DOCUMENT RESUME

ED 091 441 95 TM 003 656

AUTHOR	Light, Judy A.; Lindvall, C. M.
TITLE	The Method of "Strong Inference" in the Design of Evaluation Studies.
INSTITUTION	Pittsburgh Univ., Pa. Learning Research and Development Center.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C.
PUB DATE	74
NOTE	29p.; Paper presented at the Annual Meeting of the American Educational Research Association (59th, Chicago, Illinois, April 1974)
EDRS PRICE	MF-$0.75 HC-$1.85 PLUS POSTAGE
DESCRIPTORS	Educational Experiments; *Evaluation; Evaluation Methods; *Formative Evaluation; *Hypothesis Testing; *Instructional Improvement; Instructional Materials; Instructional Programs; *Research Design
IDENTIFIERS	*Strong Inference

ABSTRACT

The objective of this study was to adapt the ideas of "strong inference" in developing a design procedure which can be used in the evaluation of an instructional system in such a way as to identify and correct specific weaknesses within a system. This method allows the evaluator to consider many hypotheses as possible causes of system malfunction and to identify which components need modifications and how these modifications can be made to improve the instructional system. The results of using an adaptation of strong inference demonstrated that designs based on strong inference were effective in establishing causal relationships between variables. (Author)

The Method of "Strong Inference" in the Design of Evaluation Studies

Judy A. Light and C. M. Lindvall

Learning Research and Development Center
University of Pittsburgh

1974

The work reported herein was supported by the Learning Research and Development Center, supported in part by funds from the National Institute of Education (NIE), Department of Health, Education, and Welfare. The opinions expressed do not necessarily reflect the position or policy of NIE and no offical endorsement should be inferred.

The Method of "Strong Inference" in the Design of Evaluation Studies

Judy A. Light and C. M. Lindvall
University of Pittsburgh

The educational researcher working in natural settings (e.g., in on-going classroom situations) is frequently involved in situations where the type of control required in true experimental designs cannot be exercised or where the types of questions answered through the use of true experimental designs are not the important questions. In most true experimental designs a key step in carrying out the study is to randomly assign pupils and teachers (and perhaps even schools and communities) to experimental conditions. As is pointed out in explaining such designs (Campbell & Stanley, 1963) this randomization is employed to control for variables whose influence cannot be controlled in other ways and to eliminate the possible influence of causal conditions of which one is unaware.

There obviously is a need to seek out and use designs other than "true" experimental ones in certain situations where randomization is neither practical nor useful. Campbell and Stanley (1963) recognized this need for identifying "quasi-experimental" designs which could be used in these situations. Campbell (1963) feels that these quasi-experimental designs can establish causal relationships under two conditions: that the interpretations made from the collected data seem plausible, and that other plausible rival hypotheses can be eliminated.

Guidance as to one possible procedure to be used in these quasi-experimental situations may be derived by noting that the situation of eliminating rival hypotheses is similar to that described by T. C. Chamberlin as involving multiple working hypotheses:

> In developing the multiple hypotheses, the effort is to bring up into view every rational explanation of the phenomenon in hand and to develop every tenable hypothesis relative to its nature, cause, or origin, and to give to all of these as impartially as [illegible]ble, a working form and a due place in the invest[illegible]on (Chamberlin, 1944, p. 160).

Platt (1964), pointing to the thoughts of Chamberlin as providing much of the basis for his thinking, has suggested a framework for testing each of a number of possible hypotheses which he has named "Strong Inference". He has made an impressive case for the claim that the use of research procedures based on this approach has been a major factor in spectacular advances in research within certain areas of biology. Strong inference procedures require the experimenter to consider all possible explanations for a given outcome, to plan the most effective sequence for studying such explanations, and to then carry out systematic investigations to eliminate as many of these as possible. Those which cannot be rejected are accepted as establishing cause and effect relationships until they can, if ever, be disproven. Platt's opinion is that scientists should be designing experiments which systematically investigate such rival hypotheses and that the results should be based on the elimination of alternative explanations. This allows the scientist to explore the unknown at the fastest rate since there is a minimum sequence of steps to be followed; and conclusions are reached rapidly by eliminating all possibilities except one.

It appears that research designs based on strong inference can offer much help to the educational researcher working in the natural setting because he is better able to reject possible causes of effects than to directly establish specific cause and effect relationships. For example, during the formative evaluation of instructional materials, an evaluator can locate and improve inadequate materials more readily than he can establish why certain materials are effective. The result would be the quick replacement of poorer instructional materials with more adequate materials.

Designs based on strong inference seem to satisfy Campbell's two requirements: interpretations made from the data would appear plausible (e.g., the inadequate materials would be replaced by materials shown to be effective) and other rival hypotheses would be eliminated (e.g., systematic investigations would be used to eliminate as many rival hypotheses as possible).

Objective

The objective of this study was to adapt Platt's suggestions for using strong inference and inference trees to develop a design procedure which could be used in the evaluation of an instructional system. The use of inference trees seems particularly adaptable to the kinds of formative evaluation activities required during the development and tryout of complex instructional systems. First the evaluator must consider and test hypotheses concerning the effects of all components of the systemunder development including tests, lesson materials, teacher behavior, pupil behavior and classroom management rather then select only some of the variables which can effect any instructional system. Second, the procedures should be effective in that they can quickly identify and correct specific weaknesses within the system.

Method

A rather typical problem faced by the person attempting to carry out formative evaluation activities in conjunction with the development of a new educational program is that of identifying the specific causes of given instances of system failure. An example of this type of problem is the situation where there is rather consistent pupil failure on a test within a new curriculum. Faced with this particular result, the formative evaluator must determine if the cause for this failure is inadequate lesson materials, an invalid test, poor classroom instruction, the improper placement of instruction within the overall curriculum sequence, or some other unknown cause This problem of identifying the cause of some specific evidence of inadequacy in a new instructional program can be considered as a problem of investigating a number of hypotheses, each identifying a possible cause of this inadequacy. Such a situation would seem to require the use of some type of quasi-experimental procedure in investigating each such plausible hyopthesis.

The procedures developed in this study were aimed at attempting to adapt Platt's method of strong inference and the use of inference trees into procedures for establishing cause and effect relationships between one dependent variable and the multiple possible causes for that dependent variable in a complex instructional system (the independent variables). For example, one approach to the analysis of an instructional situation is to view each possible cause of failure as representing the independent variable in a hypothesis. Each such hypothesis would express its presumed influence on the dependent variable, e.g., test performance. While the approach of viewing this problem as one of investigating specific hypotheses appears to be an obvious one, it leaves the evaluator with the related question of how to design procedures for carrying out such investigations.

Basically, the evaluator attempting to provide information for the improvement of an educational program and the researcher designing a true experiment are faced with the same task. Both must design their studies in a way which will permit them to draw valid conclusions concerning the effect of a given treatment. To develop the needed design both must (1) specify the dependent variable, or exact effect, that is of concern, (2) identify the treatment, or independent variable, being studied and (3) establish control conditions that permit conclusions to be drawn concerning the effect of the variable being studied by eliminating the plausibility of other possible causal explanations. To understand the design problems of the evaluator it may be useful to examine the similarities between his task and that of the experimenter.

