DOCUMENT RESUME

ED 070 866 VT 018 409

AUTHOR Wall, James E.
TITLE Validation of Curriculum in Vocational-Technical Education.
PUB DATE Oct 72
NOTE 39p.; Presentation at the Institute for Curriculum Personnel Development (Ft. Collins, Colorado, October 26, 1972)

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Charts; *Criterion Referenced Tests; *Curriculum Development; Formative Evaluation; *Instructional Design; *Instructional Systems; Models; Summative Evaluation; Systems Development; Test Reliability; *Validity; Vocational Education

ABSTRACT

Validation in curriculum development is the "check-and-balance" dimension of any instructional system, in the broadest sense almost synonymous with evaluation and accountability. This paper relates validation to individual, formative, and summative evaluation. Validation measures to be applied to instructional systems are outlined according to a 12-point model reported by F. Coit Butler. Curriculum development is concerned with criterion-referenced tests (CRT) and the CRT is central to all validation efforts. The paper discusses validity of the curriculum generally and of the CRT specifically with reference to reliability and other factors. The appendix consists of instructional systems development charts from various sources. (MF)

VALIDATION OF CURRICULUM IN VOCATIONAL-TECHNICAL EDUCATION

by
James E. Wall, Assistant Dean (R&D) and Director
Curriculum Coordinating Unit-Research Coordinating Unit
for Vo-Tech Education, College of Education
Mississippi State University

INTRODUCTION

This paper looks at validation as one of the many aspects of the comprehensive and complex process of curriculum development in vocational and technical education. Validation is referred to here as an aspect of the curriculum development process as opposed to its being a step in the process. The reason for this is that validation measures are taken in more than one of the steps in the process. Also, validation measures are taken for more than one reason. Hence, validation as a concept is rather pervasive. Lee J. Cronbach suggests that validation is more than the process of examining the accuracy of a specific prediction or inference from a test score; validation means to demonstrate the worth of; to validate is to investigate.

Validation may be viewed as the "mortar" that holds together the "bricks" of curriculum development. It is the "check-and-balance" dimension of any instructional system. Validation is important to both job/task analysis and to deriving behavioral-performance objectives. Certain types of validation measures are taken during design and tryout of materials, during the conduct of training after implementation, at the end of a training action, and even on the job after the trained individual enters work.

A paper presented to the Institute for Curriculum Personnel Development, sponsored by the Department of Vocational Education, Colorado State University, Fort Collins, Colorado, October 26, 1972.

In the broadest sense, validation is almost synonomous with "evaluation" and "accountability." Validation as used in this paper is directly related to individual evaluation, formative evaluation, and summative evaluation. Distinctions between these three are as follows:

Individual evaluation is the monitoring of the performance of each learner as a basis for making decisions about his further vertical progression in a particular sequence, or his transfer laterally to other sequences. This is usually done by frequent monitoring of learner performance.

Formative evaluation is conducted during the experimental period and provides feedback for improvement of the instructional package. Formative evaluation is the collection of appropriate evidence during the construction of a new curriculum in such a way that revisions of the curriculum can be made on evidence. It is ongoing and is carried out concurrently with the instruction. It is distinct from individual evaluation in that the focus is on the instructional system itself rather than the learner. It seems then that "validation" is "formative evaluation." Formative tests have two purposes: (1) to find out how much students have learned in a restricted area of content, and (2) to assess whether instruction has been properly designed and conducted. Design as used here refers to appropriate content, sequence, and method of instruction.

Summative evaluation is the overall assessment of a final instructional package. Summative evaluation is the collection of data at the end of a training program to determine its effectiveness. It does not occur during the design, but rather subsequent to development, refinement, and implementation.

The primary purpose behind conducting validation procedures is to determine if the planned curriculum will achieve the responses for which it was designed. Validation searches for evidence to indicate that the curriculum can cause individuals to achieve its predetermined behavioral objectives. Validation measures should, for the most part, show that when objectives are not achieved, the fault lies with the instructional system, not with the student.

Without validation measures being applied, curriculum design is greatly handicapped. The curriculum designer usually uses only subjective and/or personal judgment in job/task analysis, formulation of behavioral objectives, and even in the construction of a criterion test. Indeed, in the early stages of instructional design, personal opinions and judgments of experienced persons as to what ought to be included in a course are valuable and necessary at the outset. However, observation, intuition, interviews, expertise, and historical precedent serve to support the designer just so long. They can serve to set early patterns for the designer. The instructional systems concept, on the other hand, demands empirical evidence that is derived through objective evaluation of content before further course design takes place. The criterion test is constructed to objectively validate course content, but its items must be validated before it can be so used.

Historically speaking, the judgment of the instructor many years ago was regarded as a criterion. Many instructors continue to rely on personal judgment as the criterion for constructing test items, and claim that when completed, they have an "obviously valid test." The "obviously valid test" cannot be validated by correlating it with something else.

However, it can be improved by applying the factor analysis technique to criterion variables.

The writer of this paper does not wish to convey the idea that all curriculum designers should become "empiricists." Neither should we go in the other direction and become total "impressionists." The empiricist validates in a completely objective, formal manner. He insists on having numerical scores, no matter how crude his instrument, to be interpreted within minimal errors of measurement, and with predictions having indexes showing how likely it is to come true. The impressionist, on the other hand, bases his analyses on observations and informal measures and estimates. He intuitively compares impressions based on one procedure with impressions gained from another. He can be classified as being somewhat more casual in his validation efforts.

Validation in vocational-technical curriculum design requires the designer to be somewhere on the continuum between the empiricist on the one hand and the impressionist on the other. He should be intermediate between obsession with scores and unrestrained use of intuition. Formal objective procedures should be combined with informal judgment in all validation efforts.

ASSUMPTIONS

For purposes of this paper, several assumptions should be made before proceeding with further discussion about validation. These are stated here to get us on common ground so that we might begin thinking along similar lines.

First, the assumption should be made that the curriculum designer has decided to utilize some type of systems approach or model in the development of his curriculum and its components. Leonard C. Silvern defines a system as: "the structure or organization of an orderly whole, clearly showing the interrelationship of the parts to each other and to the whole itself." Curriculum development involves a step-by-step process, and a system best accommodates a process, or cycled steps. The use of a systems approach implies comprehensiveness of steps, as well as interdependence of stages, components, and concepts. Systems analysis techniques enable the designer to better select the stage (time sequence) of the program operation he must validate, i.e., to identify the relevant curriculum components with the outcome changes being measured.

The use of a systems approach assures that all the necessary assessments will be made. For example, some measure should be made of student performance or entering behavior before (pre-assessment) he begins a new curriculum. Similarly, two types of assessment should be continually made while the student is undergoing training (during-instruction). One of these types of assessment serves as feedback for reinforcement, the other assures acquisition of behaviors that are prerequisite to lateral movement to other experiences. Likewise, two types of assessments should be made after completion of a lesson or a unit (post-assessment). One aids in determining if a student is prepared for vertical progression to related or advanced experiences; the other serves as a type of summative evaluation, as well as a predictor of success on the job or in more advanced courses.

Constant improvement of curriculum is a worthy goal. Hence, the influence and use of pre-assessment is an important variable for validation since it is not the terminal criterion behavior alone which dictates required instructional manipulation, but the differences between entering and terminal behavior.

Apparently there are as many systems models as there are curriculum designers. However, most all of them have some common features. These commonalities usually resemble the following:

1. Job specification or analysis.
2. Specification of objectives.
3. Development of preliminary system design.
4. Development--test--revise cycle applied to the system.
5. Implementation and field testing the system.
6. Follow-up and/or summative evaluation.

Included in the appendices are some excellent examples of models that persons have prepared to graphically depict curriculum development processes. Note the similarities. Some are specific concerning validation measures; some are not. It should be made explicit here that no one model fits all situations. Models are necessarily going to be different for secondary programs, for post-secondary programs, for industry-based programs, for government agency-based training, for military-based training, etc. Hence, it behooves the curriculum designer to become knowledgeable in principles of systems development if he is to achieve assurances that his curriculum will be valid. One of these models will be used later in this paper in an attempt to show where various validation measures will be taken, and to show how scores taken at one point in the system are correlated with those taken at another.

There are other "stage-setting" assumptions which need to be made at this point in the paper. A second major assumption is that the curriculum designer has selectively applied a number of principles of learning, because different kinds of learning require different sets of conditions. The important factors which influence learning are: motivation, organization, participation, confirmation, repetition, and application. One type of learning may require emphasis on one factor, whereas another type of learning may require two or more factors in concert.

A third assumption is that the designer intends to build the curriculum so that the sequence of learning progresses from the simple to the complex. The sequence or hierarchy should resemble the following: specific responses and associations which are prerequisite to verbal and motor chains which are prerequisite to discriminations which are prerequisite to concepts which are prerequisite to principles which are prerequisite to higher-order principles or strategies for problem-solving.

Fourth, it is assumed that enabling or interim objectives for lessons, modules, or units have been appropriately derived, and are of a degree of specificity such that the materials can validated accordingly. Likewise, it is assumed that terminal (course or training program) objectives also have been appropriately stated in such a manner that validation measures also contribute to overall or summative evaluation and accountability in programs and projects. One striking advantage of precisely stated objectives is that when one is completely clear about the nature of the terminal behavior, it is possible to arrange for appropriate practice opportunities during the instructional sequence.

A fifth assumption is that we can at least tentatively agree that the essence of education focuses on preparing persons so they might be enabled to attack all their problems by bringing knowledge and action to bear on them in a unified and integrated rather in a fragmented

7

manner. Does the student who finishes your curriculum merely possess a large bewildering body of unrelated facts? Or, can he articulate knowledge and skills learned so that he can perform? Validation through the use of criterion referenced tests assures performance.

Such all-encompassing assumptions may be misleading. They may cause some people to feel that further discussion of validation is unnecessary since so much has been "assumed." This is not the case because these assumptions touch only a small portion of the elements of the entire instructional system.

Validation involves measurement. Appropriate validation measures do not allow wide fluctuation in attainment of objectives, nor do they bring about perfect stability. Validation does aid in better control of achievement of objectives. When combined with appropriate definitions of behavior changes sought, validation provides the curriculum designer with a thermostat to assure the achievement of the instructional objectives. Hence, the curriculum designer "controls" the growth or behavior change of the student. The designer usually starts with a comprehensive description of the desired dependent set of events, i.e., a finished product or process derived by job/task analysis. Then he works backward through his analysis to the set of independent events most likely to produce the product or process.

