DOCUMENT RESUME

ED 064 327 TM 001 492

AUTHOR Abedor, Allan J.
TITLE Development and Validation of a Model for Formative Evaluation of Self-Instructional Multi-Media Learning Systems.
PUB DATE Apr 72
NOTE 45p.; Paper presented at the annual meeting of the American Educational Research Association (Chicago, Illinois, April 1972)

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Achievement Gains; *Autoinstructional Aids; Control Groups; Criterion Referenced Tests; Data Collection; Evaluation Methods; Experimental Groups; Field Studies; *Flowcharts; *Formative Evaluation; Lesson Plans; Material Development; *Models; *Multimedia Instruction; Questionnaires; Research Methodology; Statistical Analysis; Student Attitudes; Tables (Data); Testing
IDENTIFIERS MK II Model; MK I Model

ABSTRACT

The purpose of this research was to develop and field test a new flowchart model prescribing specific operations for tryout and revision of prototype multi-media self-instructional treatments. The methods and results of the following four phases of the study are described: (1) design of a MK I flowchart model based on a review of the literature; (2) assessment of the MK I model based on interviews with seven developers of multi-media instructional treatments; (3) development of a MK II model featuring a small group tryout and debriefing procedure as the main method of identifying instructional problems and developing appropriate revisions; and (4) empirical test of the revised (MK II) model involving three separate experimental-control group comparisons. Conclusions of the study are that the MK II model is highly valid in terms of both identification and remediation of major instructional problems in prototype multi-media lessons, and that it is effective in terms of facilitating statistically significant differences in student achievement and attitudes favoring the revised versions of multi-media lessons. An appendix is a copy of the Student Reactionnaire. (Author/DB)

DEVELOPMENT AND VALIDATION OF A MODEL FOR FORMATIVE EVALUATION OF SELF-INSTRUCTIONAL MULTI-MEDIA LEARNING SYSTEMS

Prepared for:

1972 Annual Meeting of the
American Education Research Association

Division C Session 30.3
Instruction and Learning

by:

Allan J. Abedor, Ph.D.
Assistant Director
Educational Development Program
Michigan State University

April 7, 1972

DEVELOPMENT AND VALIDATION OF A MODEL FOR FORMATIVE EVALUATION OF SELF-INSTRUCTIONAL MULTI-MEDIA LEARNING SYSTEMS

RATIONALE FOR THE STUDY

Formative evaluation may be conceptualized as the process wherein developers of prototype instructional systems collect and analyze information for purposes of correcting system deficiencies. To operationally define this concept, techniques must be available which answer three types of questions: (1) how to identify major discrepancies in the prototype via data collection; (2) how to analyze these data and develop revision hypotheses; and (3) how to design, integrate and evaluate the revisions.

Techniques for answering these three questions are likely to differ as a function of the stimulus complexity of the prototype instructional treatment. That is, specific procedures for formative evaluation of a relatively simple stimulus configuration, such as a programed text, are likely to differ from procedures necessary for tryout and revision of a more complex instructional treatment consisting, for example, of 35mm slides, audio tape, a workbook, a film, and actual laboratory equipment.

The rationale underlying the present study was the conviction that available models and techniques for formative evaluation were inappropriate for tryout and revision of instructional treatments of greater scope and complexity than simple programed texts. Theoretic models of instructional systems development such as Barson (1965), Smith (1966), Hamreus (1968), Paulson (1969), or Briggs (1970), or models of programed text development such as Markle (1967), Kaufman (1964) or Deterline (1967) invariably mention formative evaluation (e.g., tryout and revision) as an integral part of the development process. However, the formative evaluation procedures in such models are either too general for direct application to complex multi-media instructional treatments--or, if specific, provide techniques applicable to simple stimulus configurations such as textual programed instruction. The result is that, with few exceptions, the formative evaluation component of the instructional development process tends to be ignored by all but the most sophisticated practitioners when dealing with complex multi-media treatments. New tryout and revision procedures are needed to enable developers of prototype multi-media systems to more effectively identify problems, develop revision hypotheses, and design, integrate, and evaluate revisions.

OBJECTIVES OF THE INQUIRY

The purpose of this research was to develop and field test a new flowchart model prescribing specific operations for tryout and revision of prototype multi-media self-instructional treatments. [A multi-media self-instructional treatment was defined as one in which: (1) instructional stimuli were presented by means of 35mm slides, audio tapes, student workbooks, and direct interaction with specimens or equipment; (2) rate of presentation was controlled by the student; and (3) interaction with a human instructor was not necessary for learners to achieve the lesson objectives.] Specifically, the study was to develop techniques for: (a) identification of instructional deficiencies through data collection; (b) analysis of these problems leading to revision hypotheses; and (c) design, integration, and evaluation of revisions.

METHOD OF INQUIRY

The present study was conducted in four phases: (1) design of a MK I flowchart model based on a review of the literature; (2) assessment of the MK I model based on interviews with seven developers of multi-media instructional treatments; (3) development of a MK II model featuring a small group tryout and debriefing procedure as the main method of identifying instructional problems and developing appropriate revisions; and (4) empirical test of the revised (MK II) model involving three separate experimental-control group comparisons. The remainder of this paper will describe the methods and results from each phase of the study.

PHASE I: DESIGN OF THE MK I MODEL OF FORMATIVE EVALUATION

The initial model (MK I version) was developed by reviewing a number of research studies and theoretic papers relating to tryout and revision procedures. Emerging from this review of the literature was recognition that three general models of formative evaluation were in current use: (1) tutorial; (2) large group; and (3) a combination of tutorial and large group. The tutorial model requires the prototype designer, or a surrogate, to observe one student using the new treatment and provide tutorial instruction whenever needed. The tutorial instruction may then be incorporated into a revised version and the process repeated until a number of students are able to achieve the criterion without tutorial assistance. Unfortunately, the tutorial model is time consuming and subject to the idiosyncratic responses of individual learners and tutors. There is the constant worry that revisions based on data from a single learner may be unnecessary or even counterproductive. Therefore, a number of writers suggest that prototype lesson tryouts be conducted with a larger sample, more representative of the intended target group. Using this technique, the entire group interacts with the prototype lesson and revisions are based on post-hoc analysis of errors on posttests, intermediate quizzes, or data from an attitudinal survey. While the large group model reduces the problem of idiosyncratic revisions, the lack of face-to-face interaction between learner and tutor makes it very difficult to pinpoint the exact cause of instructional problems. For example, posttest errors can be caused by poor test items, deficient instruction, lack of student entry skills, insufficient practice, or other factors.

Since the post hoc analysis of large group data does not clearly identify the cause of treatment deficiencies, and since tutorial processes are subject to idiosyncratic results, a number of authors advocate an iterative sequence of both tutorial and large group techniques. Using this model, prototype treatments are tried out with individual students on a tutorial basis and revised until the major problems are alleviated. Then, the revised treatments are tried out with large groups and revised again if major discrepancies are revealed. As an additional precaution, several authors recommend a technical review wherein the prototype treatment is reviewed by "technical experts" for errors in content, up-to-dateness, quality of the AV presentation, and appropriateness of the evaluation instruments. Since the technical review, tutorial and large group data all appeared necessary for effective formative evaluation, all three techniques were included in the MK I model (Figure 1).

PHASE II: ASSESSMENT OF THE MK I MODEL

The MK I model was assessed by interviewing seven developers (university and community college faculty from different disciplines) who had previously designed and revised five or more self-instructional multi-media lessons. These interviews consisted of 18 questions which were structured to: (1) assess the willingness of these developers to apply the MK I model to formative evaluation of their own prototype multi-media lessons; and (2) assess the congruence, or lack of it, between the MK I model and formative evaluation procedures actually used by these developers.