1 Specifying the dependent variable. In many research studies the logical starting point for planning the investigation is the identification of the variable that one wishes to affect, e. g. reading achievement, pupil self-concept, teacher satisfaction, etc. The researcher must specify this variable quite exactly, typically in terms of the instrument or procedure that is to be used to measure or describe it. In a like manner, the formative evaluator must identify the specific program outcome that is of concern. For example, if pupils are failing certain tests, the first task of the evaluator is to specify what they are doing incorrectly.

2. Identifying the independent variable. The experimenter wishes to study the effects of a certain treatment. To do this he defines that treatment or treatments with which he will compare it (one of the latter may be the "no treatment" condition). The evaluator's role is to identify and find which probable cause helped produce a given effect. That is the evaluator must identify the specific program components which can effect the dependent variables under investigation. The evaluator wishes to be able to say "This specific program component is the cause of the poor performance of the system which we are investigating." To do this he will have to identify a number of program components whose failure could be plausible reasons for the poor performance.

3. Controlling experimental conditions. One aspect of experimental design is to establish controls so that certain conditions are common to both experimental and control groups. For example, in an experiment comparing instructional methods, both experimental and control groups may be taught by the same teacher in an effort to eliminate teacher effectiveness as an alternative explanation of any differences in results. The evaluator, too, must be concerned with [illegible] control of certain conditions. However, evaluation does not typically involve the comparison of two groups (although in certain situations it obviously could involve this). Usually it involves gathering data as a program operates within some one

context. However, this makes even more real the possibility that "uncontrolled conditions" are the actual causes of poor program performance. How, then, are such conditions to be controlled? In attempting to answer this question it is important to take into account that while the formative evaluator maybe examining one component of some type of "program" or "system," such a program is influenced by many other components and procedures. For example, if the evaluator is assessing the effectiveness of some type of lesson material, it is likely that the total program specifies certain procedures to be followed in using these materials. Such procedures serve to specify some of the things that are to be controlled. This means that if the evaluator is studying the effectiveness of given lesson materials, conditions must be controlled to the extent that teachers are using the lessons by following the specified procedures. Without this type of control the evaluator has no way of eliminating such things as "improper teacher procedure" as being the cause of lack of achievement. The experimenter, concerned largely with isolating the effect of one independent variable, controls the effect of certain other variables by equating the experimental and control groups with respect to these variables. The formative evaluator, concerned with investigating the effectiveness of some specific program component, controls the operating program so that other relevant components are functioning in the intended manner.

4. <u>Controlling for individual subject variables</u>. In a true experiment, subjects are typically assigned to treatment and control groups through some type of random assignment. This procedure, together with the use of tests of statistical significance, helps one to eliminate concern about a variety of factors associated with individual subject differences as the causes of any effects produced. The evaluator, assessing the components of a program operating within the context of an on-going school program, typically cannot employ randomization. Despite this, there must be some control for this type of individual difference factors. The evaluator

attempts to partially negate the effects of some of them by basing the evaluation of the performance on as representive a sample as possible. Other variables of this general type may be ones that the instructional program is designed to control. For example, students differ in the care with which they study and complete a lesson. This type of pupil carelessness cannot be permitted to affect pupil performance on lesson materials and the results then be interpreted as indicating some inadequacy in the materials. Pupil variables of this type must be identified and their presence or absence noted in the case of any given pupil performance. Control of the variable, in the above case, might be achieved by requiring such pupils to re-study the lesson under proper conditions.

In the present study, the foregoing steps in design were delineated in terms of their application to the formative evaluation of lesson materials being given a try-out within the context of a program for individualized instruction. In this specific application these steps can be described as:

1. Defining the dependent variable, that is, selecting what specific evidence will be used to identify a breakdown in the instructional system.
2. Defining the independent variables, that is, listing the multiple plausible hypotheses that might account for the specific breakdown.
3. Defining and controlling the "experimental conditions" by specifying and then monitoring key aspects of the instructional environment.
4. Defining and examining a number of student performance variables that have to be accounted for in attempting to clarify cause and effect relationships between independent and dependent variables.

Platt's description of the use of strong inference in establishing cause and effect relationships suggests the worthwhile use of this procedure in developing and carrying out the foregoing steps in evaluation. He defines strong inference as:

> applying the following steps to every problem in science, formally and explicitly and regularly:
>
> 1) Devising alternative hypotheses;
> 2) Devising a crucial experiment (or several of them) each of which will as nearly as possible, exclude one or more of the hypotheses;
> 3) Carrying out the experiment so as to get a clean result;
> 1') Recycling the procedure, making subhypotheses or sequential hypotheses to refine the possibilities that remain, and so on... (Platt, 1964, p. 347).

It seems that strong inference should provide the evaluator with a formal structure for developing and carrying out these steps. Adapting Platt's procedures to use in the formative evaluation of instructional systems has as its goal the need to establish that a specific system malfunction was a consequence of an identifiable inadequency within a system component (Light & Reynolds, 1972). Thus the required procedure of strong inference becomes one of formulating "multiple hypotheses" and using inference trees as a basis for specifying and eliminating rival hypotheses so as to identify the exact system component that must be changed. One major reason for assuming the effectiveness of strong inference in these types of studies is that strong inference requires the evaluator to identify and test each cause of poor test performance whether the cause is an inadequacy within the lessons (the independent variables) or an individual student inadequacy that must be corrected.

Application of the Method

Data Source

This investigation was carried out within the context of an elementray school serving as the laboratory for the development and tryout of modifications in the Individually Prescribed Instruction program in mathematics. Data concerning system malfuctions were obtained from an intensive analysis of test results and student performance on lessons and by observing classroom behavior during an entire school year.

The instructional system under development was the 45 units consisting of 264 objectives normally used by the students in the fourth grade during the course of the school year. The system components being investigated as possible causes of system failure included the lessons, the tests, teacher behaviors, and pupil behaviors.

Procedures

The procedures used in this study followed the four step outline previously presented.

