DOCUMENT RESUME

ED 071 330 EM 010 475

AUTHOR	Tosti, Donald T.; And Others
TITLE	Introduction to Psychology and Leadership. Leadership Course; Research and Evaluation Plans, Part I.
INSTITUTION	Naval Academy, Annapolis, Md.; Westinghouse Learning Corp., Annapolis, Md.
SPONS AGENCY	National Center for Educational Research and Development (DHEW/OE), Washington, D.C.
REPORT NO	TR-6-3a
BUREAU NO	BR-8-0448
PUB DATE	31 Jan 69
CONTRACT	N00600-68-C-1525
NOTE	74p.; See also EM 010 418 and EM 010 419
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Autoinstructional Aids; Cost Effectiveness; Curriculum Design; *Curriculum Development; Data Processing; Individualized Curriculum; *Individualized Instruction; Individual Psychology; Instructional Design; Leadership; *Leadership Training; Management Education; Military Training; Models; Multimedia Instruction; Psychology; Research; *Research Design; Research Methodology; Social Psychology; Statistical Analysis; Technical Reports

ABSTRACT

In the first of a two-part document, the Westinghouse Learning Corporation delineates the major aspects of research to be carried out in the experimental leadership course instituted at the United States Naval Academy (the final report appears under EM 010 418, EM 010 419, and EM 010 484). Specific topics covered in this part include validation of materials, development of evaluative measures of achievement, development of time-cost effectiveness measures, student characteristics, statistics summary, and data processing requirements. Part II, concerned with experimentation design consideration for research on media and presentation design, appears under EM 010 487. EM 010 418 through EM 010 447 and EM 010 451 through EM 010 512 are related documents. (RH)

Westinghouse Learning Corporation

LEADERSHIP COURSE

RESEARCH AND EVALUATION PLANS-PART I

Contract No. N00600-68-C-1525

TR-6.3a January 31, 1969

TR-6.3a

January 31, 1969

LEADERSHIP COURSE

RESEARCH AND EVALUATION PLANS - PART I

Contract No. N00600-68-C-1525

ABSTRACT

This report is the first of a two-part document which includes the research procedures and methodologies to be employed in evaluating the instructional system and learning modules of the Leadership Course instituted in the United States Naval Academy.

Prepared by:

Donald T. Tosti
Edwina Hubert
Jules Shrage
Rosie Bogo

Approved by:

Donald T. Tosti

Project Manager
Leadership Course

Westinghouse Learning Corporation
1840 Lomas Boulevard, Northeast
Albuquerque, New Mexico 87106

TABLE OF CONTENTS

PART I

PART II

(to follow at a later date)

I. OVERVIEW

This report is the first of a two-part document representing Westinghouse Learning Corporation's delineation of the major aspects of research to be carried out in the Experimental Leadership course instituted at the United States Naval Academy. Part II will be submitted as a separate document at a later date. The total document is primarily a discussion of the research procedures and methodologies to be employed during the initial phase of the three-year project. It is expected that the research procedures employed in subsequent phases will be determined as outgrowths of the initial research.

Major aspects of research outlined throughout this report are validation of the instructional system, development of evaluative measures of achievement, development of evaluative measures of time-cost efficiency of learning modules, and research on student characteristics.

Procedures for the validation and evaluation of total instructional system effectiveness, topic unit effectiveness, and segment or module effectiveness will be presented in Section II. The total instructional system refers to all media, media-mixes, and variations in presentation forms used to communicate the content and objectives of the entire course. Topic unit effectiveness refers to the media, media-mixes, and presentation forms used to communicate the content and objectives of specified topics or in-depth learning units within each chapter. Segment effectiveness refers to the instructional methods used to communicate the content and objectives contained within a single learning module of approximately forty minutes of instruction. Procedures for

validation will take the form of (1) statistical evaluation based on gain score ratios and test-content-objective Tables of Specifications and (2) subjective evaluation based on subject-matter expert, instructor, and student ratings of instructional materials.

Section III, the Development of Evaluative Measures of Achievement, includes the development of administrative tests, cumulative post-tests, and progress checks. Procedures for the development of these tests are outlined with reference to test validity, reliability, objectivity, item analysis, administration, and scoring. All procedures to be followed in test development are standard procedures for standardized achievement test construction.

Student characteristics to be studied and the research methodology to be employed are presented in Section IV. Specifically, the areas stressed are:

(1) the isolation of student variables which bear relationship to learning through specific media and presentation design forms.

(2) the isolation of student variables which predict academic success in the Leadership course.

(3) the assessment of student preference for specific media and presentaton design forms.

Student characteristics or student variables will be studied primarily by correlation methods.

Time-cost criteria measures are discussed in Section V. Time will be determined for each module by simply providing a time response blank at the end of each progress check answer card.

Cost effectiveness will be determined by application of the present cost accounting system to each module.

Section VI contains a summary of the probable statistics to be applied in both Parts I and II and a description of research implications for subsequent phases of the project. Procedures for processing data generated throughout the project in all phases of research are outlined in Section VII.

Part II is concerned with the experimental design considerations for research on media and presentation design. This section includes discussion of the rationale for stating several hypotheses which are felt relevant to overall instructional systems. Although all of the hypotheses may not be tested in the initial phase of the project, they are thought to be worthwhile considerations for inclusion at some point.

The stated hypotheses have grown out of an intensive library study of experts' statements of problems associated with media and instructional presentation research. The following quotation from the Journal of Educational Research is representative of the direction that leaders in the field of educational technology feel researchers should be taking.

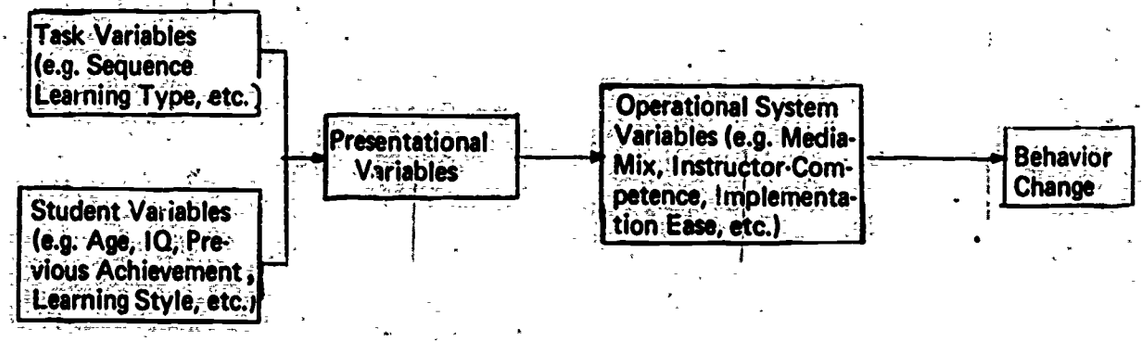
> ...in the future we will see more studies in which the purpose is to determine the relative effectiveness of various methods, techniques, or conditions of programed instruction. Through systematic study of different programing methods, principles and conditions, it will become possible to indicate the im-

portant conditions that determine the effectiveness of a program and/or machine.

> The main staying quality of programed instruction that will be recognized more and more is its capability of controlling conditions which heretofore it was not possible to control. With programed instruction and machines, it is possible to be quite explicit about either a method or a teaching sequence. Added to this advantage is that of reproducibility of the conditions. They make it possible to study teaching itself in a way that we could not do in the past. Involved is the possibility of doing research on methods independently of the teacher's personality, later on we can study the methods when combined with different personalities to determine what happens to their effectiveness. While there has been considerable interest in this problem in the past, up to now, the capability for studying it did not exist. Since it does now exist, the prediction is that we will see studies of how these two sets of variables interact with one another. This will make a science of teaching a genuine possibility. (Stolurow, 1962)

The explicit rationale for the selection of variables to be studied has been derived from A Behavioral Approach to Instruc-

tional Design and Media Selection (Tosti and Ball, 1968). In designing a behavioral change system, the several classes of variables recommended for study are illustrated in the following diagram:



In studying these several classifications, major hypotheses are grouped around three considerations:

(1) the distinction between medium and presentation

(2) the dimensions of presentation

(3) types of learning

As mentioned, student characteristics, student preference, and time and cost will also be studied in relation to these considerations.

II. VALIDATION OF MATERIALS

A. INTRODUCTION

Evaluation of the instructional system instituted in the Naval Leadership course will be continuous throughout the project; it will be aimed at overall system effectiveness, the effectiveness of the topic unit, and the effectiveness of the lesson. Evaluation will take two major forms. One is an objective or statistical evaluation based on measurement of criterion objectives. The second is a subjective or personal evaluation based on reports by subject-matter experts, students, and the instructor.

B. OVERALL SYSTEM EFFECTIVENESS

Ellis (1964) has indicated four major types of studies which are typically conducted to evaluate the overall effectiveness of instructional systems. These are:

(1) a comparison of some existing instructional procedure and teacher against the program.

(2) a comparison of some existing instructional procedure and teacher against the combination of the same instructional procedure and teacher, plus a program.

(3) a comparison of one type of program with another type of program dealing with the same subject matter.

(4) studies of pre-test to post-test gain.