Validation procedures have value in many of the stages of the system. However, their greatest value probably occurs when employed in the design stages of materials development, in which they are applied to both interim as well as terminal objectives. Hence, validation becomes the prime focus in checking out the objectives and the criterion tests, as well as later in pilot tests and field trial assessments of materials. However, after validation is completed during design-development and the training is installed in the classroom, testing for attainment of enabling or interim objectives may become a matter of self-testing by the student. Formal testing for attainment of terminal objectives is conducted by the instructor and these scores might be used to further validate the curriculum.

In the design stage the curriculum developer may wish to use a simple cycle such as: design -- test -- revise -- retest. The use of this cycle tends to upgrade the effectiveness of the instructional materials through repeated revision. Most of the models shown in this paper contain some type of design -- test -- revise -- retest dimension or component.

VALIDATION IN INSTRUCTIONAL SYSTEMS

Validation in the design of curriculum materials demands a systems approach which will ensure that testing is conducted at the right steps in the overall process. Some curriculum experts have designated these efforts with a generic term: "developmental validation."

For purposes of ordering our thinking about validating curriculums, we need to utilize a systems model which sequences the events that are necessary to produce a valid curriculum. The model recently reported by F. Coit Butler will serve this purpose.

Butler's system is briefly given as follows:

1. Conduct feasibility study --- This requires an analysis of trends with regard to job markets and occupational patterns; trends in economic, business, agricultural, and industrial

expansion; types of jobs and worker competencies needed; availability of training programs and facilities, and their costs; etc.

2. Conduct task analysis --- After the decision has been made that a specific training program or course is needed, an occupational or job analysis is conducted to determine skills and knowledges required; kinds and levels of performances demanded by the job, etc.

3. Develop training objectives --- At this point the designer must derive explicit statements about what a student, upon completion of the training program, must be able to do; the conditions and standards of his performance; etc. Both terminal (unit, course, program) objectives and interim or enabling (lesson, activity, module) objectives must be specified. These may be directly coupled to broad goal statements and possibly even broader educational or philosophical constructs.

4. Develop criterion tests --- These are used in the early stages of design to determine validity of the objectives, and later to help perform summative evaluations of the entire course or training program. Additional information about criterion tests is included later in this paper.

5. Validate the criterion test --- This is done by administering the test to an untrained-unskilled group and to a trained-skilled group and correlating the scores to obtain validity and reliability coefficients. Test item analysis at this point calls for interpretations similar to the following: (a) if, for a given test item, the majority of untrained group responses are correct, the item has little or no validity or reliability;

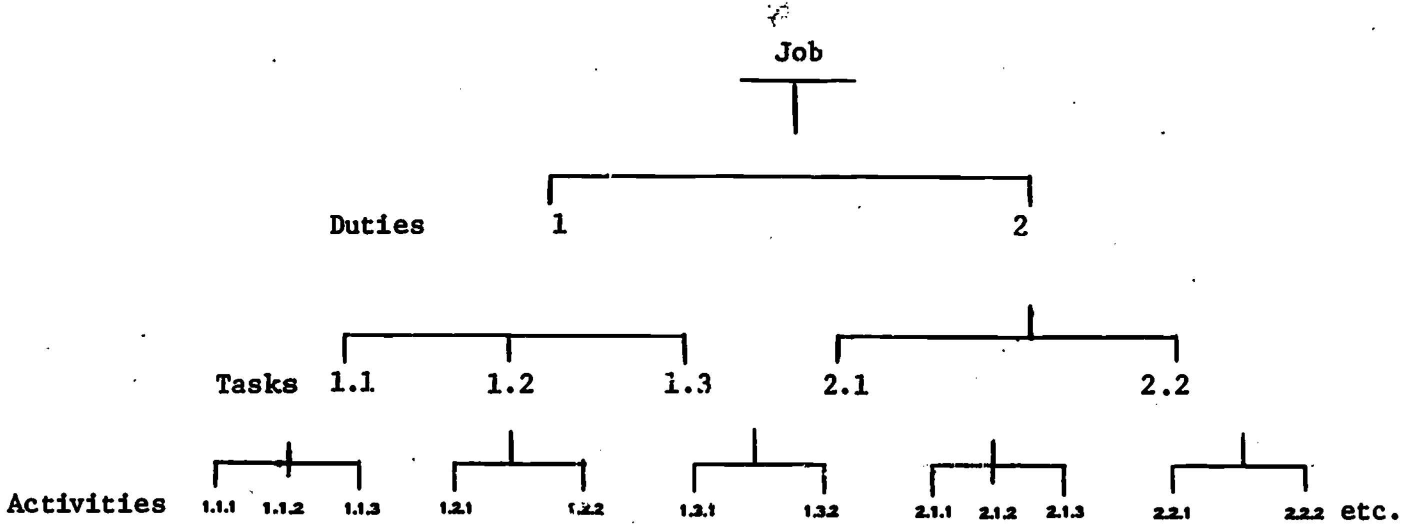
and conversely, (b) if, for a given test item, the majority of trained group responses are incorrect, the item likewise has little or no validity or reliability. Test measurement and statistics texts listed in the References section of this paper contain procedures for the test item analysis. This analysis is conducted to improve the test as a measuring instrument. This is the most important validation step in the process.

6. Validate training objectives --- The test should contain at least one item for each objective, and possibly not more than five items for each objective, otherwise the test becomes too long for practical purposes. Validating the test in Step 5 above and validating training objectives can be accomplished concurrently, provided the test item itself is not at fault. Interpretations similar to those made in Step 5 are employed in this step; e.g., (1) if, for a given test item and its companion objective, the majority of untrained group responses are correct, there may be no need to include that objective in the curriculum; and, (2) if, for a given test item and its companion objective, the majority of trained group responses are incorrect, there may be no need to include that objective in the course because, apparently, the worker can perform on the job without that knowledge or skill.

(These types of interpretations may need to be reviewed in light of some estimates concerning the possibilities and probabilities that a worker may be required to "transfer" skills and knowledges to a different work situation. However, if this becomes probable then the situation may warrant a new re-training or up-grading instructional program which calls for it to undergo the same validation procedures outlined here.)

According to Butler's model, the initial design phase has been completed at this point, but the remaining phases also require validation considerations.

7. Develop learning sequence and structure --- This is done according to the duties, tasks, and activities provided in the job/task analysis. The following sketch shows a pyramidal form of learning structure and sequence.



Activities, tasks, and duties are structured (and learned) in both a vertical and horizontal sequence. The learning of one is dependent upon accomplishment of those which precede it. Most curriculum experts recognize that sequencing must be approached with a great deal of flexibility. The general guideline of efficiency should influence sequencing.

Butler sets forth a matrix analysis technique for preparing the course outline in which supporting knowledges and skills for activities, tasks, and duties are listed. He indicates that the learning sequence can be plotted by starting with the terminal

12

objective and working backward through each preceding prerequisite--in essence, from the complex back to the simple. Butler suggests listing all terms, concepts, rules, and principles which pertain to each objective. List them as single-fact statements and assign each a number. Each number is then placed in a two-dimensional matrix (discrimination-association) along a diagonal line from top left to bottom right. Associations then are marked in the common squares above the diagonal, and discriminations are marked in the common squares below the diagonal. By shuffling and reshuffling, a rearranged matrix can be plotted which depicts an optimum clustering of discriminations and associations around the diagonal, which results in the best sequencing. The clusters tend to depict broad concepts in the curriculum.

Validating the sequence also is accomplished with the criterion test which has been validated and revised. The test is given to a group (30 to 50) of trained individuals, i.e., as a post-test to persons who just completed the program, or to those who have been on the job about six months. In the analysis of these scores, one looks for the dependency and interdependency between and among units, lessons, or fairly large blocks of curriculum content.

Butler indicates that the test data should be analyzed with two basic questions in mind: (1) Did the majority of those students who correctly performed a subordinate unit (Unit No. 1) also correctly perform the following and supposedly dependent unit (Unit No. 2)? And, (2) Did the majority of those who correctly performed the higher unit (Unit No. 2) also perform the subordinate unit (Unit No. 1) correctly? If, for a tested trained sample, the answers to both questions are affirmative, then the sequence is valid. If, for only 85% of the sample, the answers are affirmative, then the sequence is probably valid. See "Validating Content Sequence" chart on the following page for implications when 50% or less of the trained sample perform incorrectly.

The foregoing procedure is used only on a pair of tasks in a hierarchy. Suppose the hierarchy consisted of three or more tasks and validation is still required. Recent research has gone in the direction of trying to discover such hierarchies and their properties, and validation procedures are under study, using factor analysis techniques. The curriculum designer may wish to refer to "A Method for Validating Sequential Instructional Hierarchies," by P. W. Airasian, in the December 1971, issue of Educational Technology Journal. This method is based on calculation of conditional item difficulty indices and facilitates the pinpointing of sequential levels within a hierarchy which require revision.

8. Develop learning strategies --- This step has no feasible validation procedures which are not too costly and time consuming to use. Media are selected according to those that will do an effective job for the least cost. Combinations of the different media usually should be considered.

Validation is influenced by the media. Test scores may be low for students with reading problems, but the same test scores

Validating Content Sequence

Summary of procedure for analyzing criterion test data from a sample trained population when validating content structure and sequence

Trained Sample (only correct performance)	Performance	Implications
Performs unit (100%)	85% perform sub-unit	Possible correct sequence
Performs sub-unit(100%)	85% perform unit	Possible correct sequence
		Taken together, a certainty the sequence is correct.
Performs unit (100%)	85% perform sub-unit	Possible correct sequence
Performs sub-unit(100%)	50% fail to perform unit	Possible incorrect sequence
		Taken together, indicates bad test item.
Performs unit (100%)	50% fail to perform sub-unit	Possible incorrect sequence
Performs sub-unit (100%)	85% perform unit	Possible correct sequence
		Taken together, indicates bad test item.
Performs unit (100%)	50% fail to perform sub-unit	Possible incorrect sequence
Performs sub-unit(100%)	50% fail to perform unit	Possible incorrect sequence
		Taken together, a certainty the sequence is incorrect.

Source: F. Coit Butler. Instructional Systems Development for Vocational and Technical Training, Educational Technology Publications, Englewood Cliffs, New Jersey, 1972.

may be improved by using audio media instead of printed media. The objectives and student learning styles are the prime determinants in developing the learning strategies.

9. Develop instructional units (lessons) --- This is the point where a test model of the instructional system is produced. Two documents are needed: (1) the system development plan, and (2) the instructor's manual or guide.

