DOCUMENT RESUME

ED 071 569 HE 003 700

AUTHOR Pascale, Pietro J.; Dittenhafer, Clarence A.
TITLE Application of an Instructional Evaluation for University College.
PUB DATE 72
NOTE 24p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Evaluation Methods; *Evaluation Techniques; *Higher Education; *Teacher Evaluation; *Teacher Improvement

ABSTRACT

In this document, two evaluations of college instruction models are proposed. Model A, a very objective summative evaluation is made utilizing behavioral objectives. The evaluating committee secures copies of the instructor's course examination. The evaluation committee will analyze the examination from the standpoint of instructional emphasis, i.e., content, critical thinking, etc. After the examination is given, the instructor will score the students' responses and assign grades. The instructor will pass the examination on to the evaluation committee where student responses will be analyzed in relation to the course objectives as inferred by the committee from examination of the test. The committee would then determine whether or not the instructor was teaching the objectives being tested. In Model B, evaluation would be built into the course planning. The instructor would consult with the evaluation committee and explain the desired outcomes for students in his course. The committee would use the information obtained from the instructor to formulate tentative behavioral objectives for the course. After the course instruction has begun, the evaluation committee could use the instructor's student evaluation procedures to evaluate the instructor. The evaluation would then take the form of Model A in determining the degree of congruency between course objectives and student performance. (HS)

APPLICATION OF AN INSTRUCTIONAL EVALUATION FOR UNIVERSITY COLLEGE

Pietro J. Pascale
Youngstown State University
and
Clarence A. Dittenhafer
Rutgers University
1972

Abstract

Two evaluation of college instruction models are proposed. The proposed summative evaluation Model A is developed and examined in a real situation at University College. The situation is a Small Business Management course.

Part I

Context of the Evaluation Problem

Many instructors closest contact with evaluation is the assignment of a grade at the end of a semester. Even the grading procedure is simplified when a pass - fail system is adopted. Other instructors may evaluate their instruction by overall subjective inferences based on papers submitted by students at the end of the semester. Still other instructors use various unsystematic schemes and label them evaluation. But are these activities evaluation? Few university college instructors would submit to one or all of the above procedures as a basis for promotion decisions.

Unfortunately evaluation precipitates anxiety in instructors because it is usually tied to goals of merit, worth, and value (Scriven, 1967). Thus, the major objection to systematic evaluation is that value judgments must be made and these are essentially subjective and not scientific. However, there is no support for the conclusion that educational evaluation is less than a fully appropriate goal for applied science (Atkin, 1963).

If as "professional" educators we believe scientific inquiry can be applied to our evaluation endeavors, what

avenues of attack are available? To attack the problem from a somewhat theoretical base might provide some insights into the two models developed later in the paper. For purposes of evaluating instruction, evaluation procedures can be viewed as representing a continuum of objectivity.

At one end of the continuum are the proponents of evaluating instruction (and instructors) by applying a set of absolute standards which will somehow be obvious to the educator (Scriven and Stake, 1968). The absolutist approach would probably provoke the largest outcry from potential clients. However, the absolutist's position removes the necessity of the evaluator to consider the individual instructor's or the institution's educational goals. Any instructor who feels threatened by evaluators holding absolute standards would be ignored or at least discounted. The absolutist's model would use a team of specialists to determine educational goals and practices where individual evaluators are unable to do so.

The absolutist's model is not without it consequences, the most obvious being that many instructors would be unwilling to cooperate and work with evaluators (not having formulated evaluative standards). Most educational institutions look rather unfavorably (irrespective of the economic efficiency of the evaluation) on the absolute standard's model. Any results from such an approach are certain to be inconclusive and meaningless, primarily due to the lack of general agreement on educational standards (Stake, 1968).

The absolute standard's model does not appear appropriate for evaluation in University College.