The first study is often referred to as the "control group" versus "experimental group" comparison. This study assumes that all of the characteristics of the existing instructional system and teacher can be defined, and that only one variable is varied

for the experimental group (Holland, 1961). It is felt that this assumption is too gross to be accepted in the present project. There is little reason to believe that all present instructors of the Naval Leadership course employ exactly the same teaching techniques or principles of learning and that all of the variables can be controlled across classes. Without these stipulations, any comparisons of the experimental class with ongoing instruction would not be "controlled" comparisons.

A second consideration in experimental versus ongoing teaching comparisons is the need for a common examination which is appropriate to both classes. To the extent that individual instructors differ in the educational objectives they set for their students, the references they use, the sequence of presentation, examples used, and other content-related aspects, a common examination for any two classes set by one instructor is doubtlessly unfair to the other.

A third consideration in experimental versus control or ongoing instructional comparisons is the possible Hawthorne and Rosenthal effects which may bias experimental results. These two experimental effects are respectively the tendencies (1) for students to realize they are in an experiment and perform beyond typical expectations (U.S. Department of HEW, 1964), and (2) for teachers to realize they are being compared and alter their typical patterns of instruction (Rosenthal, 1966).

In other words, if differences are found, they can be attributed to a multitude of factors such as different teachers,

materials, objective tests, students, teaching methods, motivational techniques, experimental influence, etc. The lack of similarity between possible control classes limits any conclusions drawn from comparisons of the experimental course and traditional course to the particular courses being compared (Stolurow, 1962).

Despite a strong indication that experimental versus control comparisons are not desirable methods of system validation, there have been a number of such studies conducted. The Office of Education, Department of Health, Education, and Welfare (1964), has reported 36 experimental studies which have compared programed instruction with conventional classroom teaching, with the following results:

> Of the 36 comparisons, 18 showed no significant difference when the two groups were measured on the same criterion test, 17 showed a significatn superiority for students who worked with the program, and only 1 showed a final superiority for the classroom students. Eight of the experimenters mentioned a time advantage for the students who worked with the program, and only 1 (an industrial user), a cost advantage.

These results seem to indicate that almost any experimental course which emphasizes the use of programed materials can be expected to at least compare favorably with ongoing classroom

instruction. Even so, such studies are nonanalytic in that they do not isolate the particular factor which may produce no effect.

The second type of study (Ellis, 1964) of comparing some existing instructional procedure, in conjunction with a program against the effects of the same instructional procedure alone, is subject to many of the same criticisms as the first. Although the same instructor can be used, there are nevertheless multiple variables which cannot be controlled, and if differences are found, the significant variables accounting for the difference cannot be identified. Also, to the extent that conventional classroom instructional procedures such as lectures and discussions will be used in the multimedia course, there will be, in effect, internal control classes within the course.

The third type of study (Ellis, 1964) is that of comparing two programs employing different presentation forms simultaneously. In order to use this method in validation of instructional materials, a program covering the same content has to be compared with the experimental materials developed by WLC. Since such a program does not exist, i.e., specifically covering leadership objectives, the programs to be compared have to be developed. This will, in fact, be done to a certain extent. Various programs will be developed over certain segments of the same content area and presented to different students in the form of parallel modules. However, these programs will not be compared for purposes of overall instructional validation, but rather for purposes of determining the most effective media or presentational design

forms for the program.

The fourth type of study (Ellis, 1964) is the pre-test to post-test gain over the same program. The comparisons made in this manner evaluate the amount of learning that has actually taken place as the result of an instructional sequence. The student is given a pre-test to determine entering knowledge and a post-test to determine knowledge gained as a result of instruction. This type of study is susceptible to the least criticism. Consequently, the procedure to be used for evaluation in this project will be similar to the pre-test to post-test gain. It will, however, involve more than the simple raw score difference between the pre-test and post-test. (Stolurow, 1968; Ellis, 1964). A detailed description of the procedure to be used is presented in the next section.

1. Statistical Evaluation

The derivation and analysis of the gain score ratio for individual students will be used for objective analysis of the instructional system's effectiveness. Essentially, this method involves the development of tests which evaluate how well students have attained the task-level objectives (HumRRO, 1966; Stolurow, 1968). In assessing the overall effectiveness of the Leadership course, at least one major criterion measure will be used. (See Section III, Development of Evaluative Measures.) This will take the form of an administrative test which will be given in two parts: at the middle and end of the semester. Additionally, both parts of the test will be given at the

beginning of the course to determine the students' entering level of knowledge. The administrative test will be divided into two testing periods in order to increase the reliability of measures used for assigning course grades. Mid-term and final examinations will give a more reliable index of a student's performance than a single test. In addition, it is believed that this will best fulfill the administrative needs of the Naval Academy.

After the administrative pre-test is given, pre-test scores will be used to determine the maximum possible gain each student can make as a result of instruction. At the end of the mid-term test, the actual gain will be computed for each student by subtracting his scores on the corresponding half of the pre-test from his score on the mid-term test. The ratio of the student's actual gain to his maximum possible gain will provide an index of that half of the course's instructional effectiveness for that particular student. For example, if the mid-term test consisted of 50 items, and a student scored 15 on the first half of the pre-test and 45 on the mid-term, his gain score ratio would be 30/35 or roughly 85 percent. The same procedure would be followed for the final exam (Stolurow, 1968; Ellis, 1964).

To obtain an index for the overall system effectiveness for all students, the actual gain which is made by all students will be compared with the maximum possible gain which could be made by all students (Ellis, 1964). An alternative method for evaluation of course effectiveness

may also be used. This method considers the proportion of objectives successfully attained by the students, i.e., A/BC, in which A is the total number of objectives attained for all students, B is the total number of objectives measured by the test, and C is the total number of students (HumRRO, 1966). The advantage of the gain score ratio over this method is that it provides a way of estimating the efficiency of learning by controlling for differences in the incoming knowledge of the students. The ratio of gain to total possible gain takes into account how much it is possible to learn from the program and provides an objective index of the program's subsequent efficiency (Ellis, 1964).

2. Criterion Performance

As objectives are developed and approved by the subject matter expert, test questions will be developed to cover these objectives. Test questions may also be synonymous with objectives (Evans, 1968). To the extent that the test questions adequately measure the attainment of objectives, performance on the test will provide further indication of overall course effectiveness. In order to evaluate this aspect of program effectiveness, a table of objectives and test questions which measure those objectives will be developed (Stolurow, 1968). In this way, one can determine from test items missed which educational objectives are not being met. This type of table will be developed for the class as a whole rather than for the individual student. The percentage of students who miss each test

item related to an objective will indicate whether or not the instruction has been adequate.

3. Subjective Evaluation

In addition to evaluating the instructional objectives by criterion performance and statistical procedures, subjective evaluations will be made by subject-matter experts, students, and the instructor (Ellis, 1964).

Although it may be shown that learning takes place and specific objectives are mastered, subject-matter experts must agree that the content to be learned is related to the educational objectives set by the Naval Academy. In other words, it must be agreed that the materials developed have content validity.

Student evaluation will take the form of general attitudes toward the instructional materials. (See Section V, Research-Student Characteristics.)

C. TOPIC UNIT EFFECTIVENESS

The effectiveness of instructional materials for content topics will be determined in essentially the same manner as overall system effectiveness. The chief difference will be in terms of the smaller number of objectives covered and the length of the evaluative measure. The test covering units of instruction is referred to as the cumulative post-test (CPT). (See Section III, Validation of Evaluative Measures - CPT.)

The CPT will be keyed to the same behavioral objectives as the administrative test. The appropriate CPT will be administered at the beginning and end of each topic unit, and the gain score

ratio computed. A table of specifications will indicate which educational objectives are not being met by a majority of students. Based on these findings, materials will be revised to better teach the specific objectives.

Subject-matter expert, instructor, and student evaluations will be made with regard to specific materials over topic units.

D. SEGMENT EFFECTIVENESS

Segment effectiveness will be determined in a manner similar to that of the topic unit effectiveness. The number of objectives will be fewer. The objectives may be more specific, but the length of the test will be much shorter. Specific lessons covering approximately one class period or outside class work will be evaluated by progress check tests, and criterion performance will be assessed. (See Section III, Validation of Evaluative Measures - Progress Checks.) As in the previous two sections, it will be possible to pinpoint specific areas of difficulty within the materials through the use of a table of specifications. (See Table 1 on page 15 and HumRRO, 1966.)

Progress checks will be given at the end of each lesson to determine the number of objectives attained. Subjective evaluation of lesson materials will be made by spot-checks over a randomly selected number of lessons.

A second method for assessing the effectiveness of individual segments considers the degree to which the learning module is effective in accounting for individual differences in the entering ability level of students. This estimate is the correlation coefficient of pre-test scores with post-test scores, or more

specifically, the correlation of CPT pre-test scores with module progress checks. To the extent that the learning modules within the instructional system are effective in minimizing the initial effects of individual differences, correlations between CPT pre-tests and progress checks should approach zero. That is, regardless of the variation in student performance on the pre-test, all students should perform at the same level of 90 percent criterion on progress checks. The lack of variation in progress check scores would, therefore, yield a near zero correlation with pre-test scores.

Although correlations will be made between CPT pre-tests and progress checks, the correlation coefficients will not be considered as indices of segment validity. A zero correlation may indicate segment effectiveness, but it might also be accounted for in terms of a small number of subjects or the limited possible range of scores on progress checks. Since this could be the case, a zero correlation would not necessarily be an index of segment effectiveness.