The system development plan contains: (1) task analysis summary forms; (2) validated objectives in validated sequence, supported by a summary of the validation data; (3) validated criterion test items in validated sequence, supported by a summary of the validation data; (4) outline of instructional strategies with associated content (objectives) identified; and (5) production and testing plans for the system.

The design and format of the individual learning units may vary greatly, but each should contain the following: (1) the performance objectives; (2) the knowledges and skills to be gained; (3) a list of tools, equipment, supplies, references, etc., needed for the unit; (4) a learning activity guide; (5) interim progress checks and student self-evaluations; and (6) an instrument to serve as a pre-test and/or a post-test for evaluations by the instructor.

10. Validate learning units --- At this point each unit is tested and revised until 85% of sample trainees reach the criterion.

Revision may require resequencing and adoption of new learning strategies. Initial testing is done on an individual or one-to-one basis, with two or three sample trainees who have upper-level ability. Minor revisions may be made at this point; however, if major revision is indicated, two or three more individual tryouts should be conducted.

Small-group tryout is then conducted on 6 to 10 students who represent the range of ability and background of the target population. Criterion test data are again used to locate trouble spots and revision is made. At this point, 85% of the students should be performing correctly on the criterion test.

Final tryout is made on a large group of 30 to 50 students under conditions which approximate actual training. This tryout is conducted by the curriculum designer along with the instructor(s). A group this size is needed to verify (or validate) previous design results. Final revision is made following this tryout.

11. Implement and field test the system ---- This is done under actual classroom conditions. The instructor's role in the instructional system is explicated at this point, and the Instructor's Manual is developed. His role becomes that of manager and facilitator of learning. His tasks are as follows: (1) diagnose individual learning needs; (2) prescribe learning experiences; (3) provide proper materials and equipment at right time; (4) test and evaluate individual progress; (5) compile individual and group progress records; (6) provide tutorial and counseling help; (7) provide motivational reinforcement;

(8) provide supplementary materials and experiences; (9) coordinate individual, small-group, and large-group learning activities; (10) coordinate use of learning materials and equipment; and (11) evaluate feedback data on effectiveness of learning.

The Instructor's Manual should contain: (1) course description; (2) student population description; (3) performance objectives; (4) criterion tests; (5) system performance data; and (6) suggestions for administering the system.

Field testing is the final phase of the systems development process. This means the program is monitored, evaluated, and subsequently revised continuously for as long as it is in use. This phase may be more appropriately referred to as system "institutionalization." Constant monitoring and analysis of criterion test data will continue to point the way for needed revision.

Butler points out that a training system is never "finished," rather, it is constantly "in process."

12. Follow-up graduates --- At this point, effective guidance and placement are brought into play. Longitudinal planning for follow-up at 1-year, 3-year, 5-year, or 10-year intervals may be started at this point. Follow-up to obtain details of occupational patterns, changes in needed competencies, job adjustment problems, and work satisfaction indices, all can be used to improve the instructional system.

CRITERION-REFERENCED TESTS

Much has been said in the foregoing material about criterion tests, or criterion-referenced tests (CRT). The CRT is central to all validation efforts.

In curriculum development we are concerned with criterion-referenced tests, whereas, in traditional test development, the concern has been and still is with norm-referenced tests (NRT). A simple distinction should be made here between the CRT and the NRT.

The NRT is the more traditional type of test and is used to identify an individual's performance in relation to the performance of others on the same measure. Hence, the NRT is viewed as a relative measure. The CRT, on the other hand, is used to identify an individual's status with respect to an established standard, or criterion, of performance. The CRT, therefore, is viewed as an absolute measure. Curriculum developers are concerned with getting an individual person trained proficiently according to a predetermined set of absolute criteria, rather than training him relative to the performance of other individuals.

CRTs can be devised for use in making decisions both about individuals and instructional programs. Concerning decisions about individuals, one might use a CRT to determine whether a student had mastered a criterion skill that is prerequisite to starting a new training program, or a new sequence within a program. Concerning instructional programs, a CRT could be designed that would reflect attainment of objectives based on a hierarchial sequence. The CRT could be administered to learners after they completed the sequence, and, after analysis, the efficacy of the sequence might be determined.

The CRT and Reliability --- Validity of the curriculum generally, and of the CRT specifically, cannot be considered apart from reliability. This implies that the CRT must be internally consistent, i.e., CRT items must be similar to stated behavioral objectives in terms of what they are measuring. Although traditional statistical procedures for determining reliability coefficients are not necessarily appropriate for CRTs, it is thought at this time that coefficients which are derived by considering both a pre-instruction assessment and a post-instruction assessment as part of the same extended phenomenon might yield more meaningful reliability estimates.

An ideal curriculum component, package, program, unit, lesson, etc., should result in perfect learning on the part of all individuals. While individuals may differ in the amount of time required to go through a curriculum component, once they have completed it, all should have mastered the content. From this point of view a good program should result in little variability for a measure of learning. One might suppose, then, that variability of G scores (gain between pre-test and post-test) would be a criterion that could be used in assessing programs such that the less the variability, the better the curriculum component. (It should be recalled by the reader that correlation coefficients derived by traditional statistical methods rely on variability.)

The above is merely mentioned here in the event the curriculum designer has the time and inclination to work toward impirical reliability estimates.

In the initial design stages, the designer takes the objectives and recasts them as items on the CRT. If the objectives have been derived from accurate job analyses, then they should have job validity, and consequently, test items geared to these objectives also should have true validity. Reliability of the CRT will depend upon job-objective-test behavior that is observable and measurable. To improve test reliability, a preliminary check can be made on two different groups of 30 to 50 persons: (1) untrained-unskilled persons, who might be entering students, and (2) trained-skilled persons who might have been on the job for less than six months. (This procedure was outlined in one of the steps in Butler's model.) Comparisons of scores of the two groups will yield a correlation coefficient of reliability. The reliability check may require major revision of the entire test, but each item should be treated separately, since a composite test reliability coefficient will not pinpoint the specific items that need revision, whereas an item-by-item analysis will.

In the case where a curriculum is being developed for a new or emerging job or career, non-availability of trained-skilled persons for purposes of determining an in-design system reliability estimate would prevent the use of the above approach. On the other hand, for those ongoing curriculums that are being subjected to continuous revision and study, the above approach to determining reliability would seem to be a tenable one. This technique is suggested for consideration despite the fact that it may be time consuming and somewhat costly. At the risk of sounding trite, the curriculum designer is reminded that funds and time expended early in the blueprint or design stages may result in larger savings later on as training takes place.

The CRT and Validity --- Procedures used in assessing the validity of the CRT suffer similar difficulties found in assessing reliability. Validity of NRTs is based on correlations, hence on variability, the search being usually for coefficients that exceed +.60. However, CRT coefficients of less than +.60, and even those with negative coefficients, are not necessarily devastating. CRT items are validated primarily in terms of the adequacy with which they represent the criterion stated in the objective. Adequacy of content is especially important for tests that measure outcomes of education or training. Hence, content validity approaches may have some application to CRT test validation. In content validity we determine skills, knowledge, and understanding that comprise the correct behavior we are seeking in students, then translate these to objectives and construct a test or tests to measure attainment or achievement. Finally, we match the analysis of test content against the analysis of instructional program content and objectives and see how well the former represents the latter. To the extent that our objectives are represented in the test, the test is valid.

The major focus of validating the CRT is to show that its items are a representative sample of all aspects and facets of the behavior prescribed in the objective. This means there may be pencil-paper items pertaining to skills and knowledges. It also means that there may be items which measure performance. Responses to pencil-paper tests are easier to obtain than responses to tests of performance.

Performance tests usually call for responses that require all the dimensions of behavior, such as speed, accuracy, judgment, etc.

The CRT and Item Analysis --- Item analysis procedures have traditionally focused on pinpoint test items on NRTs that do not discriminate among persons who take the test. Faulty items would be those which are too easy, too hard, or ambiguous. Both positively and negatively discriminating items for CRTs may pinpoint areas of instruction which need revision. However, negatively discriminating items are the ones which should be identified, but identifying them will depend on the ease with which analyses can be conducted. This usually demands sophisticated data-processing techniques.

Webster and McLeod present an excellent technique for item analysis of a module test which can be used to perform item analyses on CRTs.

SUMMARY

The foregoing material has attempted to present a rationale for validating curriculums in vocational and technical education. A systems approach was used to present an orderly approach to validation discussions.

The curriculum designer may have concluded that validation efforts are extraordinarily time consuming and require test and measurement expertise not ordinarily found among curriculum staff members. Nevertheless, validation procedures as outlined in this paper proceed in an orderly fashion, building on each preceding step. The result is a curriculum package which can be identified as being sound and productive of persons who can perform.

23

SELECTED REFERENCES

Banathy, Bela H. Instructional Systems. Palo Alto, California: Fearon Publishers, 1968.

Block, James H., ed. Mastery Learning: Theory and Practice. New York: Holt, Rinehart and Winston, Inc., 1971.

Bloom, Benjamin S.; Hastings, Thomas J.; and Madaus, George F. Handbook on Formative and Summative Evaluation of Student Learning. New York: McGraw-Hill Book Company, 1971.

Burns, Richard W., and Brooks, Gary D., eds. Curriculum Design in a Changing Society. Englewood Cliffs, New Jersey: Educational Technology Publications, 1970.

Butler, Coit F. Instructional Systems Development for Vocational and Technical Training. Englewood Cliffs, New Jersey: Educational Technology Publications, 1972.

Cronbach, Lee J. Essentials of Psychological Testing. 3rd ed. New York: Harper & Row, 1970.

Espich, James E., and Williams, Bill. Developing Programmed Instructional Materials: A Handbook for Program Writers. Belmont, California: Fearon Publishers, 1967.

Ghiselli, Edwin E. Theory of Psychological Measurement. New York: McGraw-Hill Book Company, 1964.

Kapfer, Miriam B. Behavioral Objectives in Curriculum Development: Selected Readings and Bibliography. Englewood Cliffs, New Jersey: Educational Technology Publications, 1971.

Kemp, Jerrold E. Instructional Design: A Plan for Unit and Course Development. Belmont, California: Fearon Publishers, 1971.

McGuigan, F. Joseph. "The G Statistic: An Index of the Amount Learned," Journal of National Society for Programed Instruction, vol. 6, (1967) pp. 14-16.

Mager, Robert F. Goal Analysis. Belmont, California: Fearon Publishers, 1972.