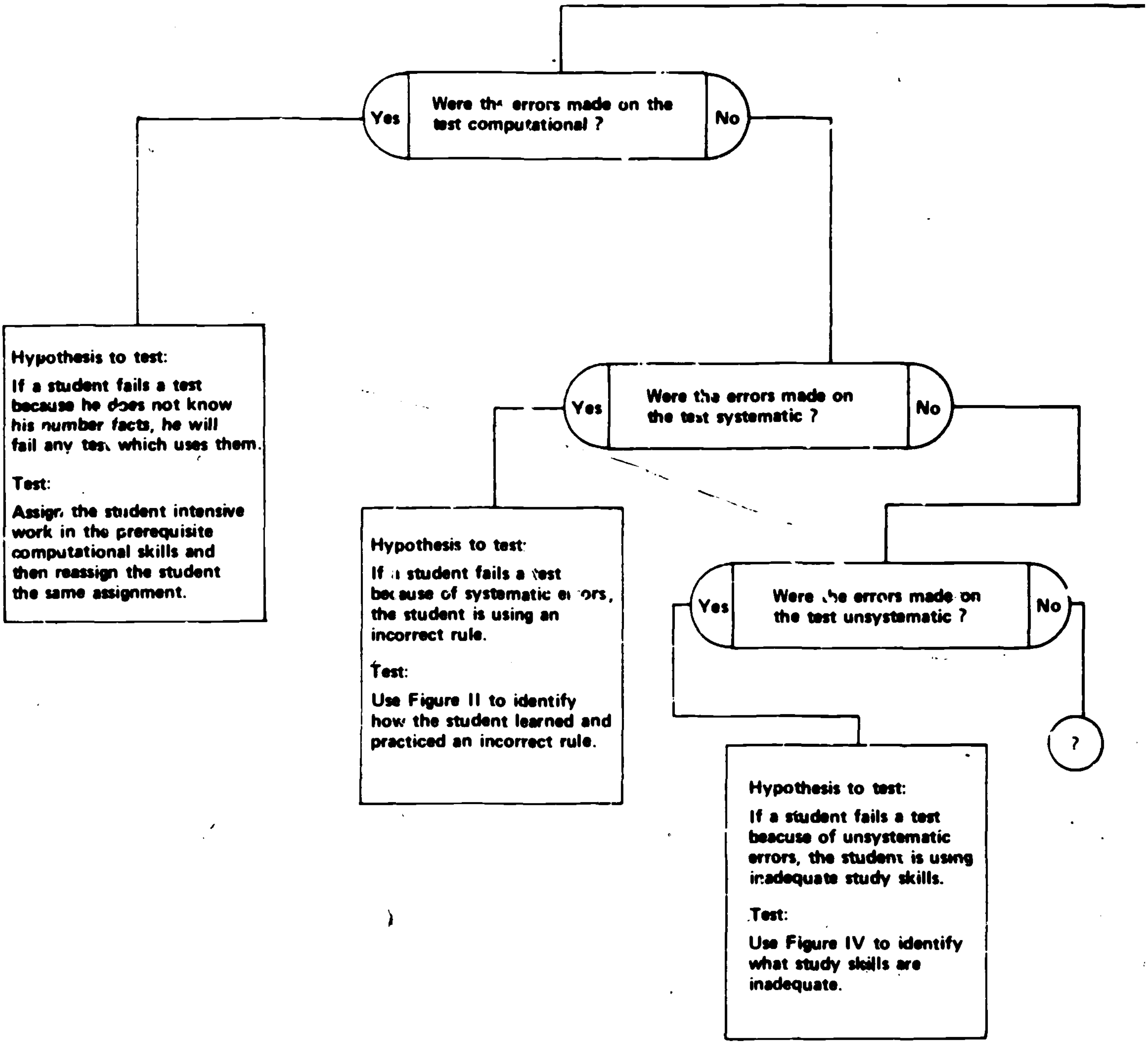
1. Specifying the dependent variable. In general, the evidence of system failure used in this particular application of the procedure, was pupil performance on a criterion test. However, to investigate the cause of any such failure it was necessary to obtain a very specific description of the exact nature of the failure. This was facilitated by the use of a form of inference tree as shown in Figure I. This tree involves answering questions by examining the student's responses on the test in order to identify as specifically as possible the type of error the student made and which branch of the tree would be the most logical to use to pin-point the specific system failure. As can be seen this inference tree was designed to pin-point the type of error made on the test. After analyzing many tests, four major types of errors were found:

1. Process errors which were defined as errors resulting from the student not carrying out the exact process being taught.

2. Computational errors which were defined as errors resulting from the student writing an incorrect sum, product, quotient, or difference in a problem.

Figure I

Identifying Student Test Errors



Were the errors made on the test computational ?

Hypothesis to test:

If a student fails a test because he does not know his number facts, he will fail any test which uses them.

Test:

Assign the student intensive work in the prerequisite computational skills and then reassign the student the same assignment.

Were the errors made on the test systematic ?

Hypothesis to test:

If a student fails a test because of systematic errors, the student is using an incorrect rule.

Test:

Use Figure II to identify how the student learned and practiced an incorrect rule.

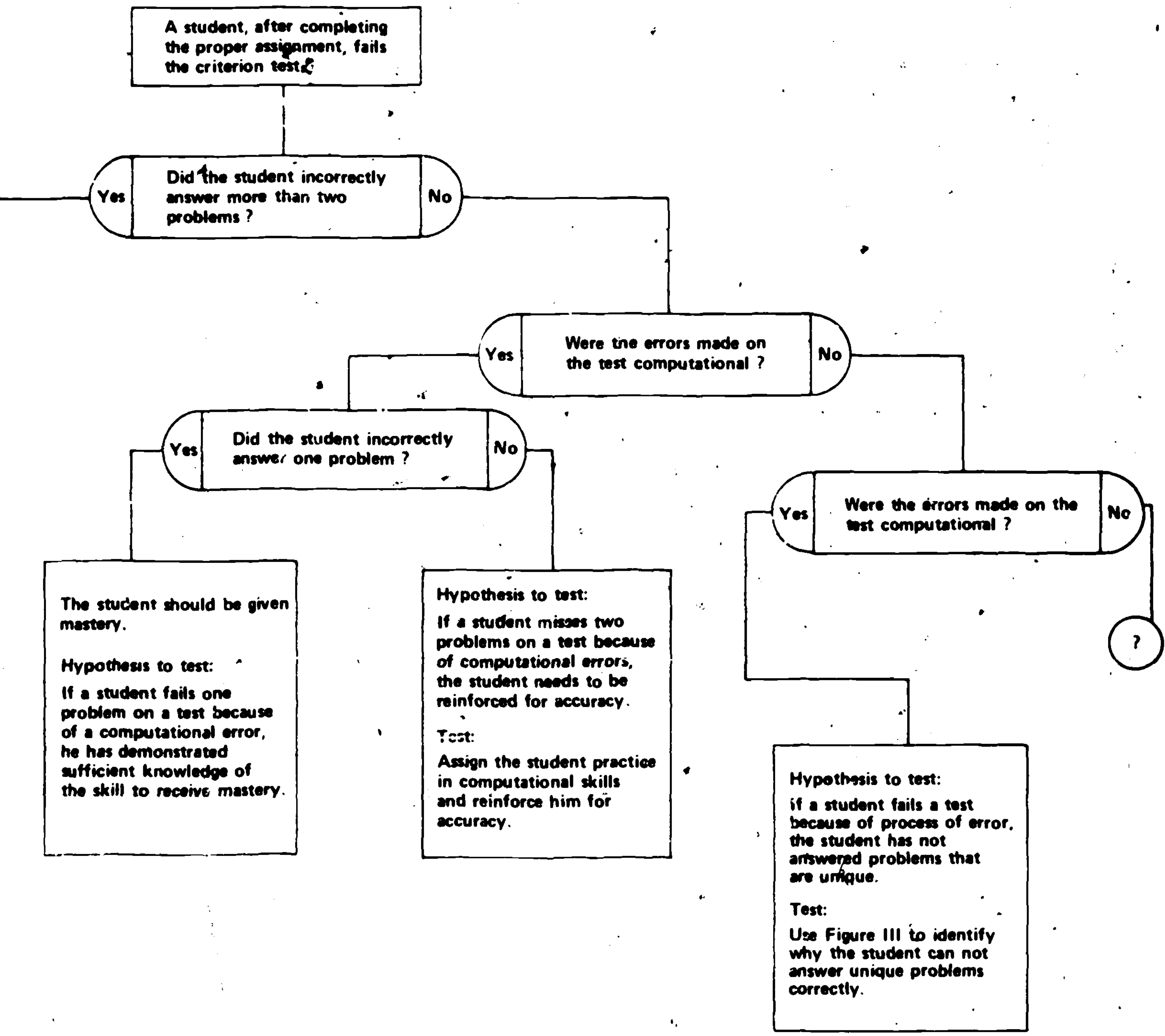
Were the errors made on the test unsystematic ?