Objectives	#1	#2	#3	#4	#5	#6	#7	#8	etc.
Test Items 1	X								
2	X								
3		X							
4		X							
5			X						
6			X						
7				X					
8				X					
9					X				
10					X				
11						X			
12						X			
13							X		
14							X		
15								X	
16								X	

Table 1.

Table of Objectives –
Test Questions Specifications Illustration

III. DEVELOPMENT OF EVALUATIVE MEASURES OF ACHIEVEMENT

A. INTRODUCTION

The basic evaluative measures to be developed for the project are administrative tests, cumulative post-tests and progress checks. Administrative tests will be represented by a sample of questions covering the entire course content. Administrative tests are actually one test divided into two parts, administered at the middle and end of the course. Cumulative post-tests will be keyed to the behavioral objectives and administered at the end of topic units. Progress checks will also be keyed to objectives and administered at the end of each segment. Specific steps for the development of these measures will be presented in this section of the report.

In general, the administrative tests and cumulative post-tests will be developed according to basic principles for achievement test construction. Basic characteristics of the tests to be considered are content validity, reliability, objectivity, test or item analysis, administration, and scoring. Progress checks will be developed with these characteristics in mind, although the exact statistical analyses for all characteristics will not be the same (Section D).

B. ADMINISTRATIVE TESTS

Administrative tests will be developed to provide a basis for evaluating total course achievement and for evaluating the effectiveness of the overall instructional system. As stated above, these tests will actually be one test which samples the most basic and important aspects of the entire course, and which

is administered in two sections at the middle and end of the semester. Additionally, the entire test will be presented at the beginning of the course to assess students' entering familiarity with course content. Differences between pre-test and post-test scores will provide the basis for the gain score ratio discussed in Section II, Validation of Materials.

1. Validity

A test is said to be valid if it measures what it purports to measure. How well it measures what it is supposed to measure can be determined statistically by correlating the test with another test of the same content or with some other external criterion measure, or it can be determined subjectively by consensual agreement of experts (Levitt, 1961; Lyman, 1963; Loree, 1965).

In the first case, validity could be determined on the basis of how well the test differentially predicts those students who make good leaders and those students who make poor leaders. However, predictive validity depends on a quantifiable criterion measure of good and poor leadership which may not be available. In addition, it would be a number of years before this type of validity could be established. Concurrent validity, or validity determined by correlation with a criterion measure obtained at about the same time (such as an external test of the same material), is

also not feasible because of the lack of such external criteria, i.e., there are no standardized tests of leadership, and the leadership rating scales which do exist cover more variables than academic ability.

Therefore, the validity which will be established will be content validity based on subject-matter experts' agreement of the correspondence between test items, content, and the stated behavioral objectives. Content validity refers not only to a matching of topics covered in the course, but also includes a matching of the type of behavior implied in the objective to the type of behavior measured by the test item (Loree, 1965; Stolurow, 1968). To the extent that test items will be developed directly from behavioral objectives, it is felt that test items will have the highest possible degree of content validity. This assumption will be further verified by agreement between subject-matter experts. Subject-matter expert approval will be solicited for purposes of determining correspondence of test items to educational objectives and relevant examples.

2. Reliability

A test is said to be reliable if it is accurate and consistent in measuring what it purports to

measure. The reliability of a test can be estimated either in terms of its stability of measurement over time or its internal consistency.

The stability of a test is typically determined by Test-Retest correlations, i.e., by administering the same test to students on two occasions separated by a short time interval. In this way, scores on both tests are correlated and the resulting coefficient is taken as an estimate of the test's ability to consistently measure the same behavior. The obvious problem with this method in the present project is that between test administrations, instruction will be given which is geared toward the objectives measured by the test. To the extent that the instructional materials themselves are valid, Test-Retest correlations, or in this case pre- and post-test correlations, should approach zero because individual differences are minimized by the instructional materials. Therefore, Test-Retest correlations would not reflect the reliability of the test, due to the intervening instruction.

An estimate of internal consistency as an index of reliability is possible and will be made by the split-half correlation method. Total scores on odd-numbered items will be correlated with total scores on even-numbered

items of the same test. In this way it can be estimated whether all test items have been drawn from the samc population of test items. That is, since all test items included in the test represent only a sample drawn from all the items which could be used to measure the behavior, some estimate of the degree to which representative sampling has been made must be included. The coefficient of equivalence or split-half method of correlation will yield this information.

3. Objectivity

A third major characteristic to be considered is the objectivity of achievement tests. A test is said to be objective if two competent judges scoring the test independently arrive at comparable scores for each paper graded. The maximum objectivity of scoring that can be obtained is that derived from objective tests as opposed to essay and short-answer tests. Objectivity is important to the extent that it is necessary to be unbiased in the assigning of grades or other evaluative indices, and to the extent that maximum reliability of scores is desired (Wood, 1960; Loree, 1965; Levitt, 1961). In the present project, it is necessary to obtain both unbiased estimates of achievement and highly reliable, consistent estimates of achievement.

Other major considerations of objectivity which have influenced the selection of the project test format are scoring economy and adequacy of content sampling.

Scoring economy is the second feature of objectivity of the administrative tests to be used in the project. By using objective tests, scoring can be done by an administrative clerk. Most important is the fact that test results can be made available to students and instructors shortly after test administration. This feature of immediate feedback may have important implications for maintaining a high student motivation level.

The third feature of objective tests is the increased probability of adequate content sampling. In a 50-minute administrative test period, more content can be covered by fifty or sixty objective questions than would be the case if essay exams were given. With the objective test, the student is not likely to be asked the two or three questions he has not studied instead of the several questions he has studied in detail. Objective tests sample the entire content area the student is responsible for knowing, and consequently by comparison, is more fair to the student who has studied appropriately. Also objective tests

do not penalize students who lack the ability for written expression.

The particular format for objective administrative tests will be multiple-choice selection of items. The literature comparing the multiple-choice format with true-false, matching, and completion formats seems to indicate that multiple-choice seletions have most of the advantages of the formats without their disadvantages (Wood, 1960; Levitt, 1961; Loree, 1965). A further advantage of the multiple-choice format which has implications for the present project is that it lends itself to item analysis and item validity assessment more readily than the alternative formats.

4. Item Analysis

The process of item analysis provides information on how well students have performed on each item of a test. Poor performance may be due to inadequacy of student learning or to faulty construction of the item.

a. Item Validity

Procedures for item analysis will be based on the assumption that the total test is a valid measure of student competency. Since this assumption is made, the validity for a single item is estimated by correlating a single item with the total score of the test for each student.

A similar method to be employed will compare the performance on the item of the students who score high and the students who score low on the total test. The item will contribute to whatever is measured by the total test if a significantly higher proportion of the top group of students, as opposed to the bottom group of students, gets the item right. Item-total test correlations will provide an index of how well each item measures what it is supposed to measure.

Steps for determining item validity for administrative tests will be taken following the first institution of the experimental course. It will not be possible to determine item validity during a pre-testing of students outside the Naval Academy because the validity of the test is based on how well the test measures instructional objectives. Therefore, unless all pre-tested students are given the entire course sequence, item-total score correlations would reflect only chance relationships.

b. Item Discrimination

The discrimination power of items will be determined for all items included

in administrative tests. Discrimination power refers to how well a particular item differentiates good students from poor students. If an item can be answered equally well by students who do well and students who do poorly on the total test, it does not discriminate among students and should be improved.

Discrimination power will be assessed in two ways. The first method is to compare the performance on each item of the top and bottom group of students. Top and bottom groups will be represented by the upper and lower 33-1/3 percent of students on the total test (Loree, 1965). The discrimination index will be determined by consulting a table which presents minimum contrasts required between the top and bottom groups of students on a test item in order to be statistically significant at the 5 percent level (Mainland & Murray, 1952).

The second method of analysis will be to compare the proportion of students in the top and bottom

groups who pass the item on the pre-test to the proportion of the same students who pass the item on the final test. This type of comparision will yield information on items designed to measure growth. The desirable discriminating item then will be the item in which students perform better after instruction than before instruction (Loree, 1965).

Both methods of analysis are important in order to determine: (1) if the item does, in fact, discriminate between good and poor students, and (2) if the item is one which allows sufficient room for growth. For example, four students in the top group and two students in the lower group may answer an item correctly on the pre-test. Since there are eight students in each group, there would be room for growth on the item as a result of instruction. If, on the pre-test, seven students in the top group and five students in the lower group answer the item correctly, and if this contrast is significant, the item can be said to provide room for growth in addition to discriminating among students.

c. Item Difficulty

Item difficulty will be expressed simply as the percentage of students who answer the

item correctly. This difficulty index will be determined at two points in the administration of administrative tests. The first index of item difficulty will be derived from the pre-test administration. If a very high percentage of students answer the item correctly on the pre-test, the item is too easy and does not allow room for growth. For a multiple choice question of four or five alternatives, it is expected that only 20-25 percent of the students would choose the correct answer by chance alone; therefore, a good item would be one which is answered by only 25-55 percent of the students.

The second index of item difficulty will be derived from the final test given after instruction. The difficulty index at this point will serve to rule out items which are too difficult for inclusion as well as those items which do not discriminate among students.

d. Administration and Scoring

Administrative tests will be given at the beginning, middle, and end of the course by the course instructor. They will consist of approximately 50 to 60 multiple choice questions with four or five alternatives. Tests will be hand-scored by an administrative clerk, and students will be given knowledge

of results immediately after mid-term and final exams. No feedback will be provided for pre-tests given at the beginning of the course.