At the other end of the objectivity continuum, one is confronted with the subjectivist's model. The rationale is that evaluation is conducted by the teacher with the help of the class and of their joint effort, however subjective, is successful to the extent it results in decisions which turn out to have educative consequences (Thelen, 1969). The model would probably find much favor with sociologists because much emphasis is placed on group processes. The student becomes a non-individual in the aspects of the classroom regarded by the instructors as being relevant.

The model becomes a cooperative feedback evaluation loop. Proponents contend the subjective procedures by the group get at the same basic discipline of evaluation as vigorous testing, i.e., the nature of discrimination, identification of prototypical situations. The instructor in the subjective model addresses himself to diagnosing and trouble-shooting the role of the student. More precisely, he is defining, delimiting, correcting, activating, and reinforcing that role (Thelen, 1969). Because of the declared purpose of the proposed evaluation in University College, the subjective model is not appropriate.

The Proposed Models

After discussion with evaluation committee members on the University College staff, it became evident that "middle-

of-the-road" (relative to the objectivity continuum) models might be appropriate. The models involve the use of behaviorally stated objectives. The models require the statement of the instructor's course objectives in terms of measurable student behavior. The evaluation becomes the process of relating antecedent conditions to outcomes and outcomes to objectives (Stufflebeam, 1966).

Proposed Evaluation Model A

In Model A, a very objective summative evaluation is made utilizing behavioral objectives. The evaluating committee, as designated by University College, secures copies of the instructor's course examination (probably final or mid-term). The instructor will have noted the desired student responses and scoring for each question. The evaluation committee will analyze the examination from the standpoint of instructional emphasis, i.e., content, critical thinking. The approximate percentages of instructional emphasis should provide the evaluators with insights into the instructor's course objectives.

After the examination is given, the instructor will score the students' responses and assign grades. The instructor will pass the examination on to the evaluation committee. The student responses will be analyzed in relation to the course objectives as inferred by the committee. The rationale for the model lies in the assumption that the student scores on the examination items is a reflection of instructional emphasis. The degree of congruency between the inferred

objectives of the evaluation committee and those reflected in the students' scores represents the evaluative data for the course. From the data, several conclusions are feasible: (1) high congruency indicates agreement between instructional emphasis and course objectives, and (2) low congruency indicates the test is not measuring the course objectives; the instructor is not testing the objectives being taught; or the instructor is not teaching the objectives being tested.

The evaluation committee would use the objective data from the above model to make an evaluative judgment of teacher effectiveness. There are inherent problems in the model. One argument could be advanced that use of the model would stultify the curriculum to a set of behavioral objectives. A more common argument when objective data is sought is the eventual defining of instructional goals around a test, i.e., teaching the test. These and others are legitimate arguments against the proposed model. However, the model does not attempt to remove individual instructor initiative and creativity. The model attempts to provide a systematic procedure for removing a portion of the subjective judgment inherent in all evaluation.

Proposed Evaluation Model B

The evaluative procedures in the second model are probably oriented toward a formative evaluation. In other words, evaluation would be built into the course planning. The in-

structor would consult with a committee (possibly the evaluation committee) and explain the desired outcomes for students in his course, i.e., content, critical thinking.

The committee would use the information obtained from the instructor to formulate tentative behavioral objectives for the course. Since course content varies among courses, the individual instructor would supply the measurable criterion for the committee's proposed objectives. A second meeting with the instructor might result in mutual agreement on course objectives. The instructor individually or in cooperation with the committee would formulate the instructional procedures necessary to accomplish the course objective.

After the course instruction has begin, the evaluation committee could use the instructor's student evaluation procedures (test papers, etc.) to evaluate the instructor. The evaluation would take the form (as in Model A) of determining the degree of congruency between course objectives and student performance.

Objectifying the data would provide information for the evaluation committee's consideration. More importantly, the information (especially when gathered before the end of the course) would provide feedback to the instructor and make instructional revisions possible before the course terminates.

Because of the behavioristic approach used in both models, the same arguments advanced against the use of Model A could be applied to Model B. However, the instructional

improvement possibilities implicit in the latter merit consideration.