Hypothesis to test:

If a student fails a test beacuse of unsystematic errors, the student is using inadequate study skills.

Test:

Use Figure IV to identify what study skills are inadequate.


A student, after completing the proper assignment, fails the criterion test.
Yes
Did the student incorrectly answer more than two problems ?
No
Yes
Were the errors made on the test computational ?
No
Yes
Did the student incorrectly answer one problem ?
No
Yes
Were the errors made on the test computational ?
No
The student should be given mastery.
Hypothesis to test:
If a student fails one problem on a test because of a computational error, he has demonstrated sufficient knowledge of the skill to receive mastery.
Hypothesis to test:
If a student misses two problems on a test because of computational errors, the student needs to be reinforced for accuracy.
Test:
Assign the student practice in computational skills and reinforce him for accuracy.
?
Hypothesis to test:
If a student fails a test because of process of error, the student has not answered problems that are unique.
Test:
Use Figure III to identify why the student can not answer unique problems correctly.

3. Systematic errors which were defined as errors resulting from the student using an identical but incorrect rule to answer all items.

4. Non-systematic errors which were defined as errors resulting from the student answering items incorrectly but for different reasons.

It was also apparent that the number of items a student missed offered helpful information in identifying the specific type of failure. If a student failed one or two items on a test, the type of error was usually the result of a process error or a computational error. When a student only missed one or two items because of a process error, it usually meant that the items missed were unique in their content. For example, a student only failed the items on a test in subtraction with borrowing when the problem contained a zero in the tens place, but passed all other subtractions items. Once the uniqueness has been identified, the evaluator can use the appropriate tree to select a testable hypothesis.

Note that the hypotheses derived from this analysis are of two major types. One type provides for improving the individual pupil's command of pre-requisite skills and then having him use the lesson again. Taking such a step provides a form of control on individual pupil differences on certain crucial variables. If testing this type of hypothesis shows that improving command of pre-requisites leads to the students passing the test, this student's lesson performance will not be subjected to further analysis. The second type of hypothesis shown in Figure I involves those that can only be examined through further analyses that can, in turn, be facilitated by the development of additional inference trees. Such trees serve to identify other types of variables that must be controlled or tested as possible causes.

The tree shown in Figure I contains several questions whose negative answer leads the evaluator to a question mark. These questions marks symbolize situations which have not yet arisen, but are included as a reminder that all the inference trees are working models which may need to be expanded as the trees are used in on-going situations. If a student fails a test for reasons other then those already explored the evaluator would build onto the tree.

2. Specifying the independent variables. Once the exact nature of criterion test failure is identified it becomes possible to generate hypotheses that identify probable causes (the independent variables) of each failure. At first these hypotheses were generated by analyzing each pupil's materials, and answering these kinds of questions:

1. What was similar about the problems missed on a test?
2. How did the problems missed differ in form or content from those items passed on a test?
3. Where in the instructional materials were these types of problems presented?
4. What in the instructional material could have caused the student to fail the test?
5. Did the student use the material in the designated manner?
6. How can the hypothesized cause of failure be experimentally investigated?

The result of this procedure was an extensive list of possible causes of test failures. Examples of thes causes are represented by the following hypotheses.

If a pupil fails a test, then:

a. the pages may not teach and provide practice on the tested content.
b. the pages may not teach and provide practice on "unique" properties.
c. the pages may not require adequate practice.
d. the prescription may not contain pages which are equivalent in form and content to the test.
e. the pupil may not have learned from the teaching pages.
f. the pupil may have demonstrated poor work skills.
g. the pupil may have done the assignment incorrectly.

h. the pupil may not have the appropriate prerequisite behaviors for a given lesson page.
i the pupil may not be motivated to do accurate work.
j. the pupil may not be "attending to task" while doing his work.
k. the pupil may not be checking his work.
l. the pupil may not be able to use self-evaluation skills to decide if he had learned the required skills.

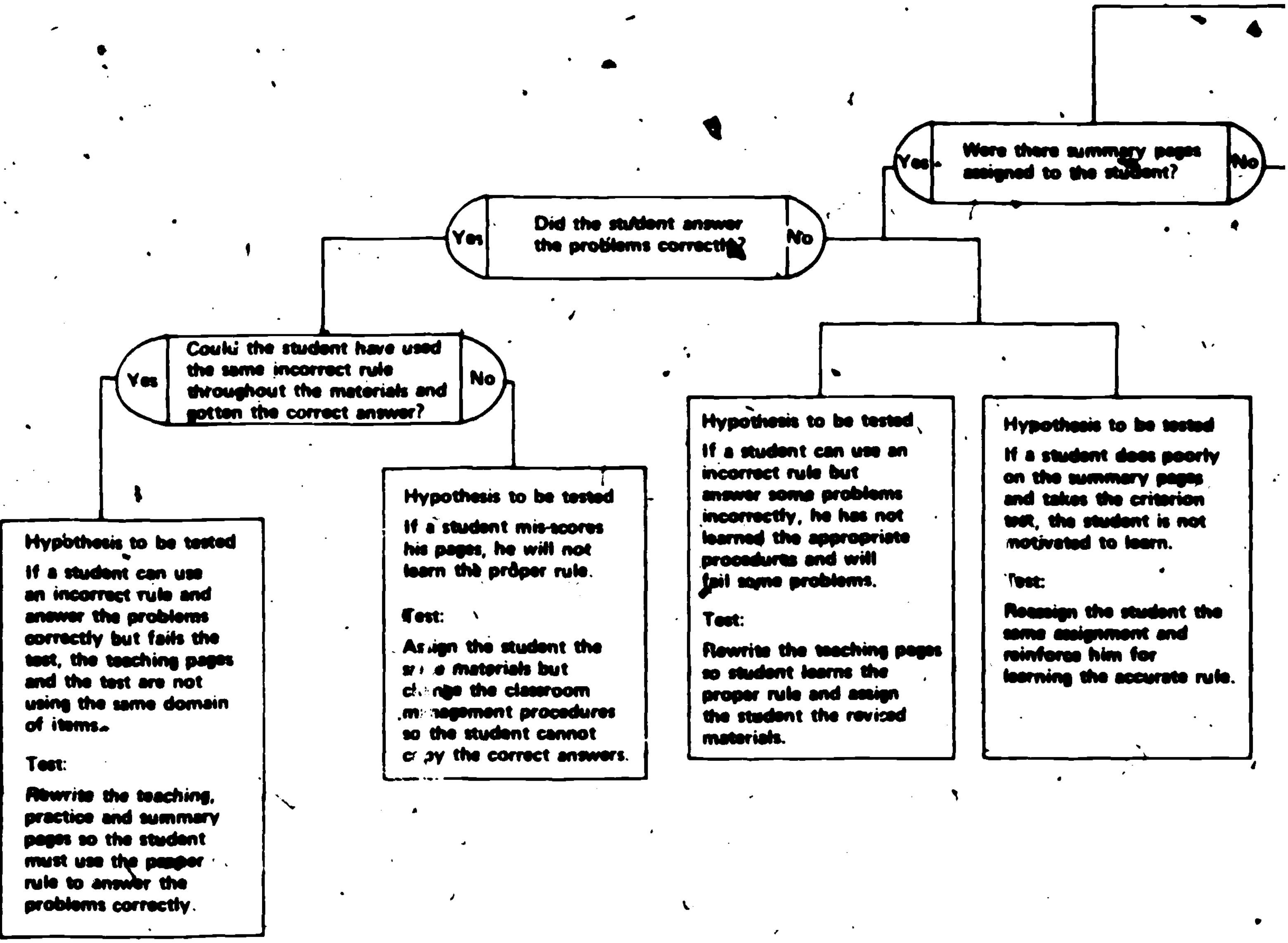
(A demonstrations of how this was carried out can be found in Light and Reynolds, 1972).