C. CUMULATIVE POST-TESTS (CPT)

CPT will be developed to provide a basis for evaluating student achievement over topical units and for assessing instructional effectiveness over those units. CPT will be keyed to terminal objectives and administered at the beginning and end of topic units, which will cover approximately five to ten lessons. The number of CPT will be designated as the number of topic units is specified.

CPT will be developed and administered in the same manner as administrative tests except that they will be shorter and more numerous. Results of these tests will be used as measures of effectiveness of mixed-media presentational forms and for research purposes rather than as bases for evaluating student performance and assigning grades.

The validity and reliability for CPT will be derived in the same manner as administrative tests. CPT validity will be established on the basis of subject-matter experts' agreement of the correspondence between test items, content, and the stated terminal objectives. Since the test items will be derived directly from the terminal objectives, the highest possible content validity is expected.

Reliability for CPT will be estimated in two ways. First, is by split-half method of correlating odd- and even-numbered items from the same test for all students. By this method, it

is possible to estimate if all test items have been drawn from the same population of test items. Second, since the CPT will be shorter than administrative tests, the reliability will also be estimated from the mean and variance of scores from each test using the Kuder-Richardson "formula 21" (Gulliksen, 1950). The Kuder-Richardson formula will be applied following an item analysis of difficulty of items, since the formula is based on the assumption of equal item difficulty.

CPT will be objective tests of the multiple-choice variety. Multiple-choice items will be used to insure objectivity, reliability, and ease of scoring. Other advantages of this format are ease, practicality of administraiton, and actual testing considerations. Multiple-choice items also allow for item analysis in the same form as administrative tests.

Item analysis will be conducted to assess item validity, item discrimination, and item difficulty. Item validity will be determined by both item-total test correlations, and subjective consensual agreement among content analysts as to the correspondence of items to content and objectives.

Item discrimination and item difficulty for CPT will be assessed by statistical analysis of responses to items made by midshipmen taking the first experimental course. It is yet undecided whether CPT will be pre-tested by a sample of students drawn from a population similar to Naval Academy students. Reasons for the indecision lie in the excessively large number of items contained in the CPT. At best, if pre-testing of CPT items was made, only item difficulty could be assessed, since item discrimination is based on the discriminating power of an item following instruc-

tion. The advantages of pre-testing CPT items will be assessed in relation to time and cost considerations.

CPT will be administered in the classroom by the instructor. Answer sheets will be provided which can be both machine and hand scored. It is expected that students will be given knowledge of results on tests shortly after the class testing period.

Advantages of the CPT are that:

(1) they provide a means of assessing student achievement over topic units and diagnosing areas of student difficulty.

(2) they provide a means of assessing instructional effectiveness over topic units.

(3) they provide a criterion measure which can be used for research purposes in evaluating the effectiveness of specified mixed presentation or media designs.

(4) they provide a review session and evaluation of long-term retention over specific lessons.

D. PROGRESS CHECKS

Progress checks will be developed to provide a basis for evaluating the effectiveness of the presentation of a segment, to evaluate student achievement over specific modules, and to evaluate different instructional strategies in presenting the same segment. Progress checks will be generally keyed to a specific objective covered in a single segment. In this way, when a student meets the criterion score of approximately 8 out of 10

points on the progress check, it is safe to assume that he has also met the objective. If he does not meet the criterion score, it is possible to assess the area of his difficulty and prescribe some form of remediation to insure that he will eventually attain the objective.

Progress checks may be developed in a manner somewhat similar to administrative tests and CPT, with the exception that some of the statistical procedures used in developing the latter will vary. For example, validity for the tests will be determined by consensual agreement among subject-matter experts and content analysts, the same as for the administrative and CPT. However, total progress check scores for each of the lessons within a topic unit will be correlated with CPT scores. This maneuver is equivalent to item validity where total progress check scores are viewed as items which are correlated with total scores.

The reliability for progress checks can only be estimated from the mean and standard deviation of group scores on individual tests using the Kuder-Richardson "formula 21" (Gulliksen, 1950). Reliability can also be estimated by correlations for validity between progress checks and CPT. (It will follow from this comparison that the reliability can be no greater than the assessed validity.)

To insure objectivity, progress checks will take the form of multiple-choice questions or specific constructed responses. Where constructed responses are used, care will be taken to avoid eliciting alternative responses which could be considered correct. That is, questions will be worded in such a way that only one response will be correct so that maximum objectivity in scoring can

be obtained. An additional precaution in scoring constructed responses will be that scoring key will be prepared prior to the course and tests will be scored by independent graders. Correlations for inter-grader scoring will be made.

Item analysis for item difficulty will be made following the institution of the course in the Naval Academy. Item difficulty indices will insure only that no items are so difficult that they cannot be answered by more than the theoretical percentage of students, i.e., 20% of students over five alternatives. Items will not be eliminated from progress checks simply because they are answered by a high percentage of students; the very nature of the instructional system in accounting for individual differences is that most students will be able to answer most of the questions. Instruction will be strictly directed at teaching objectives measured by the progress checks so that it would be considered a weakness of the instructional system if most students did not answer the majority of progress check questions.

The same reasoning which governs the exclusion of item deletion on the basis of high percentage-correct, also governs the exclusion of an item analysis for discrimination power of items. It is felt that progress checks, as opposed to CPT and administrative tests, should not consist of items which discriminate between students but rather should consist of items which students are expected to know as a result of instruction.

Silberman (1962) verifies the present position on item analysis by stating that since the purpose of a program evaluation test is to measure the behavior that should have been produced in all students receiving the program, the items may be easy, therefore resulting in low item discrimination indices. In this in-

stance, traditional item analysis data may not be particularly useful for program-evaluation tests. Effective programs will yield a very limited spread of scores on a post-test and consequently attenuate any coefficients which are a function of variance among the test scores. The validity of the test must be judged in terms of its relevancy to content and objectives as well as statistical indices. If progress check questions sample the essential subject matter the student has learned from the program, and if items are not eliminated on the basis of item-test discrimination, correlation between success or failure on each item and the criterion score may be zero, if everyone answers the item correctly. A progress check may not discriminate well those students who have had the identical instruction, but it may well discriminate those students who have had a programed form of instruction from those who have not (Silberman, 1962).

Administration of progress checks may be outside the class period as well as inside. Specific administration procedures have not yet been decided, but a system for self-administration and self-scoring of progress checks is being developed. Self-administration will probably occur for outside modules in which remediation and enrichment are contingent on the results. In order to insure maximum validity and reliability of outside progress checks, as well as providing students with immediate knowledge of results, specially devised answer cards will be given to students along with progress check questions at the appropriate in-class session to be answered outside of class for outside modules. These answer cards will be similar to tab cards and devised so that students discover the correct answer as soon as

they have made a response. In addition, if students have made an incorrect choice it will be possible for the instructor at a later point to make that determination. In this way, it will be possible to determine all of the students' first responses on the progress check and to have a reliable estimate of how much they knew at the exact time of the test.

IV. DEVELOPMENT OF TIME-COST EFFECTIVENESS MEASURES

In addition to the evaluation of student achievement which occurs as the result of differential instruction through learning modules, the attempt will be made to assess the efficiency of the instructional system in terms of both student time needed to complete the learning modules and the total developmental cost per module.

A. TIME AS A CRITERION VARIABLE

The rationale for using time as a criterion variable has grown out of research findings which indicate that it may in fact be the most relevant variable for making differential comparisons among multi-media (Silberman, 1962). Since the aim of instruction via any medium or presentation design is to effect criterion performance on progress check questions over specific learning modules, the resulting achievement measures for all students are clustered together. The clustered scores make it difficult, if not impossible, to find statistically significant differences between methods because of the lack of variance. Therefore, alternative considerations of the relative efficiency of methods of presentation are important (Gilpin, 1961). That is, two presentations may produce the same average amount of learning, but one may take twenty minutes and the other may take an hour. Another possibility is that the method which produces the greater amount of learning may also require more time. In such cases, it is advisable to consider the efficiency of the compared methods of instruction (Stolurow, 1962).

One possibility with regard to determining the efficiency of learning in terms of performance and time is to use an index which incorporates the two. Follettie (1961) has developed an index which does incorporate accuracy of performance, training, and test time, but the procedures he has used have not been examined as yet for inclusion in the present project.

B. COST AS A CRITERION VARIABLE

An additional important criterion for the multi-media course is the isolation of development and production costs for the various media and presentation designs used. In this way the value of each medium and design can be economically, as well as academically, determined.

This type of information is an important consideration for the Naval Academy and the Office of Education in the development of future courses. If differences in education effectiveness of two or more media are comparatively small or non-existent, differences in the cost of their development may become a relevant factor. Cost/effectiveness rates can be established for all types of materials prepared in the experimental course. The exact method by which the cost effectiveness study will be conducted will be contained in a forthcoming document, T.P. 6.5.

A cost criterion will not only be compared against immediate educational effectiveness, but also against other dependent variables. Long-term retention, learner time, administrative case, and student preference are a few which can be used.