A Supplementary Evaluative Procedure

Many instructors teaching non-empirical courses claim extensive knowledge of student performance because of classroom participation. In many classes, this very subjective information becomes a part of the student's final grade. It is very difficult to question the instructor's judgment because usually the instructor maintains no written record of the evaluation.

The rating scale below might be one way of validating or invalidating such procedures. Before giving a major test, the instructor would rate each student on the attached (or similar) scale. After administering the test, the instructor would score the individual tests. However, the identity of each student's test would remain unknown (to the instructor) until after the tests were scored. A comparison could be made between the instructor's pre-test ratings and the student's actual test performance. High agreement between the two measures would tend to confirm the subjective judgments of the instructor about his students. Low agreement would indicate the invalidity of the procedure.

The above procedure might be incorporated into the proposed models and provide additional information for an evaluation of instructor effectiveness.

A SPECIMEN SUBJECTIVE RATING SCALE FOR COURSE EVALUATION HISTORY OF PHILOSOPHY (UNIVERSITY COLLEGE)

Subjective Rating Scale

Rate _____ on each of the statements listed
Student's Name
below. The rating is a subjective estimate of the student's understanding of the basic concepts being taught in the University College course entitled "History of Philosophy".

Directions: Please encircle the code letter best reflecting your judgment of the student's understanding of the basic concepts taught in the course. The letter code is as follows:

A - Strongly Agrees
B - Agrees
C - Undecided
D - Disagrees
E - Strongly Disagrees

1.	The student demonstrates measurable prior understanding of philosophical thought.	A B C D E
2.	The student expresses a negative attitude toward philosophical ideas.	A B C D E
3.	The student has insufficient ability to integrate and contrast historical philosophies.	A B C D E
4.	The student possesses an acceptable ability to apply philosophical thought to current problems.	A B C D E
5.	The student is able to orally present his thoughts on philosophical issues.	A B C D E
6.	The student's oral responses indicate a very limited understanding of philosophy.	A B C D E
7.	The student integrates and contrasts historical philosophies quite naturally.	A B C D E
8.	The student sees no relationship between philosophical thought and solving current problems.	A B C D E
9.	The student expresses a positive attitude toward philosophical ideas.	A B C D E
10.	The student lacks sufficient confidence to orally express his thoughts on a philosophical issue.	A B C D E

SUMMARY

The nature of the instructor evaluation effort in University College, being one part of a developing promotional apparatus, will make any evaluative procedure offensive to some. However, the proposed models as developed in this paper merit the evaluation committee's consideration. Evaluation is difficult under the most beneficial conditions. The situation in University College is certainly not the most conducive setting for good evaluation.

Most "experts" define evaluation somewhat differently. There is, however, almost universal agreement about the "value judgments" necessary at some point in all evaluation. The individual(s) given the responsibility for the judgmental part of an evaluation must have supportive evidence for their decision. Objective information removes much of the evaluator's personality from his decisions. The information is less open to challenge than is normally true of very subjective evaluation.

The proposed models present ways of getting objective information for evaluation. The University College staff must decide whether they are interested enough in evaluation to incorporate the evaluative procedures herein described. Cronbach (1963) states, "The greatest service of educational evaluation may be to identify aspects of the course where revision is desirable."

Part II

Application of the Proposed Model A

Part II of the report applies the proposed summative evaluation Model A to a University College situation. The situation is a Small Business Management Course. The authors of the report have assumed the role of evaluators.

Step I - Examination Analysis

A copy of the mid-term examination in Small Business Management was sent to the evaluators by the instructors. The committee analyzed the examination in terms of instructional emphasis as proposed in the model, i.e., content, critical thinking, transfer. The analysis follows:

The Small Business Management test is neatly divided into two parts. Part I contains items specifying the content area. Part II contains items which tap the domain of critical thinking. Part I has a maximum of approximately 40 points. Students are given a choice of any four of the five questions. The first four questions are basically content oriented questions and the fifth question has been labeled by the evaluators as one requiring critical thinking. Question 5 was worth 12 points. Most students elected not to answer the apparently difficult question 5.

