This is to certify that the

dissertation entitled

Computer Conferencing in Formative Evaluation:
The Development and Evaluation of a
New Model of Tryout-Revision

presented by

Paul David Toner

has been accepted towards fulfillment
of the requirements for

Ph.D degree in Educational Systems Development

Major professor

Date Feb. 25, 1983

COMPUTER CONFERENCING IN FORMATIVE EVALUATION:
THE DEVELOPMENT AND EVALUATION OF A
NEW MODEL OF TRYOUT-REVISION

By

Paul David Toner

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Educational Systems Development

1983

ABSTRACT

COMPUTER CONFERENCING IN FORMATIVE EVALUATION:
THE DEVELOPMENT AND EVALUATION OF A
NEW MODEL OF TRYOUT-REVISION

By

Paul David Toner

This study was carried out to obtain at least a limited answer to the question, "is it operationally feasible to use synchronous conferencing (that is, with all participants on-line simultaneously) for obtaining tryout data during the development of self-instructional material?

Two models of tryout-revision were used. One was a modification of a small group model developed by Abedor[1], which included the use of a face-to-face group debriefing of the learners. The other model was similar in all respects to the first, except that the debriefing was carried out through use of a computer conference. CONFER II, a computer conferencing system originally developed at the University of Michigan, was used, and the conference lasted for about 90 minutes.

An instructional television program relating to the functions of an information processing system was used as the prototype. Learners were assigned to one of the two tryout-revision models. This led to

the collection of two sets of prototype tryout data which was used as the basis for the development of two revised versions of the prototype. These versions were called the computer conferencing revision (CC revision) and the small group revision (SG revision). Each revision was tried out with learners comparable to the learners who tried out the prototype. Thus, three treatment groups were compared.

The data showed that although neither revision brought about general significant improvement in learning, improvement was shown favoring the CC revision with respect to learning outcomes relating to one of the program's six instructional objectives. When the two revisions were compared to each other significant results ($p = .009$) were shown for the CC revision. This finding, however, is not very important, in view of the more general findings. Also found was that learners with little or no computer usage experience can learn to participate in a computer conference on the basis of very limited training.

It was recommended that the computer conferencing model be operationalized in the asynchronous mode to make better use of the uniqueness of computer conferencing.

1. Abedor, Allan J. "Second Draft Technology." Viewpoints 48 (July 1972): 45-78.

ACKNOWLEDGMENTS

As with most other researchers, I feel indebted to a host of persons who helped "make it happen." Through my work with Dr. James J. Mullin at the University Center for International Rehabilitation, Michigan State University, I gained considerable experience and developed much interest in computer conferencing. Jim helped me perceive the innovative nature and the great potential of this new communication medium. This work had a strong influence on my choice of a dissertation topic.

To my Doctoral Advisory Committee, I owe my greatest gratitude for their guidance throughout my doctoral program and for helping me transform a general interest in computer conferencing into a finished research study in instructional development. Dr. Erling S. Jorgensen, the Chair, guided me through the various conceptual and operational stages of the study, and was very helpful during the editing and re-writing of the dissertation. Dr. Castelle G. Gentry made many helpful suggestions relating to the operationalization of the study and for explaining the details of the study in the dissertation. Dr. Stephen L. Yelon helped in the research design and in interpreting the research results.

Dr. J. David Lewis, from the Department of Telecommunication, represented my cognate area, mass communication. Unfortunately, Dr. Lewis died in July, 1982. He was an active member of the committee until a few weeks before meeting his death, and gave me many useful suggestions relating to the writing of a training manual used in the study.

Dr. Robert Parnes, the designer of the computer conferencing system, CONFER II, was an invaluable help to me during the organization of the computer conference used in the study. Through our many telephone conversations, Bob provided instant help with several problems. Dr. John Vinsonhaler provided the television program used as the instructional prototype for the study and funds for graphics and photography. Dr. Allan Abedor gave me the benefit of his dissertation research and subsequent experience in developing a new model of instructional prototype tryout-revision. Dr. Kent Gustafson helped me select the specific research topic and focus the study so that it was "do-able." Drs. James Page and James Nord and Mr. Donald Wilkening were most cooperative in providing access to their students for the collection of research data.

Jan Vredevoogd provided great consulting assistance in statistical design and data analysis. Roy Feber was of invaluable consulting assistance in computer programming. Dennis Kutzen was a terrific workmate as the director and co-scriptwriter for the production of the revised versions of the prototype television program.

I owe a very special debt of gratitude to Tena French for her invaluable assistance as the word processing operator. In addition to producing the typed dissertation, Tena was very helpful in detecting many errors and writing inconsistencies and was unusually cooperative in helping me meet my various deadlines.

Finally my brother, Ed, my Aunt Helen, and my sister-in-law, Tommye, provided various "carrots and sticks" through letters and phone calls, which helped me progress through the dissertation work. In addition, Ed provided financial assistance during the final months of dissertation writing, which was of enormous help in finalizing the study.

TABLE OF CONTENTS

LIST OF TABLES

Table		Page
4.7	Multivariate Analysis of Variance for Questionnaire Ratings of Attitude toward Instructional System	
4.8	Multivariate Analysis of Variance for Questionnaire Ratings of Attitude toward the Instructional System, for Comparing the Two Revisions to the Prototype	

LIST OF FIGURES

CHAPTER 1

THE PROBLEM

Introduction

The September, 1981 issue of Performance and Instruction was dedicated to the theme, "Instructional Development: Art, Craft or Science?" In that issue, Reigeluth, et al. (1981) identified three approaches to instructional development; namely, the artistic approach, the empirical approach, and the scientific approach. Various contributors to that issue as well as other writers (e.g. Merrill, 1975 and Reigeluth, et al., 1978) have alluded to these three approaches, though in some instances different labels are used. For example, the empirical approach is sometimes referred to as the behavioral approach and the scientific approach is often referred to as the analytic approach (Merrill, 1975), and the artistic approach is sometimes called the intuitive approach.

In any event, the artistic approach is characterised by the application of intuition and genius. This is the approach that is associated with the subject matter expert who plans a course or creates instructional material for use in the course, without recourse to instructional development expertise, or other knowledge bases.

The empirical approach is characterized by the use of trial and error rather than intuition as the knowledge base for making decisions. The scientific approach is characterized by the application of scientific knowledge for making instructional decisions or "prescriptions," as they are often referred to by proponents of this approach.

Such prescriptions have taken the form of models of instructional development and various procedures, such as task analysis and content analysis (Reigeluth, et al., 1981).

The scientific approach is most frequently used in the design aspects of instructional development at the beginning of a development project, prior to the development of the instructional prototype. This is the area of instructional development where educational psychology has been most frequently applied, and is frequently referred to as "front-end analysis."

However, Reigeluth and his associates (Reigeluth, et al., 1981) point out that the literature indicates that little is known about formative evaluation and conclude, "Therefore, we are left with a piece of the instructional development process that is largely art and trial-and-error" (p. 22). They end their article on the the following somewhat pessimistic but cogent note:

> If instructional science is to attain the precision of other hard sciences, it must produce a procedural and theoretical knowledge base for evaluation so that the scientific process of instructional development can move one step beyond art and trial-and revision.
>
> Until that day comes, we must make the most of what science currently has to offer, filling in the remainder with our individual genius and empirical scrutiny (p. 22).

The writer agrees that formative evaluation, particularly the tryout-revision aspects of it, is not based on scientific principles. However, he would like to suggest that Gropper's comprehensive treatment of this subject in his 1975 book, "Diagnosis and Revision in the Development of Instructional Materials," seems to have been largely ignored as a source that could serve as the basis for the development of diagnostic and revision strategies in tryout-revision. In that publication, Gropper provides developers of instruction with a highly articulated set of categories of types of learning failures, program design failures, student test failures, and student program failures.

He did not, however, provide procedures for making use of these categories, as he mentions in the overview of his book. It was his intent to define the variables to be used, which he hoped would provide a basis for the development of tryout-revision. He expresses this thought, as follows:

> The future of a tryout and revision technology awaits an identification of the types of errors students commit both on programs and on tests and a parallel identification of the types of program weaknesses which are responsible for them. It also awaits the formulation of diagnostic procedures which, by making the needed identifications, can lead to relevant, reliablly implemented revision (p. 9).

Commenting further on the tryout-revision process, he says:

> Revision is, after all, just the use of development procedures at a different point in time. It is diagnosis which must inform as to where revision is needed and particularly what type of revision is needed. <u>The biggest gap in know-how exists in the area of diagnosis. To narrow it, a developer can employ tools and methods which are already available to him</u> (emphasis added) (p. 9).

In Merrill's (1975) review of the book he points out that:

> Gropper has combined his comprehensive analytic approach with the empirical approach to present <u>the first</u> (as far as the reviewer is aware) <u>comprehensive statement of a systematic analytic approach to diagnosis and revision of instructional materials</u>. Those who employ either an empirical or analytical approach to instructional developement and who do not consider some of the excellent prescriptive procedures suggested in this book will be failing to use some of the best tools yet to be introduced

for doing instructional development (emphasis added) (p. 447).

Gropper refers to tryout-revision as "rear-end analysis" as opposed to the term "front-end analysis," which is often used to characterize instructional design. Commenting on the poor state of the art of tryout-revision as being the result of overemphasis on learner error rates and underemphasis on the types of errors committed by students, he makes the epigrammatic statement that this matter "translates into 'Rear-end analysis should be the mirror image of front-end analysis'." (p. 3)

Along this same line of thought, he sets the point of departure for the book in the following context:

> A larger than necessary number of cycles of tryout and revision is inevitable if revision proceeds on an ad hoc basis. Tryout students commit errors on program tasks; clerks tally up error frequencies for each task; and developers patch up the most frequently failed tasks. This patchwork approach generally results in programs which, while often effective, are longer and less efficient than need be. If the revision process is to be more systematic and more analytic, and if the programs which result are to be more efficient, a developer must generate appropriate tryout data in addition to program errors, interpret them analytically, and act on the results in consistent, reproducible ways.
>
> The point of departure for this volume is the assumption that in his search for evidence of program inadequacies, the farther away a developer gets from performance on a program per se, the more productive his investigation is likely to be. Attention to performance on post-instructional tests is a move in the right direction. More potentially promising is attention to the conceptually more distant psychological skills which underlie performance. This does not mean that the developer should ignore performance on program tasks or on post-instructional tests. . . .From such data, however, he can draw inferences about the underlying skills which students have failed to acquire, retain, or transfer. Since correct performance is dependent on these skills, analysis of skill failures is likely to be generative in identifying problems whose effects are more pervasive. . . His identification

of any of these types of underlying failures is, therefore, likely to result in a more informed revision strategy, in a reduced number of tryout cycles, and in instructional programs which remain efficient (emphasis added) (p. 13).

The literature reviewed for this study indicates, however, that no one has "picked up the baton", so to speak, from Gropper and carried it forward, even for a single lap. Dick (1980), after reviewing the tryout-revision literature, concluded that the biggest problem in tryout-revision was that of not knowing what to do during prototype revision. He says:

> . . . the greatest limitation in formative evaluation today. . . [is] the dilemma of what to do after a problem has been detected in instruction. Nearly all instructional design writers have indicated that after the data have been collected and summarized in a formative evaluation, the designer should "revise appropriately." However, in most instances, designers have already used their best knowledge of how to design the instruction, and therefore it is not always apparent what "revising appropriately" would be (p.5).

Point of departure for the study

Perhaps the reason instructional developers don't know what to do during revision is that the wrong type of data has been collected. This possibility is the point of departure for this study. The literature indicates that test data is the most common type of tryout-revision data collected. Moreover, very few instances were found where an attempt was made to find out the reasons test items were answered incorrectly. The details of the literature on the effectiveness of tryout-revision will be given in Chapter 2.

Furthermore, on a broader scale, evaluation is supposed to be oriented to decision making, as will be shown in the review of the evaluation literature reported in Chapter 2. In the case of evaluation relating to the tryout of instructional prototypes, the decision to be made can be thought of as the answer to the general question, "What shall we do to revise the instructional prototype, so that the failures that occurred in the last tryout don't reoccur?" Yet, the collection of test data without process data that point to specific failures does little more than provide possible justification for the implementation of the instructional system as a finished product, in the event that data collected is generally supportive of such a decision. However, if the data does not support such a decision, such data may be almost worthless, since the developer must try to infer from this data what the underlying causes for such failures were. Such inferences, may be little better than wild guesses.

In order to carry forward the ideas of Gropper, and others who may think along similar lines of reasoning, it seems that it is first necessary to have at least one model of tryout-revision which facilitates the collection of data that tell the instructional developer where the source of the learning problems are.

Thus, the first step that must be taken is to assess the known models of tryout revision on the basis of their suitability to identify serious learning problems.

The qualifier, serious, is used here for the simple reason that it is probably impossible to develop an instructional system that is "failure free." Thus, it seems reasonable to attempt to deal with serious failures only.

In the next section the known models of tryout-revision will be evaluated for these purposes, and the difficulty of the operationalization of the term, serious, will be discussed. Then, in the following section the question of the importance of the number and type of tryout learners their location and access to them will be discussed.

Following that point, it will be asserted that a new model of tryout-revision is needed, and design specifications for such a model will be suggested. The relative merits of applying computer conferencing will be addressed with information provided on some of its technical aspects, its general characteristics, and some predictions about its role in society. This will form the basis of the succeeding section, in which the prospects of using computer conferencing in tryout-revision will be considered. This is followed by a description of a computer conferencing model of tryout revision.

In view of the length and complexity of this chapter a summary of the problem will be given next. This presentation will be followed by a statement of the needs that the study will address. Following this statement, the purposes of the study, the research questions and hypotheses and their rationale, the limitations of the study and an overview for the remainder of the chapters will be given.

Review of the Known Models of Tryout-Revision

Overview

The instructional development literature indicates the use of three models of the tryout-revision component of formative evaluation; these are generally referred to as the tutorial model, the large group model, and the small group model. In addition, the tutorial and large group models are often used in tandem, and this could be considered a fourth model. A fifth model is the combined use of the tutorial, small group and large group models. These models will be evaluated according to the criteria indicated in the next subsection. It should be noted that the term, instructional system, used throughout this dissertation refers to any instructional product or process, in which the systems approach was used in the development process.

Criteria for evaluation

As implied in the previous discussion, the major criterion for the adequacy of the known models of tryout revision will be their ability to pinpoint various kinds of failures relating to learning from the use of the instructional prototype being tested. The second criterion will be the relative number of tryout learners that are generally involved in any given cycle during the use of the model in question. It is this second criterion which relates to the seriousness of the failure pinpointed by the data. It is recognized that failures in an instructional prototype which don't cause problems for many learners can be serious for the learners such failures do affect. However, in view of the difficulty in operationalizing a procedure for identifying such failures, it was decided to not include this issue as a criterion in evaluating the known tryout-revision models. In the following subsections, the five models indicated in the previous subsection will be evaluated.

The tutorial model

The tutorial model consists of trying out the instruction on learners one at a time. The interaction between learner and author/developer has been characterized by some writers as being similar to that of a clinical relationship. The author/reviser observes the learner and intervenes whenever the learner seeks help or appears to need help. Corrections are made in the instruction, if possible, so that the learner may continue the learning process. If corrections in the instruction cannot be made immediately, they are made after the tutorial session. At the completion of the learning, the author/reviser may also debrief the learner in order to obtain additional feedback on the instructional system.

The major limitation of the tutorial model is that it is very time consuming and generally utilizes a very small number of tryout learners. As S. Markle (1967), Paulson (1967) and Abedor (1971, 1972) have pointed out, this makes possible the undue influence of the idiosyncratic behavior of a few learners on the revisions made. The major advantage of this model is that much process information (i.e. information about the learner's interaction with the instructional system) is obtained from each learner. Thus, much data can be collected on the failures of the system. But, unless it is tried out on many learners there is no easy way to determine whether these failures are serious failures on the basis of the limited number of learners who experienced a given failure. Of course, there is no limit on the number of learners who can participate in the tryout of the instructional system being developed. But the literature indicates that the time and effort required for each tutorial tryout discourages the use of many tryouts.

The research literature does not indicate that this model is still used alone. This literature mentions its use in combination with other models. Most of the literature describing this model was published between 1963 and 1971.

For examples of the use of this model see D. Markle (1967) Fleming (1963), and Silberman, et al. (1964)

The large group model

The large group model consists of trying out the instructional system with a large number of learners and collecting outcome data (i.e. test results and questionnaire data) from them. The major advantage of this model is that revisions are based on a large body of data from learners. Therefore, the undue influence of idiosyncratic behavior of a few learners, is reduced.

However, the nature of the data is such that the use of such data to pinpoint failures in the instructional prototype is often quite limited. Test results can be used for such purposes only to the extent of pinpointing a specific portion of the prototype's content which is causing learning failures. But this is true only if criterion-referenced testing was used. However, such analysis still does not reveal the specific nature of the failures. For that type of data, process data is needed to specify the failures related to that portion of the prototype's content. But, reports in the literature do not always clearly indicate whether criterion-referenced testing was used. Generally speaking, such testing is not used, on the basis of the literature reviewed for this study.

Though it is usually test data that is collected when the large group model is used, questionnaire data is also collected in using this model. Global questions are usually asked regarding such matters as interest level, and the audio and visual qualities of the instructional materials, among other things. But, such questions are not usually combined with criterion check lists, open ended questions asking for examples of deficiencies found by the learner, and other aids for specifying the meaning of learner responses.

Therefore, the large group model does not, in general, lead to the pinpointing of failures in an instructional prototype.

For examples of the use of this model, see Kandaswamy (1976), Gropper (1967), Gropper and Lumsdain (1961), and Nathenson and Henderson (1980).

The combined tutorial and large group model

This combined model is probably the most commonly used model. The tutorial model is used first to eliminate the major deficiencies. Then, the large group model is used to eliminate minor problems. It thereby combines the advantages of these two models; however, it means two tryout-revision phases, which some author/developers may find objectionable. However, the limitations of the large group model, are not overcome, but some of the limitations of the tutorial model are overcome, by virtue of using more learners. That is, the changes made on the basis of a few learners through use of the tutorial model are subjected to verification through use of the large group model. But the use of the large group model still does not provide the process data required for pinpointing specific types of failure revealed when this model is used in the second testing phase.

For examples of the use of this model, see D. Markle (1967), Anderson (1967), and Short (1968).

The small group model

A study involving the use of the small group model was carried out by Abedor (1971, 1972). His operationalization of the model made use of small tryout groups consisting of 6 - 12 learners. The author/developer observed the learners while using the instructional system, an audio-visual program, and intervened when a learner requested help or appears to need help. This help was in the form of tutorial assistance that enabled the learner to continue the learning process, even though the instructional system was not necessarily changed. The author/developer recorded these incidents on audio tape, and also noted the coincidence between observed non-verbal negative feedback from specific learners and the particular instructional stimuli to which such learners were responding.

Upon completion of the instructional systems, the learners complete a posttest and a questionnaire. The learners were then given a "break," during which time the test papers and questionnaires were scored. Following this, the author/developer and the learners engaged in a group debriefing. The agenda was based on these test and questionnaire items for which 30% of the learners scored incorrectly or rated unfavorably. The purpose of the debriefing was for the identification by the learners of deficiencies in the instructional system and the group's revision strategy recommendations. The author/developer acted as a group discussion facilitator.

One limitation of this model is the relatively small number of tryout learners, as compared to the number of learners that the large group model can accomodate. Abedor (1981) has suggested that 6-8 learners is an ideal number and that 10 learners are too many. These learners are also, presumably generally at the same institution. Though the group may be representative of the population of ultimate learners for the system being tried out, the number of learners who can effectively participate in the group debriefing is so low that there is a high likelihood for estimation error, since each learner represents a relatively large number of learners in the population. Thus, unless the tryout and debriefing is replicated at numerous sites, the use of this model seems limited to the development of instructional systems for local use only. However, it could be combined with the large group model, in a manner similar to that of the combined tutorial and large group model.