Odiorne, George S. Training by Objectives: An Economic Approach to Management Training. London: The MacMillan Company, 1970.

Popham, James W., ed. Criterion-Referenced Measurement: An Introduction. Englewood Cliffs, New Jersey: Educational Technology Publications, 1971.

Popham, James W.; Eisner, Elliot W.; Sullivan, Howard J.; and Tyler, Louise L. Instructional Objectives. Chicago: Rand McNally & Company, 1969.

24

Rowntree, Derek. Basically Branching. London: Macdonald, 1966.

Siegel, Sidney. Nonparametric Statistics for the Behavioral Sciences. New York: McGraw-Hill Book Company, 1956.

Smith, Robert G., Jr., The Engineering of Educational and Training Systems. Lexington, Massachusetts: Heath Lexington Books, 1971.

Thorndike, Robert L., ed. Educational Measurement. 2nd ed. Washington, D. C.: American Council on Education, 1971.

Thorndike, Robert L., and Hagen, Elizabeth. Measurement and Evaluation in Psychology and Education. 3rd ed. New York: John Wiley & Sons, Inc., 1969.

Warren, Malcolm W. Training for Results: A Systems Approach to the Development of Human Resources in Industry.Reading, Massachusetts: Addison-Wesley Publishing Company, 1969.

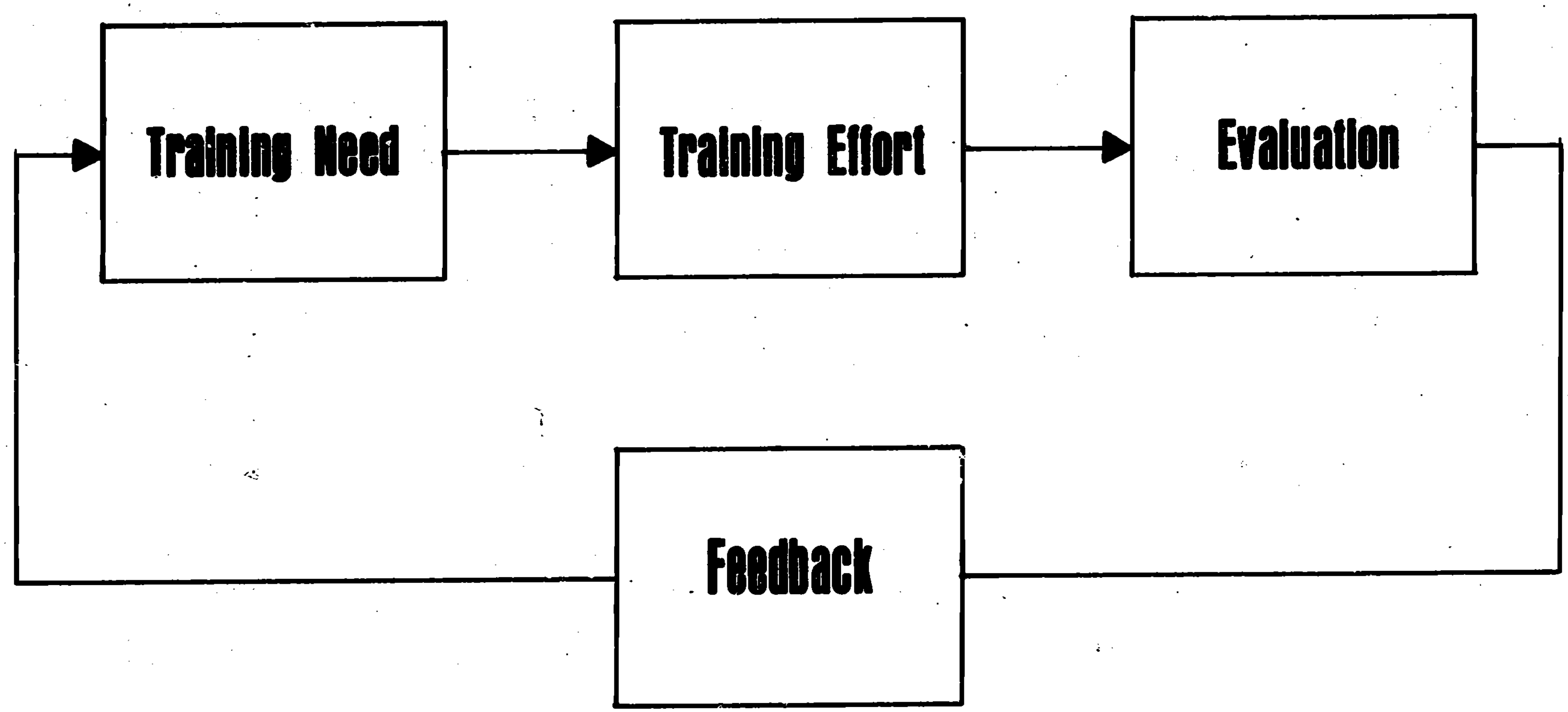
Webster, W. J., and McLeod, G. K. "An Emperical Approach to Curriculum Design." Education, XC (Feb.-Mar., 1970).

Wert, James E; Neidt, Charles O.; and Ahmann, Stanley J. Statistical Methods in Educational and Psychological Research. New York: Appleton-Century-Crofts, Inc., 1954.

Yagi, Kan; Bailek, H. M.; Taylor, J. E.; and Garman, M. The Design and Evaluation of Vocational Techincal Education Curricula Through Functional Job Analysis. Alexandria, Virginia: Human Resources Research Organization, 1971.

