the final task, then this indicates that the teaching of this operation is inadequate. Perhaps more time should be spent explaining it, or perhaps it should be taught differently or at a different time in the course.

A task analysis begins with the final task a student is expected to be able to perform as a result of having taken the course. This final task is of course stated in behavioral terms by using the method and word list presented in Chapter Two.

In conducting this task analysis, one asks of the final task,

> "Given this task, what subordinate capabilities would an individual need to have in order to do it successfully after only the presentation of instructions?"

Answers to this question are then subjected to the same type of analysis by asing the same question of them. This second level of answers is then analyzed in the same way by asking the same question of them.

Eventually, after analyzing each of these answers with the same basic question, the most basic task-relevant abilities of the individual are encountered.

To clarify this method of task analysis, its application to a data processing course is presented as an example.

Initial conferences with the course instructor and an analysis of course-relevant materials indicated that the final course objective could best be stated in the following behavioral terms:

> As a result of taking this course, the student is expected to be able to demonstrate a procedure for punching proper data into proper columns after having planned and prepared a key punch for programmed control.

This final course objective was then subjected to the following question:

> "Given this task, what subordinate capabilities would an individual need to have in order to do it successfully after only the presentation of instructions?"

It was found that the student needed to have (or know) the following subordinate capabilities:

1. He can activate the proper functional control switches at the proper times.
2. He knows the rule that the first card in a series whould be punched and proofed before the machine is put under programmed control.
3. He can punch a program card from punching and verifying instructions.
4. He can punch a card from an original source document.
5. He can mount and remove a program card from a program card drum.

These subordinate capabilities where then translated into behavioral terms to facilitate additional task analysis and subsequent test item development.

Thus they became:

1. The student can demonstrate a procedure for activating the proper functional control switches at the proper times.

2. The student can state a rule dealing with the punching and proofing of the first card in a series before the machine is put under programmed control.

3. The student can demonstrate a procedure for punching a program card from punching and verifying instructions.

4. The student can demonstrate a procedure for punching a card from an original source document.

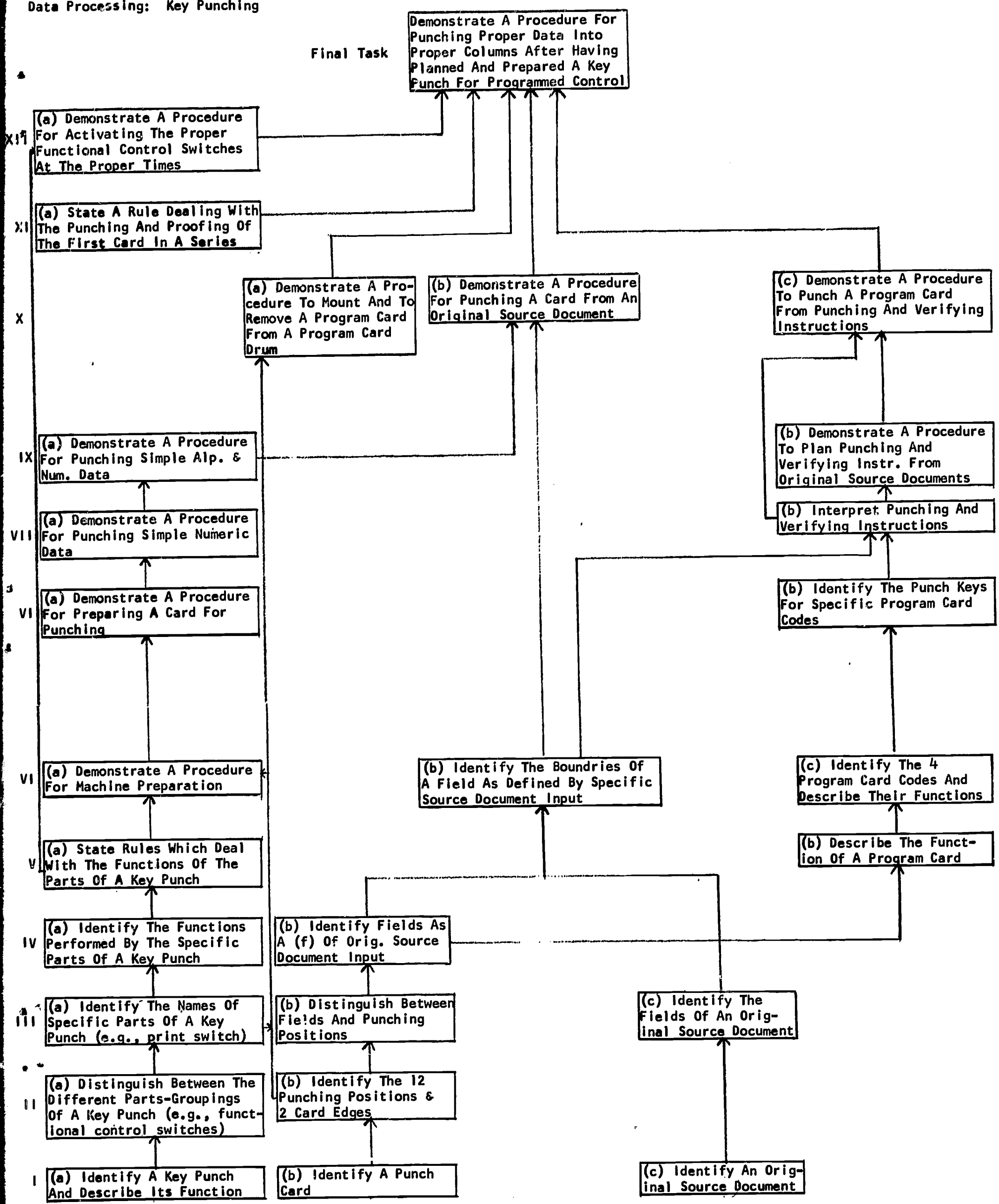
5. The student can demonstrate a procedure for mounting and removing a program card from a program card drum.