Part II has a keyed total of 60 points. As in Part I, the students are given a choice of two of the three questions (case studies). The student is required to answer three of four questions relative to each case study. Thus, each case

study is worth a total of 30 points and the maximum total for Part II is 60 points. An analysis of the mid-term examination reveals that 40% of the emphasis is on content and 60% is on critical thinking.

The use of case studies in Part II is a clever device to break out of the content domain into the realm of critical thinking. Since only 40% of the mid-term assesses content skills, the instructor has avoided the most serious pitfall of a teacher designed test, viz. too many items requiring only rote recall of information (Thorndike and Hagan, 1969).

The instructor has three case studies in Part II and the student chooses two. The only caveats are: (1) to make sure there is equal weighting (which there was in Part II but not in Part I) and (2) all options are tapping the same cognitive domain. In Part I, for example, the direction to the student is to select four of the five questions. Four of the five questions are directed at content and one is directed at critical thinking. The student may avoid without penalty question 5 and consequently not give an answer for probably the most intellectually difficult item in Part I. It should be noted that the evaluators did not feel any instructional emphasis was directed specifically toward transfer learning.

The level of difficulty of this exam was generally assessed by comments from several psychometricians and several business majors. The strategy employed was simply one of consensual validation - the consensus being that the exam

was moderately difficult. The more crucial point is whether the level of difficulty of the exam matched the potential of the students required to take the exam. No comment is made concerning the last point due to lack of information.

There are several very commendable features of this exam: (1) directions are very explicit, (2) format is orderly, (3) questions are framed in such a way as to elicit a short answer response, (4) questions are not too long and there is no ambiguity, (5) excellent use of the case study approach requiring critical thinking, and (6) reasonably fine scoring key.

Step II - Analysis of Student Responses

Tables I and II present a statistical analysis of some of the more relevant characteristics of student responses on the Small Business Management test. Additional analysis revealed that the percent of total test mastery in Part I is 72. The percent of mastery in Part II is 78.

From an inspection of the distribution of scores, it seems that there is a negatively skewed distribution. A X^2 goodness of fit test can settle the hypothesis of normality, but we have not applied the test and we are making our decision on a graphical interpretation. The negative skew usually occurs when the mean < median < mode. Classical psychometric theory would state that the test was too easy if there was severe negative skew. However, our