One advantage of the small group model is that it combines, to some extent, the advantages of the tutorial and large group models by obtaining both outcome and process data.

Abedor had developed an earlier model during this study which combined the tutorial and large group tryouts. However, the results of a small survey of author/developers carried out during this study indicated that the respondents were strongly opposed to going through the process twice and were not inclined to make revisions in an audio-visual program on the basis of the behavior of only a few learners. He therefore devised the small group model as a way of simplifying this earlier model.

A second advantage of this model is that it provides two new types of tryout data; viz., learner debriefing data on deficiencies in the instructional prototype and on revision strategies for eliminating such deficiencies. What Abedor has contributed here is to apply a group problem solving approach to tryout-revision. The assumption is that a group of learners who have just had a learning experience, including the completion of a posttest and a questionnaire will know, at least intuitively, what aspects of the instructional system are preventing the typical learner from scoring better on the posttest or rating the attitude scales on the questionnaire more positively.

Abedor (1980) feels that it is important to "capture the synergistic effect" that occurs when such a group of learners discuss the deficiencies of an instructional prototype. The job of the debriefer is to facilitate the process so that the learners might develop a consensus on the deficiencies in the instruction and on their recommendations for eliminating these deficiencies.

Other instructional developers have used the small group model in conjuction with the large group model, but Abedor is the only one who has documented a specific operationalization of this model. For example, Nathenson and Henderson (1980) report that group interviews are used at the Open University in Great Britain for pursuing special topics. Unfortunately they don't specify how these group interviews are conducted, though its quite possible that the group interview does not follow any special format.

The literature does not reveal any other use of the small group model used alone, although its use has been known for some time, especially in the development of television programming. For example, Jorgensen (1982) mentioned that he used the small group process during the early 1960's when developing programming for the Midwest Program for Airborne Television in Instruction (MPATI). During such group interviews the group would be encouraged to give feedback on what they thought were deficiencies in the program. If there were specific questions for which feedback was needed, about which there had been no discussion after the first 20 minutes or so, the group would be asked to respond to such questions.

<u>The combined tutorial, small group, and large group model</u>

This model is mentioned by Dick (1977, 1980) and Briggs and Wager (1981) and implied by Baker (1974) in their theoretical writings. The model is the same as the combined tutorial and large group model, except for a small group tryout-revision step between the tutorial tryout-revision and the large group tryout. Dick (1977) suggests that 8 to 24 students be included in the small group tryout. He mentions that after the tryout and testing "it is helpful to hold 'debriefing sessions' in which students are asked to describe their reactions to the learning materials" (p. 316). Although Dick uses the same term, debriefing session, to describe a small group discussion following testing as does Abedor, he does not give any details on how the students are debriefed.

Anderson (1968) is the only author revealed by the literature who has applied this combined model in an empirical study. His use of small groups does not include any type of group process for obtaining group discussion or debriefing data.

The models in a broader perspective

In Figure 1.1, below, a matrix is shown with the models listed in the vertical dimension and the types of tryout data collected shown in the horizontal dimension. As previously indicated, process data consists of data relating to learning difficulties met by the learner while interacting with the instruction; test data is the pretest and posttest data; opinion/attitude data is learner questionnaire and/or interview data relating to opinions about and attitudes toward the instruction; debriefing data consists of the deficiencies in the instruction identified by the learners and their revision strategy recommendations. Each cell of the matrix indicates for a particular model whether the number of learners upon which the tryout data is based is relatively large or small, except where the type of data is not applicable to that model. In such cases, "NA", is indicated, meaning "not applicable."

There is no apparent reason why the tutorial model could not be operationalized to collect opinion/attitude data. But, since the literature does not give clear evidence that this is usually done, a question mark is shown in the appropriate cell of the figure.

It can be seen from Figure 1.1 that the small group model and the tutorial + small group + large group model are the only models that includes the collection of all five types of data.

TRYOUT DATA

MODEL:	(a) Process Data	(b) Test Data	(c) Opinion/ Attitude Data	(d) Debriefing Data Deficiency Identification	(e) Debriefing Data Revision Recommendations
Tutorial	small	small	?	NA	NA
Large Group	NA	large	large	NA	NA
Tutorial + Large Group	small	large	large	NA	NA
Small Group	small	small	small	small	small
Tutorial + Large Group + Small Group	small	large	large	small	small

NA = not applicable
? = uncertain on basis of the literature

FIGURE 1.1
Relative Sizes of Numerical Bases for Types of Tryout Data Collected in Using Each Model of Tryout-Revision On the Basis of the Literature.

It is now possible to evaluate each of the models of tryout-revision, with respect to the criteria set forth at the beginning of this section. This evaluation is presented in Figure 1.2, below.

	(a) Prototype deficiencies pinpointed?	(b) Relative seriousness of deficiencies ascertained?
Model:		
Tutorial	High	Low
Large Group	Low	High
Tutorial+ Large Group	High+Low	Low+High
Small Group	High+	Low+
Tutorial+	High+	Low+
Small Group+	High+	Low+
Large Group	Low	High

Figure 1.2
Evaluation of Tryout-Revision Models, According to Relative Degree to Which They Meet Criteria, as Indicated in the Literature.

In Figure 1.2, the relative terms "high," and "low" are used to represent the degree to which a criterion has been met by a model. These categories are very arbitrary estimates, based on the literature. As can be seen from Figure 1.2, the tutorial and small group models pinpoint deficiencies to a high degree, but they are able to ascertain the relative seriousness of these deficiencies to only a low degree, due to the fact that these models are based on the use of small numbers of learners. The large group, on the other hand, does just the opposite; that is, it has a low ability to pinpoint deficiencies, but due to the large number of learners that it can accomodate it can ascertain the seriousness of those deficiencies to a high degree. Neither of these alternative sets of conditions is desirable as a basis for improving a prototype.

The tutorial + large group model combines the advantages of each of these criteria for the two phases of the model. Thus, as long as it is able to eliminate the major deficiencies during the tutorial phase and verify that to be the case during the large group phase, the model can be used successfully. However, should some major deficiencies not be eliminated during the tutorial phase, the large group is not well designed to pinpoint such deficiencies. Thus, it is rated differently for its two phases. Hence, although this model is very popular it has serious limitations, unless several cycles of tryout-revision are carried out.

The tutorial + small group + large group model is very similar to the tutorial + large group model. Having the extra step provides an additional opportunity to detect deficiencies. But, again, if any major deficiencies have not been eliminated before it reaches the large group phase, the difficulty in detecting them is similar to that of the case of the tutorial + large group model. This model is rated separately for each of its three phases.

As mentioned above, this basis for evaluating these models is very arbitrary, and the writer does not wish to make any great claims for its general utility. However, its value for this study is that its use suggests the need for a better model of tryout-revision. This will be addressed later in this chapter. But first, a discussion of the number and type of tryout learners, their importance, and access to them needs to be discussed.

The Importance of the Number and Type of Learners, Their Location and Access to Them

Though some writers suggest that large numbers of learners are not needed for tryout-revision (e.g. Baker, 1974), it can also be argued that data based on a large number of learners can be more convincing to author/developers than data based on a small number of learners, when making changes in a prototype are being considered. This is what Abedor's (1971, 1972) survey seemed to indicate, as discussed in the previous section. In addition, the statistics are more reliable, mathematically speaking, as the the number of subjects increases. On a practical basis, greater confidence in data based on a larger number of learners may be more germane to the issues involved in the case of non-print instructional media than in the case of other types of instructional system components, due to the costly and time consuming aspects of making such revisions.

On the other hand, there is the practical matter of gaining access to learners for a prototype version of an instructional system. Baker (1974) further suggests that unless one is in a large metropolitan area one can quickly use up the available sources of learners and the cooperation of instructional personnel by insisting on the use of many learners for an instructional system that is only at the prototype level of development. Therefore, she suggests using only a few learners but obtaining as much information as possible from them.

Another aspect of the problem is that of whether one can and/or should rely solely on local sources of learners. If the instructional system is being developed for strictly local use, then there need not be a concern for obtaining tryout data from learners from outside the local environment. However, if the instructional system is being developed for more general use, then the tryout learners should, ideally, represent the population of the intended users of the system, so as to avoid the possibility of basing revisions on the behavior of a sample of learners who do not represent the population of intended learners.

A further consideration is that of using statified random samples for prototype testing. Some developers will argue that such samples are not needed in order to obtain good tryout data. However, the use of such samples can lead to a better basis for deciding on the most appropriate set of revision strategies. But, the use of stratified samples requires a larger number of learners than is otherwise required. Thus, the issue, again, is that of using larger samples of learners.

Even if the system is being developed for local use only, there is still a need for a sufficient body of tryout data as a basis for making revisions in the prototype. For such purposes there is no need to use local learners only. Non-local learners who are similar in all significant respects to local learners can certainly be used and should be used when an insufficient supply of local learners exists.

The main problems with using non-local learners are the logistical difficulties encountered in obtaining tryout data from them. For outcome data there is no problem in using non-local learners since a prototype system can be mailed to tryout sites and the outcome data can be returned by mail from these sites. But, in the case of using the tutorial and/or small group models, there are serious data collection problems.

One strategy would be to send tryout personnel, to the non-local tryout sites to collect data. This strategy, however, could be very time consuming and costly.

A second strategy would be to use tryout personnel who are already located at the tryout sites, such as instructional personal, graduate students and other mature students. The major problem with this solution is that they would have to be trained so as to follow standard procedures. In view of the fact that they would be short-term employees/associates, it would probably be very difficult to control variation in their performance.

In the case of both solutions, the collection of debriefing data through use of the small group model presents another problem; namely, if there is inconsistency between the consensus arrived at by the different groups it would be difficult, if not impossible, to arrive at general conclusions.

Thus, it seems that even if one wants to or needs to make use of non-local learners there are serious practical problems in doing so. Perhaps this is why it isn't done except in some uses of the large group model. A possible solution to this problem will be suggested in the next section.

The Appropriateness of a New Model of Tryout-Revision

Perhaps a review of what has been said so far is in order. First, it was pointed out that Reigeluth and his associates (Reigeluth, et al.,

1981) contended that although the scientific approach to instructional development has been applied to other areas of the instructional development process, the formative evaluation area has, so far, not benefited from this approach. They then suggested that instructional science, to use their term, needs to produce a "procedural and theoretical knowledge base for evaluation" (p. 22).

Next, it was pointed out that Gropper (1975) contends:

> The future of a tryout and revision technology awaits an identification of the types of errors students commit. . . and a parallel identification of the types of program weaknesses which are responsible for them. . . . It also awaits the formulation of diagnostic procedures which. . . can lead to relevant, reliably implemented revision (p. 9).

Furthermore, Gropper has taken the first step by providing the field with a comprehensive taxonomy of categories of student and program failures. As Merrill (1975) points out, Gropper's suggestions imply the combination of the empirical approach and the scientific approach to instructional development. Specifically, he suggests that empirical data can be collected pertaining to various types of student and program failures, using his taxonomy as a guide for such data collection. This, of course, is in line with the traditional scientific practices found in older disciplines, whereby theory guides research and research modifies theory, which then suggests new research (Merton, 1968). Moreover, as Merton suggests, empirical data can suggest gaps in theory. In this case, such gaps could be additional categories of student and program failures.

The separation of the empirical and scientific approaches in instructional development does not appear to be sound, other than to highlight the "trial and error" aspects of the current phase of the empirical approach to instructional development. But, then, new disciplines are often "trial and error" oriented, at first until theories develop, since that's one approach to theory development when little theory exists.

In any event, it seems that in order to implement the ideas of both Gropper and Reigeluth and his associates in the development of a "procedural and theoretical knowledge base" in tryout-revision a tryout-revision model which facilitates the collection of such student and program failure data is needed. For that reason the known models of tryout-revision were reviewed, to determine whether such a model already exists. After reviewing these models, it was concluded that such a model does not exist.

It was also pointed out that another problem in tryout-revision may be that of access to a sufficiently large source of tryout learners. It was suggested that part of this problem concerns access to learners at a distance from the site where the instructional development project is being carried out. This is due to the fact that an inadequte source of tryout learners could seriously affect tryout-revision efforts.

Thus, it seems that the need for a new model of tryout-revision is quite justified. The next question that seems appropriate to ask is, "How should a new model of tryout-revision be designed so that it satisfies both the need for the right kind of data and the need for access to enough learners who may provide instructional developers with the appropriate type of data. The next step, then, seems to be to specify the criteria for a tryout-revision model that would meet such needs. In the next section such a specification will be attempted.

Design Specifications for a New Model of Tryout-Revision

As suggested in the previous sections, there seem to be at least two dimensions that need to be incorporated into the design of a new model of tryout-revision; the first dimension is the need for better learner feedback on learning problems met during prototype testing in other words, more and better process data. The second dimension is the ability to access learners from several sites so that a sufficient quantity of tryout learners can be included in prototype tryouts in a manner that does not present serious logistical problems.

With respect to the first dimension there seem to be at least four types of tryout data that are appropriate, as follows:

1. Individual process data, such as is obtained through the use of the tutorial model.

2. Group debriefing data, such as is obtained through the use of Abedor's operationalization of the small group model.

3. Test data, so as to provide indictors as to where in the instructional system learners had problems.

4. Questionnaire data, so that data about attitudes toward the instructional system may be assessed.

The large group model, as presently used is perfectly adequate for collecting the third and fourth categories enumerated above, especially if criterion-referenced test items are used. The problem lies with the collection of individual process data and group debriefing data in sufficient quantity. As was pointed out in the review of the models both the tutorial and small group model, as operationalized by Abedor, provide these two types of data, respectively, but in too insufficient a quantity.

Thus, one aspect of a new model of tryout-revision would be the ability to obtain more of these two types of data. It was also pointed out in a previous section of this chapter that access to non-local learners was a particular problem with respect to the use of the tutorial and small group models.

Hence, if suitable ways could be found for including more learners in tutorial tryouts, and increasing the group size in group debriefings the problem could be solved. The major problem with including more learners in group debriefings is that, as Abedor has pointed out (Abedor, 1981), 6-8 learners is ideal and 10 is too large.

However, sometimes technology has a way of "arriving" just at the right time for the solution of human problems. One particular technology that seems to have particular pertinence to this problem is computer conferencing, which is also referred to as computerized conferencing.

Essentially, this technology allows for human communication through the use of computer terminals. By virtue of having access to a common computer file, participants of a computer conference are able to carry on a continuing dialogue over several months or an electronic "meeting" during a shorter time interval, such as an hour. In view of the possible application of this technology to tryout-revision, the following section will provide some further details about computer conferencing.

Computer Conferencing

Technical aspects of computer conferencing

Computer conferencing, in existence since 1970 (Hiltz and Turoff, 1978, p. xxix), is accomplished through the "marriage" of computer technology and telecommunication. The computer conferencing participant connects a computer terminal to the telephone system through use of a "modem," which serves to transform outgoing digital signals into conventional audio signals used in telephone communications. Incoming audio signals are transformed into digital signals for utilization by the computer terminal. Technically the process is called modulation and demodulation, which led to the term, modem. By dialing an ordinary telephone number assigned to a time sharing computer one may use the computer in an interactive mode through use of a computer terminal and accomplish the same ends as through the use of punched cards.

Thus, it is the general availability of on-line interactive computer systems that has provided for the use of the computer as a communication tool in addition to its older functions as a calculating machine, data processor, and information storage and retrieval device. By virtue of the extensive use of computer terminals for meeting other computer needs, individuals having access to such terminals are now able to communicate through computer conferencing systems.

Furthermore microcomputers and minicomputers can be used as computer terminals, given the necessary hardware and software for converting such computers to computer terminals. Hence, with the rapid growth of microcomputers this new communication mode is of no small

Furthermore, microcomputers and minicomputers can be used as computer terminals, given the necessary hardware and software for converting such computers to computer terminals. Hence, with the rapid growth of microcomputers, this new communication mode is of no small consequence to society. Turoff and Hiltz (1981) call this type of interconnectiveness, "superconnectivity." Two main types of computer conferencing communications are commonly used; namely, private messaging between specific individuals, which is similar to electronic mail, and public communications among large groups of individuals who participate in a computer conference. To give the reader a sense of what happens in a computer conference, a typical beginning of a conference session in the Electronic Information Exchange System (EIES) is shown below. This computer conferencing system is based at the New Jersey Institute of Technology (Hiltz and Turoff, 1978, pp. 7-8).

> You dial the local number of your packet-switched telephone network service, which provides a low-cost link to the computer-host of the conferencing system. (For example, in 1978 TELENET, one such service, enabled U.S. customers to dial into a single CCS [computer conferencing system] with local calls in any of 90 major cities for an average of $3.50 per hour.) You type in a few code words to identify yourself and are then given the following sort of information:

```
WELCOME
JOHN DOE ON AT 1/25/79 11:02 A.M.
PEOPLE YOU KNOW NOW ON TERMINALS
System monitor (100)
Robert Johansen (708)
Linton Freeman (745)
Elaine Kerr (114)
Robin Crickman (727)
LAST ON 1/24/79 7:25 P.M.
WAITING:  14 private messages
          16 comments in conference 172: Therapy Group
           3 comments in conference 253: Chinese
             Recipes
```

In addition, you see the names and number of those now "on line" with you, who could receive materials immediately. It reminds you that you belong to two "conferences," which are like written group discussions on a specific topic. In one of these discussions, there are 16 new entries that you have never read, and there are three in the other. Now, you are asked what you would like to do first. Among the options you have are to

- .Accept the full text of some or all of the private messages;
- .Scan just the title line of the messages, showing author and subject, before deciding which to read first;
- .Choose to go directly to a conference and receive the new discussion entries there;
- .Send a message or enter a conference comment first;
- .Search and review earlier materials by such criteria as author and/or date, and/or subject. (Perhaps, for instance, you remember that there was a message sent to you by Elaine two days ago that you never answered. Rather than search through your file drawer, you may retrieve it by author and date, and review it before answering.)

Thus, as can be seen from this example, the EIES participant is given a number of options and a report on what has happened in his/her conferences since the last session.

General communication characteristics of computer conferencing

Some of the unique characteristics of computer conferencing are as follows:

1. Time and distance constraints are eliminated, when the system is used asynchronously, i.e., when participants are not all participating at the same time, which is the normal mode for computer conferencing (Hiltz and Turoff, 1978 and Parnes, 1981) This means that the participants are free to participate when its convenient for them

to do so. In addition, time zones do not have an effect on when individuals may be "reached" as with telephone use, since all communications are between the participant and the computer. Charges for use of the telephone lines are the same, regardless of distance from the host computer. Thus, there is no cost advantage or disadvantage in selecting conference participants on the basis of their nearness or distance to the host computer.

2. There are no place constraints (Parnes, 1981; Hiltz and Turoff, 1978) Group size can number at least 100 without causing communication difficulties. This is due to the fact that everyone can be "talking" at the same time, and reading is usually much faster than listening. There is no competition for "the floor" and there need be no concern about "talking too long" or waiting for an approprate moment to make a comment. Moreover, there is no problem in finding the right sized meeting room or the right type of seating arrangements, both of which are factors which influence the quality of face to face meetings.

3. Turoff and Hiltz (1981) point out that a greater amount of introspection may be encouraged by computer conferencing. Those who wish to "think things over" before commenting may do so without appearing "odd" to the other meeting participants. In addition, the more dynamic group members can't influence the other participants to the same degree as in face to face communication. In this respect computer conferencing is similar to the Delphi method (Linstone and Turoff, 1975) and the Nominal Group Technique (Scott and Deadrick, 1982), which are non-computer-based techniques for structuring communications through the elimination of group dynamics.