Willingness to Use MK I

In general, the developers sampled were unwilling to apply the MK I

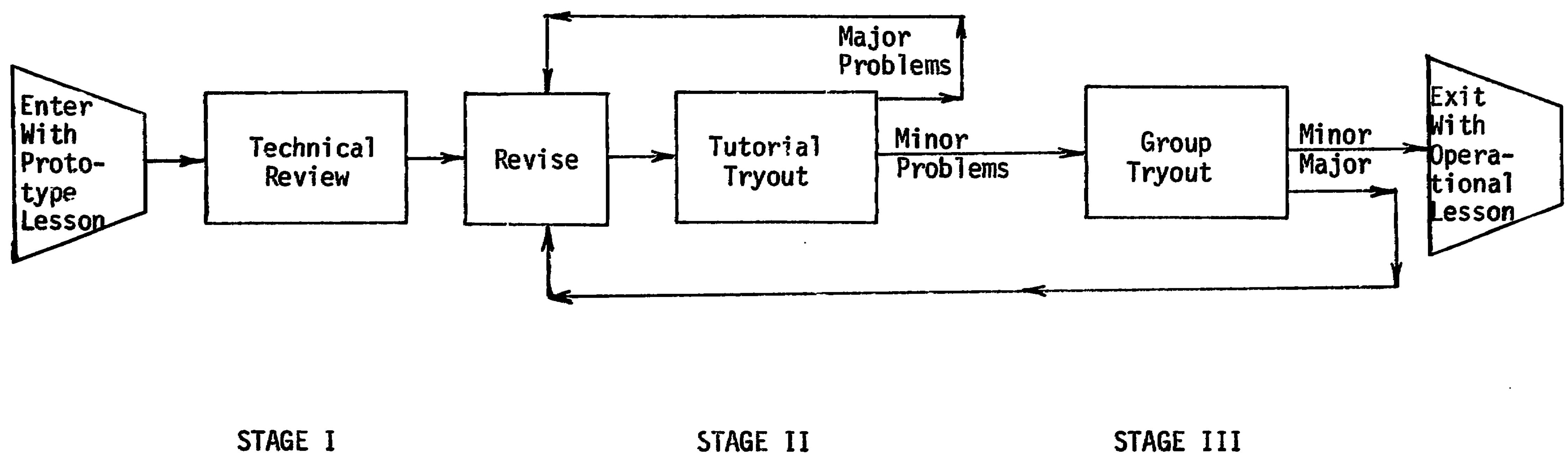

Figure 1. Configuration of the MK I Model of Formative Evaluation.

procedures. In particular, the concept of iterative revisions based on data from "experts," individual students, and then large groups, appeared totally out of the question because of: (a) the time involved; (b) the extremely high cost of producing revisions (both labor and materials); and (c) the difficulty of integrating slides, tapes, workbooks, models, laboratory exercises, directions, etc., and concomitant necessity for reorganization of the whole system when even minor revisions are made. Developers were clearly unwilling to make multiple revisions of the whole set of interrelated instructional stimuli on the basis of feedback from a single student. On the other hand, the prospect of revising on the basis of group feedback seemed more acceptable, but posed logistical and sequencing problems. For example, in many technical areas such as biochemistry, soil science, geography, engineering, and medicine, knowledge is structured hierarchically. Tryouts of new lessons must occur at that point in time when students have acquired the prerequisite skills, but are still naive with respect to the content of the new lesson. The respondents felt that it was simply too difficult to coordinate prototype production, tryouts with large groups, revision development, and course schedules.

The technical assessment component of the MK I model was recognized as potentially valuable but not worth the effort. Most respondents regarded themselves as content experts; hence, additional technical review was felt to be redundant. In addition, most felt they were capable of assessing media and evaluation instrument quality due to previous experience in teaching. Thus, the technical review, the tutorial, and group tryout components were unacceptable to the developers sampled.

Congruence Between MK I and Procedures Used by Respondents

With respect to the congruence between the MK I procedures derived from the literature and those used by the respondents, there appeared to be little similarity. The pattern of formative evaluation activity actually used was as follows. First, the lesson was designed as carefully as possible, then used in prototype form by the intact class under control of the developer. During this initial usage, random feedback was obtained via the developer's personal observations, verbal reports from lab assistants, carrel room attendants, discussion group leaders, and/or students. In some cases, systematic feedback was obtained from end-of-course evaluation of student learning and attitudinal data, or assessment of student achievement and attitudes after each prototype. (Typically, these instruments were of the "how did you like it" variety and were too general to provide specific guidance for the design of revisions.) Nevertheless, data on problems in various prototypes gradually accrued from several sources. Finally, when a critical mass of corroborative data was obtained and if time and resources permitted, revisions were attempted. These revisions were developed on an intuitive basis, often in consultation with GTAs (what should we do about "X?") but seldom, if ever, using the students as a source of design information. The most common revision reported by respondents was a reduction and simplification of subject matter content--a reduction in "coverage"-- which reduced the average instructional time by 10%-25%. (This differed from findings in programed instruction studies where revised programs are often longer than original versions.) It appeared that the major impact on the developer of typical after-the-fact feedback data was a rapprochement between estimated and actual entering student capabilities and a reassessment of objectives and content coverage in the newly developed lessons. Typically, prototypes were too ambitious; so when revisions were made, the net effect was to reduce lesson complexity. Thus, formative evaluation most often caused reformulation of prototype content and objectives as well as minor revisions in programing and/or presentation techniques.

Summary of MK I Assessment

To summarize the assessment of the MK I model, the data showed conclusively that: (1) formative evaluation as practiced by this sample of developers bore little resemblance to formative evaluation as recommended in the literature; and (2) this sample of developers was both unwilling and unable to apply the MK I procedures to their own work because the technical review, tutorial and group feedback techniques were far too time consuming, logistically complex, and costly. In short, the MK I model was of no practical use.

PHASE III: DEVELOPMENT OF THE MK II MODEL

MK II Development Heuristics

Two heuristics guided development of the MK II model. The first heuristic was that multi-media lesson developers need a critical mass of data to convince them that any revision effort is "worth it." Operationally this means that at least two data sources (such as carrel room attendants and student/developer interaction must corroborate the fact that the same problem has been encountered by five to ten students. Moreover, several problems requiring revisions must exist on the same prototype lesson before a commitment to revise will be made. In other words, the delivery vehicle must have several serious discrepancies before it warrants an overhaul.

A second heuristic suggested by assessment of the MK I model was that the concept of iterative revisions was not feasible for developers similar to those interviewed. University and community college faculty developing multi-media lessons do not have the time or resources to permit multiple revisions. Therefore, for a model of formative evaluation to be useful to this type of developer, it should be conceptualized as a "one-shot" effort.

Combining the two heuristics above, it became clear that the MK II model should provide the capability to generate a large amount of corroborative data on instructional problems in a one-shot trial of the prototype lesson.

<u>The Small Group Debriefing Model</u>

After considerable deliberation, it was determined that a more appropriate model for formative evaluation of multi-media lessons was one in which the necessary data were collected by means of a face-to-face interaction or debriefing between the lesson developer and a small group of students. The task of problem identification and design of revisions could thus become a lesson developer/student group responsibility.

<u>Rationale</u>. The rationale for the group debriefing model emerged from the research literature on small groups as problem solving agencies (Schmuck & Schmuck, 1971; Maier, 1963; McGrath & Altman, 1966). This literature seemed to indicate that a number of techniques were available for structuring a small group interaction so that problems may be quickly identified and then made the subject of an organized discussion in which the group assumes responsibility for development of solutions to each problem.

<u>Group Organization</u>. In the present study, it was determined that the prototype lesson developer would be designated the group discussion leader by virtue of his expertise in the subject matter and his responsibility as the instructor in the course. The size of the group was determined largely by research on group processes and logistic considerations. For example, Maier (1963) cited evidence that greatest productivity in problem solving groups is often obtained when the group contains between six and ten participants. Logistically, six to ten students from the target population should be readily available when the opportune time for tryout is reached. The optimal size decided upon was nine students plus the group leader (Lesson author) for a total of ten participants.

The composition of the group was guided by the desire to obtain a sample which represented as nearly as possible the spectrum of abilities in the target population. It was assumed that students of varying abilities would encounter different types of learning problems with prototype lessons and it would be valuable for the developer to be confronted with these problems. Furthermore, it was hoped that by varying the group composition between high and low ability students, the high ability students could assist the developer in determining solutions to problems encountered by themselves and the low ability students. The Scholastic Aptitude Test (SAT) was selected as the measure with which to stratify students into high, medium, and low ability groups. Thus, nine students, three students from each ability group, were the planned composition of the tryout group.