Figure II provides an example of how such possible causes can be structured into an inference tree that explains some particular type of test failure Note that this involves an analysis of how the pupil performed on certain major parts of the lesson materials. In this example such "parts" of the instructional booklets were practice pages, teaching pages, and summary pages. Answers to questions concerning how the individual student did on these types of pages were used to identify which hypotheses should be tested first. The order of these questions was determined by deciding what types of information were needed to either eliminate or establish certain conditions as the cause of failure. For example, if a student fails a test because he has missed "an unique item" this could be caused by one of several factors: the materials do not teach the student how to solve the unique type of items, the materials do not provide any practice in solving the unique items, or the student did not do the pages properly. Questions concerned with identifying if the unique items were taught or practiced should be answered before analyzing how the student did on these pages, since hypotheses based on the student's answers to certain problems presuppose that the problems were taught in the materials.

Again it can be seen that the hypotheses generated by this tree are of two general types. One type is the same as that found in Figure I. This type involves changing the student and represents a form of control of individual student performance. The second type of hypothesis deals with

Figure II

Identification of how a student learned and practiced an incorrect role.



Were there summary pages assigned to the student?

Did the student answer the problems correctly?

Could the student have used the same incorrect rule throughout the materials and gotten the correct answer?

Hypothesis to be tested

If a student can use an incorrect rule and answer the problems correctly but fails the test, the teaching pages and the test are not using the same domain of items.

Test:

Rewrite the teaching, practice and summary pages so the student must use the proper rule to answer the problems correctly.

Hypothesis to be tested

If a student mis-scores his pages, he will not learn the proper rule.

Test:

Assign the student the same materials but change the classroom management procedures so the student cannot copy the correct answers.

Hypothesis to be tested

If a student can use an incorrect rule but answer some problems incorrectly, he has not learned the appropriate procedures and will fail some problems.

Test:

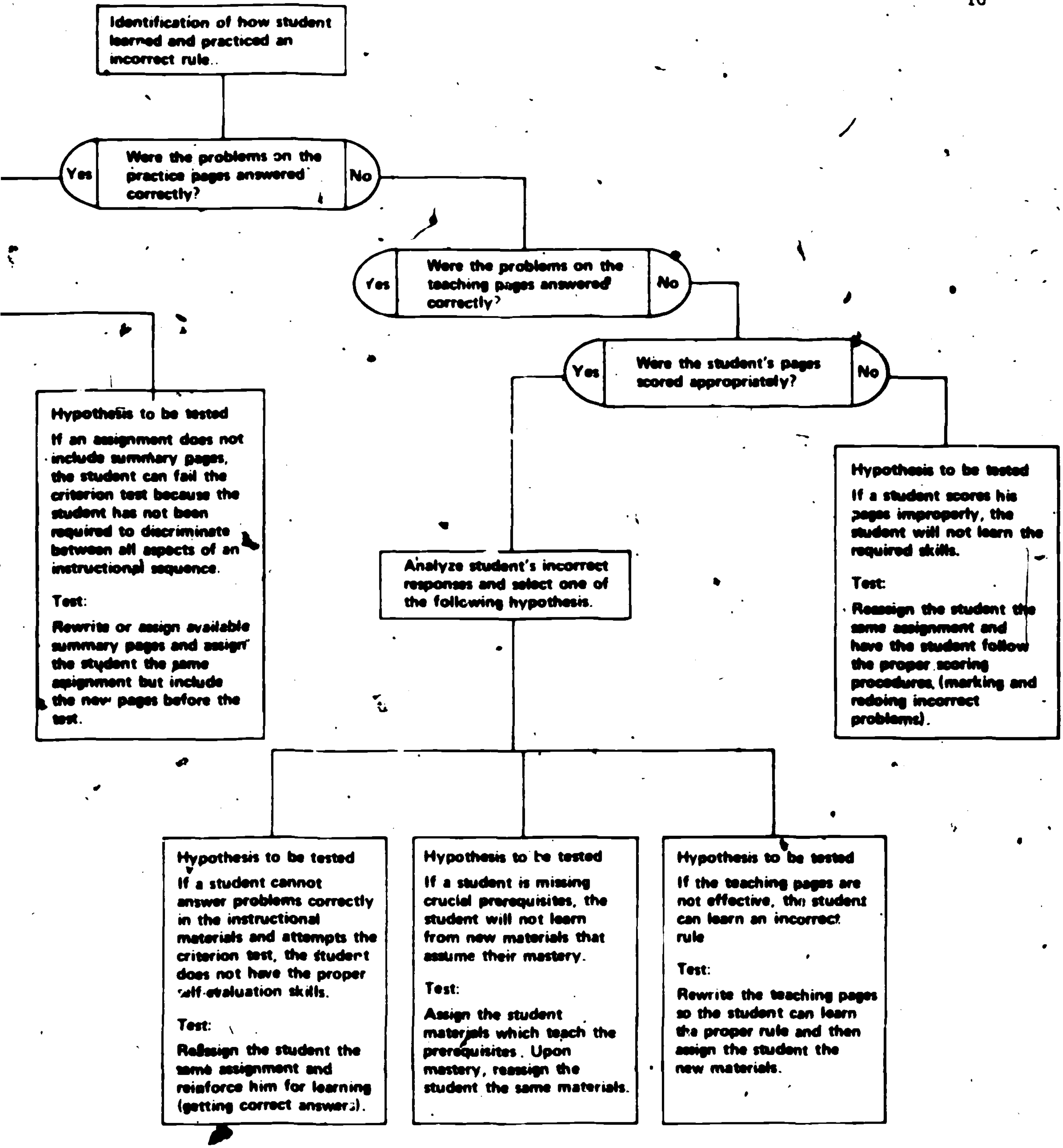
Rewrite the teaching pages so student learns the proper rule and assign the student the revised materials.