By evaluating each module with respect to all of these variables, a system can be established for the selection of appropriate media and presentation design on the basis of the priority assigned to any given set of criteria. Cost/effectiveness rates and cost/time ratios have long been used as criteria for establishing training courses in

industry. They will be of comparable value to the Naval Academy in selecting future modes and media for materials development.

C. COST FOR THE COURSE DEVELOPMENT MODEL

In addition to the resolution of costs with respect to medium and presentation design, charges will also be itemized for all major functions with respect to the various types of technological and professional requirements. This type of breakdown is actually an extension of the cost per medium analysis. Not only is it important to determine the specific costs of various media but it is also valuable to know what contributes to the variation in such costs. In this way, possible cost variations might be controlled for, or eliminated in, future projects.

The final result of this type of cost analysis is the development of a model for general course development in which all major functions and tasks can be isolated and evaluated.

D. IMPLEMENTATION OF A COST ACCOUNTING SYSTEM

There are two main objectives in the accumulation of costs for this project. They are:

(1) to provide material production costs for use in cost effectiveness studies across presentation form, media, and learner characteristics.

(2) to provide a detailed breakdown of all costs for the establishment of a baseline for a course development model.

For tasks such as these, an extensive cost accounting system must be established to provide for the accurate collection of cost data. This section will generally outline the procedures that

have been developed for accumulating the costs for the Naval Academy project.

The basis of the present cost accounting system is the course development model. This model consists of nine major "functional" areas. At present, these are:

(1) Project management.

(2) project administration.

(3) research design.

(4) validation.

(5) analysis and materials preparation.

(6) presentation design.

(7) production and control.

(8) implementation.

(9) data processing and computer analysis.

Each of these functions is broken down into a series of tasks. Many of the tasks are iterative in nature, especially in the production of various units of course material. Each function and each task is assigned a specific number. Costs are accumulated on the basis of these numbers.

All labor and non-labor charges to the Naval Academy project use this numbering structure. The format of the accounting number for segregating costs is shown on the following page, and each portion of the format is subsequently explained.

Number "9" is used to identify non-standard number	(9)
Function Code	8
Task Number	33
Chapter	6
Segment	07
Letter "9" used as a separator	(9)
Element	A
Sub-Element	C
Budget Center	A99

Figure 1 — Accounting Number Format

The Function Code is a unique digit assigned to each of the nine functions of the course development model.

The Task number is the identifying number for a specific task under the function heading.

The Chapter number is the number of the particular chapter to which the work pertains. Chapters are numbered from one.

The Segment number is the number of the particular segment within a chapter to which the work pertains. Segments are numbered from one.

The number "9" appears next as a separator.

The Element letter is the identification of a particular element within a segment.

The Sub-Element may be used by department managers for their internal needs or by the research design.

The Budget Center code copies the budget center code of the standard budget number of which this number is an extension.

The extension number is a 13-digit number similar in format to the standard WLC number except for the identifying 9 in the first position.

Costs for non-iterated tasks belonging to specific functions will be accumulated within the framework of the standard 13-digit WLC accounting number. When iterated tasks or tasks related to a specific element of the course are involved, an extension of the standard number as shown above is mandatory. However, the extension will be structured in such a way as to be easily ignored by the standard WLC accounting system.

In summary, the general procedures for each division and each individual in the WLC Naval Academy contract are the

following:

1. Know necessary general function numbers.
2. Specify all tasks per function and update weekly.
3. Instruct all personnel in the use of the single 13-digit number.
4. Instruct personnel in the use of 13-digit extension numbers, where applicable.
5. Submit Labor Detail sheets each week.
6. Submit Non-Labor Detail sheets each week.
7. Monitor all updating of the course development model published so that task charges are appropriate.

Summaries of cost per function will be submitted to Naval Academy on a quarterly basis. Summaries of cost per media will be submitted as they become available toward the end of the project.

V. RESEARCH - STUDENT CHARACTERISTICS

A. INTRODUCTION

The primary purpose of this aspect of the research project is the determination of student characteristics which may be significantly related to different types of media and/or presentation designs and the development of a set of criteria for predicting academic success in the Leadership course. Specifically, the research program will attempt:

(1) to isolate student variables which bear relationship to learning through specific media and/or presentation design forms.

(2) to isolate student variables which predict academic success in the Leadership course.

(3) to determine student preference for specific media and presentation design forms.

Many educators have hypothesized that there may be some relationship between the learning style or specific personality traits of the individual student and the specific type of medium or presentation form which is most effective for that student. This possibility has important implications for the entire field of educational technology. For example, it may be important to know that two students of the same general ability, but differing in anxiety level, learn at different rates when taught by an instructor, programed texts, movies, and so on. If significant differences are found among individual students when taught by one method or another, it may be possible to prescribe learning

modules which will maximize individual learning.

To investigate this hypothesis, it is necessary to study all student variables which have been found to relate or interact with the learning environment.

B. SELECTION OF STUDENT VARIABLES

The student variables to be studied throughout the research project have been selected in a variety of ways. Procedures and criteria for these selections are as follows:

(1) Variables are selected which may bear relationship to specific media and/or presentation design forms. These are reading aptitude (speed and comprehension), listening ability, verbal ability, vocabulary, and selected personality factors such as group dependence versus independent personality traits.

(2) Variables are selected which are identified as significant predictors of acadenic success through general research endeavors. These are high school grade average and/or high school rank in class, English achievement, Mathematics achievement, Scholastic Aptitude Test-Verbal, and Scholastic Aptitude Test-Quantitative. (Goldman, 1961; Kring and Stolurow, 1968; Educational Testing Service, 1967).

(3) Variables are selected which are identified as bearing a relationship to classroom success although lacking in predictive power. These are authoritarianism-submissiveness, need for achievement or motivation, interest, and anxiety (Loree, 1965).

(4) Variables may be included on the basis of mere availability and studied to determine if they do bear a significant relationship to either overall performance in the course or performance on any particular unit. These would include variables which are measured in the same test and measures which are already available at the Naval Academy such as the Fiedler Leadership Scale.

C. MEASUREMENT OF STUDENT VARIABLES

Variables have been selected on the basis of their actual or potential predictive power or performance relationship. Even so, a number of variable possibilities have been deleted because of the lack of a well-developed measuring instrument. Therefore, the process of test selection in this study has included a careful analysis of test validity, reliability, and other standardization procedures reported by test authors (Buros, 1959, 1965).

1. Psychological Tests

It is felt that the following list of psychological tests are among the best possible measures of the specific traits they purport to measure: (Buros, 1959, 1965).

Variable	Test
Aptitude	Scholastic Aptitude Test - Verbal (SAT-V) Scholastic Aptitude Test - Quantitative (SAT-Q)
Achievement	English Achievement Mathematics Achievement
Reading Ability	Ohio State Psychological Examination (OSU)

Variable	Test
Personality	Edwards Personal Preference Schedule (EPPS)
	Sixteen Personality Factor Scale (16PF)
Interest	Strong Vocational Interest Blank (SVIB)

2. Tests Used at the Naval Academy

Test scores available through the Naval Academy which will be included in the student data base, are the Cornell Word Form - 2, Fiedler Leadership Scale, and The Adjective Check List.

3. Additional Variables

Other variables to be investigated are:

a. predicted grade average - which includes SAT-V, SAT-M, English Achievement, Mathematics Achievement, recommendation scores, and converted high school rank in class

b. high school rank in class

c. recommendation scores (high school)

D. STUDENT VARIABLES AND MEDIA EFFECTIVENESS

Research relating student variables to various media forms and presentation design variables has been meager and generally inconclusive. Two studies, for example, have found no correlation between IQ and aptitudes related to achievement in instructional systems using a PI presentation form with a workbook medium (Porter, 1959; Ferster and Sapon, 1958). Two other studies

report no correlation for IQ and performance on a criterion test, but do report significant relationships between general achievement level and performance (Feldhusen and Eigen, 1963; Hatch and Feint, 1962). A third type of study has been made relating intelligence to frequency of response demand within a program presentation, but with no significant results (Shay, 1961).

Studies relating personality variables to learning from programed instruction have had similarly negative results (Carpenter and Greenhill, 1963).

1. Isolation of Variables Related to Instructional Effectiveness

In the present research project, an attempt will be made to isolate student variables which may be related to learning through specific media and presentation design forms. The procedures used will be: (1) to study the relationship of variables which research indicates may be related to general learning and to learning through specific media or presentation design forms, and (2) to study the relationship of variables which are believed to contribute to learning through different media independent of previous research inquiries.

Variables which have been found to be related to general learning are need for achievement or achievement motivation and interest. These variables will be derived from the EPPS and SVIB, respectively, and correlated directly with progress checks and end-of-semester administrative tests.

Variables which are simply believed to be related to learning through specific media are reading ability levels, listening ability, and select personality traits such as expedient versus conscientious, practical versus imaginative, conservative versus experimenting, group-dependent versus self-sufficient, relaxed versus tense. These variables will be derived from the OSU, the STEP, and the 16PF, respectively, and correlated directly with both learning modules and end-of-semester achievement. Reasons for studying the variables are to see what relationships exist between reading ability and learning through conventional texts; listening ability and learning through lectures or tapes; group dependent versus self-sufficient personality traits, learning through group discussion versus independent study, etc.