In order to obtain valid information on instructional problems, students would necessarily be selected from the target population for whom the prototype lesson was intended; i.e., the lesson developer's course. Students should possess lesson prerequisites but not score higher than the chance level on the lesson pre-test. Furthermore, to ensure some degree of success in obtaining the desired interaction and feedback, students should possess a positive attitude towards the task of the group. To obtain students with a positive task orientation it was felt necessary to select students from a pool of volunteers within the lesson developer's course.

The Group Process. The process to be followed by the group was as follows. First, students were given an orientation in which the developer communicated his commitment to the principle of "no reprisals" for frank and/or derogatory comments. It was the materials which were on trial, not the students. Next, the students interacted with the prototype lesson materials and completed a lesson pre-test, posttest, and rating scale questionnaire. Students were given tutorial assistance by the developer where required and the interaction tape recorded for later analysis. After lesson completion, the instruments were scored while students took a "break." Those items which 30% or more students missed on the posttest or had given low scores on the rating scale became "agenda" items for the debriefing to follow. The debriefing itself was conducted by the lesson developer so that students not only identified and corroborated their specific problems but suggested revisions to solve these problems.

Summary of the Group Debriefing Process. In sum, the MK II model substituted a group tryout and debriefing technique for the tutorial and large group tryout procedures specified in the MK I model. The group debriefing process is shown in Figure 2 and essentially involves five elements: (a) selection and orientation of nine volunteer students who vary in their entering abilities (SAT scores), (b) individual use of the prototype lesson materials by these volunteers, (c) administration and assessment of learning and attitudinal measures to provide a basis for conducting an organized debriefing, (d) development of a debriefing agenda, and (e) participation in the group debriefing and problem solving interaction. The overt objectives of the group debriefing were twofold: (1) to identify major deficiencies/ instructional problems in the prototype; and (2) to develop feasible solutions to these problems. A covert or "hidden agenda" objective was to provide the lesson developer an opportunity to personally observe the deficiencies in the prototype and thus help overcome the natural reluctance to revise.

The MI II Model

Figure 3 illustrates the final configuration of the MK II model incorporating the technical review and group tryout and debriefing procedures. Technical review was included in the MK II because it appeared that multimedia lesson developers varied considerably in their media design and

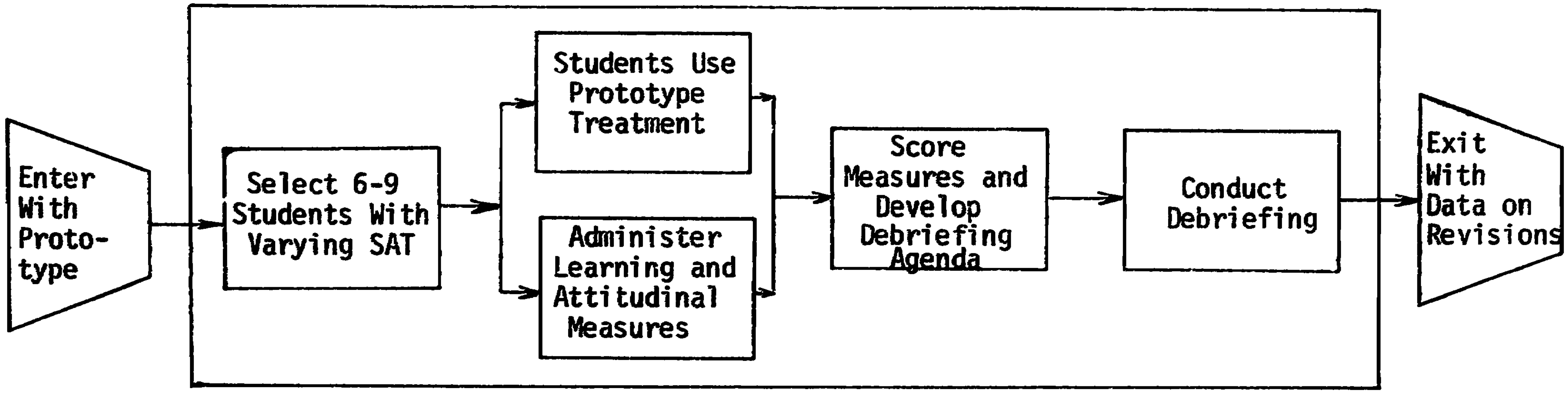

Figure 2. MK II Student Group Tryout and Debriefing Technique.

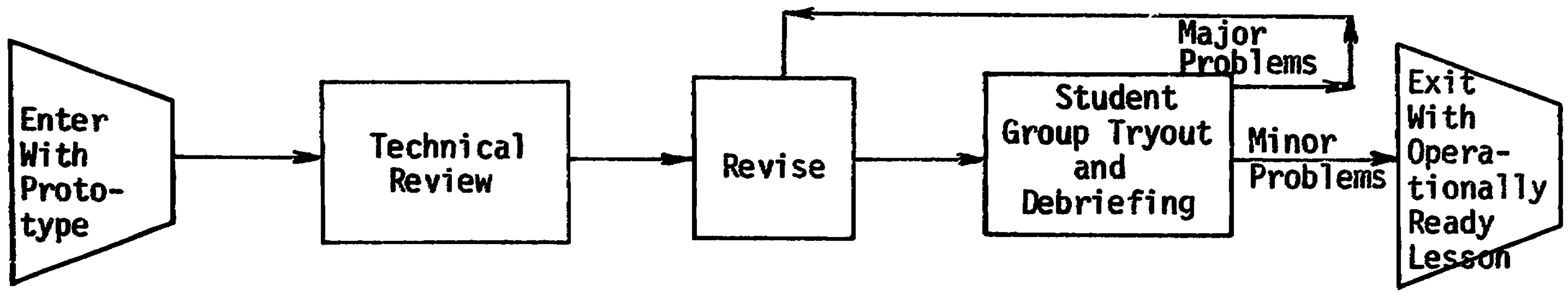

Figure 3. MK II Model of Formative Evaluation.

production skills, their knowledge of and ability to organize subject matter, and in their skill in designing evaluation instruments appropriate to formative evaluation. To preclude students' learning erroneous content, being confronted with illegible or inaudible stimuli and to avoid critical omissions in evaluation instruments, the MK II model stipulates that the lesson developer should obtain feedback from technical experts prior to student tryouts.

PHASE IV: EMPIRICAL TEST OF THE MK II (REVISED) MODEL OF FORMATIVE EVALUATION

Empirical test of the MK II model was conducted in five field experiments involving three Michigan State University faculty and students in their courses. The purpose of the experiments was to determine, insofar as possible, the validity and effectiveness of the MK II model in facilitating formative evaluation of prototype multi-media self-instructional systems. Validity was defined as the degree to which use of the model: (a) enabled the prototype lesson developer to distinguish those sequences of instruction which were unsatisfactory; and (b) resulted in development of revisions which remediated the unsatisfactory instructional sequences. (An unsatisfactory instructional sequence was defined as one in which at least one-third or more of the tryout group indicated dissatisfaction, boredom, confusion, or lack of comprehension.) Effectiveness was defined as the degree to which measures of student achievement and/or attitude on prototype and revised versions showed statistically significant differences favoring the revised versions.

Experimental Procedures and Methodology

Experimental Design. The experimental design used in this study was the before-after control group design (Campbell & Stanley, 1963), in which the control groups (N=12) used the prototype lessons and the experimental groups (N=12) used the revised versions. This design has sometimes been criticized for its use of pre-tests which may be reactive. That is, experimental Ss may become sensitized to the criterion test items and may then be responding to a combination of reminiscence of test items as well as the experimental treatment. In the present study, this sensitization effect was not considered a problem, but, quite the contrary, as an advantage. Pre-test items were regarded as sensitizing Ss to operational definitions of lesson objectives. Sensitization to objectives by means of test-like events may enhance learning (Rothkopf, 1966, 1968) so pre-tests were considered essential and integral parts of both experimental and control group treatments.

Selection of Prototype Lesson Developers. The three prototype lesson developers (A, B, and C) who participated in this study were selected on the basis of availability, willingness as well as:

1. They were currently teaching a course using multi-media lessons which they had personally developed.
2. They had developed a prototype lesson for use in their course which had not previously been used by students or undergone any formative evaluation.
3. They were willing to use volunteer students from their current course to provide feedback on their prototype lessons.
4. They had similar backgrounds and amount of experience in multi-media lesson design, but were from different academic disciplines.