4. In addition to introspection, there is also the possibility of the acquisition of additional insights, information and opinion from interacting with other persons outside the conference and using various resources between computer conferencing sessions. Broadly defined, computer conferencing is a complex of interconnecting networks of human and non-human resources. Thus, it appears to be a potentially rich environment for the exchange of ideas, opinions and information.

5. Parnes (1981, p. 672A) suggests that it is the conceptual and not the technical constraints of computer conferencing that will limit its social diffusion. These conceptual constraints "have to do with the seemingly simple but profoundly difficult matter of people understanding that computer conferencing is a genuinely new communications medium, not the augmentation of old forms of communciation. (Parnes, 1981, p. 672A)"

For a comprehensive treatment of computer conferencing, the reader is referred to Hiltz and Turoff (1978), and Parnes (1981a). Parnes is the designer of CONFER II, which is a computer conferencing system originally developed at the University of Michigan. "CONFER" is not an acronym but, rather, the first six letters of the word conference. Turoff is the designer of EIES.

Predictions about computer conferencing

The following predictions have been made about computer conferencing by Hiltz and Turoff (1978, p. xxix-xxx):

-Computerized conferencing will be a prominent form of communications in most organizations by the mid-1980s.

-By the mid- 1990s it will be as widley used in society as the telephone today.

-It will offer a home recreational use that will make significant inroads into TV viewing patterns.

-It will have dramatic psychological and sociological impacts on various group communication objectives and processes.

-It will be cheaper than mails or long distance telephone voice communications.

-It will offer major opportunities to disadvantaged groups in the society to acquire the skills and social ties they need to become full members of the society.

-It will have dramatic impacts on the degree of centralization or decentralization possible in organizations.

-It will become a fundamental mechanism for individuals to form groups having common concerns, interests or purposes.

-It will facilitate working at home for a large percentage of the work force during at least half of their normal work week.

-It will have a dramatic impact upon the formation of political and special interest groups.

-It will open the doors to new and unique types of services.

-It will indirectly allow for sizable amounts of energy conservation through substitution of communication for travel.

-It will dramatically alter the nature of social science research concerned with the study of human systems and human communication processes.

-It will facilitate a richness and variability of human groupings and relationships almost impossible to comprehend.

Turoff and Hiltz (1981) further predicted, as follows:

> There is no segment of society which will not be touched in some way or another by this technology. We may have institutions called as they are now by the names of

> universities, business, government agencies; but they will become institutions without walls, existing largely in the chips and bubbles of the network society (p. 10).

The Prospects of Using Computer Conferencing in Tryout-Revision

In view of the fact that computer conferencing provides for both private communications, between two or more individuals, and public communications among larger groups, this communication mode may be technically feasible for application to tryout-revision. Through use of computer conferencing it may be possible to carry out tutorial tryout, using private communications, with a larger number of learners than is ordinarily the case. Such a result could occur, perhaps, by virtue of having remote access to learners at many sites.

In addition, the optimal size for a computer conference is between 30 and 40 participants, according to Palme (1982). This is when the computer conference is spread out over time and not held as an "electronic" meeting. Hence, a computer conferencing version of the group debriefing might be possible. If so, it would extend the number of participants from 6-8 to 30-40, a 5-fold increase.

Furthermore, computer conferencing could provide convenient access to non-local learners, which was pointed out earlier as the second dimension that ideally, needs to be incorporated into the design of a new model of tryout-revision. Since computer terminals are found at so many institutions where likely tryout learners are located, computer conferencing could also solve the common instructional development problem of not having a sufficient supply of tryout-learners in the local area.

Additionally, computer conferencing may enhance the possibilities for using stratified random samples of tryout learners as well as special samples of learners having a particular characteristic; for example, a particular learning style, subject interest, or learning disability.

Thus, by building certain modules into a modular instructional system, it might be possible to make the system useful to a larger population than otherwise may be the case. These modules could be added to the system on the basis of the success met in trying it out with particular strata or separate samples during tryout-revision.

In this way, tryout-revision would, be taking on some of the functions of summative evaluation. During such evaluation, instructional systems are tested for the limits of the population that can utilize the instructional system. However, by testing for these limits during tryout-revision, its potential marketability could be increased while development is still in-process. Such development is not always possible, due to limitations found in the local availability of tryout learners. But with the possibility of using remote tryouts via computer conferencing, the scale of instructional development might be enlarged.

Furthermore, computer conferencing could make it possible to use random samples of tryout learners, regardless of stratification considerations. Again, due to local limitations in the availability of tryout learners, random sampling is not always possible. Thus, instead of having to "settle for whatever students can be found," instructional developers might, through computer conferencing, be able to select learners from a variety of sites, selecting a few learners from each site. This would make unnecessary the practice of selecting of entire classrooms for a sample simply because logistical reasons prevented random selection of individual learners. Statistically, the latter practice leads to more accurate results than selecting entire classrooms. Computer conferencing might also facilitate the repetition of the tryout-revision cycle more than may often be the case, due to the availability of more learners through remote access.

In summary, then, the possibility of increased access to tryout learners through computer conferencing might improve the general quality of tryout-revision, and tryout-revision research and development. In addition, such increased access might lead to the development of instructional systems that could be used with larger populations of learners, through the inclusion of extra modules and other supplementary material, through more extensive tryout-revision with stratified random samples and special samples of individual tryout learners.

In view of the far-reaching predictions made about computer conferencing for society in general by Hiltz and Turoff, mentioned previously, it seems to be a technology that instructional developers should not ignore. By applying it to the improvement of its own methodology, such as tryout-revision, it might be a way for testing the suitability of this technology to instructional development in general, without necessarily involving actual instructional development projects. This could be accomplished through limited studies in improving existing instructional systems. By doing so, instructional developers would be preparing for the day when the use of computer conferencing might be expected, by instructional development clients, to be used in actual instructional development projects, for the purpose of saving their time and adding to their convenience.

Through the development of a computer conferencing model of tryout-revision, then, the profession would gain experience in the use of computer conferencing as well as, perhaps, improve the state of the art of tryout-revision. In view of the fact that the literature reveals no previous work done in this area, this study was designed as an initial application of computer conferencing to tryout-revision. In the next section, an ideal Computer Conferencing Model of Tryout-Revison will be described. This model will serve as the basis for developing an operationalized Computer Conferencing Model of Tryout-Revision to be used in this study.

An Ideal Computer Conferencing Model of Tryout-Revision

In a previous section of this chapter entitled, "Design Specifications for a New Model of Tryout-Revision," four types of data were listed as being appropriate; namely, individual process data, group debriefing data, test data, and questionnaire data. In view of the fact that both tutorial tryouts and group debriefings are possible through use of computer conferencing, it seems that the new model could be the combination of the tutorial model, Abedor's operationalization of the small group model, and the large group model in a single model that uses computer conferencing as the primary communication mode.

These models could be carried out in different phases with revision cycles between each phase or all the data could be collected at one time from different samples of learners. However, as previously mentioned, some writers (e.g. Dick, 1977, 1980, and Briggs and Wager, 1981) mention that these three models are used in a series starting with the tutorial model and ending with the large group model. Therefore, it may make sense to follow the same paradigm.

The first phase could be called the tutorial phase. Two possibilities seem possible. One possibility would be to carry out tutorial tryouts via computer conferencing. The disadvantage of doing this is that the benefits of negative non-verbal communication normally observed during face-to-face debriefings would be lost. The second possibility would be to carry out face-to-face tutorial tryouts, which could be followed up by individualized computer conferencing debriefings with the same learners.

As was mentioned in the early part of this chapter, the point of departure for this study is the belief that, in general, the wrong type of data is collected during tryout-debriefing. In order to improve upon this situation, the writer feels that the categories of learner and program failures developed by Gropper (1975) and similar sets of categories should, ideally, be the basis for obtaining tryout data. For that reason any new model of tryout-revision should, ideally, incorporate the use of such categories in collecting tryout data.

Thus, this ideal model should incorporate the use of such categories, even though the operationalized model to be explained later in this chapter will not incorporate the use of such categories. (Such categories were not used in this study so that the complexity of this study would be kept within reasonable limits.) To use such categories, it seems quite feasible to carry out the tutorial tryout in two stages. The tutoring for the first stage could be carried out face-to-face by a tutor at the site where the learner is located. The tutor could be a subject matter expert or some other type person other than a trained debriefer.

His/her role would be to enable the learner to use the instructional system and to administer the posttest and the questionnaire. The only debriefing that would be done would be for the purpose of enabling the tutor to help the learner understand the content of the instructional system. At the end of the tutorial, the learner and the tutor would co-author a computer conference message to a remote debriefer, regarding problems met in using the instructional system. The tutor would also indicate negative non-verbal communications observed while the learner was using the instructional system and specify the coresponding parts of the instructional system with which the learner was interacting.

The second phase of the tutorial tryout would be carried out by a debriefer. This person would be trained in the use of the categories of learner and program failures used as the basis for collecting data on why the learner had difficulties while using the instructional system. This debriefing would be conducted via computer conferencing through the use of private messages while both parties were on-line with the computer; that is, they would both be using the computer conferencing system at the same time.

After all the learners designated for the tutorial debriefing had been debriefed the instructional system could be revised. The number of tryout learners for the tutorial stage would be 30 or so. Thus, the data for making changes in the instructional system would be reasonably reliable.

Following the tutorial stage, a small group tryout and group debriefing, involving, perhaps, 30-40 learners would be held. It would be very similar to Abedor's operationalization of the small group model, except that the group would be much larger. All the communications would be through the computer conference, except that the posttest and questionnaire would be administered to the learner by a person at the site where the learner is located. The responses would be sent to designated graders through the computer conference by private message by the learner.

The group debriefing would involve various types of experts, such as educational psychologists, media specialists, media technicians, instructional developers, and evaluation experts, all of whom would be trained in the use of the learner and program failure categories being used. Their role would be to develop a consensus among the learners regarding their learning problems in terms of the various failure categories. This would be in addition to the posttest and questionnaire data sent by these learners to the graders. The latter data would provide indicators as to where the learners had problems. The learners would also be asked for their group recommendations for improving the instructional system, based on the group's consensus. These recommendations would also be in terms of the categories of program failures. These recommendations would need to be interpreted by the various experts participating in the debriefing, since the learners would not necessarily understand the failure categories being used.

The purpose of the group debriefing would be to make use of the synergistic effect that occurs, according to Abedor (1981), when a group is debriefed. The group interaction has the effect of triggering responses from the individuals that doesn't occur when each individual is debriefed separately.

The instructional system would then be revised on the basis of the data collected. The revised version would then be tried out through use of the large group model. This tryout would involve the administration of a posttest and a questionnaire at the sites where the learners were located. The responses could be sent through the computer conference to graders by private message or they could be mailed to the development project office. For more efficient handling of the data, it would be best if the responses were sent through the computer conference. The number of learners could be in the hundreds.

In addition, the learners could be involved in the revision process following their debriefing, particularly in the case of the revisions following the tutorial and small group tryouts. This would provide for their further input and verification that the problems they met have been eliminated. However, motivation for such participation might be a problem, in view of the limited commitment learners might have for improving instructional systems.

The revision process would make use of the computer conference.

Thus, ideas could be shared, scripts co-written, priorities established, work assigned, and reports written and critiqued through private messaging and the use of public conference items. This process should make revision more efficient than is usually the case through the elimination of staff meetings and phone calls, and through a more efficient handling of various types of paperwork. In view of the fact that much of instructional development involves meetings and paperwork, it is quite possible that these aspects of tryout-revision could serve as a model for using computer conferencing in other areas of instructional development.

This model is, of course, not operationally defined. It is essentially a concept that needs to be worked with over time until it can be either implemented, abandoned due to being too complex, or modified and then implemented. Its primary importance for this study was its heuristic value in serving to help in the development of some specific ideas on how computer conferencing might be applied to tryout-revision. A portion of this "grand design" will be implemented for this study, as will be detailed in later sections of this chapter. But first, the problem needs to be summarized, in view of the length and complexity of this chapter.

Summary of the Problem

The major problem that this study addresses is that instuctional developers do not always know how to revise a prototype system after tryout data has been collected. It would seem that this problem could be alleviated by a) obtaining better tryout data, as suggested by

Gropper (1975) and b) improving the revision process. Intuitively, it also seems that the latter could be accomplished by improving the communication process during revision and by including tryout learners in the revision process.

One way to reach these ends would be to design a new model of tryout-revision that would accomplish the following results:

1. Include the four types of data that are collected through use of the basic known models of tryout-revision; namely, process data, group debriefing data, test data, and questionnaire data.
2. Provide for the collection of data from large numbers of learners, for each type of data indicated in "1," above.
3. Provide for the collection of data from learners at multiple tryout sites, for each type of data indicated in "1," above.
4. Improve communication during the revision process by facilitating the ability to exchange ideas and information.
5. Include tryout learners in the revision process.

One such model might be one that utilizes computer conferencing as its main communication channel, which hereinafter will be called the Computer Conferencing Model of Tryout-Revision (CCTR). As mentioned previously these five features might be possible through the use of CCTR. What remains to be seen is whether such a model is workable.

In order to implement such a model, the tutorial, small group, and large group models need to be modified, as needed, so that:

1. They utilize computer conferencing as the primary communication channel for collecting tryout data.
2. The tryout data is provided by a large number of learners located at multiple tryout sites.
3. The revision process utilizes computer conferencing as a primary communications channel.

The next question to consider, then, is, "which part of the system should be implemented first?" One approach would be to implement the most critical subsystem, since the success of any system is usually dependent upon the ability to operate its most critical component. For this system, the question becomes, "which subsystem is most dependent upon the use of computer conferencing?"

The ability to conduct a tutorial session with a learner is no different from normal electronic messaging between two parties, so this component is not critical. Collecting test and questionnaire data from learners via computer conferencing would not seem to present a problem, since its merely a matter of each learner sending a message to a designated conference participant, with the response data clearly identified in the message.

The most critical component seems to be that of how to carry out a group debriefing via computer conferencing, so that it is just as useful as a face-to-face group debriefing, even though there is no non-verbal communication and no opportunity for the immediate group interaction that is characteristic of face-to-face meetings.

Thus, before any progress can be made on developing a computer conferencing model of tryout-revision, one of the first requirements is to demonstrate that a group debriefing can be successfully carried out via computer conferencing.

Although this model is designed for use with non-local learners, it seems that it first must be demonstrated as being workable on a local level. The rationale for this approach is that since computer conferencing is, at this stage, such a novel technology we need to know whether learners with little or no experience with either computer terminals or computer conferencing can successfully communicate with the host computer in a debriefing process via computer conferencing, and whether a debriefer can successfully debrief by means of computer conferencing.

The aspect of the overall model implementation plan that this study addresses is the modification of the small group model so that it utilizes computer conferencing as the communications channel for collecting tryout data; specifically it uses computer conferencing for collecting debriefing data. Through such modification, the collection of data from large numbers of learners at multiple tryout sites is also made possible, even though such possibilities were not incorporated into the data collection plan for this study. This modification of the small group model can also be considered an early prototype of CCTR, since its the first known attempt to apply computer conferencing to tryout-revision.

Furthermore, in view of Abedor's (1971, 1972) successful use of group debriefing for obtaining group consensus data on deficiencies in the prototype and their recommendations for eliminating such deficiencies, the possibility of basing revision on debriefing data alone seemed worth pursuing. If prototypes can be revised on such a basis it would mean a simplification of the small group model through the elimination of the collection of individual process data and the analysis of individual process and outcome data during prototype revision. Such simplification would make prototype revision easier. The question is, "can a prototype be successfully revised on the basis of group debriefing data only?"

Need

On the basis of the description of the problem presented in the last section, the needs that this study addresses to are as follows:

1. To determine the technical feasibility of applying computer conferencing to some aspect of the tryout-revision of instructional prototypes.
2. To explore the possibilities of basing the revision of an instructional prototype on debriefing data only, resulting from a group debriefing of the learners who used that prototype.
3. To compare the relative effectiveness of face-to-face and computer conferencing communication in the group debriefing of the learners who tried out one particular instructional prototype.

Purposes of the Study

This study has four purposes:

1. To develop a small group model of tryout-revision (SGTR), which would be a simplification of Abedor's MK II model. (Abedor, 1971 and 1972).
2. To develop an early prototype version of a computer conferencing model of tryout-revision (CCTR), based on SGTR, but with computer conferencing substituted for face-to-face communication during the group debriefing.
3. To evaluate the effectiveness of SGTR and CCTR as a means for improving one instructional system, with respect to posttest scores and attitude toward the instructional system.
4. To compare the relative effectiveness of SGTR with CCTR as a means for improving one instructional system, with respect to posttest scores and attitude toward the instructional system.

Research Questions and Hypotheses

Six research questions, which relate to four hypotheses, are asked in this study. Half the research questions and hypotheses deal with posttest data research; the other half of the research questions and hypotheses deal with questionnaire data research. For each type of data, the research questions and hypotheses deal with two basic concerns, as follows:

1. Are the two revisions of the instructional prototype developed in this study significantly better, as measured by the learning outcomes, than the prototype.

2. Is the computer conferencing revision, that is, the revision developed through use of the Computer Conferencing Model of Tryout-Revision (CCTR), better than the small group revision, that is, the revision developed through uses of the Small Group Model of Tryout-Revision (SGTR), as measured by the learning outcomes.

In the case of the first type comparison the choice of the test statistic, analysis of variance, made it necessary to combine the learners who tried out the two revisions into one treatment group. Thus, the hypotheses for the first type comparison are stated in terms of the combined treatment group. This was necessary due to restrictions in the number of planned comparisons that are possible in the use of analysis of variance. By so combining these learners, one less planned comparison was necessary, which made it possible to abide by the limit regarding the number of planned comparisons.

However, one of the primary concerns of this study was the comparison between each of the revisions, respectively, and the prototype; therefore post hoc analysis was carried out so as to make such comparisons.

In the case of the second type comparison, better learning outcomes were predicted for the learners who tried out the computer conferencing revision than the learners who tried out the small group revision. This prediction was based on the expectation that computer conferencing would permit a higher quality of debriefing. This expectation is due to the independent nature of this mode of communication; that is, since each participant communicates by engaging in an interactive session with the computer from an on-line terminal, each participant, acts independently of the other participants.

This independent nature of the communication suggests several notions, each of which poses the possibility of a higher quality debriefing when computer conferencing is used. First, the amount of communication could be greater than in the case of a small group debriefing, because some or all of the participants can send or receive messages simultaneously. In comparison with face-to-face communication, this implies the possibility of more communication per unit of time.

However, this doesn't take into consideration the different communication rates for speaking versus typing and listening versus reading, as well as the time it takes to obtain the appropriate responses from the computer relating to message sending and receiving. Therefore, simultaneous communication does not necessarily imply a greater overall amount of communication per unit of time, but it might. If so, this greater amount of communication might imply a higher quality of debriefing; however, this is not necessarily true either.

Second, as suggested by Turoff and Hiltz, (1981), a high amount of introspection is associated with computer conferencing. This is due to the fact that each participant in computer conferencing has complete control over the communication process, with respect to the amount and timing of stimulus and response, which is not true of face-to-face debriefing. This factor also suggests a higher quality of debriefing for the computer conferencing debriefing than the small group debriefing.

Thirdly, computer conferencing is in many ways similar to the Delphi method. (Linstone and Turoff, 1975) The Delphi method consists of a series of "rounds" of opinion questionnaires mailed to the respondents. During each succeeding round, the respondents are provided with the results from the previous round and given the opportunity to change their votes. Delphis are normally used for developing consensus. For further details about the Delphi method see Linstone and Turoff (1975). Furthermore Delphis have been carried out by means of computer conferencing (Linstone and Turoff, 1975). One characteristic of a Delphi is that the strong personalities have less influence over the other participants, due to the fact that there is no group dynamics present, as in the case of face-to-face communication.