DATA COLLECTION AND INTERPRETATION

N = 31

Central Tendency

mode = 83
median = 79
mean = 77

Dispersion

range (56-89) = 24
interquartile range = 5
standard deviation = 8
variance = 64

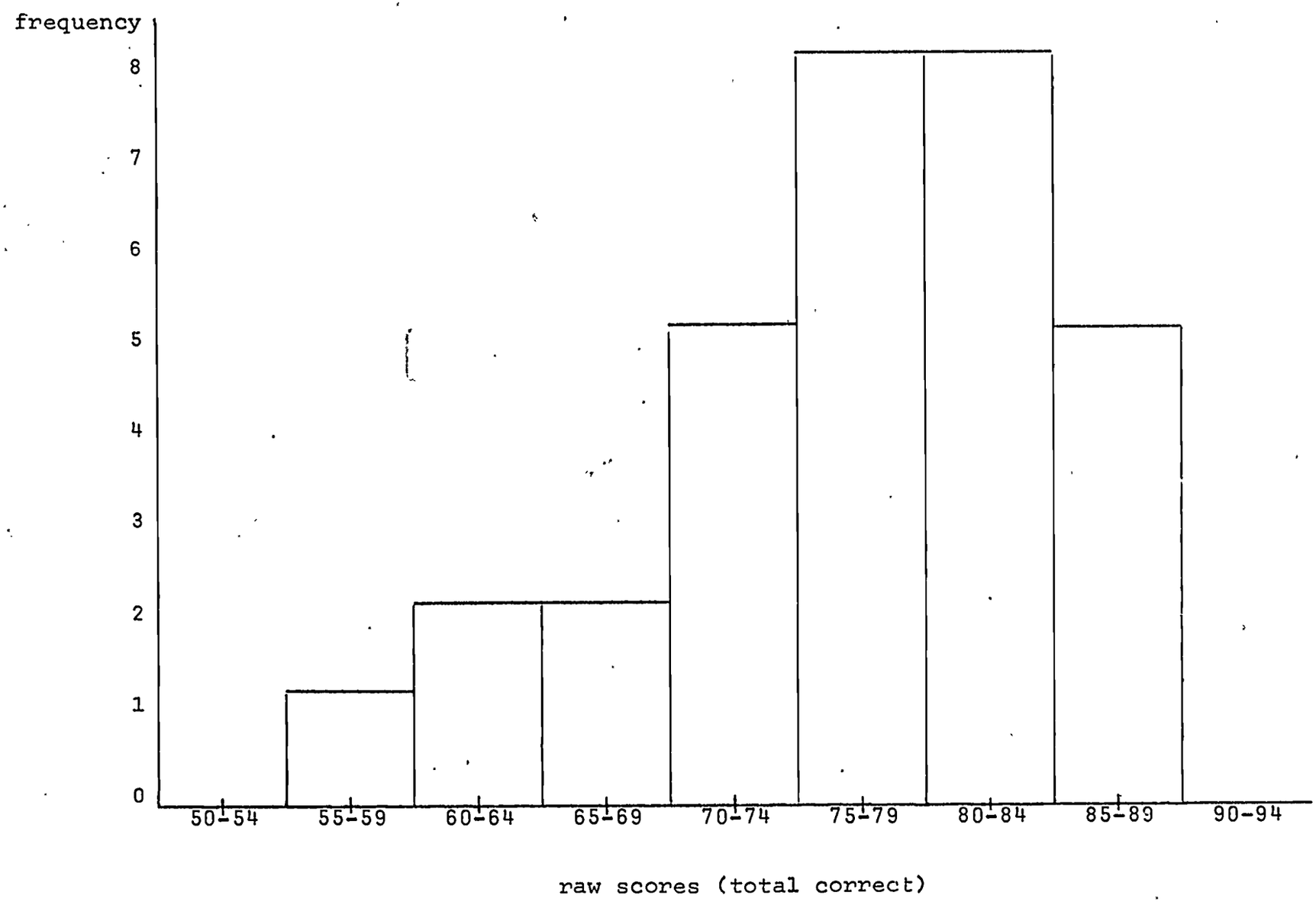
Estimates of cognitive activity required

content skills	40%
critical thinking	60%
transfer	0%

Mastery scores (percent of class correct)

content skills	72%
critical thinking	78%

DATA COLLECTION AND INTERPRETATION



evaluation model could plausibly account for the negative skew as a situation wherein the majority of students were attaining the criterion.

Step III - Inferred (By Evaluators) Overall Course Objectives

From the examination analysis, the evaluation committee inferred the following course objectives:

1. Students will demonstrate their knowledge of Small Business Management principles by describing in writing how one becomes established in a small business enterprise. The descriptions are to include the following:
   a. An analysis of the supply-demand phenomenon as it affects small businesses.
   b. How one acquires a small business.
   c. How the various community resources available to potential owner are best utilized.
   d. The part that planning must play in establishing a new business.
   e. Fiscal policy to follow in the new business.
2. The students will demonstrate critical thinking ability relating to Small Business Management by analyzing, in writing, case studies relative to the following criterion:
   a. Fiscal management (especially record keeping).
   b. Personnel management.
   c. Buying and marketing techniques.

Step IV - Summative Evaluation

Before a final course evaluation is rendered, the individ-

ual instructor should be made aware of the findings. He in turn could clarify and/or defend his position relative to the findings. However, the evaluation committee was unable to arrange an audience with the instructor in question. (Time limitations, student unrest, and distance were the primary deterrents). Therefore, the committee made an evaluative decision from available data.