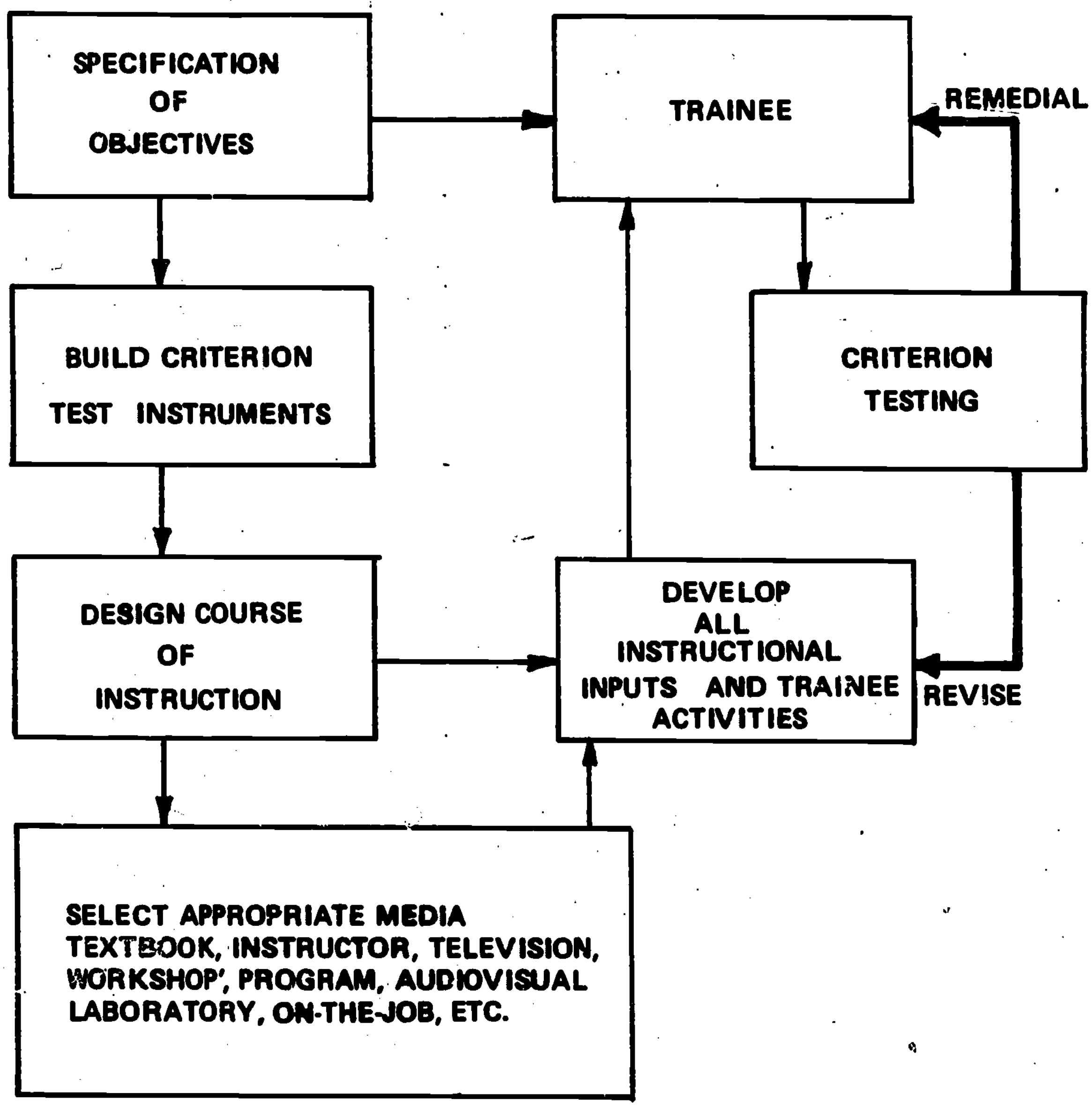
APPENDICES

CYBERNETIC SYSTEM FOR TRAINING

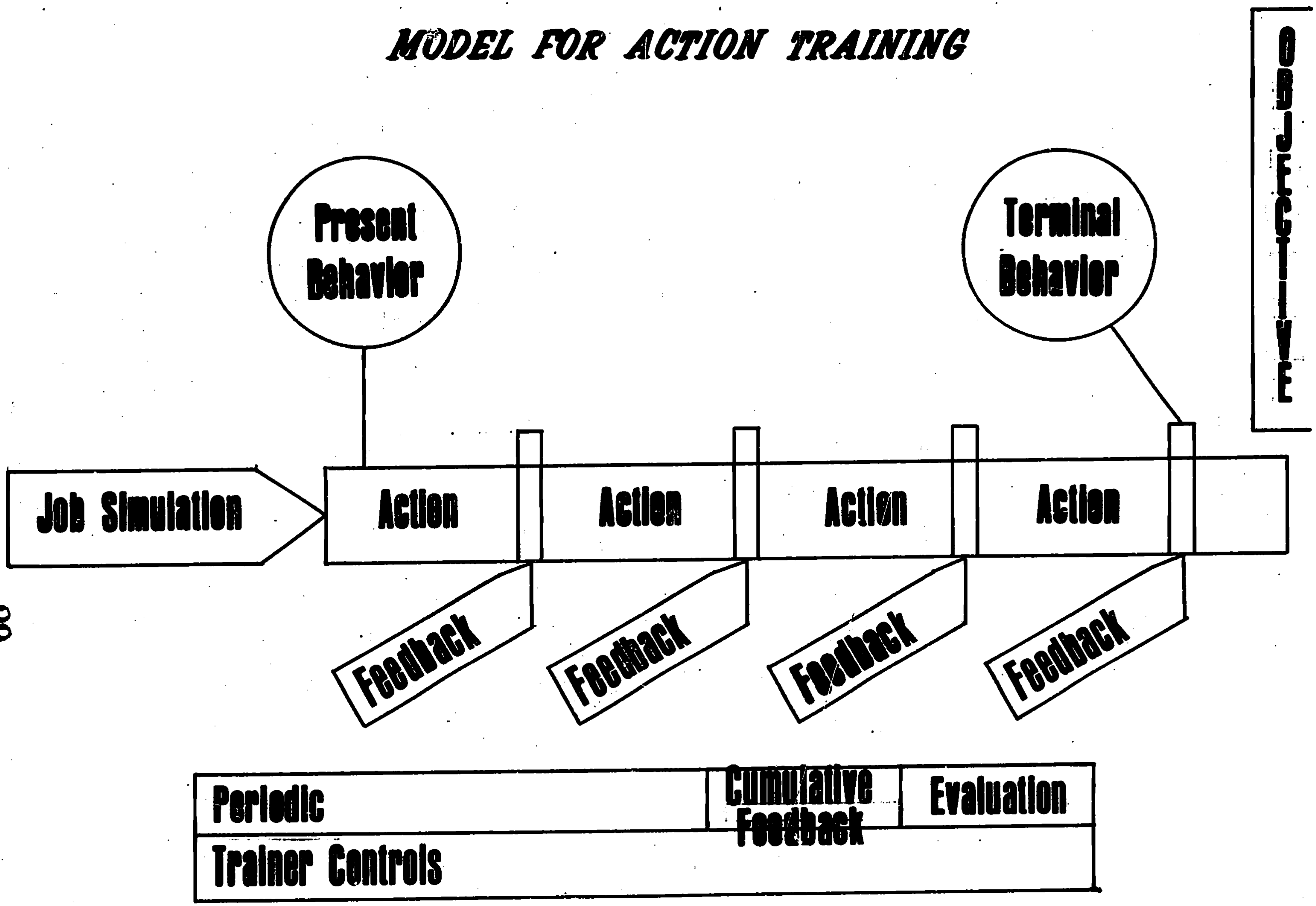


Odiorne, G.S. Training By Objectives, Macmillan Company, New York, 1970.

A SIMPLIFIED INSTRUCTIONAL SYSTEM DESIGN



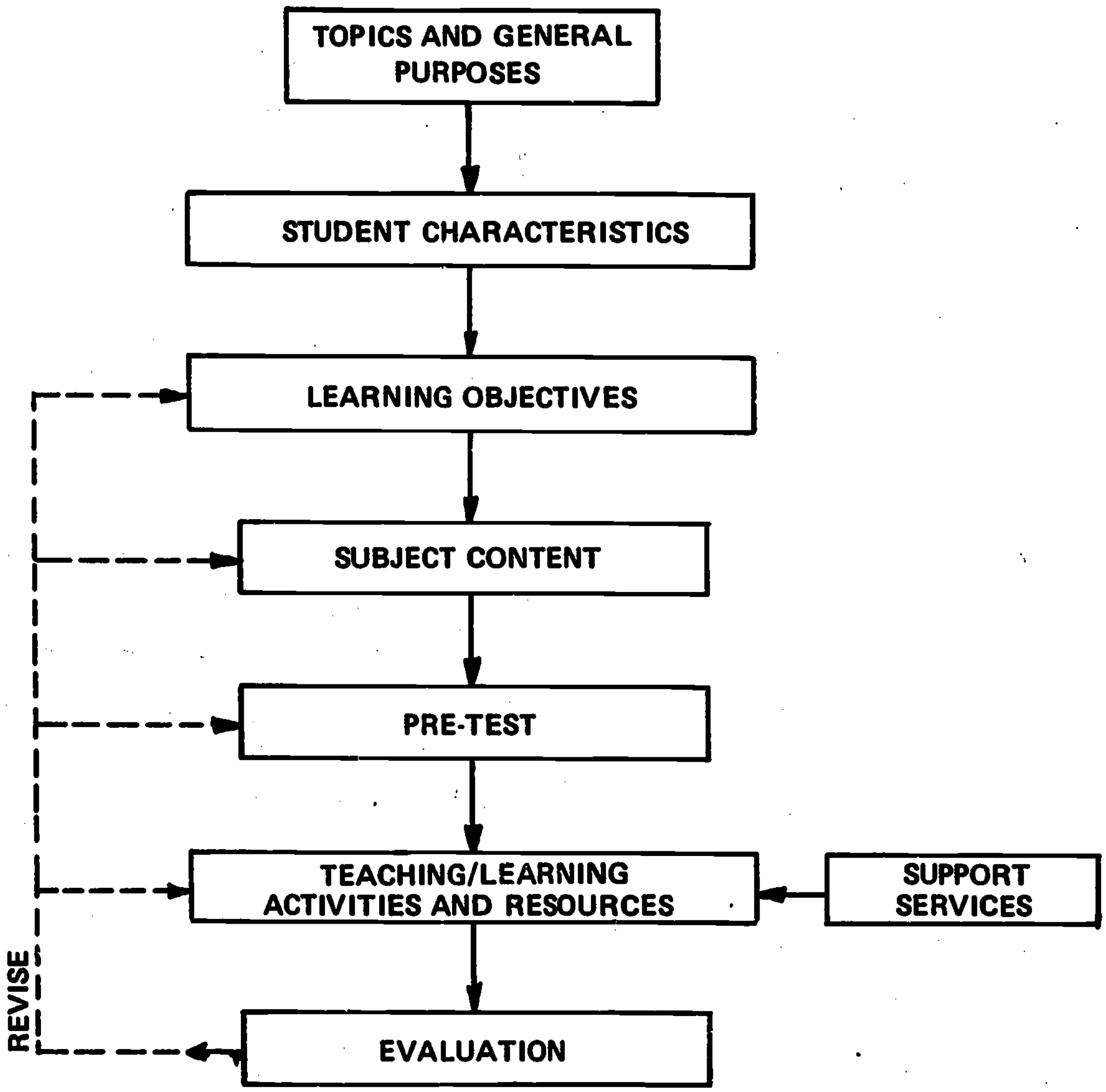
MODEL FOR ACTION TRAINING



Odiorne, G.S. Training By Objectives, Macmillan Company, New York, 1970.