2. Experimental Design Considerations in Variables Related To Instructional Effectiveness

Although most student variables will be studied by direct correlation with achievement, the anxiety variable will be used to stratify groups and the anxiety-media or presentation interaction observed. The reason for this particular treatment of anxiety is that anxiety is the one variable which typically yields a curvilinear correlation with learning, i.e., students who are very high or low in anxiety perform more nearly the same than students with moderate anxiety (Loree, 1965).

Measures of anxiety will be derived from the 16PF.

Need for achievement may also be used as a variable for stratification and subsequent study of interaction where instructional management decisions occur and motivation level is increased. An example of this stratification would be in learning modules where enrichment exercises or high probability responses are made contingent on the completion of a given task.

Interaction of anxiety and need for achievement will be studied in conventional two- or three-way classification analyses of variance.

E. STUDENT VARIABLES AND ACADEMIC PERFORMANCE

There has been much research activity on the relationship of student variables to general learning and on the predictive power of student variables in relationship to academic performance with conventional presentational forms. However, there is little evidence which indicates that various student characteristics will predict learning from specific instructional media and presentation forms. In addition, the factors which will predict performance in a personal-interactive course such as the Naval Leadership course have not been isolated.

In general, research has indicated that most variation in academic success in any conventional classroom is due to an interaction of

motivation, study habits, past learning experience, and intelligence, plus certain chance factors associated with the students' performance and the instructor's grading. Where the subjective aspects of teacher appraisal and grading have been controlled, such as in standardized achievement testing, academic aptitude is the most important variable in determining grades of students (Loree, 1965).

Other variables which have been found to be related to variations in classroom performance are socio-economic background, need for achievement or motivation, self-perception of academic ability, and anxiety (Loree, 1965).

The best single predictor of college freshman grades is high school grade point average. Variables from test scores which seem to predict early academic success in college most accurately in decreasing order of effectiveness are:

(1) achievement tests of high school course contents.

(2) general college aptitude tests such as the ACE and SAT.

(3) general scholastic aptitude tests such as Otis and Henmon-Nelson.

(4) special aptitude tests, such as verbal and numerical parts of the multi-factor tests of mental abilities (Goldman, 1961).

Although these predictors may be effective for predicting performance of college freshmen, they lose some of their predictive powers for subsequent college performance. Kring and Stolurow (1968) cite studies which indicate that precollege variables are not altogether effective for long-range prediction.

Better predictors are the most recently collected data. That is, rather than high school grade average, the better predictor would be grade average from the preceding year. Also, achievement, aptitude, and especially personality and interest data, should be collected in close proximity to the semester's performance to be predicted.

1. Isolation of Variables Which Predict Academic Success

To find that certain student variables relate to academic performance will by no means imply that students rich in those traits will make the best leaders. It will mean, however, that students high in those traits can be expected to assimilate more readily the academic knowledge requisite to the course, and consequently requisite to the theory of military leadership. The extent to which this prediction is important is proportional to the extent to which the course is necessary or important.

The steps which will be taken to isolate student characteristics which predict academic success in the leadership course are:

(1) to isolate variables which predict general academic success at the academy.

(2) to study the relationship of the general academic predictors to success in the Leadership course.

(3) to study the effectiveness of additional select variables which research indicates may be of value in predicting success in an academic personal-interactive course.

Predictors of general academic success will be determined for freshmen by multiple regression analysis of predictor variables available through the academy. These include SAT-V, SAT-Q, English Achievement, Mathematics Achievement, converted high school rank in class, and recommendation scores. These variables have been found to consistently predict academic success in a variety of undergraduate schools (ETS, 1967).

Computation of the multiple regression equation will be handled by Educational Testing Service at the request of the Naval Academy with Westinghouse Learning Corporation (WLC) serving as liaison. The regression equation for predicted grade average (PGA) will then be used to determine the entering base ability level of students within the experimental class. Predicted grade averages will also be compared with actual grade average within the standard error of estimate to determine predictive efficiency of the variables.

A second aspect of the prediction section will be to study these basic predictive variables in relation to end-of-semester success in the Leadership course. The same six academic predictors will be correlated with final course achievement, and the obtained multiple correlations for both general achievement and leadership achievement will be compared. It is expected that predictors of general achievement will result in a significantly higher correlation with freshman grade point

average than with leadership achievement, since the course is structured to compensate for individual differences in ability.

The final aspect of the prediction section will be to study the effectiveness of additional select variables which may predict success in this particular course. Variables to be included in the multiple correlation along with PGA predictor variables are interest, need for achievement, and freshman grade point average. The interest variable will be represented by select scale scores from the Strong Vocational Interest Blank (i.e., academic interest and Air Force and Army officer interest). Need for achievement will be represented by scale scores from the Edwards Personal Preference Schedule. Freshman grade point average will give additional information on entering base level of ability and also motivational level. Unlike aptitude scores, freshman grade point average may be found to relate significantly to end-of-course achievement since is is partially an index of motivation rather than pure ability.

F. STUDENT PREFERENCE

Research on student attitudes toward different modes of instruction is generally inconclusive. Using programed instruction, for example, group attitudes may be favorable and yet attitudes may be vastly different from student to student (Eigen, 1963).

General statements from research findings on student attitudes toward programed instruction and automated instruction are:

(1) Students feel that they learn more from a combination of programed instruction and conventional teaching than from either alone (Hickey, 1962; Holland, 1960; Eigen, 1963; Smith and Moore, 1962).

(2) Students feel that with the same amount of time and effort they learn somewhat more from programed instruction than from a conventional text (Holland, 1960).

(3) Students have a somewhat more favorable reaction to programed textbooks than to teaching machines (Eigen, 1963; Smith and Moore, 1962).

(4) Students' attitudes toward programed instruction appear to have no significant relationship to how much they actually learn by the method (Eigen, 1963).

Attitudes toward programed instruction among high intelligence students appear to be a function of the program itself. One expressed attitude among students in this group is that it is considered the "best method of learning" for good students, because they are not held back by the rest of the class (Eigen, 1963). On the other hand, other studies report generally favorable reactions to programed instruction, but report objections to the amount of repetition, the short steps in the program, and sustained exposure to the program (Smith and Moore, 1962; Van Atta, 1961).

1. Isolation of Variables Related to Student Preference

In this study, the attempt will be made to determine student attitudes in relation to:

(1) media, e.g., programed textbooks, films, tapes, etc.

(2) presentation design features, e.g., size of step, encoding form, duration, response demand form, amount of repetition, branching, remediation, and enrichment exercises.

(3) task variables.

(4) other student variables.

Attitudes will be determined by giving students a seven-point rating scale at appropriate points throughout the course. In addition to the rating scale, students may be asked to rank order media, or they may be interviewed individually as a check on the reliability of the rating forms.

VI. STATISTICS SUMMARY

To the extent that the exact sequence of course presentation is undecided, the specific statistics which will be employed in analyzing media-presentation design comparisons and relationships of student characteristics cannot be specified for the independent hypotheses discussed in Part II. However, it is possible to give a general indication of the several statistical procedures which will probably be used when the presentation is decided.

Anticipated statistical procedures will be standard data manipulations, which are relatively simple. These statistics are grouped on the basis of the general outline for the present report.

A. VALIDATION OF MATERIALS

The gain score ratio of actual gain over maximum possible gain will be derived from pre- and post-test discrepancies. This index will provide an estimate of overall instructional system effectiveness and topic unit effectiveness (Stolurow, 1968; Ellis, 1964).

B. DEVELOPMENT OF EVALUATIVE MEASURES OF ACHIEVEMENT

Split-half correlation methods will be employed to estimate test reliability. The Spearman-Brown formula will be applied to the split-half correlations to correct for test length (Lyman, 1963). Estimates of reliability will also be made from the means and standard deviations of tests (where the assumption of equal item difficulty can be made) using the Kuder-Richardson "formula 21" (Gulliksen, 1950).

Estimates of item difficulty will be made using simple percentages based on the number of students who respond correctly to items on the pre-test. Estimates of item difficulty will be

made based on high and low group item-total score comparisons and high and low pre-test groups and high and low post-test groups (Loree, 1965). A table for use in fourfold contingency tests will be used to assess the significance of group contrasts on item discrimination power (Mainland and Murray, 1952).

Item-total score correlations will be made to estimate the relative contribution and consequent validity of each item on the test.

C. STUDENT CHARACTERISTICS

The relationship of student characteristics to learning modules and total course achievement will be assessed primarily by correlation methods. The Pearson Product Moment correlation will be used to determine individual relationships. Multiple correlation analysis and multiple regression will be used to assess the relative contribution to achievement of a number of independent student variables simultaneously (McNemar, 1962). Where decided relationships exist, analysis of co-variance may be employed to control for the differences in treatment variance contributed by student characteristics (Lindquist, 1956).

Student characteristics may also be studied in conventional two-way or three-way analyses of variance in relationship to differences between media or presentation design forms. Special cases of student characteristics treated in analysis of variance will probably be (1) pre-determined student variables such as anxiety and need for achievement, which typically yield curvilinear correlations, and (2) post-determined student variables which have been found to correlate significantly with one or another of the treatment variables already being compared by analysis of variance.

D. EXPERIMENTAL DESIGN

Tests of the specific hypotheses proposed in the experimental section (Part II) will be primarily:

(1) T-tests for the significance of difference between treatment means (McNemar, 1962).