Developer A participated in formative evaluation of three prototype lessons designated A_1, A_2, and A_3. Developers B and C each conducted formative evaluation of one lesson each, designated B_1 and C_1.

Selection of Students. The populations from which students (Ss) were selected were defined as the target populations for which the prototype lessons were intended. Three populations were involved; specifically, the students enrolled in three courses at Michigan State University including:

(1) Animal Husbandry 111 (an introductory course for majors); (2) Education 327M (an introductory course for teachers of secondary school industrial arts, metalworking); and (3) Biology 141 (an introductory course in biology for majors). These courses were taught by the three participating developers.

Sampling procedures treated Ss from each course as essentially different populations due to differences in subject matter content and prerequisite skills involved. Selection of Ss for experimental and control groups was predicated on four criteria: (1) voluntary status, (2) stratification by SAT score, (3) randomization, and (4) Ss would possess prerequisite skills required by the prototype lesson but would be naive with respect to the lesson's terminal objectives.

One week prior to prototype (control group) tryout, developers personally solicited volunteers from their classes. The experiment was described as a learning experience in which all class members would have to participate eventually, but that some volunteers were needed immediately to provide constructive feedback on a prototype version. This feedback would be used by the developer to revise the lesson and hence improve the learning experience for those to follow. Solicitation was successful in that a sufficient number of volunteers were obtained to permit stratification and randomized assignment to treatments.

After obtaining a pool of volunteer Ss from each population, the experimenter (E) obtained Scholastic Aptitude Test (SAT) scores from University records. Volunteers not having SAT scores were dropped from the pool. A schematic of the sampling procedure used for the experimental comparisons is shown in Figure 4.

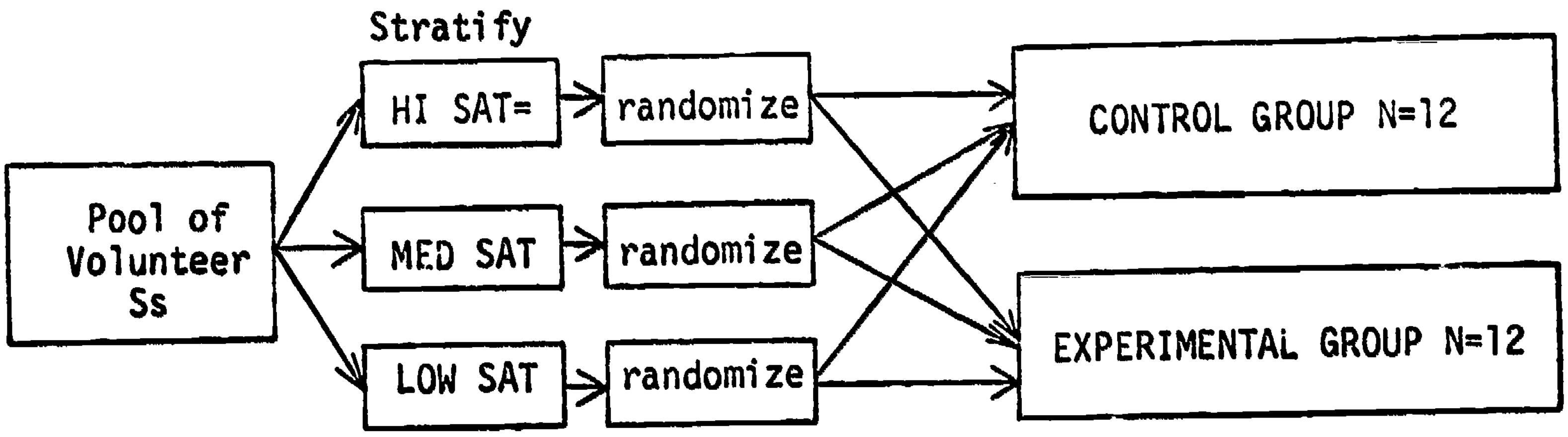

Figure 4. Procedure for Assignment of Ss to Treatments

In all three experimental comparisons, Ss were volunteers from the ongoing course, SAT scores were used as the partitioning variable, equal numbers of Ss from high, medium, and low sub-groups were represented in experimental and control treatments, and pre-experimental equivalence was substantiated by comparison of pre-test scores between experimental and control groups.

Treatments. Three 40 min. multi-media self-instructional prototypes designated A_1, A_2, and A_3, were developed by faculty member A. Faculty members B and C developed one lesson each, designated B_1 and C_1. Each field experiment consisted of the lesson developer conducting a tryout and debriefing on his prototype lesson using control group Ss. Following the control group tryout and debriefing, revisions suggested by the students were incorporated into revised versions (e.g., the experimental treatments). As revised lessons were completed, a second tryout and debriefing was initiated using the experimental group. The purpose of the second tryout and debriefing was to compare the revised version with its prototype counterpart to assess the effect of the revisions on measures of student attitude and achievement. On two trials (A_3 and C_1), however, after the control group tryout and debriefing was completed, the developers concerned felt that the initial prototype was sufficiently effective and did not warrant revision. Hence, in these two cases, (A_3 and C_1) an experimental comparison between prototype and revised versions was not possible. Therefore, the following discussion of treatments and results relate only to experiments A_1, A, and B_1, where experimental/control group comparisons were made.

All control group treatments involved Ss use of unrevised prototype lesson materials which had been reviewed by the experimenter for audio-visual and evaluation instrument quality and reviewed by author peers for content accuracy. Control treatment lessons consisted of pictorial information on 35mm slides and in student workbooks, audio information on a tape recording, printed and pictorial information in the student workbook, pre- and post-tests and a post-instruction attitude survey. In experiments A_1 and A_2, Ss used these self-instructional materials individually in learning carrels. Students thus proceeded at their own rate, controlling number of repetitions of slides and tapes and response rate in their workbooks. (Any time they repeated slides or tape, they were asked to note this activity.) Audio information was presented via headphones, and Ss were asked not to interact with one another but to direct any questions to the lesson developer who was available in the carrel room.

In experiment B_1, however, insufficient carrels were available for simultaneous use by individual students. Therefore, out of necessity, a group presentation mode was adopted instead of individual presentations. In the group mode, the lesson developer controlled a single slide projector and tape recorder, stopping or repeating the presentation at the request of any S. Ss' responses were, nevertheless, still recorded individually in their workbooks. Since the whole group was affected whenever a S stopped the presentation by asking a question, B_1 was not as close a simulation of a self-instructional environment as A_1 and A_2.

Each experimental treatment consisted of Ss using the set of slides, audio tape, workbook and pre-and post-tests which were revised on the basis of feedback from the control group. A_1 and A_2 experimental treatments again used the self-paced carrel mode and B_1 used the group presentation mode. In two cases (A_1 and A_2), the elapsed running time (no playbacks) of the revised versions was reduced 20%; on the other hand, B_1 elapsed time was increased 50% (17 minutes to 26 minutes). All experimental treatments did include a group debriefing session in order to obtain additional experience for the developers concerned and obtain additional data regarding the necessity for further revisions.

Dependent Measures. Four dependent measures were used to assess the effect of the MK II procedures.

1. Group Mean Achievement--An immediate post measure of student achievement of lesson objectives as measured on individual lesson posttests.
2. Gain Score--Mean difference between pre-test and post-test scores. Pre- and post-tests were self-scoring equivalent forms developed specifically for the formative evaluation tryouts by the individual lesson developers.
3. Percentage of Students Achieving "Mastery" --Intended as a criterion referenced measure to determine which treatment enabled a greater number of Ss to achieve a minimum acceptable level of performance, e.g., 80% or more correct on the lesson post-test.
4. Student Attitudes--An immediate post measure of student perceptions of lesson deficiencies and strengths, measured by a 27-item

Likert-type instrument developed by the experimenter (E). (Appendix A).

Feedback from students (control group) showed that numerous items on the prototype achievement measures were faulty. These items were then either deleted completely, or were replaced by new and presumably better items. Therefore, to assess the statistical significance of differences between experimental and control achievement measures, only those items common to both original and revised measures were used.