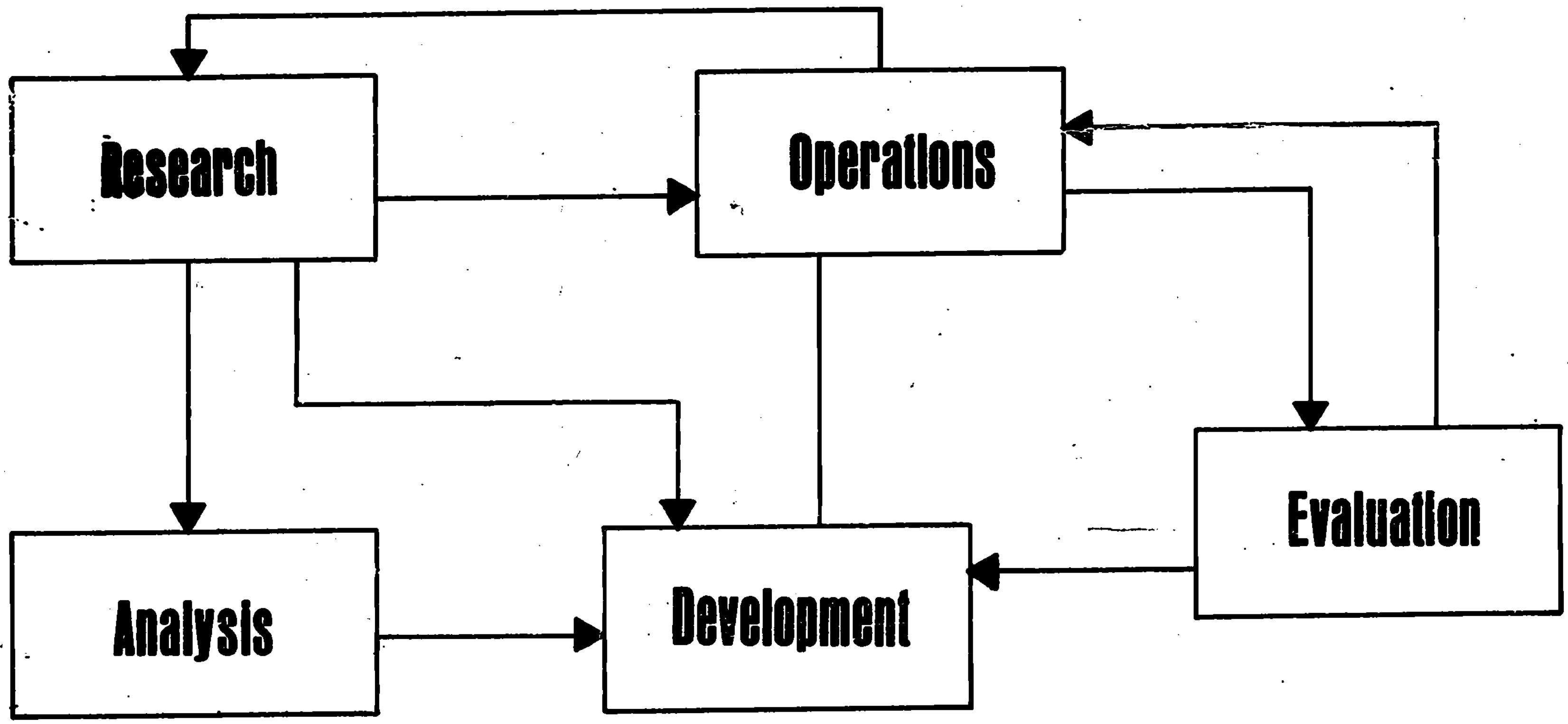
29

INSTRUCTIONAL DESIGN PLAN



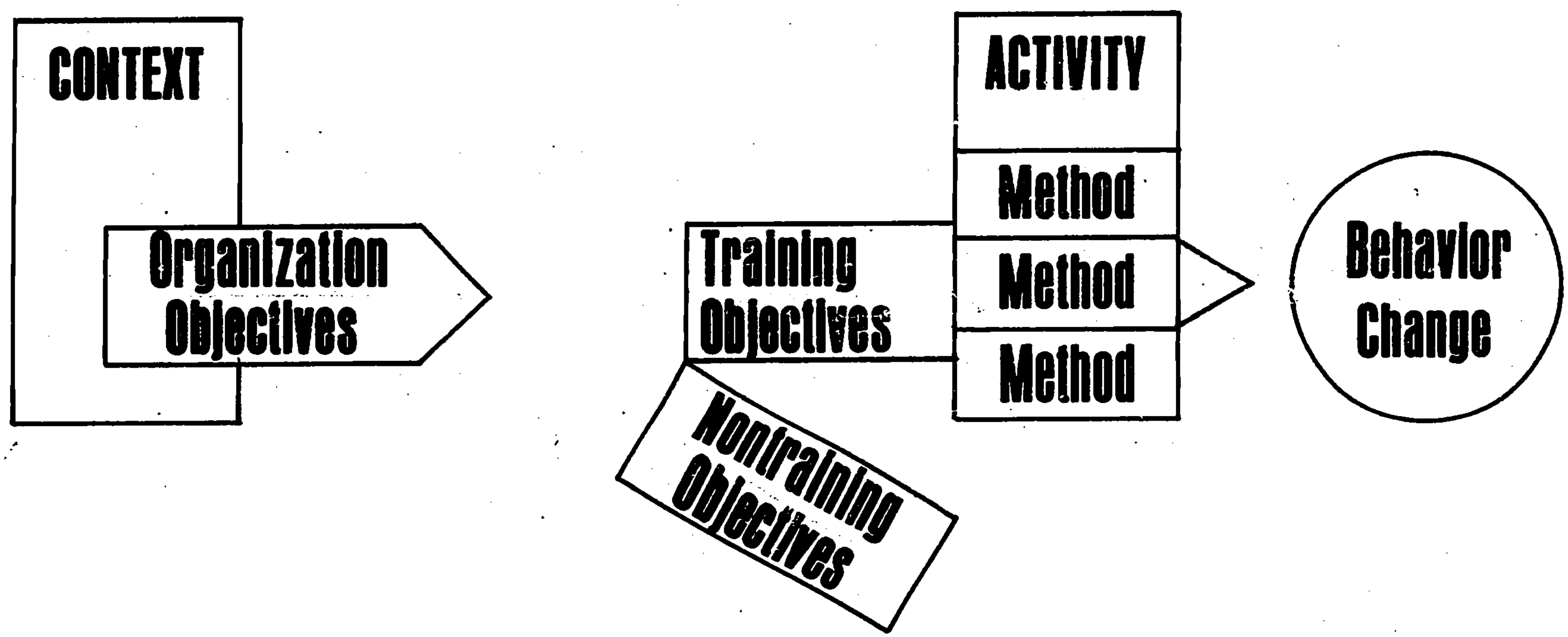
Kemp, Jerrold E. Instructional Design, Fearon Publishers, Belmont, Calf., 1971.

GENERAL MODEL of a TRAINING OPERATION

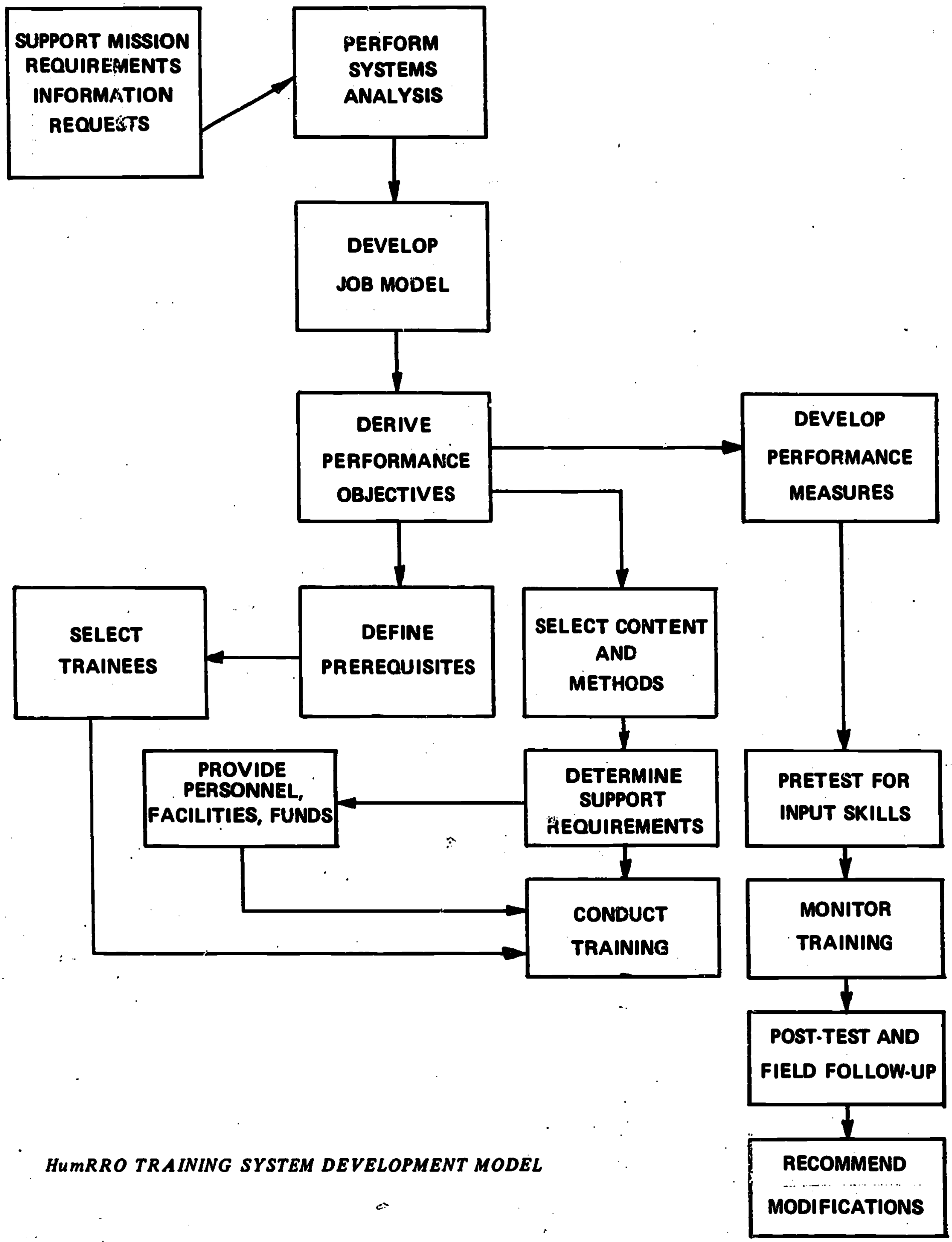


Warren, M.W. Training for Results, Addison-Wesley Publishing Co., Reading, Mass., 1969.

A SYSTEMS MODEL for MANAGEMENT TRAINING

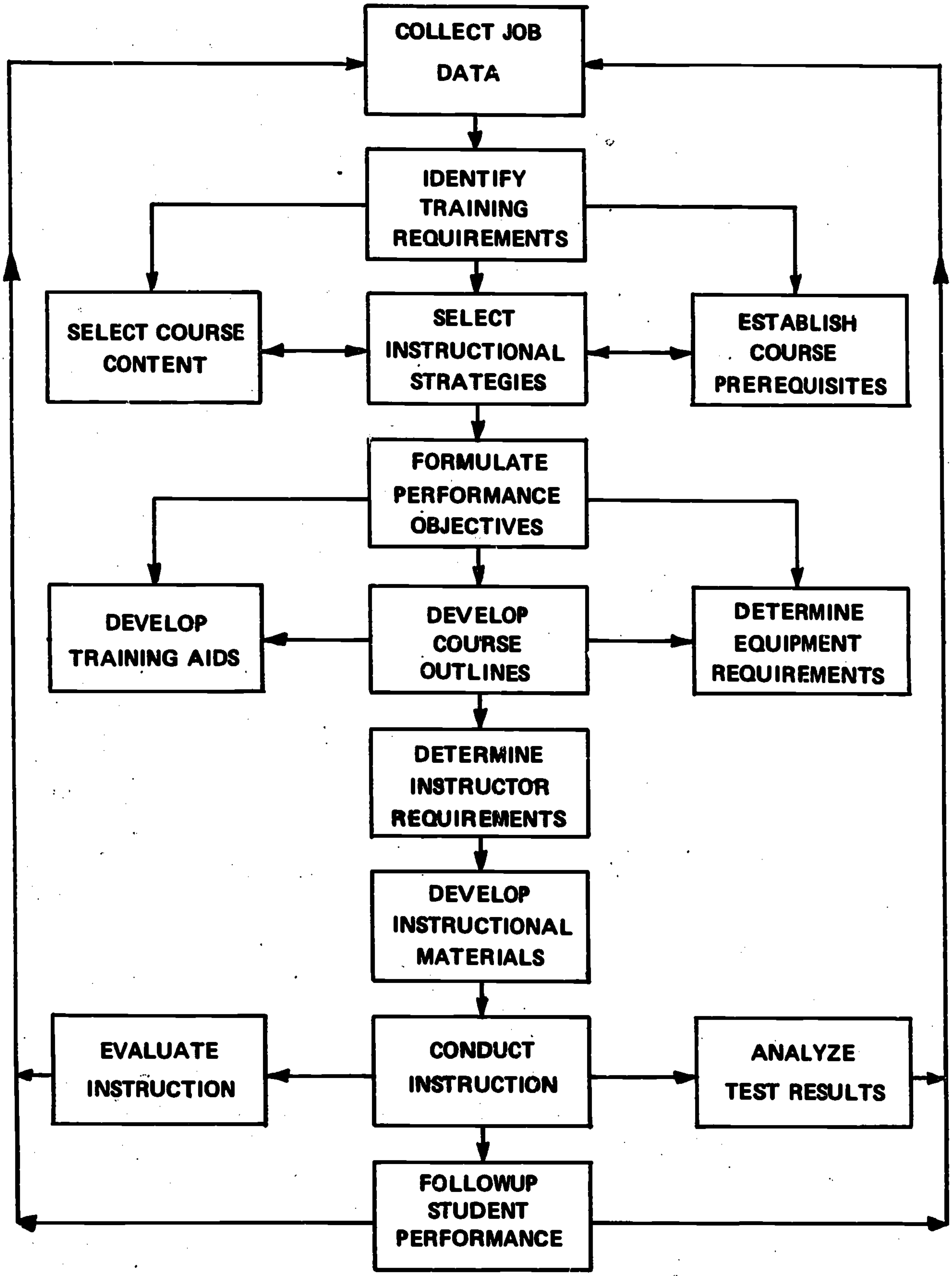


Odiorne, G.S. Training By Objectives, Macmillan Company, New York, 1970.