In view of this factor, it could mean that the absence of strong interpersonal influence results in greater equality in participation. Hiltz et al., (1978) found greater equality in participation in the case of computer conferencing when face-to-face communication was compared to computer conferencing in an experiment on problem solving. Assuming that greater equality in participation leads to more ideas being discussed, the possibility that computer conferencing debriefing is of a higher quality than face-to-face debriefing is suggested.

Therefore, it was concluded that, in spite of the unknown factors pointed out and assumptions made, there could be a basis for asserting that computer conferencing debriefing is of a higher quality than small group debriefing. On this basis, better learning outcomes were predicted for the learners who tried out the computer conferencing revision than for the learners who tried out the small group revision. This, then, is the basis for the directionality of the second type comparison indicated in the hypotheses; that is, the comparison between the learners who tried out each of the two revisions, respectively.

It is now appropriate to list the research questions and research hypotheses. The four research hypotheses are stated in general form here, but will be re-stated in testable form in Chapters 3 and 4. The research questions for the first research hypothesis is as follows:

> Research Question 1 (RE: Purpose 3): To what extent will use of a small group revision lead to a higher posttest score than use of the prototype?
>
> Research Question 2 (RE: Purpose 3): To what extent will use of a computer conferencing revision lead to a higher posttest score than use of the prototype?

These research questions lead to the following hypothesis:

> H_1: The use of a small group revision or a computer conferencing revision will lead to a higher posttest score than use of the prototype.

The research question for the second research hypothesis is as follows:

> Research Question 3 (RE: Purpose 4): To what extent will use of a computer conferencing revision lead to a higher posttest score than use of the small group revision?

This research question leads to the following research hypothesis:

> H_2: The use of a computer conferencing revision will lead

to a higher posttest score than use of a small group revision.

The research questions for the third research hypothesis are as follows:

<u>Research Question 4</u> (RE: Purpose 3) To what extent will use of a small group revision lead to a more favorable attitude toward an instructional system than use of the prototype?

<u>Research Question 5</u> (RE: Purpose 3): To what extent will use of a computer conferencing revision lead to a more favorable attitude toward an instructional system than use of the prototype?

These research questions lead to the following research hypothesis:

H_3: The use of a small group revision or a computer conferencing revision will lead to a more favorable attitude toward an instructional system than use of the prototype.

The research question for the fourth research hypothesis is as follows:

<u>Research Question 6</u> (RE: Purpose 4): To what extent will use of a computer conferencing revision lead to a more favorable attitude toward an instructional system than use of a small group revision.

This research question leads to the following research hypothesis:

H_4: The use of a computer conferencing revision will lead to a more favorable attitude toward an instructional system than use of a small group revision.

Limitations of the Study

This study is limited in various ways. Due to limitations in available resources, only one instructional system could be tried out with the tryout-revision models, and that instructional system was a self-instructional system; hence, the conclusions of this study will relate to self-instructional systems only.

Due to the fact that only graduate students in the College of Education enrolled in three instructional media courses were used as experimental subjects, the conclusions of this study are even further limited. Strictly speaking, such inferences cannot be extended beyond the limits of the type of prototype used and a population of learners consisting of graduate education students at Michigan State University. However, by making certain assumptions about the prototype and the experimental subjects, the research findings relating to tryout-revision will be extended to a larger population and a more general category of instructional systems. This will be done in Chapter 5.

Dissertation Overview

In Chapter 2, the literature relating to the application of computer conferencing to tryout-revision, the need for a new model of tryout-revision and/or the revision of one of the known models of tryout-revision, the state of the art of tryout-revision, the effectiveness of tryout-revision, the evaluation and evaluation research and the computer literature bearing on literature tryout-revision will be reviewed. Descriptions of the instrumentation, the two models, and the study design will be provided in Chapter 3. In Chapter 4, the research findings will be presented and discussed. In Chapter 5, a summary, a further discussion of the findings, the conclusions, and recommendations for further research development and practice in this area will be provided.

CHAPTER 2

REVIEW OF THE LITERATURE

Introduction

As discussed in Chapter 1, this study deals with the development of an improved model of tryout-revision. Since tryout-revision is related to formative evaluation, the evaluation literature was searched in order to identify the functions of evaluation in general. Closely related to the evaluation field is the evaluation research field. Since this study is, in different contexts, related to both fields, the literature on evaluation research field was searched, to a limited extent, in order to identify the function of that field. The results of these two searches will be reported in this chapter first.

Given that background, the literature relating to the degree to which instructional developers are satisified with the state of the art of tryout-revision and the extent to which the known models of tryout-revision have been effective will be reviewed. Since two searches lead to conclusions that a new model of tryout-revision may be in order, the question is raised as to whether the literature indicates an expressed need for either a new model or the improvement of any of the existing models.

Next the computer conferencing literature is reviewed for information that might be useful in designing a new model of tryout-revision based on computer conferencing. Finally, the instructional development literature is reviewed to determine whether any previous studies have been made in which computer conferencing was applied to tryout-revision.

Definitions of "Evaluation"

The evaluation literature reveals numerous definitions and other comments regarding the term, evaluation. As a preliminary to a discussion of the state of the art of tryout-revision and the effectiveness of tryout-revision a representative number of these definitions will be quoted and, then, discussed as a whole. This compilation is however, by no means an exhaustive inventory of such definitions.

Alkin (1972), representing the Center for the Study of Evaluation, University of California at Los Angeles defines evaluation as follows:

> Evaluation is the process of ascertaining the decision areas of concern, selecting appropriate information, and collecting and analyzing information in order to report summary data useful to decision-makers in selecting among alternatives (emphasis added) (p. 107).

Stake (1967) in summarizing his comments on evaluation says, "The countenance of evaluation should be one of data gathering that leads to decision-making. . . (emphasis added) (p. 539).

Weiss (1972) says, "The basic rationale for evaluation is that it provides information for action. Its primary justification is that it contributes to the rationalization of decision making" (emphasis added) (p. 318).

Berk (1981) reviewed definitions of evaluation in the writings of Alkin, Cooley and Lohnes, Cronback, Fink and Kosecoff, Freeman, Guba, Popham, Posavec and Carey, Provus, Scriven, Stake, Stufflebeam, Walberg, and Wolf. He concluded:

> A critical survey of these definitions revealed that there was a single common thread running through all of them: evaluation is the process of providing information for decision making. This concept of evaluation as a political decision-making tool is expressed most clearly in the definition by Stufflebeam et al.: 'the process of delineating, obtaining, and providing useful information for judging decision alternatives.' (1971, p. 36) This comprehensive and extremely popular definition was an outgrowth of the work of the Phi Delta Kappa National Study Committee on Educational Evaluation from 1968 to 1970 (p. 4).

In summary, it is apparent that from these definitions that the purpose of collecting data in evaluation studies is for the support of the decision-maing process. The problems faced in tryout-revision regarding the difficulty in knowing what to revise (cf. Dick, 1980 p. 5) suggest that the data collected does not contribute to making precise decisions as to what to change during revision. It would seem, then, that the broader aspects of the evaluation literature, as represented by the writers indicated above, have considerable implications for tryout-revision from the point of view of specifying the type of data that should be collected. From this standpoint, it seems apparent that instructional developers involved with tryout-revision should consult the more general aspects of the evaluation literature as sources of ideas in designing prototype testing and tryout-revision research studies.

More specifically, since the essential type of decision made in tryout-revision is related to making changes in the instructional prototype, it appears that the primary type of data that should be collected in tryouts is process data until it is quite apparent that the prototype is performing at a satisfactory level. It is only at that point, it would seem, that the emphasis should shift to the collection of outcome data, as verification of the workability of the instructional system.

The importance of the evaluation literature for this study was that it was supportive of the notion that a new model of tryout-revision is needed, whereby the type of data collected during prototype testing is more directly related to the type of information required to make decisions during prototype revision.

The Evaluation Research Literature

The field of evaluation research parallels that of evaluation. it is concerned with the search for scientific generalizations regarding decision making on the basis of evaluation data In addition to the provision of data to administrators and others as a basis for making sound management decisions, it is concerned with the search for scientific generalizations regarding decision making on the basis of evaluation data In this respect, evaluation research has implications for both tryout-revision and tryout-revision research. The latter will be discussed in Chapter 5 in the context of the implications of this study. However, some of the methodology of evaluation research is highly pertinent to tryout-revision, especially to the type of model suggested in Chapter 1. For these reasons two highly credible sources in evaluation research were consulted for definitions of that field and for information regarding the types of data that is collected and analyzed in that field.

In the Handbook of Evaluation Research, Volume I, sponsored by The Society for the Psychological Study of Social Issues, which is a Division of the American Psychology Association, Weiss (1975) defines evaluation research, as follows:

> By objective and systematic methods, evaluation research assesses the extent to which goals are realized and looks at the factors associated with successful or unsuccessful outcomes. The assumption is that by providing "the facts," evaluation assists decision-makers to make wise choices among future courses of action. Careful and unbiased data on the consequences of programs should improve decision-making (p. 13).

In a Russell Sage Foundation publication, Suchman (1967) defines evaluative research, which is the same as evaluation research, as follows:

> But evaluative research is generally applied or administrative research, the primary objective of which is to determine the extent to which a given program or procedure is achieving some desired result. The "success" of an evaluation project will be largely dependent upon its usefulness to the administrator in improving services (p. 21).

He proposes five categories of criteria that can be used for evaluating the success or failure of a program; namely, effort, performance, adequacy of performance, efficiency, and process. Performance criteria parallels criteria for outcome data collected during tryout-revision. Process criteria is comparable to criteria for the process data collected through the use of the tutorial model of tryout-revision, for example.

In view of the importance of process data to this study, the following quotations from this source are included here:

> 5. Process. In the course of evaluating the success or failure of a program, a great deal can be learned about how and why a program works or does not work. . . . Locating the cause of the failure may result in modifying the program so that it will work, instead of its being

> The rationale and recommendation for LVR are based more on folklore than on research. To date there has [sic] been few serious investigations on such critical questions as the following:
>
> 1. What type of learner-verification data should be collected?
> 2. How many learners should be involved in LVR procedures.
> 3. What types of learners provide optimal feedback for revision?
> 4. Should the roles of the developer, the evaluator, and the reviser be separated?
> 5. What principles should prescribe the types of revisions to be based on the learner-verification data?
> 6. At what stages of development should LVR be undertaken. In view of the justifiable politicization of the practice of LVR (i.e. state legislation demanding LVR as a requirement for state-adopted instructional materials), a more effective empirical base is urgently required (p. 317).

Thiagarajan (1978) asks the following questions about LVR:

> "14. How many learners are needed for LVR?
>
> . . . .
>
> "15. What types of learners provide optimal feedback?
>
> . . . .
>
> "16. What type of feedback is collected from learners?
>
> . . . .
>
> "17. What types of performance data are collected from LVR?
>
> . . . .
>
> "18. What types of test instruments are used for collecting learner feedback?
>
> . . . .
>
> "19. How and when is learner feedback collected?
>
> . . . .
>
> "20. How is learner feedback converted into revisions?
>
> . . . .(p. 141)."

He concludes by saying, "The field of IPVR instructional product verification and revision is full of mystique and myth--even from the initial assumption that IPVR does improve instructional products (p. 141)."

Though neither of these writers suggest the need for a new model the degree to which instructional developers are satisfied with the state of the art of tryout-revision. Kandaswamy (1976) discusses some of the limitations in the state of our knowledge about tryout-revision. He refers to tryout-revision as learner verification and revision (LVR):

> The rationale and recommendation for LVR are based more on folklore than on research. To date there has sic been few serious investigations on such critical questions as the following:
>
> 1. What type of learner-verification data should be collected?
> 2. How many learners should be involved in LVR procedures.
> 3. What types of learners provide optimal feedback for revision?
> 4. Should the roles of the developer, the evaluator, and the reviser be separated?
> 5. What principles should prescribe the types of revisions to be based on the learner-verification data?
> 6. At what stages of development should LVR be undertaken. In view of the justifiable politicization of the practice of LVR (i.e. state legislation demanding LVR as a requirement for state-adopted instructional materials), a more effective empirical base is urgently required (p. 317).

Thiagarajan (1978) asks the following questions about LVR:

"14. How many learners are needed for LVR?
. . . .

"15. What types of learners provide optimal feedback?
. . . .

"16. What type of feedback is collected from learners?
. . . .

"17. What types of performance data are collected from LVR?
. . . .

"18. What types of test instruments are used for collecting learner feedback?
. . . .

"19. How and when is learner feedback collected?
. . . .

"20. How is learner feedback converted into revisions?
. . . .(p. 141)."

He concludes by saying, "The field of IPVR instructional product verification and revision is full of mystique and myth--even from the initial assumption that IPVR does improve instructional products (p. 141)."

Though neither of these writers suggest the need for a new model of tryout-revision, as such, their suggestions and concerns indicate that there is a general research need for varying the use of tryout-revision techniques. In effect, this seems equivalent to suggesting the development of new models.

Diagnosis

One of the major strategies for improving instructional systems is the identification of system deficiencies so that appropriate corrective action can be taken during the revision of an instructional prototype. Cronback (1963) pointed this out even before the term, formative evaluation, was part of our language. "The greatest service evaluation can perform is to identify aspects of the course where revision is desirable," (p. 675) he said.

Scriven (1967), the so called "father of formative evaluation", pointed to the same thing when he said:

> The performance of the students on the final tests, as upon the tests at intermediate stages, must be analyzed in order to determine the exact locations of shortcomings of comprehension, shortages of essential facts, lack of practice in basic skills, etc. Percentages are not very important. It is the nature of the mistakes that is important in evaluating the curriculum, and in rewriting it (pp. 61-62).

Apparently, the suggestions of these early writers on formative evaluation were not sufficiently heeded by the mid 1970s, as was indicated in Chapter 1, for Gropper (1975) found it necessary to say:

> The future of a tryout and revision technology awaits an identification of the types of errors students commit both on programs and on tests and a parallel identification of the types of program weaknesses which are responsible for them. It also awaits the formulation of diagnostic procedures which, by making the needed identifications, can lead to relevant, reliable implemental revision (p. 9).

Gropper further comments on the need for the identification of learning failures:

> Instructional development, it is suggested, will more closely approximate a technology when instructional design and instructional revision work from the same blueprint. . . . Design is made adaptive to skills to be learned; revision or redesign is made adaptive to failures to learn those very same skills (p. 39).

On program design failures, Gropper says:

> To improve on the ad hoc approach to program revision, a developer must do more than identify topographically those parts of a program which have failed. He must be able on the one hand to identify specific student learning/performance failures, and on the other hand to link them with the specific program design elements which have failed (p. 43).

No literature more current than Gropper's 1975 publication was found that suggests that tryout-revision has improved to the point that Gropper's comments are not still valid. Therefore, it is concluded that the state of the art of the diagnosis process is still in need of considerable improvement. Essentially, the heart of this problem seems to be attributable to the lack of adequate tryout data.

Revision

After system deficiencies have been identified in the instruction,the logical next step is to make appropriate revisions. However, as was pointed out in Chapter 1, Dick (1980) contends that the biggest problem in formative evaluation is "the dilemma of what to do after a problem has been detected in instruction (p. 5)."

Baker and Alkin (1973) report on studies of the revision process by Rosen (1968), Sulzen (1972) and Baker (1970) and conclude:

> While the above studies provided gross information regarding the impact of formative evaluation data upon revision, the technology related to the uses of formative evaluation data in revision is still obscured. What does or should the program writer do with given formative evaluation data? In what form should formative evaluation data be provided? What rules are there to guide the way in which developers might use such data (p. 400)?

Dick (1977) comments upon the lack of research data on identifying the specific contributions of different types of formative evaluation data as aids to revision. He then says:

> We have even less research knowledge about the process which should be used to revise instruction based on formative evaluation data. At this point, it could appear that we can only follow the logic of the theories which have been used in the instructional design process (p. 329).

Though Gropper (1975) has suggested an extensive number of categories of revision options, relating to such matters as revising behavior control techniques, task content, the comprehension level of instructions on task content, transitions between tasks, and cumulative learning experiences, no literature was found reporting the empirical application of Gropper's suggestions. Until such studies have been made, his suggestions do not provide any improvement in the state of the art of tryout-revision, in spite of the potential impact of his work.

Abedor (1971) seems to be the most certain among formative evaluation researchers as to what to do in order to make effective revisions in the instruction. Among the heuristics Abedor lists, which he feels may facilitate the use of the small group model of formative evaluation, is:

> Heuristic 1: Listen to the students; they are one of the best sources of information for identification and

remediation of learning problems.

> . . .In short, the student groups provided unique and insightful solutions to their own learning problems -- a skill which authors typically were unable to achieve because of their more sophisticated conceptualization of the subject matter (p. 155).

Abedor does not specify any procedures for implementing this heuristic to resolve the "what to do dilemma", expressed by Dick. However, this heuristic is fairly suggestive as to what to do. It appears that the obvious implication of the heuristic is to ask the learners what should be done in as operational terms as possible, so that their suggestions may be implemented. This does not imply that students will be able to state their suggestions in terms that are immediately implementable. No doubt, such suggestions will, in many cases, need to be interpreted by an instructional developer, and thereby translated into operational suggestions.

This literature on the revision process suggests a need for improving this process. But, this literature also seems to point to inadequacies in the tryout data as the reason for not knowing how to revise.

Conclusions on the state of the art of tryout-revision

Though there is not a considerable amount of evidence, there seems to be sufficient evidence to support the contention that there is a general dissatisfaction with the state of the art of tryout-revision. However, the remarks of the cited authors suggest that we don't collect the right type of tryout data. This is the case in both diagnosing instructional prototypes and in revising them.

Perhaps the use of an analogy from the practice of medicine would help clarify the problem. To diagnose a medical problem, a physician needs sound data based on physical examinations and/or laboratory tests. Once the problem has been diagnosed, a treatment can be prescribed. But the selection of the treatment depends upon the pattern of results indicated in the data, which can pinpoint the exact nature of the problem when different data sources point to the same possible source of the problem. The physician plans for this by identifying combinations of such corroborative data and then collecting such data.

Thus, the pinpointing of the problem seems to lie at the heart of the tryout-revision process. Since instruction is developed for many learners, there is also a need for knowing the relative seriousness of the problem, as revision must focus on the most serious deficiencies in the instructional system.

Of all the authors reviewed, in this section Abedor seems to be the most sure about the source of data that should be used. "Listen to the learners," he says. The literature does not reveal that the MK II small group model that he developed has been applied either in a research project or in professional practice since he reported his first results from the use of this model (Abedor, 1971, 1972).

Since this model places major emphasis on learner-identified deficiencies and learner recommendations for eliminating those deficiencies, it seems that the use and further development of these aspects of the model might be a fruitful area of research and development in tryout-revision. But first, the literature on the effectiveness of tryout-revision will be examined in the next section.

The Effectiveness of Tryout-Revision

In this section, the major empirical studies on tryout-revision will be reviewed. The section will consist of a review of each of the major models of tryout revision referred to in Chapter 1, plus a summary.

The basis of selection for literature reviewed in this section is, primarily, the provision of data showing that the revised version of the instructional system was an improvement over the prototype. In addition, some other literature will be included where something unique is described, with respect to the empirical use of the model, in spite of the absence of statistical findings.

Large group model

Gropper and Lumsdaine (1961) used prototype versions of two television programs on general science for 7th and 8th grade students for obtaining tryout data for revision of the programs. The tryout data consisted of test data, in the form of responses to objective and essay type test items.