Results

A summary of the findings on the dependent measures in the three field experiments is shown in Figure 5. In two experiments (A_1 and B_1), significant differences were obtained ($p < .01$) favoring the experimental (revised) version on all four dependent measures. In the third experiment (A_2), a significant difference ($p < .05$) favoring the revised version was obtained on the post-test measure only.

	POST TEST	GAIN SCORE	PERCENT ACHIEVING 80% CRITERION	STUDENT ATTITUDES
LESSON A_1	$P < .01$	$P < .01$	$P < .05$	$P < .01$
LESSON A_2	$P < .05$	NSD	NSD	NSD
LESSON B_1	$P < .01$	$P < .01$	$P < .01$	$P < .01$

Figure 5. Summary of Findings on Dependent Measures

Post-Test Achievement. Data relative to post-test achievement is shown in Table 1. T tests were used to determine the statistical significance of differences on post-tests. These data clearly show marked improvement in student achievement on post-tests in all three field experiments ($p < .01$). This result was, of course, precisely the hoped-for effect of the tryout and revision efforts. It should be reiterated that only those post-test items

Table 1.--Comparison of Experimental and Control Treatment Post-Test Scores

	Treatment	N	Max. Points	$\bar{X}$	σ	DF	Tabled T	Calc. T	Significance
A_1	Control	12	47	37.17	3.87	22	2.508	2.842	$P < .01$
	Experimental	12	47	42.33	4.62				
A_2	Control	12	40	29.33	5.58	19	1.729	2.071	$P < .05 > .01$
	Experimental	9	40	33.44	6.60				
B_1	Control	7	15	9.86	2.92	13	2.650	3.796	$P < .01$
	Experimental	8	15	14.25	.83				
A_3	Control	12	46	43.22	4.71				
	Experimental	NA →	→	→					
C_1	Control	14	NA →	→					
	Experimental	NA →	→	→					

NOTE: These data based on scores from test items common to both prototype and revised instruments.

common to both experimental and control treatments were used to make the statistical calculations.

The degree of improvement between prototype and revised versions may be partially attributed to the fact that irrelevant information was deleted in revised versions. Furthermore, information which was critical to achieving 80% mastery on the post-test was emphasized by redundancy, voice inflection and embedded criterion test items (equivalent, not identical items) on the revised versions. Essentially the presentation was sharpened, important points highlighted and content delimited to facilitate the desired learning outcomes; e.g., achieving 80% or more on the post-test.

Gain Scores. Data relative to comparison of gain scores is shown in Table 2. Again, T tests were used to determine the statistical significance of gain scores between experimental and control treatments. With two of the three comparisons resulting in significant differences ($p < .01$), there remains positive evidence that the model and attendant procedures were capable of identification and remediation of problems in the prototype lessons.

In the case of A_2 where no significant differences occurred, feedback during the debriefing revealed that in setting up the experimental version, the lesson developer had inadvertently used two improperly exposed slides on the post-test. Several students were able to guess the correct answer on the pre-test but became confused and missed the items on the post-test, thus attenuating the gain scores. The important point, however, is that the MK II debriefing process enabled the lesson developer to pinpoint the cause of the problem so remedial action could be taken.

Proportion Achieving 80% Criterion. Data relative to the proportion of students achieving 80% criterion on post-tests is shown on Table 3. The

Table 2 --Comparison of Experimental and Control Treatment Gain Scores

	Treatment	N	Max. Points	$\bar{X}$ Pre	$\bar{X}$ Post	$\bar{X}$ Gain	DF	Tabled T	Calc. T	Signif.
A_1	Control	12	47	23.33	37.17	13.83	22	2.508	2.711	P<.01
	Experimental	12	47	21.25	42.33	21.08				
A_2	Control	12	40	14.75	29.33	14.58	19	1.729	1.024	No
	Experimental	9	40	16.44	33.44	17.00				
B_1	Control	7	15	3.72	9.86	6.14	13	2.650	2.701	P<.01
	Experimental	8	15	5.00	14.25	9.25				
A_3	Control	12	50	26.33	43.22	16.89				
	Experimental	NA								
C_1	Control	14	NA							
	Experimental	NA								

NOTE: These data based on scores from items common to both prototype and revised instruments.

Table 3.--Comparison of the Proportion of Students Achieving 80% Criterion on Post-tests Between Experimental and Control Treatments

		No. Ss Failing to Meet Criterion	No. Ss Achieving Criterion	Total Ss	% Ss Achieving Criterion	% Difference Between Experimental and Control	Standard Error	Z Score	Significant
A_1	Control	5	7	12	58.33%	33.27%	.1770	1.8790	Yes $P<.05$ ($P=.0294$) 1 Tail
	Experimental	1	11	12	91.60%				
A_2	Control	5	7	12	58.33%	8.27%	.2144	.3857	No
	Experimental	3	6	9	66.60%				
B_1	Control	4	3	7	42.85%	51.15%	.5715	2.4960	Yes $P<.01$ ($P=.0064$) 1 Tail
	Experimental	0	8	8	100.00%				
A_3	Control	2	7	9	77.70%	NA →			
C_1	Control	NA →							

significance of these differences was computed by determining the standard error of the difference between two uncorrelated proportions, converting this to a z score, and determining the probability of such a z score from the table of the normal curve (Edwards, 1950).

In two cases (A_1 and B_1), the percentage of students who achieved the 80% criterion during the experimental treatment was significant ($p < .05$ and $p < .01$). This reflects a remediation of both organizational and content emphasis problems as well as elimination of poor evaluation items. The improved student performance in B_1 was remarkable in that 100% achieved criterion in 47 minutes instructional time, as opposed to 42.85% at criterion after one and one-half hours of group discussion and instruction during the prototype tryout. (This lesson had been completely reorganized to closely follow suggestions given by students at the prototype debriefing.)

The exceptional case again was lesson A_2 which only showed 8.27% improvement in percentage of students achieving the 80% criterion. Part of this relatively poor showing could be attributed to students' confusion on the post-test items due to the improperly exposed slides mentioned earlier. Another problem with this lesson may have been transfer of training combined with satiation. Students were expected to learn a number of complex anatomical discriminations based primarily on line drawings in their workbooks. Yet they were post-tested on these concepts using 35mm color photographs of animal carcasses. Since they had been given insufficient practice in making these discriminations on 35mm color photographs, many were unable to perform this task satisfactorily on the post-test. Furthermore, there was a satiation or fatigue factor operating. Many students complained that they had seen so many animal carcasses in the lesson that they all began to look alike; hence on the post-test they just "gave up." Again, the interesting phenomenon regarding lesson A_2 was that the MK II procedures successfully provided insight into why the data showed no significant difference.

Included in Table 3 is the percentage of students achieving criterion for SLATE A_3. It can be seen that 77.7% did achieve criterion when using the prototype; hence the lesson developer felt justified in not making any further revisions.

Attitudinal Data. Data relative to differences on post instruction attitudes is shown on Table 4. T tests were used to determine statistical significance. Again, two of the three lessons showed significant differences in the mean scores on the post instruction attitudinal survey instrument. (Appendix A). Of particular note was lesson B_1, which showed the greatest change in attitudinal scores of all three experiments ($\bar{X}_C$ =88.85; $\bar{X}_E$=112.0). The relatively low initial score on B_1 could be attributed to a number of factors, primarily lack of lesson preparation, organization, and technical problems which caused undue student frustration. The revised version, however, was precisely organized and thoroughly reviewed to avoid technical problems.

The deviant again was lesson A_2, which showed very little difference in student attitudes between experimental and control versions. Note, however, that the initial rating of the prototype was unusually high (105.00). This rating approached the rating achieved by this developer in the revised version of lesson A_1 ($\bar{X}$=106.58). A mean score of 105.00 could thus be interpreted to mean that students were generally pleased with the presentation.

While the overall attitudinal rating of 105.00 was unusually high for a prototype, student achievement on this lesson was unspectacular (66% achieved criterion). The revision hypothesis drawn from these data was that the lesson instruction per se was satisfactory, but the pre- and post-tests needed revision. This hypothesis was corroborated by students during the prototype debriefing.