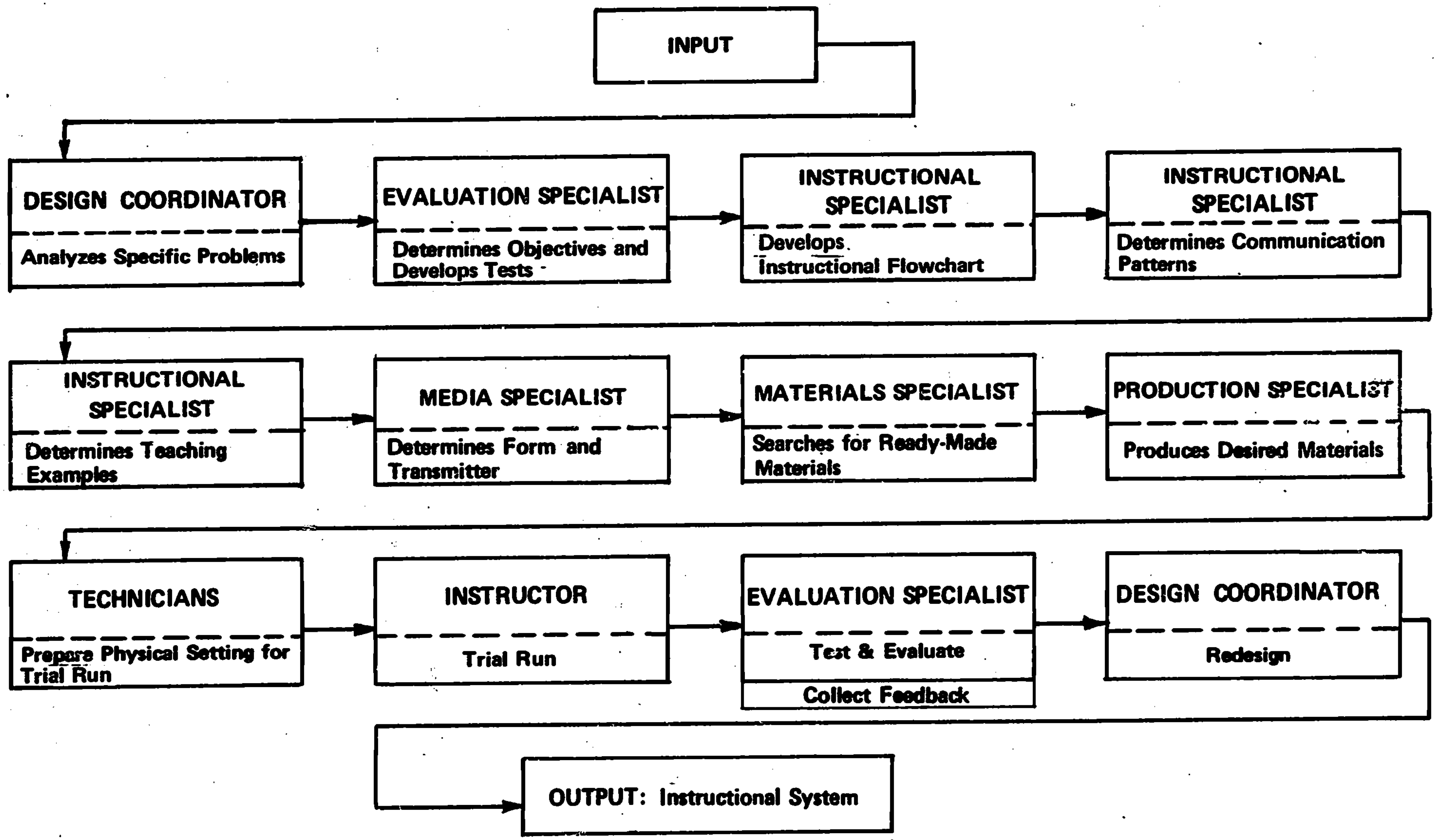
HumRRO TRAINING SYSTEM DEVELOPMENT MODEL

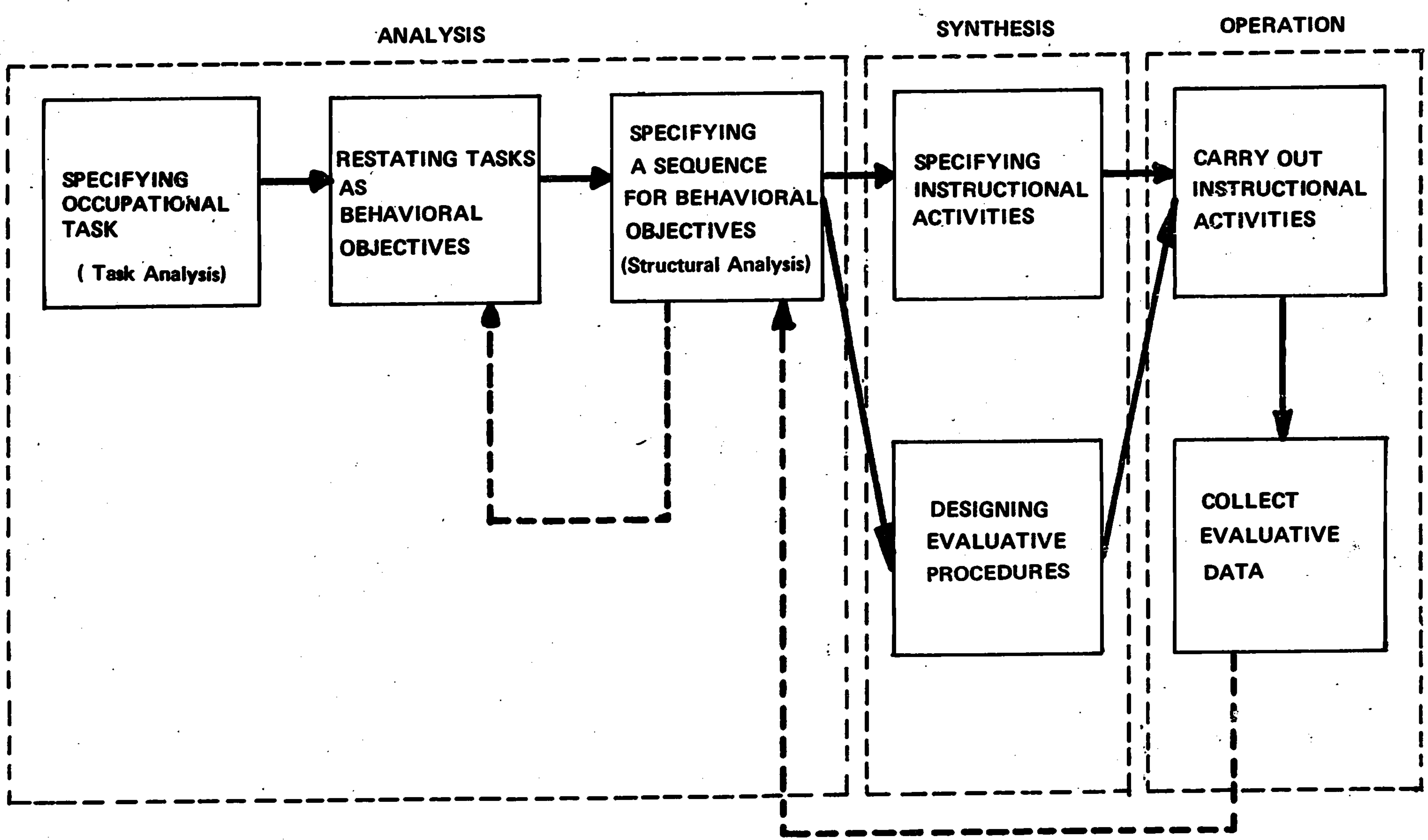
INSTRUCTIONAL SYSTEM DESIGN DIAGRAM



Bureau of Training, U.S. Civil Service Commission. An Application of a Systems Approach to Training: A Case Study, Pamphlet T-2, 1969, U. S. Government Printing Office, 30¢.