Data was collected on the average percent of correct answers for each test item for each version of the program. Improvements between 67% and 86% were found. "In general," say the authors, "there were statistically significant improvements in test results for the revised version" (p. 13). However, the criterion level for statistical significance was not given.

An average of 10% improvement is shown for essay and objective test items and 11% for items on demonstrations. No indication was given as to whether the demonstration--type questions were a subset of the former or not. The authors also claim that the test data indicated which specific parts of the lesson were "too difficult for the students to learn or understand." No details are given on the use of instructional objectives for specifying such specific parts.

Gropper (1967) carried out a later study in order to determine whether the use of programmed learning techniques for the design of the stimuli materials, but without requiring active responding, leads to effective learning. He collected tryout data through the use of a kinescope of a television program for 7th graders.

The data included diagnostic test data and the results from a questionnaire which he used to ask for identification of what they thought they were supposed to learn and reasons why they thought the lesson failed in getting any points across that they had identified.

The revised versions were developed, based on this data; one was designed to facilitate retention of the material; the other was designed to facilitate transfer--that is, the ability to learn new material based on the learning they had just acquired.

He found that the revised versions produced total test scores that were 30% greater than that produced by the prototype. Moreover, the revised versions brought about an increase of 20% in "retention" scores and 40% in "transfer" scores, though twice as much learning time was required. These differences were significant at the .01 level. No significant differences were found between the two experimental versions.

However, on a delayed posttest given one month after the "immediate" posttest, no significant differences were found among the three treatment groups. In addition, the absolute performance levels on the revised versions were quite low--they attained approximately 50% of the total possible score, on the average.

VanderMeer and Thorne (1964) carried out a study based on a filmstrip, designed for use with learners in grades 5 to 12, entitled, "The Sun and its Planets." In Study I, on the basis of test data from the first administration of the posttest, consisting of multiple choice questions, frames were revised or eliminated. This involved both the direct use of the test data and the application of the experience, intuition and imagination of project personnel in attempting to improve the filmstrip based on interpretations of the test data.

The original filmstrip and a revised filmstrip were then shown to a second sample of learners in these grade levels, with random assignment of learners to one of the filmstrips. Significant differences between the test results for the two filmstrips were not found.

In Study II (VanderMeer, 1964), the filmstrip was revised on the basis of the results of Study I, and shown to learners in grades 5, 6, 7, and 10, with random assignment of learners to the second revision and the original filmstrip. One exception made in the data collection procedure was that the learners were told that each learner's grade on the postest would be reported to his/her science teacher. Statistical significances in mean posttest scores at the .01 level were found for grades 5 and 6; significance at the .05 level was found for grade 7; no statistical significance was found for grade 10.

A second filmstrip "The Earth's Satellite, the Moon" was developed as part of Study III in this series of publications (VanderMeer & Montgomery, 1964). The prototype was shown to learners in grades 5, 6, 8, 10, 11, and 12. As in the case of the first filmstrip, a revised version of the filmstrip was developed based on posttest data relating to multiple choice questions, from the learners viewing the prototype. Statistical significant results were found for grades 10, 11, and 12 at the .01 level of significance.

VanderMeer, et al. (1965) also applied the large group model to the development of two instructional films, "Why Foods Spoil." and "Atoms and Molecules." As in the case of the aforementioned studies involving filmstrips, prototype versions of the films were shown to learners in grades 5 to 12, and revised versions of the films were developed based on the posttest data, again consisting of responses to multiple choice questions. A second revised version was produced for the film, "Why Foods Spoil," on the basis of data from the administration of first revised version.

On the basis of the first revised versions, significant results at the .05 level of significance was found for only one of the six comparisons (grade 7) made (2 films x 3 grade-level groups). For the second revision of "Why Foods Spoil," statistical significance at the .01 level of significance was found in the results for grade levels 8 and 10-12, but not for grade levels 5-6.

On a broader curriculum scale based on individual instruction through use of self-instructional materials, Light & Reynolds (1972) applied tryout-revision to the mathematics curriculum for one elementary school classroom during an entire school year. Their evaluation efforts centered on the review of results of tests for lessons, each of which was based on a mathematics skill defined by a behavioral objective. This curriculum, which was in the process of development at the time of this study, required that the learner demonstrate mastery of each skill prior to being allowed to advance to the study of the next skill in the sequence.

The authors applied a three-step process. First, they identified an incident where the learner did not perform well on a lesson test. Next, they identified the cause of the incident. The third step was the identification of a problem solution.

In order to find an appropriate problem solution, the following questions were asked:

1. What was similar about the items missed on the test?
2. How did the items missed differ from those items passed on the test?
3. Where in the instructional materials was the content presented?
4. What in the instructional materials could have caused the test failure?
5. How can the hypothesized cause of failure be experimentally tested (Light and Reynolds, 1972, p.55)?

Although this type of data collection, analysis and interpretation is much different from that followed by VanderMeer and by Gropper and Lumsdaine, it is still an application of the large group model, even though the data is collected, analyzed, and interpreted on the basis of the test results of one learner at a time.

Baker (1970) made use of the large group model in a study of the procedures of development rather than for the validation of a revised version of an instructional system. Her study is included simply as another example of the use of the large group model. In this study 10 students in a course in programmed instruction were given the assignment to develop a program and to obtain tryout data from two fifth grade learners. This data was summarized and submitted to the instructor with the program. The programs, together with the tryout data were then randomly assigned among the students in the programming class, who were told to revise the program according to a given set of rules. These revised programs were subsequently tried out with other learners. Differences in the test scores were found to be significant at the .01 level of significance.

In a study concerned with the comparison of the relative effectiveness of expert evaluation versus learner tryout data, Rosen (1968) asked each of twenty teachers to view a 25-minute slide-tape program on English money. The author then asked each teacher to prepare a 15-minute video program as a supplement to the former in order to improve the effectiveness of a total lesson based on a viewing of the basic program and the supplement. Half the teachers were told to use their teaching experience for preparing the video program, subsequent to studying the general objectives for and viewing the base program. The other 10 teachers were also given the results of a test given to a sample of sixth graders who had been shown the base program.

The base program together with one of the video programs was then shown to each of 20 sixth-grade classes, which was followed by the administration of a posttest. The results showed statistically significant results in favor of the video programs prepared by teachers who had been given the earlier test results, at the .01 level of significance.

Sulzen (1972), studied the multiple revision of five slide-tape programs and one programmed text, all of which were used for military training. The author found that the revised programs were all significantly better at the .01 level of significance, than the prototype versions, across subject matter areas, type of learners, and the number of the revision in the revision cycle. Revision was based on the use of tryout test data.

In a 1972 doctoral dissertation on the revision of programmed texts on problem solving for elementary school students, Robinson found, as reported by Nathenson and Henderson (1980), that revisions based on tryout test data led to better results than those revisions not based on tryout test data. The level of significance was .001.

In a comparative study of the large group and tutorial models, Kandaswamy (1976) found that all the revised versions of an instructional program in algebra used in an elementary school in India led to statistically significant improvements in posttest scores over the prototype at the .001 level of significance. Revisions were based on analysis and interpretation of pretest and posttest data from learners using the prototype.

Summary and discussion of the large group model

As can be seen from these exemplars of the large group model and as was pointed out in Chapter 1, this model almost exclusively involves the use of test data resulting from use of the prototype as the basis of revisions made in the instructional system. Gropper (1967), was the only study reviewed which involved the use of other type data. As was pointed out previously in this chapter, he was concerned with the identification of the difficulties experienced by the learners, as perceived by the learners and not on the basis of inferences drawn from test data.

Tutorial model

As pointed out in Chapter 1, the tutorial model involves the collection of process data, by virtue of trying out a prototype with one learner at a time and making changes in the instructional system to meet the needs of that learner.

Silverman et al., (1964) carried out a group of experiments of an exploratory nature, involving the use of an experimenter acting as a tutor who interacted with a single learner at a time. The learning materials were programmed texts in reading, arithmetic, Spanish and geometry.

The tutor-experimenter would try out a prototype with a learner and provide tutorial assistance, as needed, in response to verbal requests for such assistance, puzzled looks and other non-verbal messages, and test item errors. Records were maintained of these tutor-learner interactions and whenever the same type of tutorial help was needed by three or more learners the program was revised, accordingly.

The tutor would make a subjective decision to cease the tutoring process whenever it was felt that the program had been revised "sufficiently," a term that was not operationally defined. At this point, the prototype and the revised version would be compared. If the latter proved to be "not much longer" than the original and "statistically superior" (neither term was operationally defined) the tutorial process would be terminated; otherwise additional tutoring sessions would be held. Up to 29 iterations of the tutoring process were used during the tryout. This general process was called, "tutorial engineering."

The only program for which details of the statistical analysis was given was the Spanish program. Four tests were used, and the revised versions of the program were significantly better at the following levels of significance: .01 for listening, writing and speaking; .05 for reading. The authors asserted that for the other three programs the revised version showed statistically significant results, without providing information on the levels of significance.

Robeck (1965) developed two revised versions of a prototype programmed text on English money on the basis of two tutorial tryouts. The first revision was based on the outcome of tutorial tryout with a "bright" 6th grade student. The revised program was then tried on a second learner and was further revised as a result. The prototype and the two revisions were then tested with matched groups of learners. The first revision led to statistically significant improvement in test scores at the .05 level of significance; the second revision led to statistically significant improvement in test scores at the .01 level. There was no statistical significance between the test scores for the two revised versions.

A third empirical report on the use of the tutorial model is that of Fleming (1963), reporting on educational film production at Indiana University. In order to increase the quality of a particular film they were developing, they stated the purposes of the film in advance in the form of 40 test items that they felt a 7th grade student should be able to answer correctly if he/she had learned the subject.

This list of questions was used to check each new version of the film or storyboard version, to determine if it had improved from the previous version. The film or storyboard was shown to one student at a time, one sequence at a time. "Then," says Fleming, "we stopped and discussed, pried and probed to see what the student thought he had seen and heard and what he had made of it (p. 18)". Fleming asserts that the film was improved "a specific 20% (p. 18)" through this process.

Kandaswamy (1976) made use of the tutorial model, as was pointed out in the review of the large group model, since he was comparing these two models. As was pointed out in that sub section, he found statistically significant improvements in the revised versions of the instructional system, regardless of the model of tryout-revision that was used.

Summary and discussion of the tutorial model

Thus, the tutorial model has been shown to be a successful model on the limited basis of these four reports. The literature does not reveal the use of this model in tryout-revision practice since the late 1960s.

However, the model has some distinct advantages, which should not be overlooked. It provides an opportunity for a tutor to interact with a learner while in the process of learning. By doing so, first-hand knowledge can be gained about the learner's learning difficulties as it occurs. The specific stimuli related to the learning difficulty can then be pinpointed. This makes possible the acquisition of much critical information about the instructional stimuli, which may never be obtained in any other way.

As is suggested by some writers, such as Dick (1977, 1980) and Briggs and Wager (1981), in a theoretical context, the tutorial model is useful in the early stages of the development of an instructional system, even though there is no documentation of this in the recent literature. Moreover, the writer feels that the tutorial model could be used to advantage for the specification of learning difficulties even in the later development stages of developing an instructional system.

For example, if large group data based on criterion-referenced test items points to a specific slide in slide-tape program as being the source of learning difficulties, it would seem that one highly useful strategy, for specifying the precise nature of those learning difficulties would be the use of the tutorial model. It shouldn't matter that these difficulties have been identified fairly late in the development process.

Abedor makes use of limited aspects of the tutorial model in his MK II small group model, as will be discussed in the next sub-section.

<u>Small group model</u>

The small group model involves the use of a small group of learners for trying out the prototype. The literature reveals that the only empirical study in which this model has been used was Abedor's (1971, 1972) study. The same study is reported in these two publications.

As was pointed out in Chapter 1, Abedor combined aspects of the large group and tutorial models. The tryout data consists of test data, questionnaire data, group debriefing data, and process data collected during tutorial sessions between the author/developer and individual learners, when such help is required.

The study consisted of five field trials, each of which involved the use of a prototype multi-media instructional system. These instructional systems were in three subject areas; namely, animal husbandry, industrial arts and biology. For other general aspects of the study, the reader is referred to Chapter 1.

However, in two of the trials, the developer found that revision was not necessary on the basis of the tryout data. For the remaining three field trials, the test data showed that the revised versions led to statistically significant improvements in learning at the .01 level of significance for two of the field trials and at a level between .01 and .05 for the other field trial.

Data was also collected that pertained to the proportion of learners who attained 80% criterion (that is, the attainment of scores representing 80% of the maximum possible score). In one field trial, the proportion of learners reaching this criterion through use of the revised version of the instructional system was significant at the .05 level of significance; in a second field trial, this proportion was significant at the .01 level; and in the third field trial this proportion was not statistically significant.

The questionnaire data also showed significant results in the same two field trials where significance was found for reaching 80% criterion. In both field trials the difference in mean attitudinal scores, in favor of the revised version, was significant at the .01 level of significance.

Summary and discussion of the small group model

Thus, the small group model, as operationalized by Abedor, is a combination of the large group and tutorial models, with the addition of a group debriefing, a unique feature of this operationalized model. The literature does not indicate that any other researchers or instructional developers have operationalized this model, or replicated the use of Abedor's operationalization.

Similar to other hybrids, this model is characterized, by some of the advantages of the two "source models" from which it was derived; however, some of the advantages of the source models are lost through the hybridization process. Thus, the use of this model makes possible the limited use of process data collected through tutorial interactions with learners during the tryout of the prototype. However, it is still a small group situation, with learners interacting with the instructional system on an individualized basis but in the context of learning in a room with other learners similarly occupied. Thus, the tutorial interactions are not necessarily as good as when a tutor and a learner are able to interact without the influence and/or distractions resulting from other learners being in the room, as in the case of the tutorial model.

Similarly, the use of 6-8 learners makes it possible to collect outcome data, which can be quickly analyzed and used to set the agenda for the group debriefing. But, due to the statistical unreliability of data based on such small numbers of learners, such data is not as good as outcome data based on larger numbers of learners, as in the case of the large group model. The advantages of this hybridization are that the combined use of limited amounts of process and outcome data and the use of a small group debriefing session is made possible.

Combined models

It was pointed out in Chapter 1 that the theoretical literature on tryout-revision indicates that, most typically, combinations of the tutorial and large group models and the combination of all three of the basic models are used in instructional development. These models were said to be used during subsequent phases of the development process, with larger numbers of learners involved in the tryout-revision process as the prototype became better developed. This, however, was not completely borne out during the search of the empirical literature. Instances were found of the combined use of the tutorial and large group models, with the tutorial model used in the early stages of the development process and the large group model used in the later stages of development. However, only one example of the use of all three models was found. This literature search also revealed instances of the simultaneous use of the large and small group models.

The best known examples of the combined use of the tutorial and large group models have been reported by D. Markle. In one study (D. Markle, 1965), he used criterion questions to structure the scripts in the development of a series of films on programmed instruction for the U.S. Office of Education. An elementary school teacher was given page-long sections of a draft script to read. Upon finishing, the teacher was asked orally the criterion question for that section of the script. The teacher was also encouraged to comment freely. Three teachers were used during this process, which can be considered a unique application of the tutorial model.

On the basis of this tutorial-type tryout, the script and the criterion questions were revised. Then, another set of teachers were used for a second tutorial-type tryout. This process was followed by the production of a "rough version" of the film (that is, a prototype).

The prototype was shown to three groups of teachers, who were then asked to respond to a questionnaire. The questionnaire contained the criterion questions, open-end questions and several questions asking for their advice on how to improve the film. This second phase of the tryout can be considered an application of the large group model.

Similar to other applications of the tutorial and large group models, the second phase of the tryout led to very general comments about the film. The tutorial tryout on the other hand, led to detailed comments which proved useful in the revision process.

D. Markle (1967) used this same combined model of tryout-revision in the development of a 7 1/2 hour first aid course. He used the model for establishing course objectives, determining the learning sequence, revising print and film media and for the devleopment of the evaluation instruments.

By use of the tutorial model, he was able to add instructional materials to the course on a gradual basis until students were performing at an acceptable level. Films were developed through a similar process, by starting with a set of black and white "still" pictures as the original "film" and adding additional shots and color in response to learner feedback. Once a sequence in a film or some other unit of instruction was found to be successful on a tutorial basis, it was tried out with a larger group until they were performing at a 90% criterion level.

A good exemplar of the use of all three basic models of tryout-revision was reported by Anderson (1968). This study involved the development of a self-instructional program in population genetics for use in high school biology, which was sponsored by the Biological Science Curriculum Study.

In this study, the first version of a segment of the program was tried out with individual learners, with one of the authors "monitoring each student's performance (p. 8)". Revisions were made after every few learners completed the segment. Each segment was developed, using this tutorial approach. After all the segments had been tried out individually, the entire program was tried out with small groups of learners, with revisions made on the basis of the results. This can be considered an application of the small group model, though its much different from Abedor's (1971, 1972) operationalization of this model.

The final phase of the development process consisted of testing the program with 750 high school biology students in 30 classrooms in 2 high schools. This phase can be considered an application of the large group model.

A criterion-referenced test was used throughout the development process. This test consisted of open-end, constructed response items, problems, and concepts and principles to be defined and illustrated.

Nathenson and Henderson (1980) report on the simultaneous use of the small group and large group models at the Open University (OU) in Great Britain. OU makes use of these models in very unique ways, within the context of an instructional delivery system which is also quite unique. The following quotation gives a brief description of OU courses:

> The core of each teaching unit is a specially-written correspondence text, together with one television and one radio programme broadcast by the British Broadcasting Corporation (BBC). Associated with each unit may be readings from text books, assignments, broadcast notes and other supplementary materials of various kinds. In addition, students following the course have the opportunity of attending a weekly group tutorial at a study centre near their home, and are required to attend a one-week residential summer school. (Nathenson and Henderson, 1980, p. 84).

The OU tryout-revision system uses second drafts of authors' correspondence units, printed offset, radio programs recorded on cassettes rather than being broadcast, and television programs viewed at the local study centres, rather than being broadcast. An evaluator, who is one of the educational technologists on the course team that is developing the course, conducts the tryout. The evaluator prepares in-text feedback questions for each unit of the course, and these are incorporated into the draft correspondence text.

Responses to the feedback questions are mailed to the evaluator together with records of study-time and copies of their assignments. This constitutes an application of the large group model. The small group model is also used by virtue of the use of group interviews with one of the two tutorial groups, which meets at the study centre, participating in the tryout. These group interviews made it possible for the evaluator to investigate specific issues. No indication is given as to whether the group interviews occur before or after the large group data has been analyzed. If it is after such analysis, it implies a reversal in the usual sequence in using the large group and small group models and implies the use of large group data as the basis for deciding which specific issues to investigate. No statistical data is reported; however, the authors report that, for the case study presented to describe their tryout system, the learners, who were given the opportunity to examine the final version of the course's correspondence text, were unanimous in their opinion that the course had been greatly improved.

Summary and discussion of the combined models

This literature showes how the unique advantages of the different models of tryout-revision can be incorporated into a single mulit-phase system. By uses of the tutorial model, it is possible to eliminate some of the gross deficiencies of the prototype without necessitating the use of many learners. Then a large sample can be used to check the workability of the system, having eliminated those gross deficiencies.

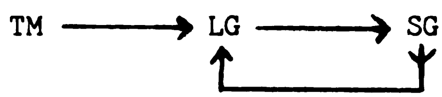
The question arises as to what to do if the prototype still shows need of revision on the basis of the outcome data from this second phase. The OU study indicates that special issues can be investigated through use of the small group model and conducting a group interview or "debriefing."