Table 4.--Comparison of Experimental and Control Treatment Mean Attitudinal Scores

	Treatment	N	Max. Points	$\bar{X}$	σ	DF	Tabled T Ratio	Calc. T Ratio	Signif.
A_1	Control	12	135	95.17	12.20	22	2.508	2.539	$P < .01$
	Experimental	12	135	106.58	8.28				
A_2	Control	12	135	105.00	9.82	19	2.539	.496	No
	Experimental	9	135	107.22	9.29				
B_1	Control	7	135	88.86	8.72	14	2.650	4.101	$P < .01$
	Experimental	8	135	112.00	11.26				
A_3	Control	12	135	103.44	10.31				
	Experimental	NA →							
C_1	Control	14	135	95.64	9.64				
	Experimental	NA →							

NOTE: 27 item 1-5 rating scale used.

Data from all trials (A_1, A_2, A_3, B_1, and C_1) indicated that when the particular attitude survey used in this study showed a mean score above 100.00, the lesson was approaching operational readiness. This heuristic was based on observations of eight tryout and debriefing sessions where this instrument was administered. Typically, when the instrument scores were over 100.00, the debriefings were not nearly as interactive or critical of the lesson as when scores were lower.

Discussion of Student Tryout and Debriefing Sessions

Quite a large amount of observational or "naturalistic" data emerged from the debriefing process which was not necessarily reflected in the objective measures reported earlier. The purpose of this section is, therefore, to discuss both the objective and observational data obtained in this study.

With respect to the objective data, it appeared that use of the MK II procedures led to development of revised lessons which were more effective than prototype versions. In three separate field experiments, statistically significant differences were obtained on nine out of twelve dependent measures. Since the MK II model prescribed the process or pattern of activities leading to identification and remediation of deficiencies in prototype lessons, and since the data strongly favored the revised versions, it is reasonable to infer that under conditions similar to those in the three field trials in this study that the MK II model could be an effective tool in conducting formative evaluation of multi-media self-instructional lessons. Furthermore, one might infer from the relative success of the revised versions that students groups, organized within the framework of the MK II model, were able to (1) identify major deficiencies in prototype lessons and (2) suggest effective revision hypotheses for most such deficiencies.

With respect to the naturalistic data, several observations may be made.

First, because of the interaction of social and psychological variables over which the model has no control, the overall effectiveness of MK II procedures is likely to vary from situation to situation. For example, in the present study three lesson developers agreed to use the MK II model to revise their prototype lessons. In actuality, only two developers did so. The precise reasons for this are unknown, but several clusters of variables appeared to influence this decision. First, was the personality and motivation of the lesson developer: specifically, how committed was he to the principle of tryout and revision, and how much criticism was he willing to endure in pursuit of this principle? In the case of developer A, he was able, repeatedly, to handle a number of derogatory comments and still not become defensive enough to impede the debriefing or to abandon the whole idea of revision based on student feedback. In the case of developer B, however, the prototype lesson was so ineffective and the derogatory comments of students so devastating that by the end of the debriefing he was simply unwilling to continue the process for the seemingly ungrateful students. Several months elapsed before developer B was willing to continue the developmental work. In the case of developer C, he appeared unwilling or unable to handle the relatively large number of derogatory comments. Thus, he closed off discussion prematurely and refused to revise the prototype lesson.

In the present study, another dynamic factor was observed to operate, namely, the perceived quality of the multi-media lesson (either prototype or revised). The lesson functions essentially as a common experiential referent for both students and developer. If the lesson seems disorganized, frustrating, and/or boring for the students, they rapidly became hostile, derogatory, and vehement in their comments. Furthermore, the groups appeared to develop a "momentum" phenomenon. If they got started on a derogatory theme, they kept going and the comments became increasingly derogatory until the author was forced to become defensive and begin justifying the lesson rather than exploring ways to improve it.

On a more positive note, it was observed that MK II procedures may possibly serve an instructional design training function which may result in improved quality of subsequent prototypes. During the present study, developer A finished three prototype lessons and revised two of them. The third lesson was not revised because on the first tryout, students met the established 80% level of performance on post-tests and showed no major attitudinal problems on the attitudinal survey instrument ($\bar{X}$=105.0). In contrast, developer A's first prototype lesson was the least effective. It has the lowest percentage achieving criterion, the lowest gain score, the lowest attitudinal rating, and the most vehement student debriefing. The second prototype lesson fell in between the first and third with respect to scores on measures of learning and attitude and attitude intensity of student debriefing. Since these lessons were developed sequentially within a two and one-half month time period, it was possible that a major variable influencing subsequent lesson design was student feedback obtained through use of the MK II procedures.

It appeared that in developing lessons A_1, A_2, and A_3, developer A learned not to make the same mistakes twice. For example, when students criticized poor exemplars, misemphasis of content, lack of practice in making discriminations, or use of line drawings where a color photograph was needed, developer A seemed able to remember these criticisms and not make similar mistakes on subsequent designs.

It should be pointed out that previous to this study, developer A had designed ten multi-media lessons which were currently used in his course.

These ten lessons were largely in prototype configuration since developer A had not previously obtained systematic feedback from students regarding instructional problems. It seemed reasonable to assume that prototype lesson A_1, his first lesson in this study, was similar in quality to his ten previous lessons. If this assumption was valid, it seemed fair to infer that some of the improvement in his design ability on A_2 and A_3 could be a result of internalizing design principles obtained through formative evaluation feedback.

Another observation which may be made is that MK II procedures may promote a serendipity effect in which spontaneous feedback from students may lead to: (1) revision of a larger instructional system than the multi-media lesson; and (2) improved student/faculty interpersonal relationships. While no formal attempt was made to gather data relative to the larger instructional system (e.g., curricular goals, perceived value of course content, or sequencing of course content) in two field experiments (A_1 and A_2), these types of data spontaneously emerged during the debriefings. In these debriefings, students continually questioned the relevancy of the content and suggested changes in sequence. This unsolicited feedback, having been strongly reiterated in consecutive debriefings, suddenly triggered in developer A the realization that the students were right--that the course and curricular goals were largely irrelevant to these students' professional and intellectual needs. The fact of the matter was that students were being taught many concepts simply to please faculty colleagues. Developer A subsequently revised his course objectives and sequence and advocated revision of the departmental curriculum. Thus through a series of fortuitous events, a much larger instructional system than the multi-media lesson was revised. Moreover, it was clear from comments offered by many students as well as developers A and B that the group debriefing was an excellent vehicle with which to become personally acquainted and promote much improved student-faculty relationships.

A final observation was that the group debriefing (face-to-face interaction) provided powerful, naturalistic data on lesson discrepancies but may have a traumatic effect on some lesson developers. For example, prior to the tryout of their prototype lesson, each of the three developers participating in this study were skeptical as to whether the group tryout procedure would be valuable. They doubted whether the nature of the information obtained would be worth their investment of time. At the conclusion of the first debriefing, however, each developer indicated that there was no question that the nature of the information was extremely valuable in terms of revising the prototype but that the experience had been somewhat traumatic. For example, when a student told a developer face-to-face such things as: "The lesson objectives were not clear" or "The lesson content emphasized one thing while the exam emphasized another"--the developers found this feedback uncomfortable but honest. Then, as additional students corroborated the point being made, the cumulative effect began to make an enormous impact on the developer. One might say, the developer began to "really believe" after a number of students told him the same thing. One could not argue with the students or somehow ignore the discrepancies which were discussed. These discrepancies were very real to the students and they became, through the interaction, very real to the lesson developer.

As the discrepancies gradually unfolded, the developer began to recognize the magnitude of his errors and a sense of frustration emerged. As the students proposed solutions to these discrepancies, the developer (who must do the work to change the unit) saw himself rapidly becoming inundated with more work, whereas he thought he was through. The net result was that a great deal of valuable data were produced by means of a somewhat traumatic experience. The degree of trauma may be a function of: (1) the developer's tolerance to criticism; (2) how ineffective the prototype was, i.e., how critical were the students; and (3) the developer's previous experience with leading problem solving groups.

In short, the nature of the group debriefing interaction was intense and frank. Developers using this technique for the first time are likely to find the data extremely valuable, but may find the overall experience traumatic. As additional experience in handling the group is obtained, however, the traumatic element seems to diminish as desensitization takes place.