(2) treatment X subject analysis of variance in which all subjects receive all treatments (Lindquist, 1956).

(3) two-way classification analysis of variance in which two dimensions of treatments or two levels of student variables are compared simultaneously with differences in treatment means (treatment X levels; Lindquist, 1956).

(4) three-way classification analysis of variance in which two dimensions of treatments and two levels of student variables are compared simultaneously with differences in treatment means.

E. GENERAL DISCUSSION

For the most part, criterion measures used in the experimental design will be test scores. Because of the small number of students participating in the initial experimental course, treatments will be repeated in such a way that all subjects will be exposed to all treatments. For example, Group A may first be exposed to a lecture and then be exposed to a taped lecture. Group B may first be exposed to the taped lecture over the same content as A

and then be exposed to a live lecture. In this way the treatments are counterbalanced over the same content and all subjects are exposed to all treatments. In such a case, total scores for lecture and for taped lecture are obtained by simply adding across lecture and then adding across taped lecture. The applied statistic would then be treatment X subject, since the same subjects appear in both groups rather than replications of treatments for the same subjects.

In cases where the total group is divided into two independent groups, a simple T-test would be made for differences between the groups. No special statistic will be applied for replication of treatments over the same subjects since replications are not made on independently drawn samples.

Stringent probability levels for acceptance or rejection of null hypotheses will not be set in the initial stages of this study, since the overall purpose of the first study is to identify trends rather than draw generalizable conclusions from the results. It is recognized that the computation of a large number of T-tests and analyses of variance may yield a number of seemingly significant mean differences which have actually occurred by chance. This possibility is recognized and has been weighed heavily in the selection of statistics. However, it is felt that since the initial study is largely exploratory, it is possible to simply note the experiments which provide significant results during the first year and repeat those hypotheses and studies in the second and third years. It is felt that in the beginning stages of a three-year project, it is better judgment to risk making a Type 1 error of rejecting a hypo-

thesis of no differences between treatment means than of failing to reject a hypothesis of no differences between treatment means. Accordingly, where hypotheses have been rejected as showing significant differences, these studies can be replicated to retest the hypotheses in subsequent courses.

VII. DATA PROCESSING REQUIREMENTS FOR DATA BASE

A. DATA BASE

The data base for the project will consist of a number of separate files that are accessible to any program, statistical routine, or retrieval query. The six files, as they are currently established, will contain:

(1) student information and identification. This file contains background information on the student, information on his performance at the Naval Academy, and scores on various psychological tests.

(2) data file for content objectives and their classifications. This file contains identification of each objective and characteristic data for each student. This data includes appropriate values as applicable to the dimensions of presentation.

(3) data file for module classification. This file, besides identifying the module by chapter and segment, specifies each presentation dimension (duration, response demand, stimulus encoding, management decision, and response demand frequency).

(4) data file for segment test. This file uniquely identifies each test and each test question, as well as recording the responses by students and the relationship between each test question and the appropriate objective in the course.

(5) the data file for dependent variables. This file includes information on module cost, time, decision

criteria, and preference rating on tests and modules.

(6) data file on classroom performance. This file will list and summarize student performance on each module and element in the course.

B. PERFORMANCE CHARACTERISTICS

Before considering the operations to be performed upon these files, certain characteristics should be noted. The first four files are fixed. That is, they will be loaded with data before the course begins and will be referenced for analysis and correlation to student performance. They will not be updated on a regular basis. The last two files will be updated on almost a daily basis with data from student tests, questionnaires, and research analysis. The key characteristic of operations on the first four files will be the requirement to access discrete and identifiable portions of the record. There will be no manipulation of string data as in files five and six.

Moreover, since the characteristic of the research will be to ask how student performance correlates to the characteristics of learner variables, presentation variables and the like, the task of file design would be to insure the discrete labeling of every characteristic or element. Correlations may also be drawn between student performance and discrete characteristics from different files. For example, a researcher may wish to know not only the relationship between a student's test score and his SAT Verbal Score, but also the performance relationship to media design and learning styles. Such correlation analysis is only viable when cross-referencing throughout the file is insured. Hence, each record must contain necessary indices to

items in the same file or to related items in another file.

The use of indices within the data items themselves are advantageous in considering another aspect of the data base concept, i.e., information retrieval. While it is obvious that retrieval data should be pertinent to the inquiry only and unencumbered by extraneous data, it should be remembered that some responses should be able to append relevant data within the necessity to detailed parameters. For example, it may be highly desirable that a response to a query of behavioral objectives should include information on test questions associated with particular objectives regardless of number or location.

C. DATA MANAGEMENT SYSTEM

With such requirements, the need for a data management system becomes clear, assuming such capabilities are either already available or can be developed. It is desirable to use an already existing information handling system rather than attempting to generate one for this unique purpose. There are file handling systems already in existence in tested software packages. One, under current consideration, is the IBM Generalized Information System, known as the GIS. The IBM GIS system expands upon the 360 operating system data management package to create, maintain, and query files. It requires the use of configuration of a 360/40 or better, with 132 K bytes.

The GIS system operates on the principle of a common data base serving multi-users or application programs. The common data base is essentially a collection of separate data files which are unified through the use of a common file descriptor

table. This file table contains a unique method of access for each file. Through the creation of synonyms for each file and each element of the file at creation time, a program can access a particular data file in the common data base by using the unique synonym as his key for the search. The data management system searches the descriptor table and loads the necessary data for the program. The descriptor table in conjunction with the common data base makes standardized programing possible and desirable, and at the same time it allows for flexible procedures and one-shot report inquiries.

Extension will include specifications for file format, data collection, I-O requirement, processing routine, and operation procedures in accordance with further definitive prescriptions of project research and requirement.

REFERENCES

Allen, H. Research in instructional media and art education. Final Report of the Uses of Newer Media in Art Education Project. National Art Education Association, Washington, D.C., August 1966.

Alter, Millicent, and Silverman, R. The response in programmed instruction. Journal of Programmed Instruction. 1962, 1 (1), 55-78.

Ash, P., and Jaspen, N. The effects and interactions of rate of development, repetition, participation, and room illumination on learning from a rear projected film. Technical Report SDC 269-7-39. Instructional Film Research Reports. Port Washington, Long Island, N.Y.: U.S. Naval Special Devices Center, 1953.

Becker, S.L., and others. Communication skills. An experiment in instructional methods. Iowa City: The State University of Iowa, August 1958.

Berelson, B. and Steiner, G.A. Human Behavior: An Inventory of Scientific Findings. New York: Harcourt, Brace and World, Inc., 1964.

Berger, R. J. A test of the Premack hypothesis with responses to audio and visual stimuli. Unpublished doctoral dissertation, Arizona State University, 1965.

Bloom, B.S. (Ed.) Taxonomy of Educational Objectives. New York: Longmans, Green, 1954.

Briggs, L.J. Self-pacing versus automatic pacing in paired-associate learning. In Lumsdaine, A.A. (Ed.) Student response in programmed instruction. A symposium on experimental studies of cue and response factors in group and individual learning from instructional media. Sponsored by Headquarters, Air Research and Development Command, U.S. Air Force. Washington, D.C.: National Academy of Sciences, National Research Council, 1961.

Brown, J.I. The importance of structured outside reading assignments. College and Adult Reading, II, Second Annual Yearbook, North Central Reading Association, 1963. (Abstracted in part in NAEB Journal 21, March 1962.

Carpenter, C.R., and Greenhill, C.P. Comparative Research on Methods and Media for Presenting Programmed Courses in Mathematics and English. University Park: Pennsylvania State University, March 1963.

Csanyi, A.P. An investigation of visual versus auditory programming in teaching foreign language pronunciation. In Investigations of the Characteristics of Programmed Learning Sequences. Pittsburgh: University of Pittsburgh, Programmed Learning Laboratory, 1961. 135-150.

Cronbach, L. J. Essentials of Psychological Testing. New York: Harper and Brothers, 1949.

Crowder, N.A. Automatic tutoring by means of intrinsic programming. In Lumsdaine, A.A., and Glaser, R. (Eds.). Teaching Machines and Programmed Learning. Washington: National Education Association, 1960, 286-298.

Educational Testing Service. Designing Validity Studies and Collecting Data. College Entrance Examination Board, 1967.

Edwards, A.L. Experimental Design in Psychological Research. New York: Holt, Rinehart and Winston, 1950.

Eigen, L.D. High school student reaction to programmed instruction. Phi Delta Kappan, 1963, 44, 282-285.

Ellis, H.C. Research Design in Programmed Learning. National Society for Programmed Instruction. August 1962, 7-8.

Evans, R.I. An examination of students' attitudes toward television as a medium of instruction in a psychology course. Journal of Applied Psychology, 1956, 40, 32-34.

Feldhusen, J.F., and Birt, A. A study of nine methods of presentation of programmed learning material. Journal of Educational Research, 1962, 55, 461-466.

Feldhusen, J.F., and Eigen, L.D. Interrelationships among attitude, achievement, reading, intelligence, and transfer variables in programed instruction. Paper presented to the Midwestern Psychological Association, May, 1963.

Ferster, C.B., and Sapon, S.M. An application of recent developments in psychology to the teaching of German. Harvard Educational Review, 1958, 28, 58-69.

Frank, J. H. An evaluation of closed circuit televisi for interceptor pilot training. Dissertation Abstracts, 1955, 15, 2060-2061.