Hypothesis to be tested

If a student does poorly on the summary pages and takes the criterion test, the student is not motivated to learn.

Test:

Reassign the student the same assignment and reinforce him for learning the accurate rule.


Identification of how student learned and practiced an incorrect rule.
Yes
Were the problems on the practice pages answered correctly?
No
Yes
Were the problems on the teaching pages answered correctly?
No
Yes
Were the student's pages scored appropriately?
No
Hypothesis to be tested
If an assignment does not include summary pages, the student can fail the criterion test because the student has not been required to discriminate between all aspects of an instructional sequence.
Test:
Rewrite or assign available summary pages and assign the student the same assignment but include the new pages before the test.
Analyze student's incorrect responses and select one of the following hypothesis.
Hypothesis to be tested
If a student scores his pages improperly, the student will not learn the required skills.
Test:
Reassign the student the same assignment and have the student follow the proper scoring procedures (marking and redoing incorrect problems).
Hypothesis to be tested
If a student cannot answer problems correctly in the instructional materials and attempts the criterion test, the student does not have the proper self-evaluation skills.
Test:
Reassign the student the same assignment and reinforce him for learning (getting correct answers).
Hypothesis to be tested
If a student is missing crucial prerequisites, the student will not learn from new materials that assume their mastery.
Test:
Assign the student materials which teach the prerequisites. Upon mastery, reassign the student the same materials.
Hypothesis to be tested
If the teaching pages are not effective, the student can learn an incorrect rule
Test:
Rewrite the teaching pages so the student can learn the proper rule and then assign the student the new materials.

the basic independent variables that were of concern in this study, namely, needed changes in lesson materials. The latter hypotheses were tested by writing or rewriting the specified lesson pages and then having the student work through the lesson again. When such changes resulted in the pupil passing the criterion test, it was assumed that the lesson inadequacy had been identified. Of course, both those hypotheses dealing with changes in the pupil and those dealing with changes in lesson materials had to be tested. In essence, such tests are equivalent to Platt's "crucial experiments" in that they provide a means for rejecting a specifically hypothesized cause of test failure.

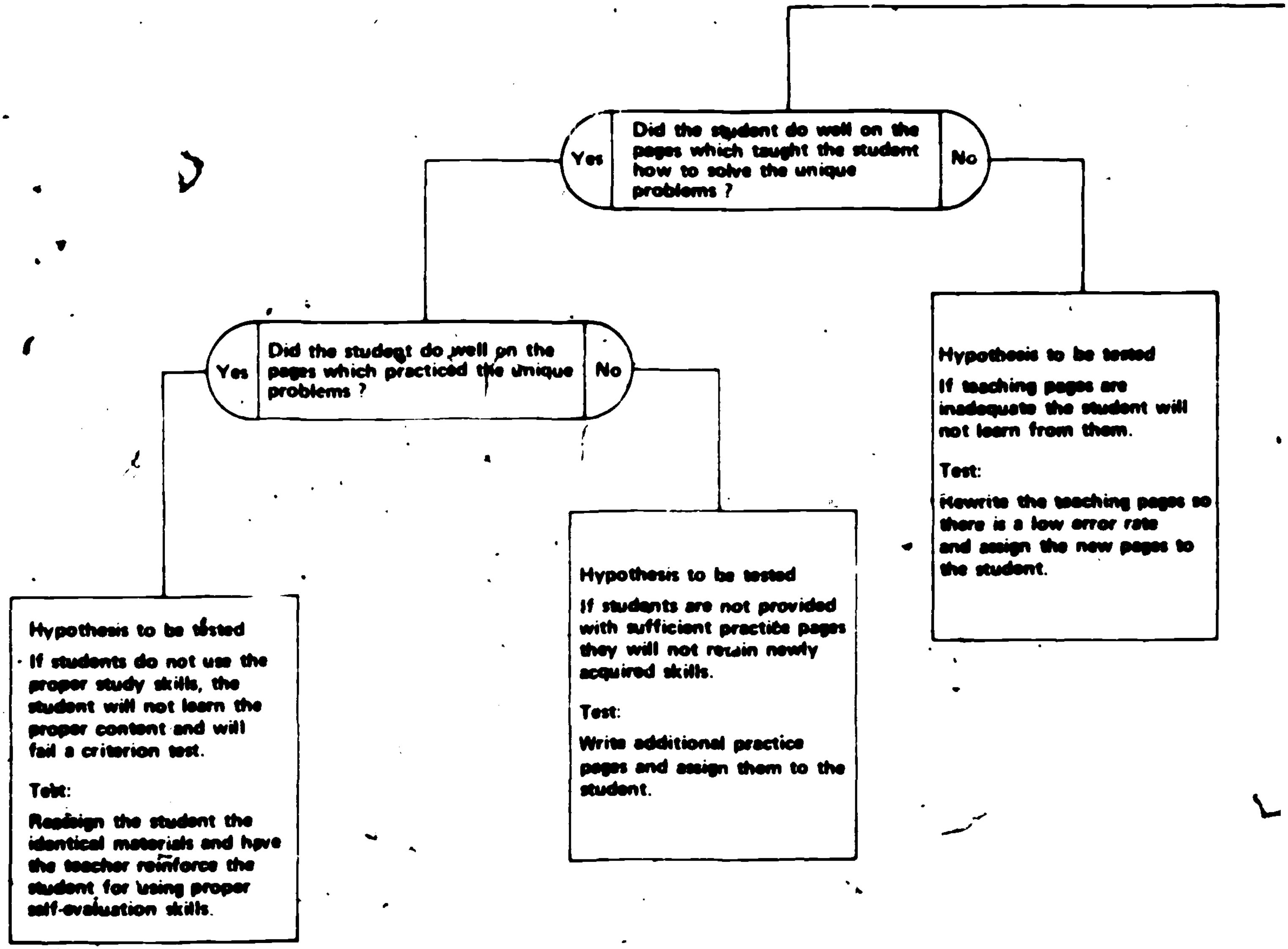
Figure III provides another example of an inference tree developed to identify the same type of variables as those identified in Figure II. Figure III starts with the specification of a slightly different type of test information than that which provided the basis for the analysis developed in Figure II.