The advantage of the debriefing, as operationalized by Abedor, is the posting of problems, which facilitates the structuring of the communications and the recording and verification of the major points made during the discussion. The individual members of the group can then be asked to respond to specific problems that have been posted and to develop a consensus based on these problems through voting and the assignment of priorities and/or weights to these problems. In addition, separate listings can be posted for the remediation of the problems. Otherwise, group interviewing might not lead to conclusive results.

The specific advantage of the OU system is that large group data can be used as the basis of setting the agenda for the small group discussion, so that the time of the group is spent on significant problems. This is an improvement over Abedor's MK II model, where the agenda is set by the outcome data of the 6 - 8 learners in the small group.

In this way, it is possible to have a tutorial model (TM) ⟶ large group (LG) ⟶ small group (SG) paradigm instead of the TM ⟶ SG ⟶ LG paradigm suggested by Dick (1977, 1980) and Briggs and Wager (1981) and as implied by Baker (1974). Moreover, it would be desireable to recycle the SG-LG iterations until all the major deficiencies have been eliminated; that is, to obtain new outcome data through additional LG tryouts that would indicate whether the deficiencies discussed in the SG interviews have been eliminated.

Symbolically this paradigm would be:



The significance of these combined models to this study was that they gave encouragement to the development of the formulation of the ideal Computer Conferencing Model of Tryout-Revision which served as the basis for the more limited computer conferencing model developed in this study.

Summary and discussion of the effectiveness of tryout-revision

In reviewing this empirical literature, it seems clear that tryout-revision is effective, even if, as has been pointed out by Baker and Alkin (1973), Dick (1980), and Henderson and Nathenson (1980), the case for tryout-revision is based more on "the cumulative evidence," (cf. Henderson and Nathenson, 1980, p. 167) than on a solidly corroborative set of research results.

The implications that this literature had for the design of the computer conferencing model used in this study were as follows:

1. Direct learner feedback in general, as opposed to the implications that can be derived from outcome data, can be quite valuable in revising an instructional system.
2. Direct learner feedback on why the learners think they

answered a question incorrectly can be particularly valuable, as Gropper (1967) suggested.

3. Group interviewing can be a useful tool for exploring particular issues with tryout learners.

4. The small group model, as operationalized by Abedor (1971, 1972) is an effective means for revising instructional systems.

Need for a New or Revised Model of Tryout-Revision?

In spite of the dissatisfaction expressed regarding the state of the art of tryout-revision and the limited effectiveness of tryout-revision through the use of these models, the literature reveals no expressed interest in the need for a new model. Furthermore, the need for revision of any of the existing models has not been expressed in the literature. Thus, it seems that in spite of the problem that appears to exist, the call for a solution has not been heard. This situation is hard to understand, in view of the fact that instructional developers tend to be so problem-solution oriented.

Computer Conferencing Literature

The computer conferencing literature was searched for applications of this technology which contained information or ideas that had relevance to this study.

The only literature found that met this criterion were the reports of two experiments carried out at the New Jersey Institute of Technology, Center for Computerized Conferencing and Communciation. Both studies relate to the use of computer conferencing for the purpose of consensus formation. These studies are the first two in a series of three experiments designed for studying the use of computer conferencing as a decision support system.

In the first experiment (Hiltz, et. al., 1980), face to face communication was compared with computer conferencing in the solving of two problems by undergraduate and continuing education students at Upsala College, East Orange, N.J. One problem called, "Forest Ranger," required group decisions regarding various aspects of a human relations problem. The second problem called, "Lost in the Arctic," first requires each member of a group to rank 15 items important for survival in the subarctic. The group is then required to reach group consensus on the rank ordering of such a list.

The computer conferencing group made use of the computer conferencing system, Electronic Information Exchange System (EIES). The conference was conducted in the synchronous mode, and the computer terminals were located in different rooms. For the total experiment, 16 groups of 5 students each were used. Each group engaged in the two problems, using a different communication mode for each problem. Thus there were four treatment groups per problem per communciation mode. Each group was allowed a half hour for the first problem and one and one-half hours for the second problem. The same time limits was set for both communication modes.

The findings that were of major relevance to the present study are as follows.

> .There is no difference in the quality of solution reached between the two modes of communciation.
> .Face to face groups are significantly more likely to be able to reach total consensus on the solution to the problem. . . .

. . . .
.There are two to three times as many communication units generated in face to face meetings as in computerized conferences, within the same time period (Hiltz, et. al., 1980, p. 2).

Also of interest to the present study was the computer conferencing training and selection of subjects. They were trained for a half an hour one week prior to the experiment and then tested on their skills in using the computer commands they would be using. For each treatment group, 6 to 7 potential subjects were trained, of which 5 were selected.

The second experiment (Hiltz, et.al., 1981) consisted of a field experiment for which staff members of various business organizations served as the subjects. The organizations included Banker's Trust, Texas Instruments, and Chemical Abstracts, Inc. Training was increased to one hour, which included two practice problems. The time allowed for the Arctic problem was increased from one and one half hours to two hours.

Other important differences in the second experiment included the use of group-elected leaders for leading the group discussion in half the treatment groups. This "human leadership" factor was used as a variable in comparing treatment groups. A second new variable included in the second experiment was the use of "computer feedback" for the rank orderings. This occurred after initial rankings but before the discussion began and after a given number of vote changes occurred.

The authors concluded that human leadership or computer feedback, but not both, significantly aided groups in reaching consensus.

Consensus was measured by Kendall's coefficent of consensus, which was used for both studies.

In summary, the findings from the first experiment listed above comparing face-to-face communication with computer conferencing communication were of particular interest for the present study. This was due to the fact that the two studies involved the comparison of these two types of communication, and that the two studies were concerned with problem solving.

Thus, the fact that no difference was found in the quality of solution reached between the computer conferencing groups and the face-to-face groups gave encouragement to the development of the computer conferencing model, as a potentially workable model for the debriefing of prototype tryout learners. However, it was somewhat disconcerting to learn that the likelihood for reaching total consensus was greater for the face-to-face groups than for the computer conferencing groups since that implied that the quality of communications was greater in the case of face-to-face communications than in the case of computer conferencing communication.

The reader may recall that one of the hypotheses discussed in Chapter 1, predicted better performance for the learners trying out the CC revision than those trying out the SG revision. This prediction was based on the presumption of a higher level of communication in a computer conference than in a face-to-face group. In part, this was based on the assumption of a greater number of communications per unit of time in computer conferencing. This assumption also was not supported by the experiment carried out by Hiltz and others.

The reason why the latter two findings were discounted in the present study was due to the fact that in the present study the plan was to use a different computer conferencing system, called, Confer. (N.B. Confer is also written, CONFER, but it is not an acronym.) CONFER has features not found in EIES, which makes CONFER a faster system when those features are used. The plan for operationalizing the computer conference for this study included the use of those features. Therefore, through there was no way to determine how much faster CONFER would be than EIES, the finding in the Hiltz study regarding the difference in the frequency of communication units between the two communication modes was considered to be open to question, if CONFER were used.

Moreover, since the ability to reach consensus is, to some degree, related to the frequency of communiction units, and since the present study would be using a debriefer to foment consensus, the finding in the Hiltz study regarding consensus was also temporarily ignored.

Another aspect of the two experiments carried out by Hiltz and others that was of interest to the present study was the use of rank orderings, which lends itself to quantitative analysis. However, the present study was designed, in part, to replicate Abedor's MK II small group model and that model doesn't use rank orderings; therefore, rank ordering was not utilized. But, in future research with this model, the use of this more precise way of handling consensus could certainly be used. In addition, the use of EIES could be considered, especially if it could provide automatic feedback of rank orderings during the debriefing, as was done in the second experiment, as noted previously?

In conclusion, the computer conferencing literature does not seem to have many ideas to offer that are appropriate to beginning a computer comferencing model of tryout-revision. This is probably due to the fact that most uses of computer conferencing so far seem to be for information sharing among colleagues rather than for collecting data, as in the case of the debriefing of prototype tryout learners.

Formative Evaluation Literature Relating to Use of Computer Conferencing

Finally, the formative evaluation literature was searched for evidence that previous work has been done in applying computer conferencing to tryout-revision. However, no such literature was found. Thus, it was concluded that this study is the first such study in this area.

Chapter Summary

The following conclusions were reached on the basis of this literature search:

1. The literature on evaluation and evaluation research indicates that in both these fields evaluation is, in general, for the purpose of supporting decision-making. Therefore, evaluation data must be relevant to the type of decision that it purportedly serves.
2. The tryout-revision literature is in general, unsatisfactory, due to the fact that it doesn't provide much guidance to instructional developers in the revision of instructional systems.

3. The known models of tryout-revision show limited effectiveness on the basis of the literature. However, the general conclusions drawn by various instructional developers is that they are generally effective. What is lacking, it seems, is a body of corroborative research findings that would make it possible to draw high level generalizations across many types of learners, instructional systems, subject matter, and learning environments, with respect to the effectiveness of different models of tryout-revision that were applied according to prescriptive principles. In other words, what is lacking is an armamentarium of tryout-revision techniques, which can be selected or prescribed on the basis of the requirements of the situation at hand, as is the case in statistical methodology, and survey research.
4. There has been no expressed need for a new model of tryout-revision or for improvement in any of the existing models.
5. The computer conferencing literature does not indicate much support of tryout-revision research and development.
6. There have been no previous applications of computer conferencing to tryout-revision.

CHAPTER 3

INSTRUMENTATION AND STUDY DESIGN

Introduction

This chapter consists of three main parts. First, the two models of tryout-revision developed in this study will be described; namely, the Small Group Model of Tryout-Revision (SGTR) and the Computer Conferencing Model of Tryout Revision (CCTR). Next, the development of CCTR prior to its initial use in the pilot study will be discussed. Lastly, the research study based on the evaluation and comparison of these two models will be presented. This will include descriptions of the various procedures and the research design.

The Two Models of Tryout-Revision

The two models of tryout-revision developed in this study are both based on Abedor's (1970, 1971) MK II model. That model consisted of a technical assessment of an instructional prototype and a learner tryout of that prototype carried out by the author/developer. During the learner tryout, three types of data were collected; namely, process data, outcome data and group debriefing data. All three types of data were used as a basis for prototype revision.

The process data consisted of observations of negative non-verbal communication expressed by the learners while trying out the prototype and the transactions of tutorial interactions between the author/developer and individual learners; outcome data consisted of pretest and posttest data and questionnaire responses; group debriefing data consisted of the transaction of group discussions (debriefings), held after the tryout of the prototype, during which the author/developer attempted to obtain information from the learners about the deficiencies in the prototype and their recommendations for removing these deficiencies. Consensus of opinion was also sought on these matters.

SGTR was essentially based on the MK II model, with the following differences:

1. The only data used for prototype revision was the debriefing data. The outcome data was not used for such purposes.
2. The group debriefings were not conducted by the author/developer of the instructional system, but by other persons trained for this task.
3. The tryouts were carried out on a group basis rather than on an individual basis.

SGTR was carried out in this manner for two main reasons. First, to ease and simplify the task of data analysis for the purpose of determining the revision strategies; second, to encourage the use of MK II, in view of its decreased dependence on the author/developer and the simplified data analysis required for determining revision strategies, while providing direct learner feedback, on revision needs.

On the other hand, other sources of revision ideas and strategies were lost. Since the tryouts were carried out on a group basis, no individual tutorial data, based on learner - author/developer or learner - debriefer interaction could be collected. Moreover, the outcome data was not used for revision purposes, as mentioned above. Due to the absence of these other sources of data, it was not possible to benefit from the corroboration of facts derived from two or more sources. Thus, opinions expressed by learners about prototype deficiencies during the debriefing could not be matched with specific learning difficulties revealed during tutorial interactions. Moreover, without an analysis of the patterns of the wrong answers in outcome data, possible clues were lost, regarding the effect of the various instructional stimuli on learning behavior and its relationship to comments made during the debriefing.

CCTR, on the other hand, is a variation of SGTR which is a variation of MK II. The only difference between the two models is that the former utilizes computer conferencing rather than face-to-face communication as the medium of communication during the debriefing process. Some of the implications of this difference are that the participants can be located at different sites and that they can communicate asynchronously; that is, they can communicate during different time periods.

In this study, however, the participants involved in the computer conferencing model used computer terminals located in the same room, and communicated during the same time period; that is, the computer conference was operating in the synchronous mode. Participants were discouraged from talking with each other and, with few exceptions, communicated by computer only.

The flowchart shown in Figure 3.1, which follows, describes the tryout-revision process for these two models. The procedures for carrying out each step in the flowchart for each model will be described in detail in the section, "Phase I Data Collection--Prototype Tryout and Debriefing." Therefore, such details will not be explained here.

Pre-Pilot-Study Development of the Computer Conferencing Model of Tryout-Revision (CCTR)

Design considerations for the group debriefing

On the basis of general knowledge about facilitating the user acceptance of an innovation it was decided that this initial prototype version of CCTR should have the following design characteristics:

1. Ease of use, with respect to amount of typing required.
2. Limited amount of training.
3. Technical assistance provided regarding the use of the hardware and the computer conferencing system.
4. Assistance provided regarding procedures to be used.

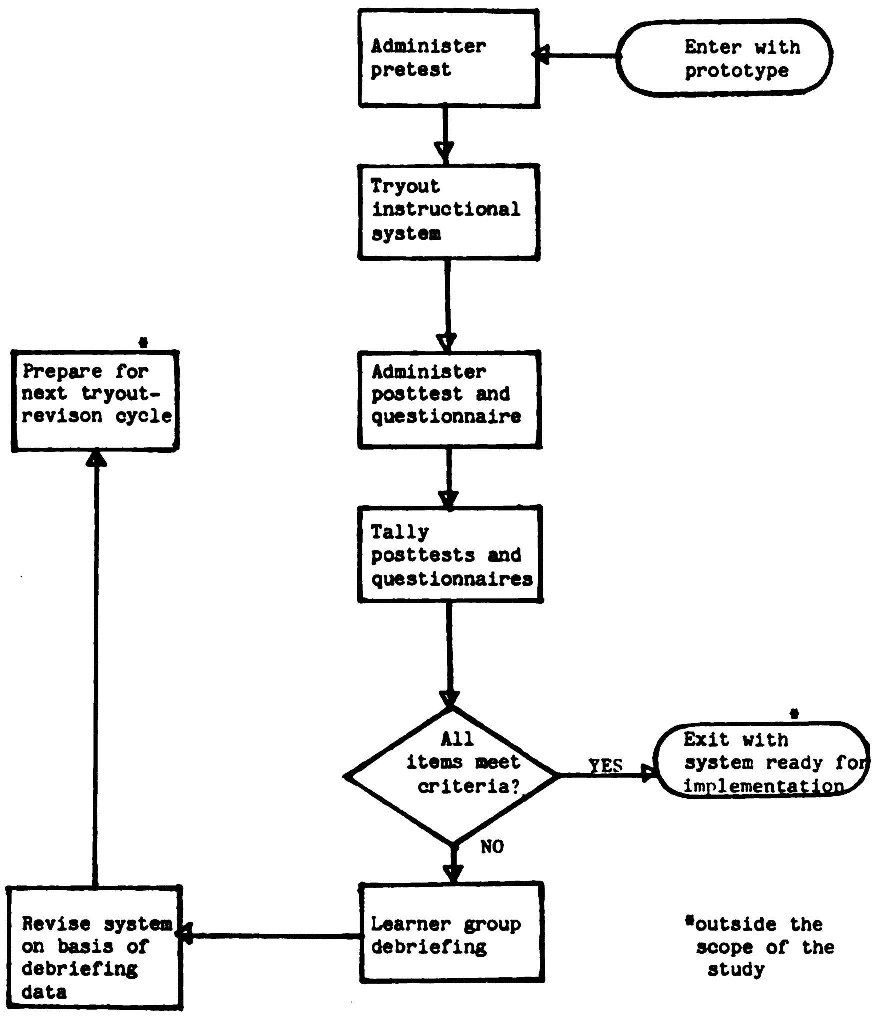


Figure 3.1
Flowchart of Small Group and Computer Conferencing Models of Tryout-Revision (SGTR and CCTR, Respectively).

Design features of the group debriefing imposed by the research plan

As indicated in Chapter 1, the plan was to modify Abedor's (1971, 1972) MK II small group model so that tryout data would be limited to debriefing data. The plan, then, provided for a comparison between this modified model MK II and CCTR in such a way that the communication mode during debriefing was the only major variable. Therefore, CCTR had to be synchronous in order to match the natural synchroneity of the face-to-face debriefing of the modified MK II model. (However, there is no general need that computer conferencing applications to tryout-revision be operationalized in the synchronous mode, since this mode does not make the best use of the unique characteristics of computer conferencing. This point is also made in Chapter 5.)

Furthermore, the learners for CCTR were gathered in one room for the group debriefing, in order to match the modified MK II model with respect to the debriefing process, as discussed earlier. However, the debriefer was located in a different room, so as to insure that all communication between the debriefer and the learners was via computer. Admittedly, this caused a minor difference between the models; however, it was felt that this arrangement was necessary for the success of the study. Moreover, this arrangement provided for a greater similarity to the way such a debriefing would be carried out in professional practice, where the learners would not necessarily be located where the debriefer would be located.

Initial design of the group debriefing

On the basis of the design considerations for the group debriefing mentioned earlier, it was decided that the debriefing process should be as easy to learn and to use as possible, and should be provided with as much supplementary assistance as possible, in order to counteract the anticipated resistence to the innovation. These basic considerations influenced the design of CCTR from its earliest stages.

The two best know computer conferencing systems available in the United States, Electronic Information Exchange System (EIES) and CONFER, which is not an acronym, were considered as potential systems for this study. However, due to relative cost considerations, CONFER was the only system that was seriously considered.

CONFER was then considered from the point of view of how to keep the debriefing process as simple as possible. Discussion voting was selected as the mechanism for debriefing. By doing so, the learner did not have to be trained how to enter a conference item, the basic unit of commnication in CONFER conferences. By the use of separate conference items for the discussion of different posttest and questionnaire items, it was possible to instruct the learners to view particular conference items and enter a discussion vote when so prompted by the computer. Another advantage of this procedure was that only a few items needed to be referred to, as opposed to the numerous items that would have been involved if each learner and the debriefer were entering items.

In addition, it was also decided to incorporate into discussion voting a notion that is found in the use of Delphi method (cf., for example, Linstone and Turoff, 1975) and the Nominal Group Technique (NGT) (cf., for example Scott and Deadrick, 1982) (e.g. Linstone and Turoff, 1975). In the use of those techniques, communication is structured so that personal influence does not affect the communication of opinion by the individual participants. In the case of the Delphi method, the individuals are located in different places and the opinion data is usually provided through use of a questionnaire. Subsequent to that they are shown information relating to how the other individuals responded. This process is re-cycled in several rounds.

The Nominal Group Technique (NGT) is, in some ways, similar except that the individuals are located in the same place in a "nominal" group; that is, they are together, but the process is structured so as to prevent the interaction that would result in the undue influence of some members over others.

It was decided to incorporate that general notion into the study by asking all the CCTR participants to view a conference item and to respond simultaneously to one of two type questions for each of the posttest and questionnaire items discussed during the debriefing; namely, "Why do you think some of the learners answered this question incorrectly?" or "Why do you think some of the learners did not rate this questionnaire item positively?" By doing so no one would see the discussion votes of any of the other participants prior to entering a first vote for a particular item, and thereby not be influenced by the votes of other participants. However, subsequent discussion voting for a particular conference item was influenced by the reading of the discussion votes of the other participants. Thus, discussion voting, as used in this study, was a combination of independent and interactive behavior.

To speed-up the CCTR process it was decided to use the command modifiers, "votes" and "votes new." These modifiers were used when requesting a particular conference item. In response, the participant is shown only the votes or only the "new" votes; that is, the votes entered since a particular conference items was last viewed by the participant. Hence, the pace of the debriefing process was increased by restricting the viewing to only the votes or new votes, after the learner has already voted on a particular item.