THE SPECIALIST MODEL FOR INSTRUCTIONAL SYSTEM DEVELOPMENT

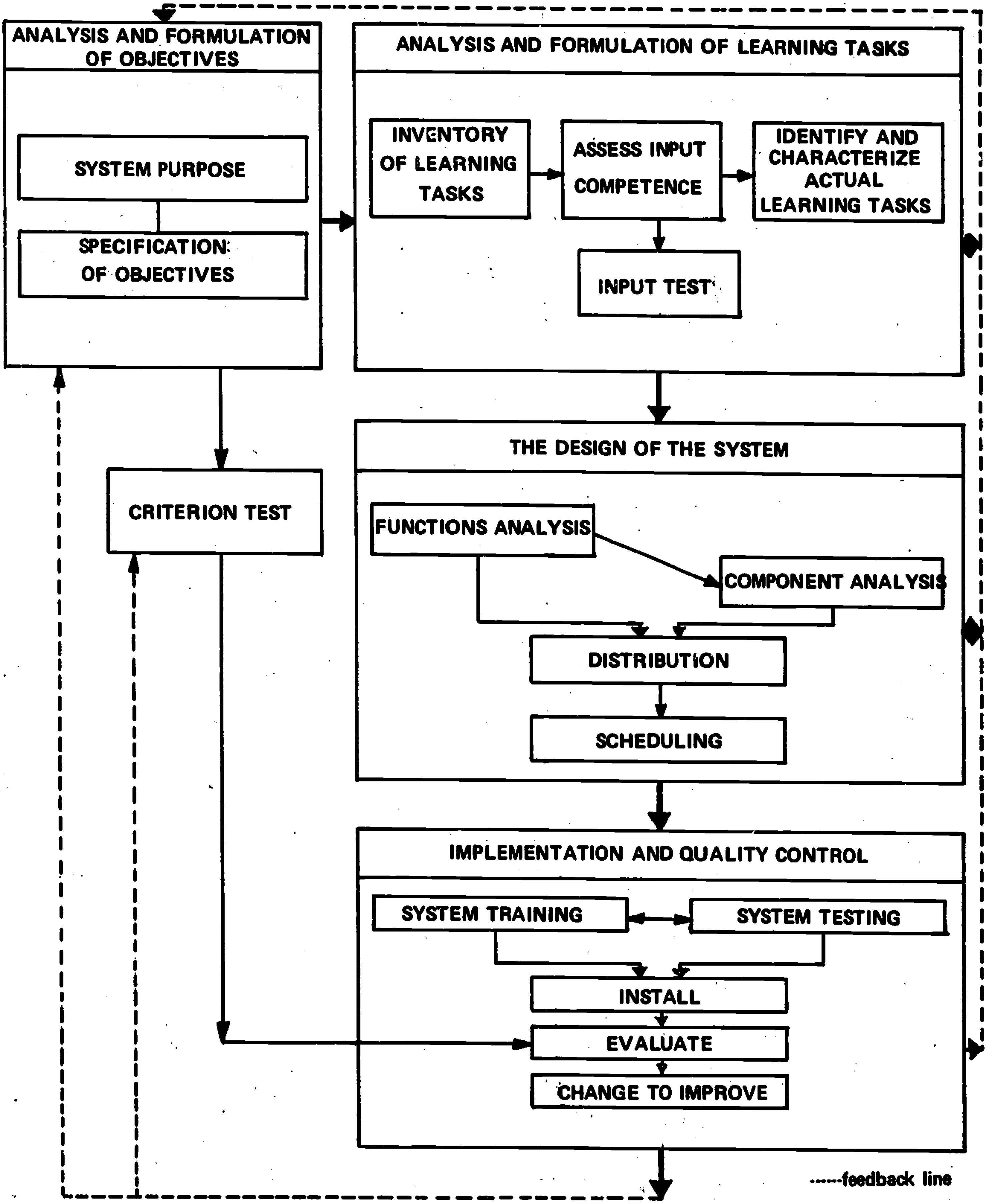




A SYSTEMS MODEL FOR INSTRUCTIONAL MANAGEMENT

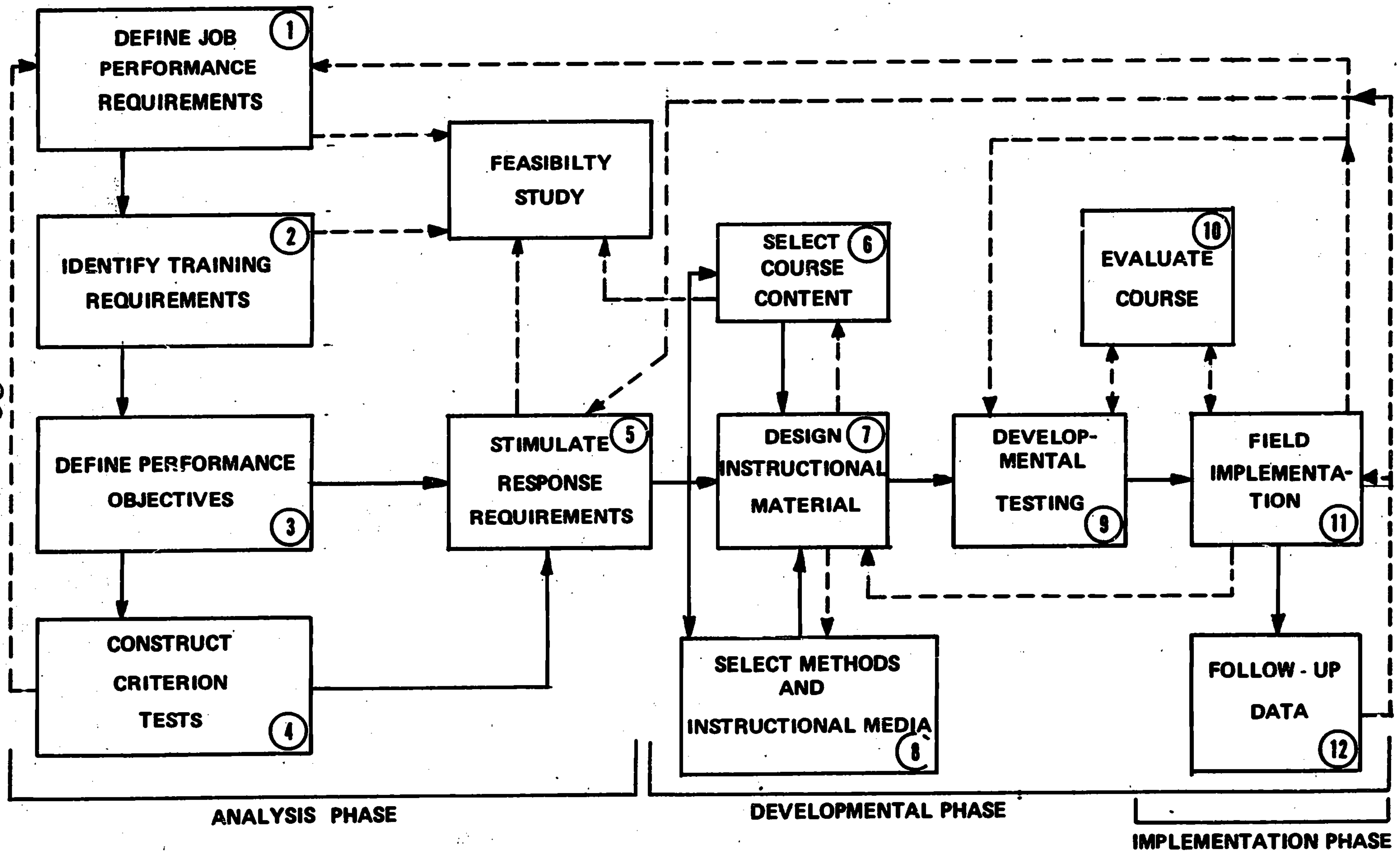
Tuckman, B. W. and Edwards, J. J. "A Systems Model for Instructional Design and Management," Educational Technology, September 1971.

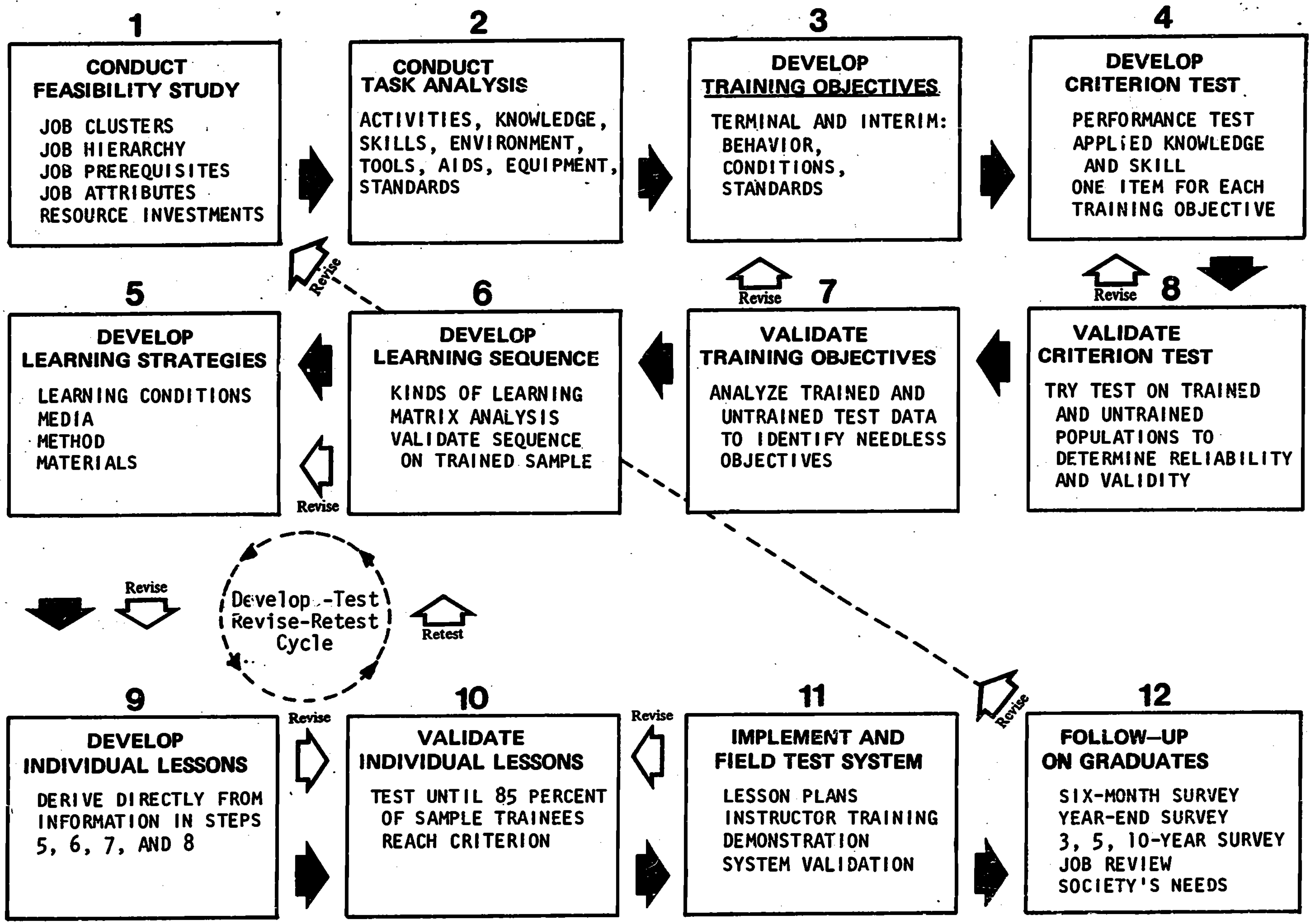
DESIGN OF INSTRUCTIONAL SYSTEMS



Banathy, Bela. Instructional Systems, Fearon Publishers, Palo Alto, Calf., 1968.

A SYSTEMS APPROACH TO INSTRUCTION





Flow Chart of Training System Development Process

SOURCE: F. Coit Butler, Instructional Systems Development for Vocational and Technical Training, Educational Technology Publications, Englewood Cliffs, New Jersey, p.53.