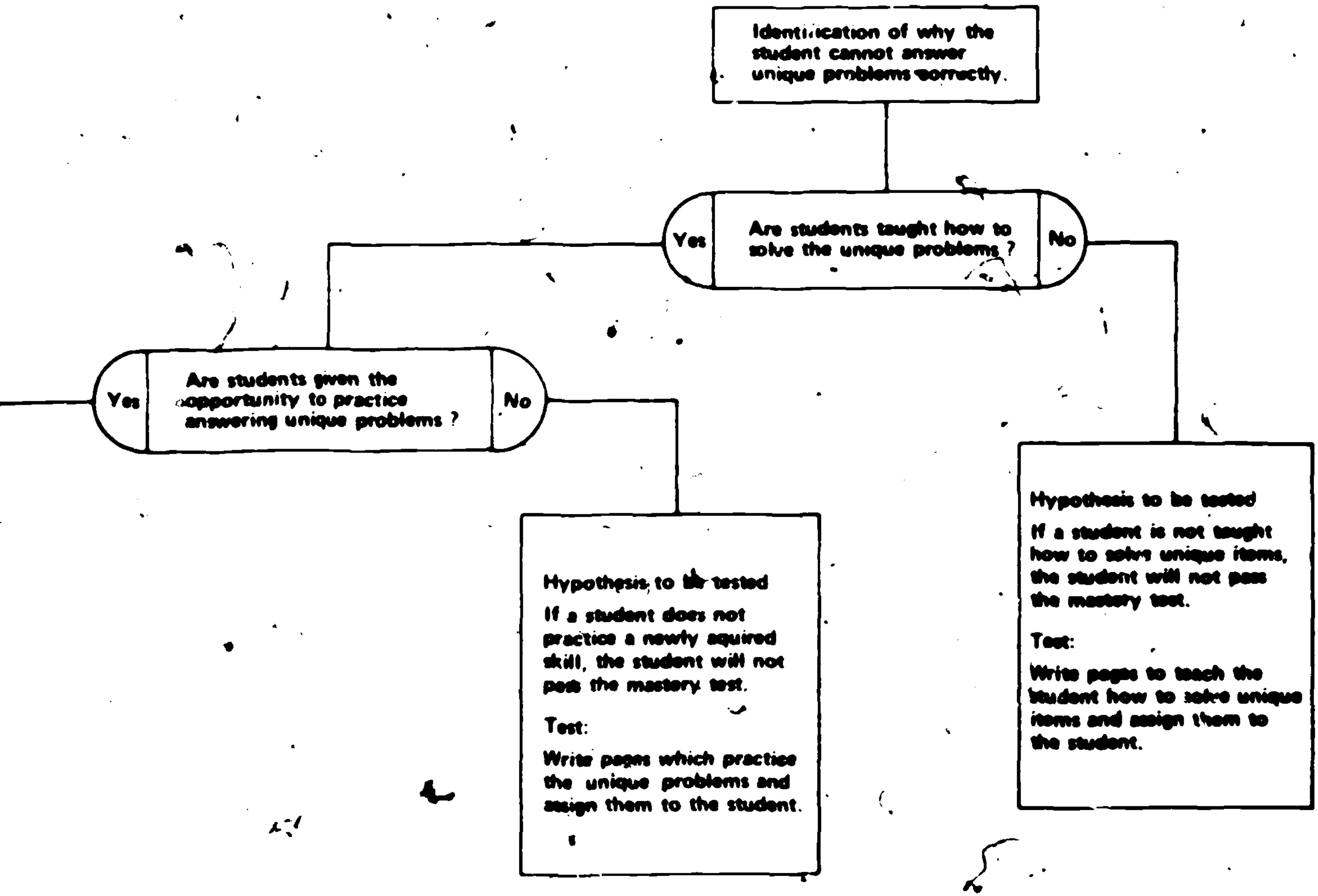
There are several points within the trees shown in Figures II and III where the answers to certain questions lead to several hypotheses rather than just one. Presently questions, which can discriminate among these hypotheses have not been developed. Certain rules have been found useful in deciding which hypothesis should be tested first. Examples of these include the following. If one hypothesis is easier to test than another, choose the easier one first. If a student has previously demonstrated poor study skills, choose the hypothesis concerning inadequate study skills first. If the student has been observed as not "attending to task" by the teacher or evaluator, choose the hypothesis concerning motivation first, etc.

3. Controlling experimental conditions. Extensive work in curriculum development and evaluation and classroom observation by both writers suggested the necessity of first carefully specifying the desired classroom procedures for the use of the given lesson materials (Light, 1972).

Figure III

Identification of why the student cannot answer unique problems correctly


Did the student do well on the pages which taught the student how to solve the unique problems ?
Yes
No
Did the student do well on the pages which practiced the unique problems ?
Yes
No
Hypothesis to be tested
If teaching pages are inadequate the student will not learn from them.
Test:
Rewrite the teaching pages so there is a low error rate and assign the new pages to the student.
Hypothesis to be tested
If students are not provided with sufficient practice pages they will not retain newly acquired skills.
Test:
Write additional practice pages and assign them to the student.
Hypothesis to be tested
If students do not use the proper study skills, the student will not learn the proper content and will fail a criterion test.
Test:
Reassign the student the identical materials and have the teacher reinforce the student for using proper self-evaluation skills.



Identification of why the student cannot answer unique problems correctly.

Are students taught how to solve the unique problems?

Are students given the opportunity to practice answering unique problems?

Hypothesis to be tested

If a student does not practice a newly aquired skill, the student will not pass the mastery test.

Test:

Write pages which practice the unique problems and assign them to the student.

Hypothesis to be tested

If a student is not taught how to solve unique items, the student will not pass the mastery test.

Test:

Write pages to teach the student how to solve unique items and assign them to the student.

Two methods were found to be effective in controling for the effects of such classroom proceduce variables. They were either eliminated or stabilized. In order to eliminate the effects of a variable, rules were constructed which prohibited their effects from occuring. For example, in order to insure that student's test performance was only the result of what was learned from the instructional material rather than being the result of another student's or the teacher's assistance, rigorous testing rules were designed to insure valid testing procedures. If any rule was broken, the student's test was voided, and an equivalent form had to be taken by the student.

The other effective method for controlling the effects of some variables which could not be eliminated was to stabilize their effects. For example, teacher behavior is known to influence student performance. The teacher's role in the class was therefore explicitly defined as to what she could and could not do when interacting with her pupils. This interaction was then observed by the evaluator and was monitored through cooperative planning involving the evaluator and the teacher who was taking part in this formative evaluation program. This type of "control" of conditions is analogous to the experimental control imposed by the researcher.