The committee believes the achievement of criterion by most of the students (as reflected in the negatively skewed distribution) reflects acceptable congruency between instruction emphasis and objectives. We have some reservations about the narrow spread (72% - Part I, 78% - Part II) between the two parts when our test analysis indicated 40% content emphasis (primarily Part I) and 60% critical thinking (primarily Part II). A higher mastery level on critical thinking skills would have been desirable.

Summary

The writers of the report have tried to assume the role of evaluators for a Small Business Management course in University College. The essence of Part II was to apply evaluation Model A developed in Part I of the report. In applying the model, four steps were employed, namely; (a) Examination analysis, (b) Statistical analysis of student responses, (c) Inferred objectives, and (d) Summative evaluation.

The report writers believe the approach to summative

evaluation herein described can be an effective tool in an evaluative information arsenal. Evaluation results using the model should certainly be supported by additional information.

What can be done if an evaluator using the model finds a discrepancy between instructional emphasis and student responses? Suppose the discrepancy is traced to a poor test. The following pages illustrate how the instructor might consider developing new test-making skills (Part III). Similar incongruences in evaluation results (i.e., statement of objectives) could be rectified by helpful advice from evaluators or specialists.

Part III

Suggestions for Teacher-Made Tests

A crucial part of the evaluation process is the design of an appropriate test by the classroom teacher. In this paper, we refer to these classroom tests as custom made tests. The need for a highly sophisticated psychometric tool is obviously unwarranted but surely the validity of the evaluation process is a function of adequately designed tests. On the following pages are several diagrams (Cook, 1966) which map the sequence of steps in preparing and designing a custom made test. A list of references for test-making follows the diagrams. The texts are recent and readable. The Gorow (1966), Lindvall (1967), and ETS (1969) references are in paperback. The ETS booklet is available free of charge from Educational Testing Services and contains an excellent annotated listing of test construction books.

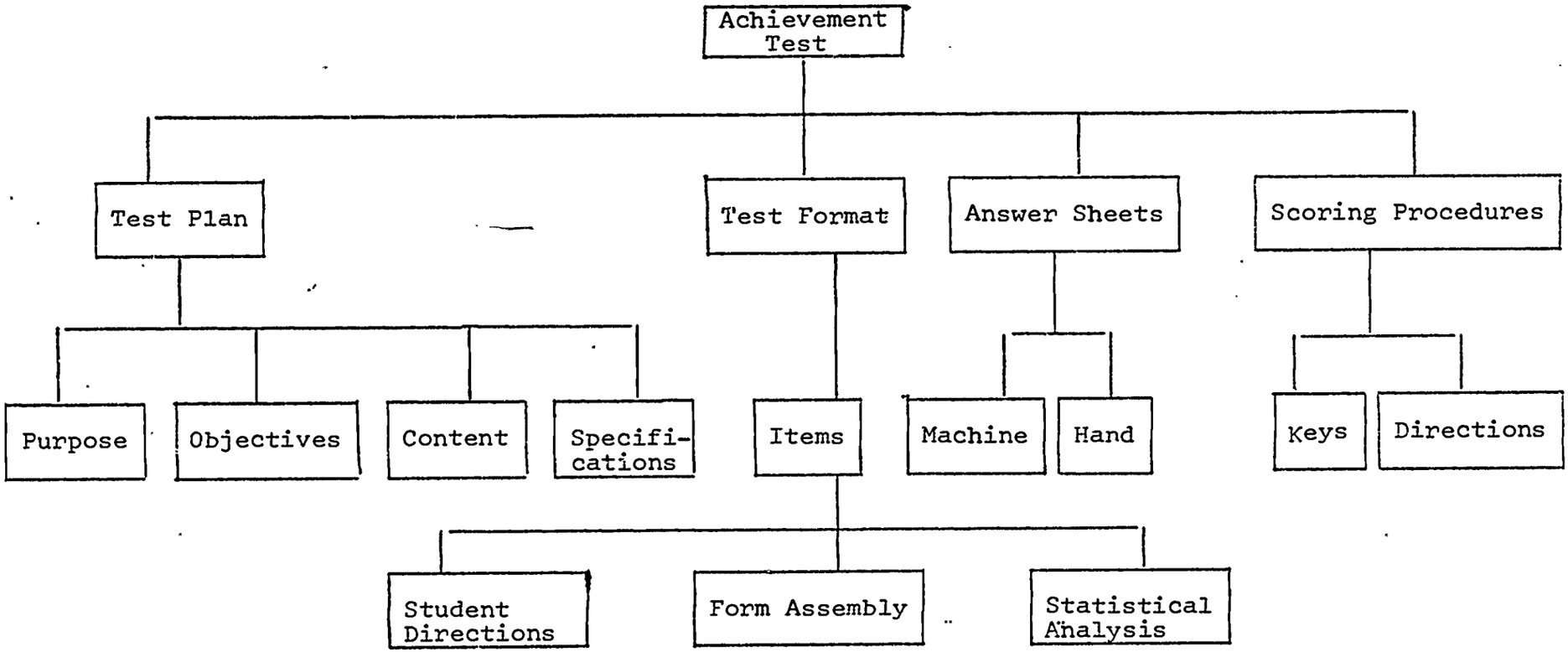
We have omitted in the PERT diagram the events of statistical analyses. If a test is viewed as an element in formative evaluation, then statistical analyses should be done and this information could be used as a guide to revision of either teaching strategies or of testing strategies.