Once these subordinate capabilities were translated into behavioral terms, they were considered to be sub-tasks, and were subjected to the same analytical question in order to determine the subordinate capabilities required to perform each of them. Thus the same analytical question was asked for each of the five sub-tasks, and the subordinate capabilities of these sub-tasks were translated into behavioral terms, then considered sub-tasks, and subjected to the same analysis. Eventually, the analysis was ended when the most basic task-relevant subordinate capabilities were defined. The results of this complete analysis are presented in Figure Two.

Curriculum Hierarchy for Evaluation of Course Knowledge

Tuckman

Data Processing: Key Punching



Notice that lines connect sub-tasks with their subordinate capabilities. For example, to be able to identify the 12 punching positions and two card edges of a punch card (sub-task), the student must first be able to identify a punch card (subordinate capability of this sub-task). As another example, before a student can identify the boundries of a field as defined by specific source document input (sub-task), he must first be able to identify the fields of an original source document (subordinate capability of this sub-task), and also be able to identify fields as a function of original source document input (another subordinate capability of this sub-task). All but the most basic abilities at the base of the hierarchy require at least one or more subordinate capabilities. Of course, all sub-tasks are subordinate capabilities of the final task which encompasses all course objectives.

In conclusion, a behavioral hierarchy of this type which is developed by asking the same analytical question over and over again is useful both to the instructor and to the evaluator. It is useful to the instructor in that it indicates the nature and proper sequencing of sub-tasks in order to achieve the final goal. As as previously mentioned, it is useful to the evaluator in that it enables him more easily to develop the content valid test items which are required for evaluation. How to develop these items is the topic of Chapter Four.

CHAPTER FOUR

THE DEVELOPMENT OF TEST ITEMS

The behavioral hierarchy developed in Chapter Three should include all the behaviors which a student must perform before he can successfully perform the final task. The next step in the process of evaluation is to determine whether the student can indeed perform these behaviors. This step involves the creation of test situation in which these behaviors can be demonstrated. A more familiar name for these situations is "test items." Thus, the purpose of test items is to provide an opportunity for the students to demonstrate the behaviors defined in the behavioral hierarchy as being sub-tasks which the individual must perform before he can successfully complete the final task.

To develop content valid test items, one must first consider the specific behaviors he is attempting to measure. One must then create test situations in which these specific behaviors can be demonstrated. For example suppose one wants to measure the following sub-taks of the hierarchy in Chapter Three:

The student can <u>identify</u> a punch card.

To measure this behavior, the student might be given a picture of four objects, one of which is a punch card. He might then be instructed to draw a circle around the picture of the punch card. This is a content valid test item which requires the student to <u>identify</u> a punch card by drawing a circle around a picture of one.

As another example, suppose that one wants to measure the following sub-task:

> The student can <u>identify</u> the functions performed by the specific parts of a key punch.

In this instance, the student might be given a picture of a key punch with numbers indicating its specific parts. He might also be given a list of functions performed by these specific parts. Instructions might then require him to match the specific part numbers with their proper functions. In this way, the student can demonstrate his ability to identify the functions performed by the specific parts of a key punch.

Although both examples represent test items of a paper-and-pencil variety, only infrequently can all behaviors defined in a hierarchy be measured with paper-and-pencil tests alone. Consider for example the following behavior:

> The student can demonstrate a procedure for mounting and removing a program card from a program drum.

In this instance, it is difficult to create a paper-and-pencil test item to appropriately measure this behavior. One might therefore prefer to test students individually by giving them both a program card and a program drum, and by instructing them to mount and remove the program card from the program drum. In other words, actual equipment rather than paper-and-pencil test items would better measure this specific behavior. However, when possible, it is desirable to write test items that can be performed on a group rather than an individual basis. This allows entire classes to be tested at once. The use of pictures and diagrams can sometimes substitute for actual physical demonstrations.

In the three examples given above, it is important to note that two behaviors were appropriately measured by paper-and-pencil test items, while a third behavior required the use of actual key punch equipment. The point being made is that the behavior required by the test item is the primary concern. If, for example, identifying behaviors can be elicited by paper-and-pencil test items, than these items

are content valid ones for these specific behaviors. Such group test items should be sought after. However, if paper-and-pencil items will not measure the behavior in question, then they should not be used. Some other materials must be used in their place.

A set of test items to measure the tasks of the Data Processing-Key Punch hierarchy shown in the previous chapter appears below.

The criterion of a "good" test item is that it measures the behavior specified in an objective or sub-task. To the extent that sub-tasks have been properly written in behavioral terms, the job of developing sub-task test items is a simple one.

Once all test items have been developed and the test is in its final form, a procedure for test administration must be carefully planned. Instructions to students should be clearly stated and placed in writing. At this time, students' questions should be anticipated and answers to these questions should be included in the instructions.

It is felt by the writer that the instructions need not mention the specific course-evaluational purpose of the test. Instead, it might better be introduced as a "regular" achievement test. This should guarantee student motivation while taking the test, and thus, produce more accurate test results.

Of course, the procedures of administrating any achievement test require that each student do his own work, and that the instructor avoid inadvertently providing students with clues which might help them in answering specific test items.