4. Examining student performance variables. The inference trees developed in Figures II and III illustrated that a logical analysis of the causes of a specific test failure, involving a detailed examination of how the student performed on the related lesson booklet, results in two types of hypotheses. The first type related to needed changes in student performance and capabilities, largely the pre-requisite skills which he did or did not possess. The second type related to needed changes in lesson pages. This second type was discussed at some length in a foregoing section under " 2. Specifying the independent variable." The first

type was described as involving hypotheses dealing with the control of student differences in performance capabilities. An additional form of such performance capabilities is represented by hypotheses related to the type of study skills which the individual pupil uses in studying the lessons. An example of an inference tree involving such variables is provided in Figure IV. The hypotheses identified by this tree should be tested in the same manner followed with respect to the hypotheses identified in Figures II and III, that is, "crucial experiments" must be performed. Such experiments help the evaluator to rule out certain individual pupil qualities as explanations for poor test performance. In this way they provide a "control" of pupil variables when one is attempting to identify those portions of a lesson that need to be changed

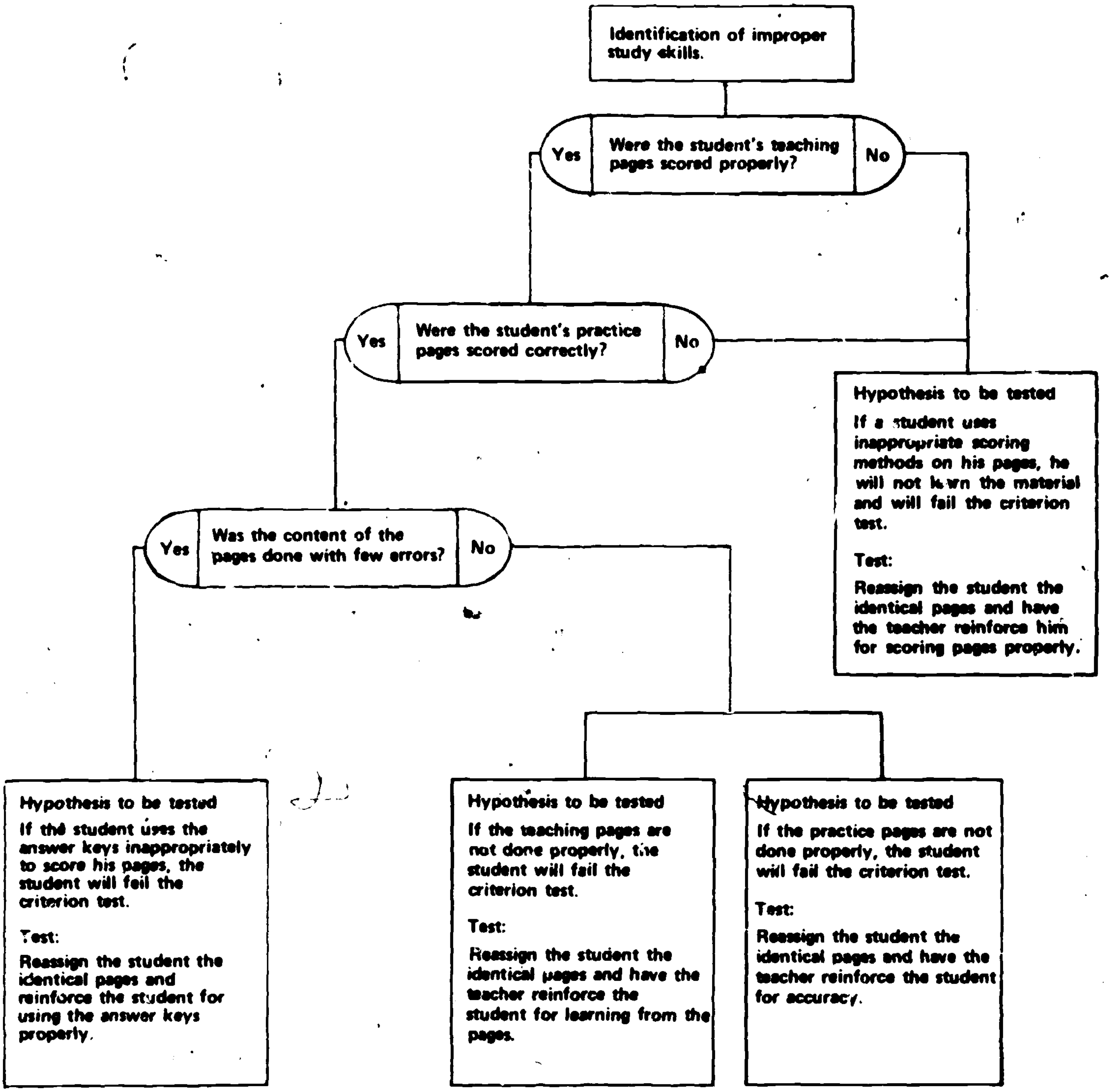


Figure IV

Identification of Improper Study Skills

Results and Conclusions

The purpose of this study was to investigate procedures for using strong inference and inference trees in the evaluation of an instructional system in order to establish the causes of, and to improve, inadequate instructional materials.

The procedures described in this paper were effective in identifying a cause of failure for every test failed during an entire school year in two classrooms. It is difficult to report exactly how effective these procedures were in improving the instructional materials because, for many objectives, only a few students used the revised lessons, making it difficult to evaluate the effectiveness of some revisions. Gross analysis does indicate that improvements were made in the instructional materials during the school year: student performance, measured by passing tests on the first attempt, was improved on 82% of the objectives studied during the school year.

The writers of this report were encouraged about the use of strong inference procedures and feel that they can be effective in certain settings in establishing causal relationships. One of the major impacts strong inference can have is its requirement that the evaluator must identify and test multiple hypotheses until a hypothesis is identified that cannot be rejected. Because of this, the evaluator will locate previously unknow independent variables which are affecting the instructional system.

The purpose of developing inference trees is to provide a formal structure for using strong inference. The trees, once constructed, provided a listing of possible causes of test failure, a description of how to carry out a test of each hypothesis, and questions whose answers have a high probability of leading the evaluator to those hypotheses that should be the most difficult to reject first. If the hypothesis selected is rejected, that is the student fails an equivalent test, other hypotheses must be tested until one cannot be rejected. The criterion for successful identification of cause of test failure is always student mastery of an equivalent test; if the student does not master the test, the entire process is begun again.

Summary

The use of procedures based on strong inference were found to be effective in establishing cause and effect relationships during the formative evaluation of an individualized instructional program. The construction of inference trees that were applicable in evaluating any lesson in this program provided a set of efficient guides for identifying weaknesses in lesson materials. The authors feel that the use of designs based on strong inference can be of value in many settings where the type of control required by true experimental designs cannot be exercised or where the types of questions answered through the use of true experimental designs are not the important questions.

References

Campbell, D. T. From description to experimentation: Interpreting trends as quasi-experiments. In C. W. Harris (Ed.), *Problems in measuring change.* Wisconsin: University of Wisconsin Press, 1963.

Campbell, D. T. & Stanley, J. C. *Experimental and quasi-experimental designs for research.* Chicago: Rand McNally and Company, 1963.

Chamberlin, T. C. The Method of Multiple Working Hypotheses. In W. J. Gephart and R. B. Ingle (Eds.), *Educational research.* Columbus, Ohio: Charles E. Merrill, 1969. Pp. 155-164.

Light, J. A. The development and application of a structured procedure for the incontext evaluation of instructional materials. Unpublished Master's thesis, University of Pittsburgh, 1972.

Light J. A., & Reynolds, L. J. Debugging product and testing errors. In T. M. Schwen (Ed.), *Four views of formative evaluation in instructional development.* Bloomington, Indiana: School of Education, Indiana University, 1972. Pp. 45-78.

Platt, J. R. Strong inference. *Science,* 1964, *146,* No. 3642, 347-353.