In the tabular diagram, machine-scoring refers to the IBM Optical Scanning device which will score selection-type (multiple choice, true-false) responses. The optical scanner can be set up with a small machine (534) to give punched output. From the punched output a TESTAT program can be run which will give a complete item analysis.

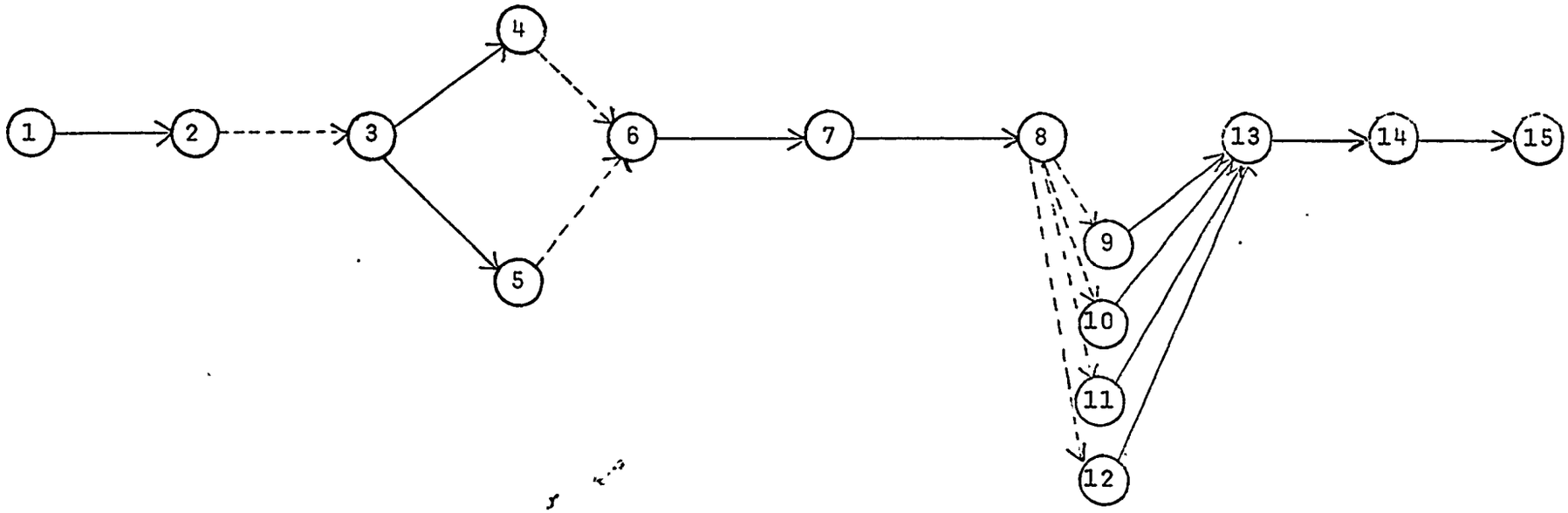
Custom-Made Achievement Test

LEVEL 0	LEVEL 1	LEVEL 2	LEVEL 3
Achievement Test	Test Plan	Purposes Objectives Content Specifications	
	Test Format	Items	Student Directions Form Assembly Statistical Analysis
	Answer Sheet	Machine Score Hand Score	
	Scoring Procedures	Keys Directions	

Custom Made Achievement Test



Custom Made Achievement Test



Event Identification

1. Start Specification	6. Complete Test Plan	11. Start Answer Sheet
2. Complete Purpose	7. Complete Item Review	12. Start Scoring Procedures
3. Start Objective/Content	8. Start Form Assembly	13. Start Final Form of Test
4. Complete Objectives	9. Start Directions (Students)	14. Complete Final Form
5. Complete Content	10. Start Administrator Directions	15. Project Complete

References for Test-Making Procedures

Adams, G.S. Measurement and Evaluation in Education, Psychology, and Guidance. New York: Holt, 1966.

Ahmann, J.S. and Glock, M.D. Evaluating Pupil Growth. Boston: Allyn and Bacon, 1967.

Cook, D.S. Program Evaluation and Review Technique: Application in Education. Washington, D.C.: U.S. Government Printing Office (OE012024), 1966.

Cronbach, L.J. Essentials of Psychological Testing. New York: Harper, 1970.

Educational Testing Service. Making The Classroom Test. Princeton: ETS, 1969.