CONCLUSIONS

The purpose of the experimental phase of the present study was to determine, insofar as possible, the validity and effectiveness of the MK II model in facilitating formative evaluation prototype of multi-media self-instructional systems. Validity was defined as the degree to which the MK II process enabled the prototype lesson developer to: (1) identify those sequences of instruction causing major instructional problems; and (2) develop revised instructional sequences which remediated the major instructional problems. Effectiveness was defined as the degree to which measures of student achievement and/or attitudes showed significant differences favoring the revised versions.

Validity

It was concluded that in the present study, the MK II model was highly valid in terms of both identification and remediation of major instructional problems in prototype multi-media lessons. In all three field experiments, the developer/student group debriefing process enabled lesson developers to positively identify the major instructional problems (including deficiencies in lesson objectives and evaluation instruments), determine the cause(s) of the problems, and develop revised instructional sequences which remediated these major problems. It may be inferred that under conditions similar to those in the three field experiments in the present study, the MK II model is likely to be a highly valid means of identification and remediation of major instructional deficiencies in prototype lessons.

Effectiveness

Based on the evidence from three field experiments conducted in the present study, it was concluded that the MK II model was effective in terms of facilitating statistically significant differences in student achievement and attitudes favoring the revised versions of multi-media lessons. In two of the three field experiments, data showed statistically significant differences ($p < .01$) favoring the revised versions on all four dependent measures (post test, gain score, percent achieving 80% criterion, and student attitudes). In the third experiment, significant differences ($p < .05$) were noted on post test achievement only. Thus, statistically significant differences were obtained on nine out of twelve dependent measures in three separate field experiments. It may be inferred that under conditions similar to those in the present study, the MK II model is likely to be effective; e.g., capable of facilitating statistically significant improvements in student achievement and attitudes through revised instructional sequences.

SOME OBSERVATIONS REGARDING FORMATIVE EVALUATION

As a consequence of conducting the present research, several observations may be made concerning the general process of formative evaluation of multi-media self-instructional treatments.

OBSERVATION 1: DEVELOPERS ARE UNWILLING TO REVISE PROTOTYPES UNLESS A LARGE AMOUNT OF DATA CORROBORATES THE EXISTENCE OF SEVERAL MAJOR INSTRUCTIONAL PROBLEMS.

The developers interviewed during assessment of the MK I model and the developers who conducted actual tryout and revisions in the experimental phase of this study were extremely reluctant to change their prototype lessons. This reluctance to revise was due, in large part, to: (1) the time, energy and dollars already expended to produce the prototype; (2) the expectation (often unjustified) that because considerably more time and effort were spent in lesson development that the lessons were therefore superior and needed no revision; (3) the demand for additional time, effort, and dollars required for revision development; and (4) the highly interdependent nature of multi-media instructional stimuli; where,for example, revision of pictorial stimuli normally necessitates a revision of audio and print stimuli as well.

Given the high reluctance to revise prototypes, formative evaluation techniques must generate a large amount of data which corroborates the existence of several major deficiencies before a revision effort will be undertaken. Thus, formative evaluation must generate a "critical mass" of data to convince prototype developers that revision is cost effective, or worth the effort.

OBSERVATION 2: A FORMATIVE EVALUATION PROCESS BASED ON MULTIPLE (ITERATIVE) REVISIONS IS NOT USEFUL TO SMALL SCALE DEVELOPMENT PROJECTS BECAUSE OF THE LIMITED RESOURCES AVAILABLE.

Most of the developers involved in the present study were teaching faculty at universities or community colleges who were developing multi-media lessons for use in their own courses. In most cases, these developers were severly pressed for resources (time and money) to develop prototype lessons, consequently development of revisions to prototypes (formative evaluation) was regarded as a low priority task in the overall development process. Given the relatively low priority of formative evaluation, it is not surprizing that developers such as these simply rejected the notion of multiple iterative revisions as being totally out of the question. On the other hand, professional program development agencies such as regional laboratories, R & D centers, etc., having more resources available and a potentially larger target population against which specific program development costs may be amortized, appear more likely to be able to implement the concept of iterative revisions. For small scale development projects, however, a "one-shot" formative evaluation process appears to be the most practical strategy.

OBSERVATION 3: A "ONE SHOT" DEBRIEFING MODEL OF FORMATIVE EVALUATION IS APPROPRIATE: (1) TO IDENTIFY AND REMEDIATE THE MAJOR INSTRUCTIONAL PROBLEMS; AND (2) AFTER TECHNICAL ASSESSMENT IS COMPLETED.

The "one-shot" debriefing model of formative evaluation seems best suited for identification and remediation of the major descrepancies in prototype lessons. This is because the highly interactive and unstructured nature of the group debriefing usually produced information overload during the debriefings; so nuances of instructional problems wer- lost and only the major, or gross deficiencies were thoroughly conceptualized by the lesson developer.

However, technical assessment (review by content, AV and evaluation experts) appears prerequisite to identification of major instructional problems for two reasons. First, if the unit being evaluated appears sloppily put together, unorganized, or technically poor, the student comments are likely to become so derogatory that the developer will become humiliated and wish to terminate the whole formative evaluation effort. Second, if technical descrepancies are too numerous, students will tend to focus their comments on the rather obvious technical flaws and will ignore the more substantive but subtle descrepancies.

Thus it appears that when employing a "one-shot" debriefing model of formative evaluation, the prototype lessons, like a prototype aircraft, must be as carefully engineered and executed as humanly possible--preferably achieving some minimal level of sophistication prior to student tryouts. "Sophistication" in this sense means attention to technical details, organization, and continuity of the presentation. In short, one should not use the student debriefing model of formative evaluation until technical assessment is complete and pedagogical, technical, or organizational details have been completely worked out.

OBSERVATION 4: STUDENTS ARE CAPABLE OF NOT ONLY IDENTIFYING DEFICIENT INSTRUCTIONAL SEQUENCES, BUT OF PROVIDING INSIGHTFUL AND PRACTICAL DESIGN INFORMATION TO REMEDIATE THE DEFICIENCIES.

In the three cases in the present study where prototype lessons were revised, the students both clearly identified and provided strategic level solutions to major instructional problems. For example, in lesson A_1, students suggested a major reorganization, a major change of emphasis, new objectives which clarified what was to be learned, and suggested a new presentation sequence to present the information in what they perceived to be a more logical sequence. With reference to lesson A_2, students suggested the deletion of a large amount of extraneous information which was hindering their learning of important content and suggested major changes in the slides and student workbook. as well as revision of a number of pre and post-test items. In the case of lesson B_1, students suggested a simple analogy which provided an organizing structure to relate a number of disparate and confusing concepts. In short, the student groups provided unique and insightful solutions to their own learning problems--a skill which lesson developers were usually unable to demonstrate because of their more sophisticated conceptualization of the subject matter.

OBSERVATION 5: FACE-TO-FACE FEEDBACK FROM STUDENTS CAN BE THREATENING, EVEN DEVASTATING TO DEVELOPERS; THEREFORE, FORMATIVE EVALUATION USING A FACE-TO-FACE DEBRIEFING PROCESS IS MOST EFFECTIVE WITH DEVELOPERS WHO HAVE A HIGH TOLERANCE FOR CRITICISM

Students were blissfully unaware of the enormous effort required to develop a prototype multi-media lesson. Consequently, when asked to criticize the product, they did so quite willingly if they perceived the developer was genuinely interested and there would be no reprisals for telling it "the way it is." In providing their feedback, students were brutally frank, which meant the developer had to listen while his product was critically dissected by a panel of judges. To maximize the interaction, it was necessary for the developer to try and understand why the students encountered their problems rather than defend the unit. This was a difficult task unless the developer made a conscious attempt to separate himself, as it were, from the fruits of his labor and accepted the criticism as it came. Developers who did not have a high tolerance for criticism tended to defend the unit rather than understand why the students had their problems; hence they were often unable to remediate the difficulty. It may be possible to desensitize developers by allowing them to hear a tape recording of another developer's debriefing session.