Since all the participants were to be on-line at the same time, it seemed quite feasible to require the use of these modifiers so as to increase the rate of flow of communications during the debriefing.

In order to facilitate a feeling of comfort about the debriefing process, in general, it was decided to provide for the sending of messages between the debriefer and the learners. In addition, there were some more practical reasons for maintaining this type of communication. It was decided that in order to lighten the amount of training needed to make successful use of the computer conferencing system, that the procedure for registering each participant into the conference would be handled by the writer. However, this required knowlege of the name of each participant and such information would not be known until the time of the prototype tryout when learners would be randomly assigned to one of the two tryout-revision models.

To resolve this problem, it was decided to register the participants into the conference with false names. Then, at the time the participant signed-on (i.e. made connection with the host computer and received authorization to begin interacting with the computer conference) the participant could change his/her name to the real name. Following this, the participant could inform the debriefer that he/she had signed-on and had changed his/her name.

This procedures would let the debriefer know when each participant was ready for participation in the conference, as well as to serve as a mechanism for the debriefer to inform the participants as to the first conference item to view. The learners could be advised at the end of the debriefing on a given posttest or questionnaire item as to the next conference item to view, by means of a discussion vote by the debriefer in one of the conference items currently being discussed. (For example, "please view item 6 next.")

Another notion that was considered basic to the debriefing was that a learner would continue to return to conference items already voted on in order to view the votes not previously seen and to add new votes to the discussion. It is in these situations that the learner would be told to use the "votes" or "votes new" modifer in using the computer command for viewing an item.

It was decided to provide the participants with a training guide which would be discussed with them at the beginning of the computer conference. This will be discussed in a later sub-section of this section.

These, then, were the basic design components to be incorporated into the initial version of CCTR.

Selection and training of the debriefers

Two persons were selected for participation in the experiment as debriefers. Fortunately, two persons were found who appeared to be potentially equally skillful as debriefers. One of them had had some experience in the group facilitation process, but had never been involved in prototype tryout-revision. This person also had much experience helping students in as an audio-visual technician at Michigan State University, who thereby was familiar with the problem solving process. However, he had not had any formal training relating to tryout-revision, but had been involved with numerous television production projects and was familiar with the basic concept of tryout-revision.

He had "audited" the computer course in which the instructional prototype for the project had been used, and was thereby familiar with the instructional prototype. However, he did not feel comfortable in the use of a computer terminal, since he had not performed any of the practice exercies for that course.

The other person had had a course in instructional product design and was familiar with the tryout-revision process. She had had some experience with the individual learner debriefing process, but not with the group debriefing process. She had also completed the computer course in which the instructional prototype for this project had been used. Moreover, she felt quite comfortable in the use of a computer terminal.

Both of these individuals seemed interested in the project at about the same level. Though it was difficult to tell whether they would be equally skillful as debriefers, each seemed to have the requisite skills for one of the two types of debriefing to be used in the project.

A training session was held with the debriefers two days before the pilot study. The purposes of the training session were as follows:

a) To explain the procedures that the debriefers would be involved in and to clarify their duties and responsibilities.

b) To explore in an open communication environment how the debriefers might best facilitate the consensus formation process during the group debriefing of the learners.

c) To model and practice the debriefing process.

In preparation of the training session the writer prepared two memoranda to be used as a means of initiating discussion. The memoranda were discussed and the debriefers were encouraged to ask for further explanations and to suggest alternative procedures at any point in the discussion.

The writer then modeled what he thought would be a desireable way for the small group debriefer (SG debriefer). The debriefers acted the roles of learners being debriefed. After about five minutes of the simulation, it was critiqued. The SG debriefer then practiced the debriefer role and the writer assumed the role of one of the learners. The same subject matter was used for the second simulation. After about 10 minutes, the performance was critiqued. A large drawing pad was used for posting problems identified by the learners in the simulation.

Using a prototype version of the training guide, the writer and the computer conferencing debriefer (CC debriefer) engaged in a training session, using 2 computer terminals. After practicing the "logging-on" procedures a simulation of a debriefing session was carried out, applying the same process skills taught during the first training session, and making use of a practice conference item.

Following the pilot study, individual meetings were held with the debriefers in order to provide them with additional information and to obtain from them formative evaluation information. The CC debriefer and the writer spent considerable time in reviewing the training guide and the computer conference procedures.

Development of the prototype version of the training guide

One of the most important tasks that had to be carried out prior to the pilot study was the writing of a guide for training the participants assigned to the computer conferencing model how to make proper use of the computer terminals and how to participate in the computer conference. A prototype training guide was produced and given to the CC debriefer for review. This guide was revised several times prior to its final use during the study, in response to the debriefer's suggestions and in adjusting to various problems that were met during the course of its use with learners during the study. These changes will be alluded to in subsequent sections of this chapter. For a copy of the final version, see Appendix A.

Selection of an instructional prototype

In order to carry out the study, a self-instructional system was needed which was either an instructional prototype in the process of being developed or a developed instructional system, which was in need of improvement. In either event, it was necessary to obtain permission from the author/developer of the instructional system to use both the "prototype" and revised versions of it developed during the study, if the revised versions contained portions from the prototype.

The primary criterion for evaluating instructional systems as possible prototypes was the extent to which the author/developer was willing to allow the writer to use the instructional system as he wished during the study. Several systems were considered, but the one that was finally selected was one developed by a Michigan State University College of Education faculty member. This system is a television program entitled, "The History of Computing Machines and Basic Concepts" and is part of a series on computers in education. This author/developer regards the series as being in "the public domain" and therefore available for use in any way desired. For this reason, this system was rated very high on the criterion mentioned above. Moreover, in view of the fact that this television series was developed for a course offered by the Michigan State University College of Education, the use of this instructional prototype would make possible the further development of this system for the same population as for the original program. In other words, the development process could continue without changing the population of intended learners.

The combination of these two factors, then, made this instructional system a highly desirable one to use as the instructional prototype for this study, and it was therefore selected for that purpose.

Sample

The selection of the sample was determined, primarily, by the instructional prototype that was selected for use in the study. One of the major concerns was to insure that the experimental subjects be as similar to each other as possible with respect to the following variables, in view of their important relationship to learning:

1. Educational level.
2. Motivation to learn from the instruction.
3. Interest in the subject matter of the instruction.
4. Previous knowledge about the subject matter of the instruction.

Since the instructional prototype was originally developed for a graduate education course, it was appropriate to carry out this study with graduate education students. However, it was anticipated that there would be some practical difficulties in selecting experimental subjects for this study, for the following reasons:

1. Only one course in computer applications to education was offered at Michigan State University at the time of the data collection for this study. This class alone was not large enough to provide a sufficient number of experimental subjects.
2. Students taking other graduate education courses were not likely to be interested in computers in education and unlikely to volunteer for participation during non-class time.
3. The subject matter of the instructional prototype was not related to the instructional objectives of courses, other than

the computer course offered by the College of Education at Michigan State University. Therefore, the likelihood of finding faculty members who would be willing to make their classes available for the study during regular class time was limited.

4. The anticipated difficulties of finding experimental subjects at other academic institutions was considered too great for such a strategy to be practical.

In view of this appraisal of the situation, one strategy for obtaining a suitable sample seemed to be in soliciting the cooperation of instructors of graduate eduction courses that relate to the use of instructional media. Since media evaluation is an important aspect of education in instructional media, it seemed that the use of class time for viewing a television program could be justified as an appropriate use of class time in instructional media corses. Three such courses are offered each term in the College of Education at Michigan State University. The cooperation of the instructors for these courses was obtained for the purpose of collecting data for the study. The approach used in soliciting their cooperation was based on the potential opportunity for their using this class activity as an exercise in media evaluation.

Thus, strictly speaking, the sample for the study is drawn from a population of instructional media students enrolled in graduate education courses in instructional media at Michigan State University. In Chapter 5 it will be argued that the conclusions of the study can be extended to a more inclusive population.

In view of the limited size of enrollment in the three classes and the inability to include in the sample students who had previously viewed the instructional prototype, the total number of experimental subjects was anticipated to be quite small. Therefore, no special effort was made to restrict the sample to those who were at the same level for the four variables mentioned at the beginning of this section.

However, the experimental subjects were all graduate students and, thus, at the same general educational level. In addition, the statistical tests that were used included an adjustment for pretest scores, thereby eliminating any effects due to possible differences in previous subject knowledge. No effort was made to equalize levels of interest and motivation among the experimental subjects.

As previously indicated, the study involved the collection of data during two time periods separated by a period during which the revisions of the instructional prototype were developed. In view of the relatively small class enrollments of these instructional media courses and the inability to use the same class twice in the same term for an study unrelated to the objectives of these courses, the collection of data necessarily had to be spread over two academic terms.

Although this provided for a larger sample frame, it did not necessarily alleviate the problems of obtaining an adequate sample size. This was due to two facts. First, students who take one of the instructional media courses tend to take two or all three of the courses in the series. Second, the computer course which uses the instructional prototype was given Fall, Winter and Spring terms. Therefore, the use of the instructional media classes for this study during Spring and Summer terms necessarily ment the high likelihood of using classes with large proportions of students who had either taken the computer course or had already participated in the study in another instructional media class.

Thus, in order to insure an adequate sample size it was decided to use the outcome data provided by the students who would provide the debriefing data as the outcome data for the prototype, rather than obtain such data during the same time interval when outcome data for the two revisions would be collected. Though, at the time this decision was made the primary concern was that of obtaining a sufficient sample size, other implications of this decision came to bear on the interpretation of the results. These matters will be discussed in Chapter 5.

Development of Instruments

Pretest and posttest

The process of developing questions for the pretest and posttest began with a viewing of the television program several times while making notes on the content. An audiotape was also made of the program's audio content for future reference when writing instructional objectives and test items. Twelve instructional objectives were subsequently written after listening to the tape and reviewing the notes. On the basis of the objectives, 25 multiple choice test items were written.

The test items were sent to 3 subject matter experts in the area of computers in education for rating for face validity. Two were Professors of Education in the Educational Systems Development Program at Michigan State University, College of Education, at that time. One of them was scheduled to teach the course, Computer Applications in Education, for which the prototype had been developed, during Spring Term, 1981. The other Professor was scheduled to teach that course during Fall Term, 1981. Thus, both were preparing themselves to teach the course, and were provided with a copy of the television program to view prior to rating the questions. The third rater was a graduate student who was the co-instructor for the above mentioned course during Fall Term, 1980, Winter Term, 1981, and scheduled for same during Spring Term, 1981.

The 3 raters were asked to use the following rating scale:

5 - valid	2 - fairly invalid
4 - fairly valid	1 - invalid
3 - undecided	

They were given copies of the pretest and posttest but asked to rate the pretest only. The posttest consisted of the same multiple choice questions, but the items were sequenced differently than in the pretest and their alternative answers were sequenced differently than in the pretest. The raters were also provided with a copy of the answer sheet for each instrument and were asked to check them for correctness.

Of the 25 items listed on the instruments only 9 were rated by all 3 raters as being "fairly valid" or "valid". In order to obtain more test items a second rating form was sent to the raters with 10 new multiple choice test items, plus a few questions from the first rating form which had been reworded in accordance to feedback the raters provided on the first rating form. This resulted in 2 more test items being rated at least "fairly valid" by all three raters, for a total of 11 test items.

In considering the number of test items to be used, two considerations had to be kept in mind. On the one hand, there was the matter of test reliability, and test validity. In order to insure test validity, high test reliabiity is required, and the later is, in part, dependent upon a sufficient number of test items.

On the other hand, one of the purposes of the test was to provide the basis for selecting agenda items for the debrifing session during the tryout of the prototype. As pointed out later in this chapter the test results during the tryout of the prototype were analyzed and those items that were incorrectly answered by at least 30% of the learners was considered eligible for the debriefing agenda. A self-scored test was used to add to the efficiency of the analysis of the test results. However, the length of the test was the crucial factor, in view of its bearing on the amount of time needed for the testing, scoring and analysis of the results.

This was due to the fact that only a minimal amount of time was being made available by the instructor of the course in which the prototype was to be tried out. Hence, to keep the total time for the prototype tryout and debriefing within the amount of time being made available, it was decided to limit the posttest to 20-25 items.

In view of the difficulty in getting test items validated by the raters, the decision was made to write 11 true/false items based on those multiple choice items for a total of 22 items. It was assumed that the true/false items would have face validity since they represented alternative answers from the multiple choice questions.

The final versions of the pretest and posttest items were written after learner feedback from the pilot study had been analyzed. Split half tests of reliability were carried out with the pretest and posttest data for the experiment. These values were .511 and .777, respectively.

A copy of the pretest and posttest will be found in Appendices B and C, respectively.

Questionnaire

The questionnaire evolved over a period of several months. Eventually, a semi-final version was sent to three faculty members of Michigan State University, as face validity raters for the questionnaire. One rater was a Professor of Education and Educational Psychology in the Educational Psychology and Educational Systems Development programs; a second was a Professor of Telecommunication and Education in the Telecommunication Department and the Educational Systems Development program; and the third, was an Associate Professor in the Department of Telecommunciation.

The three raters completed a rating form, using the same rating scale as used for the pretest and posttest. They made various suggestions for re-wording many of the questionnaire items. These were considered along with comments from the learners who participated in the pilot study. For the latter the writer reviewed the audio tape and the sheets of drawing paper from their debriefing session. He also read their comments on the pilot study version of the questionnaire, which invited them to edit any of the questionnaire items, so that they became clearer, in their opinion.

The questionnaire used in the pilot study also listed two questions about the questionnaire; namely, "Were the questions clear?" and "Were the right questions asked?" The former question provided space for them to explain which questions were not clear to them and why. The latter question asked them to specify "Which questions should be eliminated?" and "What other questions should be asked?"

The statistics based on a sample of 15 learners for the closed-end questions were as follows:

Were the questions clear?

Yes	14
No	1

Were the right questions asked?

Yes	13
No	0
No answer	1
Didn't know	1

A second validation form was sent to the three raters shortly before the final version of the questionnaire was typed. Though they made more suggestions for re-wording, they all rated every questionnaire item at least "fairly valid."

Thus, on the basis of the validation forms and the learner feedback, it was concluded that it was a reasonably valid form. Numerous suggestions of the raters on the second validation form were incorporated into the final version of the questionnaire. A copy of the final version of the questionnaire will be found in Appendix D.

Research Design

The research design used in this study is a variation of what is referred to by Campbell and Stanley (1963) as "the nonequivalent control group design" (pp.47-50). This design, they point out, is "well worth using in many instances" when random selection of experimental subjects is impossible. The control group and the experimental do not have "pre-experimental sampling equivalence. Rather the groups constitute naturally assembled collectives such as classrooms, as similar as availability permits but not so similar that one can dispense with the pretest." (p.47)

In this study two experimental groups and one control group were used. The experimental groups were the groups using one of the two revised versions of the instructional system, and the control group was the group that used the instructional prototype. The treatment is, thus, the version of the instructional system; that is, the prototype or one of the revised versions.

This research design is illustrated below:

	O	X_1	O
R	O	X_2	O
R	O	X_3	O

The symbol, O, represents observations; that is the pretest, the posttest and the questionnaire, which were given before and after the treatment, which is represented by X. X_1 represents the prototype, X_2 and X_3 represent the two revisions. The line of dashes symbolizes the fact that the control group, which is represented by the row of symbols shown above the line, were not equated with the two experimental groups represented by the rows of symbols shown below the line.

Normally, this research design involves intact classes assigned to each treatment. However, in this study the two experimental groups were formed through the random assignment of learners in three intact classes to the treatments assigned to the two experimental groups. For that reason, the symbol, R, indicating random assignments, is shown before the "O" representing the pretest for the two experimental groups.

Moreover, the control group is made up of learners from two classes. The two experimental groups were randomly assigned from intact classes except for students who had previously used the prototype or who were absent from class on the night of the tryout and debriefing.

Thus, the research design used in this study does not exactly fit the non-equivalent control group design described by Campbell and Stanley (1963). However, the treatment groups are non-equivalent, but are very similar, as stipulated by Campbell and Stanley (1963).

Pretest score was used as a covariate in the statistical analysis to insure equality between the groups at the beginning of the experimental treatment.

As was pointed out, earlier in this chapter in the section, "Sample," the anticipation of the difficulties relating to the random sampling of the population of users of this prototype prompted the consideration of arranging for the use of intact classes. The decision to pursue this strategy determined, in large measure, the use of the non-equivalent control group design, in view of its appropriateness to the sampling situation.

Variables

Dependent variables

Two dependent variables were used in this study; namely, mean posttest score, and mean rating for attitude toward the instructional system. Each posttest score was adjusted, according to the subject's pretest score, by the appropriate computer program in the Statistical Package for the Social Sciences, Second Edition (Nie, et. al., 1975). The questionnaire consisted of ten 5-point scales, ranging from strongly agree to strongly disagree. Each scale was stated in the same direction, and their points were numerically coded in a similar way. The two dependent variables are defined, as follows:

> Mean posttest score -- the total of the posttest scores, after adjusting for pretest scores, for the subjects in a given treatment group, divided by the total number of subjects in that treatment group.
>
> Mean rating for attitude toward the instructional system -- the sum of the total questionnaire ratings for the subjects in a given treatment group, divided by the number of subjects

in that treatment group.

Independent variable

The only independent variable is the version of the instructional system; namely, the instructional prototype (often referred to in this study as "the prototype"); the small group revision (SG revision) and the computer conferencing revision (CC revision). The instructional prototype consisted of a videocassette, which was taken from a set of instructional materials forming an instructional module. The other module materials are not used in this study.

Research Hypotheses

In Chapter 1, the research questions and hypotheses were stated so as to provide the reader with a general orientation to the nature of the research involved in this study. For that reason the hypotheses were stated in a general way. The testable hypotheses are shown below. The following abbreviations will be used in these hypotheses:

data are as follows:

a. Hypotheses for comparing the prototype to the two revisions:

H_0: The mean posttest score, after adjusting for pretest score, for learners who try out a CC revision and for learners who try out an SG revision of an instructional system will be less than or equal to the mean posttest score, after adjusting for pretest score, for learners who try out the prototype version of that instructional system.

H_1: The mean posttest score, after adjusting for pretest score, for learners who try out a CC revision and for learners who try out an SG revision of an instructional system will be greater than the mean posttest score, after adjusting for pretest score, for learners who try out the prototype version of that instructional system.

b. Hypotheses for comparing the two revisions against each other:

H_0: The mean posttest score, after adjusting for pretest score, for learners who try out a CC revision of an instructional system will be less than or equal to the mean posttest score, after adjusting for pretest score, for learners who try out an SG revision of that instructional system.

H_2: The mean posttest score, after adjusting for pretest score, for learners who try out a CC revision of an instructional system will be greater than the mean posttest score, after adjusting for pretest score, for learners who try out an SG revision of that instructional system.

The null and research hypotheses used in analyzing the questionnaire data are as follows:

a. Hypotheses for comparing the prototype to the two revisions:

H_0: The mean rating for attitude toward the instructional system for learners who try out a CC revision of an instructional system and for learners who try out an SG revision of an instructional system will be less than or equal to the mean rating for learners who try out the prototype of that instructional system.

H_3: The mean rating for attitude toward the instructional system for learners who try out a CC revision of an instructional system and for learners who try out an SG revision of that instructional system will be greater than the mean rating for that attitude for learners who try out the prototype of that instructional system.

b. Hypotheses for comparing the two revisions against each other:

H_0: The mean rating for attitude toward the instructional system for learners who try out a CC revision of an instructional system will be less than or equal to the mean rating for that attitude for learners who try out an SG revision of that instructional system.