Gorow, F.F. Better Classroom Testing. San Francisco: Chandler, 1966.

Gronlund, N.E. Measurement and Evaluation in Teaching. New York: Macmillan, 1965.

Lindvall, C.M. Measuring Pupil Achievement and Aptitude. New York: Harcourt, Brace and World, 1967.

Noll, V.H. Introduction to Educational Measurement. Boston: Houghton Mifflin, 1965.

Thorndike, R.L. and Hagan, E. Measurement and Evaluation in Psychology and Education. New York: Wiley, 1969.

References

Atkin, J.M. "Some Evaluation Problems in a Course Content Improvement Project." Journal of Research in Science Teaching. 1, 1963, pp. 129-132.

Cronbach, Lee J. "Course Improvement Through Evaluation," Teachers College Record. 1963, 64:8, pp. 672-683.

Scriven, Michael. "The Methodology of Evaluation." AERA Monograph Series on Curriculum Evaluation No. 1. Chicago: Rand McNally and Company, 1967. Chapter 4, pp. 39-84.

Scriven, Michael and Stake, Robert E. In G. Sorenson's, "A New Role in Education - The Evaluator." UCLA Evaluation Comment (Newsletter). Center for the Study of Evaluation of Instructional Programs. University of California, Los Angeles, California. Vol. 1, No. 1, January, 1968.

Stake, Robert E. "The Countenance of Educational Evaluation." Teachers College Record. 1967, 68:7, pp. 523-540.

Stufflebeam, Daniel L. "A Depth Study of the Evaluation Requirement." Theory Into Practice. 1966, 5:3, pp. 121-133.

Thelen, Herbert A. "The Evaluation of Group Instruction." Education Evaluation: New Roles, New Means. Sixty-eighth Yearbook, Part II, National Society for the Study of Education. Chicago: University of Chicago Press, 1969. Chapter 7, pp. 115-156.