OBSERVATION 6: IT IS VITAL TO OBTAIN STUDENT CONSENSUS ON: (1) WHETHER A GIVEN INSTRUCTIONAL SEQUENCE WARRANTS REVISION; AND (2) WHAT SPECIFIC REVISION ALTERNATIVES SHOULD BE IMPLEMENTED

In their role of providing feedback to the lesson designer, students often disagree as to whether a particular instructional sequence warrants revision, and/or how to revise a deficient sequence. Students sometimes would state a concern with some segment of instruction or point out some specific problem which could easily be the result of an idiosyncratic response to the instruction. To maximize the probability that a given sequence does, in fact, warrant revision, or that a particular suggestion for revision will be effective, a student consensus decision model should be used. Thus, for example, when a particular descrepancy is under discussion, the group leader can quickly determine student agreement on the seriousness of the problem by asking"How many agree that 'X' should be revised?" After a consensus is reached on whether to revise 'X,' then consensus can again be used to select the best revision alternative from among those which emerge during the discussion.

OBSERVATION 7: THE FOLLOWING GROUND RULES WILL FACILITATE INTERACTION BETWEEN STUDENTS AND LESSON DEVELOPER.

(A) THE TONE OF THE DEBRIEFING MUST BE KEPT OPEN, POSITIVE, AND FACTUAL

(B) THE DEBRIEFING MUST BE NON-THREATENING AND NON-INTIMIDATING FOR STUDENTS

(C) ALL STUDENTS MUST BE ENCOURAGED TO MAKE A CONTRIBUTION

(D) THE DEBRIEFING SHOULD BE ORGANIZED AROUND OBJECTIVE DATA

Concluding Remarks

The process developed and tested in this study provides an operational definition of formative evaluation of mulit-media self-instructional systems. This process should enable systematic feedback from students to readily be used as an integral part of the instructional development process, thereby improving the efficiency and/or effectiveness of newly developed instructional treatments.

REFERENCES

Barson, John. Instructional Systems Development: A Demonstration and Evaluation Project. Contract No. OE-5-16-025. East Lansing: Michigan State University, June, 1965.

Briggs, Leslie J. Handbook of Procedures for the Design of Instruction. Pittsburgh: American Institutes for Research, Monograph No. 4, 1970.

Briggs, Leslie J., et.al. Instructional Media: A Procedure for the Design of Multi-Media Instruction, A Critical Review of Research, and Suggestions for Future Research. Pittsburgh: American Institutes for Research, Monograph No. 2, 1967.

Campbell, Donald T., and Stanley, Julian C. "Experimental and Quasi-Experimental Designs for Research on Teaching," Handbook of Research on Teaching, ed. N. L. Gage. Chicago: Rand McNally & Co., 1963.

Deterline, William A. "Practical Problems in Program Production." Programed Instruction. Sixty-sixth Yearbook of the National Society for the Study of Education, Part II. Edited by Phil C. Lange. Chicago: University of Chicago Press, 1967.

Edwards, Allen L. Experimental Design in Psychological Research. New York: Rinehart & Company, Inc., 1950.

Hamreus, D. G. "Prototype Development." National Research Training Institute Manual. Monmouth: Teaching Research Division of the Oregon State System of Higher Education, 1969.

Hamreus, D. G. "The Systems Approach to Instructional Development." The Contribution of Behavioral Science to Instructional Technology. Monmouth: Teaching Research Division of Oregon State System of Higher Education, 1968.

Kaufman, Roger A. "The Systems Approach to Programing." Trends in Programed Instruction. Edited by Gabriel D. Ofiesh and Wesley C. Mierhenry. Washington, D.C.: Department of Audiovisual Instruction, National Education Association and National Society for Programed Instruction, 1964.

Maier, Norman R. F. Problem Solving Discussion and Conferences: Leadership Methods and Skills. New York: McGraw-Hill Book Company, 1963.

McGrath, J. and Altman, I. Small Group Research. New York: Holt, Rinehart, & Winston, 1966.

Markle, Susan M. "Empirical Testing of Programs" (Chapt. V). Programed Instruction. Sixty-sixth Yearbook of the National Society for the Study of Education, Part II. Edited by Phil C. Lange. Chicago: University of Chicago Press, 1967.

Paulson, Casper F. "Evaluation of Instructional Systems." National Research Training Manual. Edited by Jack Crawford. Monmouth: Teaching Research, Division of the Oregon State System of Higher Education, 1969.

Rothkopf, E. Z. "Learning from Written Materials: An Exploration of the Control of Inspection Behavior by Test-Like Events." American Educational Research Journal, Vol. 3, 1966, pp. 241-249.

Rothkopf, E. Z. "Two Scientific Approaches to the Management of Instruction." Learning Research and School Subjects. Edited by R. M. Gagne and W. J. Gephart. Itasca, Illinois: Peacock Publishers, 1968, pp. 107-149.

Schmuck, Richard A., and Schmuck, Patricia A. Group Processes in the Classroom. Dubuque: William C. Brown Co., 1971.

Smith, Robert G. The Design of Instructional Systems. Alexandria: The George Washington University Human Resources Research Office, 1966.

APPENDIX A

STUDENT REACTIONNAIRE

STUDENT REACTIONNAIRE

NAME ______________________________ DATE ______________

LESSON TITLE ____

Please be frank and honest in answering the following questions. Remember, you are our prime source of information regarding what needs to be revised.

KEY: 1 means you strongly agree; 2 means you agree; 3 means you are uncertain; 4 means you disagree; and 5 means you strongly disagree.

1. I had sufficient prerequisites to prepare me for this lesson.	1	2	3	4	5
2. I was often unsure of what, exactly, I was supposed to be learning.	1	2	3	4	5
3. After completing the lesson, I felt that what I learned was either directly applicable to my major interest, or provided important background concepts to me.	1	2	3	4	5
4. Manipulating the equipment, or equipment breakdowns often distracted my attention.	1	2	3	4	5
5. Listening to the tapes and watching the slides became tedious, or boring.	1	2	3	4	5
6. This lesson was very well organized. The concepts were highly related to each other.	1	2	3	4	5
7. A professional speaker (announcer) should be used to make the tapes.	1	2	3	4	5
8. The audio tape moved too fast for me, there was too much information.	1	2	3	4	5
9. There was too much redundancy. I was bored by the repetition of ideas.	1	2	3	4	5
10. There was a lot of irrelevant information in this lesson.	1	2	3	4	5

11. The workbook was excellently designed. I could easily follow the instructions and perform the exercises. 1 2 3 4 5

12. Frequent reference to and use of the workbook was distracting. 1 2 3 4 5

13. Often the tape and slides seemed unrelated to each other. 1 2 3 4 5

14. This lesson had very serious gaps and lacked internal continuity. 1 2 3 4 5

15. The examples used to illustrate main points were excellent. 1 2 3 4 5

16. The vocabulary used contained many unfamiliar words. I often did not understand what was going on. 1 2 3 4 5

17. The pre-test and final exam questions did a good job of testing my knowledge of the main points in the lesson. 1 2 3 4 5

18. The questions during the lesson gave me valuable feedback on how I was doing. 1 2 3 4 5

19. Many of the things I was asked to do, or questions I was asked to answer during the lesson seemed like needless busy work. 1 2 3 4 5

20. At the end of the lesson I was still uncertain about a lot of things and had to guess on many of the final exam questions. 1 2 3 4 5

21. I believe I learned a lot, considering the time spent on this lesson. 1 2 3 4 5

22. I would recommend extensive modifications to the lesson before using it with other students. 1 2 3 4 5

23. For you, what was the most difficult part of the lesson? __________

____

24. What was the easiest part of the lesson? __________________

____

25. What were the three worst things about this lesson? ____________

____

____

____

26. I understood most of the concepts and vocabulary immediately after completing the lesson. ___ ___ ___ ___ ___
1 2 3 4 5

27. I think this whole procedure of trying out new materials with students is a waste of time. ___ ___ ___ ___ ___
1 2 3 4 5

28. I would prefer a textbook or lecture version of this lesson rather than the slide/tape/workbook version. ___ ___ ___ ___ ___
1 2 3 4 5

29. I often needed to go back over a portion of the lesson to fully understand it. ___ ___ ___ ___ ___
1 2 3 4 5

30. After completing the lesson, I was more interested in and/or favorably impressed with the general subject matter than I was before the lesson. ___ ___ ___ ___ ___
1 2 3 4 5

31. Please write below any comments, suggestions, or changes which you believe will improve this lesson. Thank you.