Frye, C.H. Group vs. individual pacing in programmed instruction. In Schramm, W. The Research on Programmed Instruction. Washington, D.C.: U.S. Department of Health, Education and Welfare, 1964. (OE-34034) (Abstract).

Fulton, W.R., and Rupiper, O.J. Selected vicarious experiences versus direct observational experiences of pre-service teachers in the foundation areas of professional preparation at the University of Oklahoma. USOE Project No. 192. Norman: College of Education, University of Oklahoma, 1961.

Gagne, R.M. The Conditions of Learning. New York: Holt, Rinehart and Winston, 1965.

Gagne, R.M., and Brown, L.T. Some factors in the programming of conceptual learning. Journal of Experimental Pyschology, 1961, 62, 313-321.

Glaser, R. and Schaefer, H. Principles of Programming Printed Materials. Pittsburgh: University of Pittsburgh, 1961. Final report on Contract AF 33(616)-7175.

Goldman, L. Using Tests in Counseling. New York: Appleton-Century-Crofts, Inc. 1961.

Goldstein, L.S., and Gotkin, L.G. A review of research: teaching machines versus programmed textbooks as presentation modes. Journal of Programmed Instruction, 1962, 1 (1), 29-36.

Greenhill, L. P. Review of trends in research on instructional television and film. In: Reid, J.C., and MacLennan, P.W. Research in Instructional Television and Film. U.S. Department of Health, Education, and Welfare, Bureau of Research. Catalog #FS-5.234:34041, 1967.

Gropper, G. L. Learning from Visuals: Some Behavioral Considerations. AV Communication Review. Spring, 1966.

Gropper, G. L. Why is a picture worth a thousand words? AV Communication Review, 1963, 11, 75-95

Gropper, G. L. and Lumsdaine, A.A. An experimental comparison of a conventional TV lesson with a programmed TV lesson requiring active student response. Studies in Televised Instruction, Report No. 2. Pittsburgh: Metropolitan Pittsburgh Educational Television Stations WQED-WQEX, and American Institute for Research, 1961a.

Gryde, S.K. The feasibility of programmed television instruction. AV Communication Review. 1966, 14 (2), 71-89.

Gulliksen, H. Theory of Mental Tests. New York: John Wiley & Sons, Inc., 1950.

Hatch, R.S. and Flint, L.L. Programed Learning: A Comparative Evaluation of Student Performance Variables Under Combinations of Conventional and Automated Instruction. New York: U.S. Industries, Educational Sciences Division. 1962.

Hickey, A.E., Autor, S.M., and Robinson, E.J. Programmed Instruction Integrated with Broadcast ETV. Boston: Northeastern University. 1962.

Holland, J.G. Evaluating teaching machines and programs. Teachers' College Record, October 1961, 56-65.

Holland, J.G. Teaching psychology by a teaching machine program. In Berelson, B. and Steiner, G.A. Human Behavior: An Inventory of Scientific Findings. New York: Harcourt, Brace and World, Inc., 1964.

Homme, L.E. (principal investigator). A demonstration of the use of self-instruction and other teaching techniques for remedial instruction of low-achieving adolescents in reading and mathematics. Submitted to U.S. Office of Education, August 1965.

Kasten, D.F., and Seibert, W.F. A study of televised military science instruction. TVPR Report No. 9. Lafayette, Indiana.: Purdue University, July 1959.

Kimble, G.A., and Wulff, J.J. The effects of 'response guidance' on the value of audience participation in training film instruction. Human Factors Operations Research Laboratories, USAF, Report No. 35. Audio-Visual Communication Review, 1953, 1, 292-293. (Abstract).

Kring, W.K., and Stolurow, L.M. Predictive Variables. Technical Recommendations No. 5.2 USNA Contract No. N00161-7339-4781, July 25, 1968.

Levitt, E. Clinical Research Design and Analysis in the Behavioral Sciences. Springfield, Illinois: Charles C. Thomas (publisher), 1961.

Lindquist, E.F. Design and Analysis of Experiments in Psychology and Education. Boston: Houghton Mifflin Company, 1956.

Logan, F.A. Micromolar behavior theory and performance speed in education. Harvard Educational Review, 1963, 33(2), 178-185.

Loree, M. R. Psychology of Education. New York: Ronald Press, 1965.

Lyman, H.B. Test Scores and What They Mean. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1963.

Maccoby, N. and Sheffield, F.D. Theory and Experimental Research on the Teaching of Complex Sequential Procedures by Alternate Demonstration and Practice. In symposium on Air Force human engineering, personnel, and training research. Washington, D.C.: National Academy of Sciences National Research Council, 1958.

Macomber, F., and others. Experimental Study in Instructional Procedures. Oxford: Miami University, October 1, 1957.

Mainland, D., and Murray, I.M. Tables for use in fourfold contingency tests. Science, 1952, 116, 591-594.

Margolius, G.J. and Sheffield, F.D. Optimum methods of combining practice with filmed demonstration in teaching complex response sequences: Serial learning of a mechanical assembly task. In: Lumsdaine, A.A., (ed.), Student Response in Programed Instruction: A Symposium on Experimental Studies of Cue and Response Factors in Group and Individual Learning from Instructional Media. Washington, D.C.: National Academy of Sciences, National Research Council, 1961.

McNemar, Q. Psychological Statistics. New York: John Wiley & Sons, Inc., 1962.

Moore, J.W., and Smith, W.I. Knowledge of results in self-teaching spelling. Psychological Reports, 1961, 9, 717-726.

Parry, S.B. To err is human ... and sometimes desirable. Programed instruction newsletter, 1963, 2 (4), 4-5.

Porter, D. Some effects of year long teaching machines instruction. In E. Galanter (ed.), Automatic Teaching. New York: Wiley. 85-102.

Reid, J.C. and MacLennan, P.W. Research in Instructional Television and Film. U.S. Department of Health, Education, and Welfare, Bureau of Research, Catalog #FS-5.234:34041, 1967.

Schramm, W. The Research on Programed Instruction. Washington, D.C.: U.S. Department of Health, Education, and Welfare, 1964. (OE-34034)

Shay, C.B. Relationship of intelligence to step size on a teaching machine program. Journal of Educational Psychology, 1961, 52, 98-103.

Silberman, H.F. "Self-teaching devices and programed materials." Review of Educational Research, 32, 179-193.

Silverman, R., and Alter, Millicent. Response mode, pacing, and motivational effects in teaching machines. Port Washington, N.Y.: U.S. Naval Training Devices Center. Tech Rep. 507-03, 1961.

Smith, R.G. Jr. The Design of Instructional Systems. The George Washington University Human Resources Research Office. Technical Report 66-18. November 1966.

Smith, W.I. and Moore, J.W. Programed Materials in Mathematics for Superior Students in Rural Schools. Lewisburg, Pa.: Bucknell University, 1962.

Stolurow, L.M. Instructional Systems Validation Planning. Technical Recommendation No. 2. USNA Contract No. N00161-7339-4781, March 29, 1968.

Stolurow, L.M. Implications of current research and future trends. Journal of Educational Research, 1962, 55, 518-27.

Taber, I.I., Glaser, R., and Schaefer, H.H. Learning and Programmed Instruction. Massachusetts: Addison Wesley, 1965.

Tosti, D.T., and Ball, J.R. A Behavioral Approach to Instructional Design and Media Selection. Behavior Systems Division, Westinghouse Learning Corporation: Albuquerque, New Mexico, 1968.

Twyford, L. Film profiles. Instructional Film Research Reports. Technical Report SDC 269-7-23. Port Washington, Long Island, N.Y.: U.S. Special Devices Center, 1951.

Van Atta, L. Behavior in Small Steps. Contemporary Psychology, 1961, 10, 378-81.

Weiss, W., Maccoby, N. and Sheffield, F.D. "Combining Practice with Demonstration in Teaching Complex Sequences: Serial Learning of a Geometric-Construction Task." In Lumsdaine, A.A., (ed.), Student Response in Programed Instruction: A Symposium on Experimental Studies of Cue and Response Factors in Group and Individual Learning from Instructional Media. Washington, D.C.: National Academy of Sciences, National Research Council, 1961.

Westover, F.L. A comparison of reading and listening as a means of testing. Journal of Educational Research., 1958, 52, 23-26.

Wood, D. A. Test Construction. Columbus, Ohio: Charles E. Merrill Books, Inc., 1961.

ADDENDUM

Buros, O.K. The Fifth Mental Measurements Yearbook. Highland Park, New Jersey: The Gryphon Press, 1959.

Buros, O.K. The Sixth Mental Measurements Yearbook. Highland Park, New Jersey: The Gryphon Press, 1965.

Evans, J. Behavioral objectives are no damn good. Technology and Innovations in Education. New York: Frederick A. Praeger, 1968.

Follettie, J.F. Effects of Training Response Mode, test form, and measure on acquisition of semi-ordered factual materials. Research Memorandum 24. Fort Benning, Georgia: U.S. Army Infantry Human Research Unit, April 1961.

Gilpin, J. Design and Evaluation of Instructional Systems. Paper presented in a symposium at the 1961 meetings of the American Psychological Association. Richmond, Indiana: Earlham College, 1961.

Rosenthal, R. Experimenter Effects in Behavioral Research. New York: Appleton-Century-Crofts, 1966.