H_4: The mean rating for attitude toward the instructional system for learners who try out a CC revision of an instructional system will be greater than the mean rating for that attitude for learners who try out an SG revision of that instructional system.

Justification of Analytical Procedures

Hypotheses 1 and 2 were tested by use of analysis of covariance and hypotheses 3 and 4 were tested by use of multivarate analysis of variance. In the case of hypotheses 1 and 2, pretest scores was used as a covariate in order to eliminate the effect of pretest differences from the total variance. In situations where more than 2 comparison groups or more than one dependent variable are involved, analysis of variance is the most appropriate test of significance that can be used, in view of its capability to handle such data analysis in one test. By doing so, there is a minimal loss in degrees of freedom. This research involved three comparison groups; namely, learners who tried out the instructional prototype; learners who tried out the CC revision, and learners who tried out the SG revision. Therefore, analysis of variance was used. In the case of hypothesis 1 and 2, which deal with the posttest data, multivariate analysis was also used to determine whether the hypotheses might hold for certain of the instructional objectives only. Multivariate analysis is particularly suited to this type of analysis.

In the case of hypotheses 3 and 4, which relate to the questionnaire data, it was of interest to see if any combination of questionnaire items as well as the total questionnaire ratings (i.e. the combined ratings based on all the questionnaire items for each learner) showed significant differences. For this reason, multivariate analysis of variance was used rather than (univariate) analysis of variance.

Pilot Study

The pilot study was carried out on April 15, 1981. Previous arrangements had been made with the instructor for the graduate education course, Graphics Design and Use in Instruction (ED831B). The instructor used the first few minutes of the class and then introduced the writer to the class. The remainder of the class period was made available to the writer for the pilot study data collection.

A brief explanation of the research project in general was given, with mention made that it was for the purpose of collecting data for a doctoral dissertation. He asked for their cooperation and gave a brief overview of the events that would take place during the remainder of the class period. A few students who had already viewed the television program to be used in the study were then excused from participation in the study.

Pretest

The pretest was then administered to the remaining 15 students.

The learners were asked for their feedback about the questionnaire by writing on the questionnaire itself and by answering two questions "about this questionnaire." Some of their comments were about the instructional prototype. The questionnaire is shown in Appendix D.

Setting the debriefing agenda

The debriefing agenda for each tryout-revision model was determined by the outcome of the tallying of the incorrect posttest answers and the unfavorable questionnaire ratings for each of the two debriefing groups. A criterion was set for determining the eligibility of an item for the debriefing; namely, the item had to be incorrectly scored or unfavorably rated by at least 30% of the learners assigned to a given debriefing group.

However, if was anticipated that there would be only about one hour available for the debriefing. Moreover, Abedor (1971, 1981) has found that after one hour of debriefing interest in the process begins to lag. Thus, it was decided to carry out the debriefing on a worst posttest item and worst questionnaire items basis, with alteration between posttest items and questionnaire items, starting with the worst posttest item. The learners were randomly assigned to the two debriefing groups by alternating the pretest forms given to them. The learners assigned to computer conferencing group were also given a training guide. (See Appendix A)

Tryout

After all the pretests were collected from the learners the television program was played back.

Posttest and questionnaire

At the conclusion of the tryout, the learners were re-seated according to debriefing group assignment. Each debriefer then distributed the posttest to one of the two groups. Upon completion of the posttest, each learner was given a self-scoring form and a questionnaire to complete by the debriefer. Meanwhile, the debriefers tallied the incorrect answers on the posttests and the ratings on the questionnaires. The favorable ratings were defined as either "agree" and "strongly agree" or "disagree" and "strongly disagree." Therefore, the ratings that were tallied were the remaining three ratings.

Since there were seven learners assigned to the small group model and eight learners assigned to the computer conferencing model, the criterion number of wrong posttest items or unfavorable questionnaire items for each model was 3, when rounded to the next highest interger (7 x 30% = 2.1; 8 x 30% = 2.4). For the small group model, posttest items 1, 5, and 7 and questionnaire items 2, 4, 5, 7, 8, 9, and 10 met the criterion. For the computer conferencing model, posttest items 1, 5, 6, and 7 and questionnaire items 2, 4, 5, 6, 7, 8, 9 and 10 met the criterion.

Group debriefing

Small group debriefing

As soon as the SG debriefer derived the information from the outcome data needed to set the agenda, the debriefing began. A drawing pad was used for posting the major comments made by the learners. This served to keep track of what was said for the benefit of the other learners and to facilitate the verification of what had been said, so as to avoid distorting the comments made.

The final 10 minutes of the one hour debriefing session was devoted to obtaining feedback on the debriefing process and the questionnaire and posttest as data collection instruments. Many critical comments were made during this part of the session, which were very helpful for revising the instruments and other aspects of the data collection process.

Computer conferencing debriefing

As soon as the debriefer for the computer conferencing debriefing group set the agenda, the writer joined that group and began an orientation on the computer conference in which they were about to participate. He first gave each learner a sign-on ID to be used in signing-on to the University of Michigan computer. He then gave a brief overview of what is involved in computer conferencing and the events they would be engaged in during the remainder of the class period. There were also reassured that lack of prior computer experience would not be a particular problem and that a demonstration would be given prior to their use of the computer terminals.

The debriefing group and the writer then walked to the Computer Center where 9 computer terminals with paper output capabilities (hard-copy terminals), located in a special laboratory, had been reserved. The debriefer and the writer each demonstrated the use of a terminal in carrying out the procedures to be used in the computer conference, as described in the training guide. The group was split in half so that four learners gathered around each terminal being demonstrated. They were asked to follow the procedure in their training guide.

Immediately following the demonstration and the answering of questions each learner sat down at one of the terminals, except for one learner who left the group for unexplained reasons. This left, a total of seven learners.

As mentioned earlier in this chapter the writer registered all the sign-on IDs into the computer conference, prior to the tryout and debriefing, by using a fake name, This was done to save time during the debriefing. Therefore, the first thing required of the learners after "signing on" was to change the registration record for the name they would be using in the computer conference. The next procedure they were asked to carry out was to send a message to the debriefer that they changed their name. This procedure also had the purpose of informing the debriefer that that learner had signed on. The debriefer then informed that learner as to the first conference item to view and enter a discussion vote.

Unfortunately, some of the learners figured out that the messaging facility could also be used for sending messages to each other. Thus, the debriefer never received messages from some learners that they had changed their names and, thus, ready for a message as to which conference item to view and enter a discussion vote.

Additionally, some of the learners had much difficulty in learning how to "log on" and change their names. The writer, acting as a technical assistant, spent much time helping these learners in overcoming their difficulties. Meanwhile he did not become aware of the learner to learner messaging that was being carried on until late in the session.

Thus, only five of the seven learners participating in the conference entered any discussion votes. Only two learners entered discussion votes on all six of the items that were the subject of debriefing. One learner entered votes on three of the items, and the remaining two learners entered votes on only one item.

At the end of the session, the debriefer entered an item for obtaining feedback from the learners about the debriefing process. The comments came from three learners. One learner was critical of the debriefing because "the students were not well taught the procedure of the computer." Another learner said she was totally confused, but wanted to do more than she was able to do. The third learner said it was very interesting and that everyone should know about it. However, he only entered one discussion vote.

Evaluation of the Pilot Study

The pilot study was evaluated on the basis of the following data:

1. The tape recording of the small group debriefing.

2. Feedback given on the questionnaire.

3. Observation data from the computer conference.

4. Feedback from the debriefers.

<u>Modifications Made in the Instrumentation and Procedures on the Basis of the Evaluation of the Pilot Study</u>.

1. In order to provide better instructions to the learners assigned to the computer conferencing model on how to participate in the computer conference, the following revisions were made for the Phase I data collection.

   a. The writer, acting as the instructor, would lead learners through a training session for each of the procedures explained in the "training guide" as a substitute for the demonstrations. In this way, all learners would have to perform each step of each procedure.

   b. Anyone who had difficulties was to raise a hand and either the debriefer, acting as the instructor's assistant, or the instructor would help the learner overcome the difficulty. Meanwhile, the other learners were to stand-by and wait for the next instruction.

   c. A conference item serving as a combination of a model item and a practice item was entered into the conference. Fake participant names were used in modeling the discussion of the instructional prototype. During the training session, the instructor would ask the participants to view that item and then enter discussion votes on how they liked the conference so far. After every participant had entered at least one discussion vote, the debriefer would terminate her role as the instructor's assistant and enter a discussion vote telling the participants to view a particular conference item, thereby beginning the debriefing process.

2. Directions from the debriefer to the learners would be provided through discussion votes for the conference item currently being discussed. The messaging facility was not used at all.

3. In order to keep the learners assigned to the computer conferencing model "together", only one conference item was made the subject of debriefing at a time. Thus no information would be given about conference items to be discussed later in the conference.

4. In order to relieve the debriefer for the computer conferencing model of the task of entering conference items, a conference item was entered prior to the beginning of the experiment, for each posttest and questionnaire item. Each conference item gave complete instructions on how to enter a discussion vote and how to view the other discussion votes. Thus, by entering such information into each conference item the learner would be able to see these instructions as soon as he/she read the posttest or questionnaire item that was the subject of that conference

item.

5. Various changes were made in the wording of the instruments on the basis of learner feedback from the pilot study. These included, but are not limited to the following:
    a. Elimination of wordiness.
    b. Changing the direction of the questionnaire items so that they are all positive statements.
6. A greater effort would be made to collect consensus data on prototype deficiencies and data relating to the elimination of such deficiencies even if it meant covering fewer posttest and questionnaire items.
7. In order to provide for more time, an additional fifteen minutes was obtained from the instructor of the class to be used.
8. The learners would be encouraged to complete the test and questionnaire instruments faster than was the case during the pilot study, in order to save some needed time.
9. A shorter introduction would be given prior to the pretest than in the pilot test, and the orientation given to the computer conferencing group was eliminated in order to save time.
10. The training guide was modified in various ways. This included adding information that if the computer did not respond appropriately to the CONFER command modifier, "votes new," to use the command, "votes." The computer often responded to "votes new" by sending the message, "NO ITEM

SATISFIES THIS REQUEST". It was clear that this part of the CONFER software appeared to be unreliable and in need of correction. (This point is discussed at length in Chapter 5).

11. The learners would be alerted to the computer's apparent inconsistency noted in "10," above, during the training session and again at the beginning of the debriefing session and on an individual basis when a learner appears confused in using that command.

Phase I Data Collection -- Prototype Tryout and Debriefing

In spite of the advance planning and the benefits derived from the pilot study, a number of problems eventuated during the course of attempting to collect data for the first phase of the study.

First, the instructor of the class, was not able to provide the additional 15 minutes that had been promised, due to an unexpected delay in their class activities. The writer tried to make up for the lost time in various ways, including the elimination of most of his planned introduction, which may have led to some of the instances of non-cooperation from the students.

The tryout was carried out exactly as in the pilot study, and the small group debriefing worked out quite satisfactorily, with greater emphasis placed on the obtaining of consensus data and data relating to the elimination of the deficiencies in the prototype identified by the learners than was the case in the pilot study.

Most of the difficulties occurred during the computer conferencing debriefing session. Some technical problems arose in attempting to log-on with the University of Michigan computer for two of the computer terminals, including the terminal assigned to the debriefer. Eventually, these problems were resolved, but much time was lost.

As planned, the writer led the group through a brief training session on the procedures to be used during the computer conference. The major problem that arose during this session was that the students who learned the procedures quickly grew impatient with the delays brought about by the learning problems of some of the other students. This resulted in the group not "staying together", as planned, as the fast learners began "experimenting with the conferencing system by trying out various options offered by the system that were not part of the planned debriefing process. In some cases, this led to even more delays when the writer tried to bring everyone back to a common point in the computer software.

Another source of difficulty was the problem with the CONFER command modifier "votes new" met during the pilot study. As was pointed out previously in this chapter, this command appeared to be very unreliable. In spite of the warnings given to the learners regarding its unreliability they still seemed confused by its unreliability.

The result of these various delays and problems was that, the amount of time available for the debriefing was only about forty minutes. As the learners were unwilling to stay beyond 10:10 (their class normally ended at 9:50), it was not possible to have a debriefing of the same time length as for the small group model, which lasted approximately one hour.

Brief Analysis and Review of the Debriefing Data

The small group debriefing session resulted in the discussion of two posttest items, two questionnaire items and the generation of a set of general recommendations. (N.B. The specific items will be discussed in Appendix E, "Comparison of the Outcome and Debriefing Data for the Two Model").

There was general consensus with regard to the general recommendations as well as to many of the prototype deficiences and the recommendations for their elimination during the discussion of individual posttest and questionnaire items. Where general consensus was lacking among this group, the consensus was, in many instances very close to general consensus.

By contrast, the learners assigned to the computer conferencing model were much less productive. They discussed only one posttest item and two questionnaire items. Only five of the seven learners assigned to this model made comments bearing on the substance of the matter under discussion. The total number of substantive comments made by these learners for the three items discussed were 12, 8, and 13. The average number of comments per item is, 11 and the average number of comments made for item per active participant for these items wer 2.4, 1.6, and 2.6 respectively. The overall average number of comments made per active participant per item was 2.2.

In view of the fact that the small group debriefing lasted approximately 50% longer than did the computer conferencing debriefing (58.5 minutes versus 40 minutes), as well as the relatively small amount of discussion that was carried on during the computer conferencing debriefing, it was decided that new data was needed; that is, either the entire tryout and debriefing should be carried out again with another class or that the tryout and debriefing for the computer conferencing model only should be carried out again with another comparable class of students.

Further Development of and Planning for the Computer Conferencing Model

In order to restrict the sample to graduate education students in instructional media classes the only other class appropriate for the study was the class in Effective Use of Instructional Media (ED831A). However only twelve students were registered in the course, which meant that there would be only enough students for a second tryout and debriefing for the computer conferencing model. The cooperation of the instructor of this class was obtained, and it was decided to use this class for re-doing the Phase I data collection for the computer conferencing model only.

It was further decided that several changes in procedures needed to be made in order to insure the success of the computer conferencing debriefing. The following needs, relating to these changes, were identified.

1. Greater emphasis needed to be placed on the importance of the task prior to the time of the tryout and debriefing, so that the learners would participate in a mature and serious manner.
2. Greater motivation needed to be used to enlist the cooperation and serious participation of the learners.
3. Greater effort needed to be made in training the learners.

To accomplish these goals, the following strategies were decided upon:

1. To distribute the training guides two weeks prior to the night of the tryout and debriefing.
2. To provide an individualized training session in computer conferencing to each learner during the week prior to the week during which tryout and debriefing would take place.
3. To make a presentation to the class about the experiment at the time the training guides were distributed, placing emphasis on how they can benefit from their participation in it and why it's important to participate seriously.
4. To involve the instructor in motivating the students to participate in the experiment in a serious manner.
5. To revise the training guide so that it is completely self-instructional, and provides useful information about the nature of the computer conferencing.
6. To hire an extra technical facilitator who was familiar with CONFER in order to help the writer in assisting the learners in using computer terminals and CONFER.
7. To enter another practice item, so that the learners would have a model item to view (item 11), which the learners voted on during the April 20th session, and a practice item.

These stratgies were carried out, and the second tryout and debriefing was set for May 12, 1981. However, it was necessary to arrange with another organization on campus for use of computer terminals.

Second Tryout and Debriefing for the Computer Conferencing Model

This session went quite smoothly, relatively speaking. Since the entire class period was available and since the learners had already met and had worked with the writer in the individualized training sessions there seemed to be a much more relaxed and friendlier attitude between the learners and the writer than in the previous two sessions. This was in spite of the fact that the video playback equipment arranged to be used failed to operate and had to be replaced by other equipment. Moreover, the connections to the University of Michigan computer was lost for about ten minutes at the beginning of the debriefing.

The debriefing session lasted about 1 hour and 30 minutes, after the connection with the computer was regained. This included time at beginning for logging in, viewing the model items and the practice item, and for entering a discussion vote for the practice item. They did not have to change their names since their real names were known at the time the writer registered them into the conference. Therefore, the actual debriefing session was probably about 1 hour and twenty five munutes or approximately 50% longer than the small group debriefing.

However, computer conferencing is a much slower mode of communication than face to face communication. Therefore, it was allowed to continue until 10:00 (10 minutes beyond the normal end of the class period) in order to attempt to equalize the amount of debriefing data for the two models.

In Appendix E, some data will be given, relating to the comparison of the Small Group and Computer Conferencing tryouts and the discussion voting during the second computer conferencing debriefing.

The Debriefing Data

In this section, the debriefing data that resulted from the use of the two models of tryout-revision will be discussed separately. Then, some comparisons will be made between them. These discussions will include information on how the data was used in the production of the revisions.

Small group debriefing data

The deficiencies in the instructional system identified by the small group debriefing data centered around three main themes, as follows:

1. The lack of adequately defined terminology.
2. The poor performance of the presenter.
3. The overall poor quality of the television production.

The general recommendations that were made by the group at the end of the debriefing are as follows:

1. A better, more professional production is needed.
2. A better script is needed.
3. A longer script is needed, so that there are four separate programs: one for the overall idea presented in the prototype and one program for each of the three information processing systems discussed in the prototype.
4. The use of a better instructional design, so that the program(s) is/are more attractive and more fun to watch.

The information shown in the tables that follow indicate the relationships between the deficiencies identified by the learners, their recommendations for eliminating those deficiencies and the corresponding changes made in the instructional system as it is represented in the small group revision. In addition the posttest or questionnaire item instigating the discussion of the various program deficiencies will be given.

Furthermore, the degree of consensus on the issue under discussion will be given in cases where that information was provided by the debriefer. This information will be shown in the form of an arithmetic fraction, indicating how many of the 7 learners agreed (e.g. 6/7 = 6 of the 7 learners agreed). The information has been arranged so that information on similar-type deficiencies can be grouped together. Thus, the original chronology of the debriefing sessions have not been retained here. However, the interested reader could reconstruct that natural chronology by rearranging the information according to the following order in which the debriefing occurred:

1. Posttest item 2A.
2. Questionnaire item 10.
3. Posttest item 4B.
4. Questionnaire item 8.
5. Posttest item 6B.
6. General recommendations.

The following abbreviations will be used in these figures and in the discussion of the general recommendations:

P = the television presenter

L = the learner

PT = posttest item

Q = questionnaire item

SG = small group

CC = computer conferencing

ITEM	DEFICIENCIES	RECOMMENDATIONS	CHANGES MADE
PT 2A	Not enough emphasis on the definition of processing. (7/7)	1. Change the word "alter" to "manipulate" in the test item. (4/7)	None. According to expert opinion, "alter" is the correct term for that test item.
PT 2A		2. More detailed definitions of key concepts. (7/7)	1. More emphasis placed on terminology in script. 2. Graphics used for definitions and terms "keyed over video" through use of special effects generator during TV production.
PT 4B	Ambiguous question.	1. Option "d" could also be a correct answer. (6/7)	None. According to expert opinion, "d" is not correct.
		2. Better definition of key concepts. (6/7)	1. More emphasis placed on terminology in script. 2. Graphics and "keys" used.
PT 4B		3. Calculators can do more than implied in program. (5/7)	None. Implementation strategy not clear.
PT 4B		4. Present information slower so that L can take notes. (7/7)	1. Used a more professional type presenter, who spoke slower. 2. Script written for presentation of ideas at a slower rate.
PT 6B	Better definition of concepts. (6/7)	1. Explain jargon. 2. Spend more time on key concepts.	1. Graphics and "keys" used. 2. More emphasis placed on terminology in script.

Figure 3.2
SG Debriefing Data and Changes Incorporated Into
the SG Revision, Relating to Coverage of Terminology.

Figure 3.2 (continued).

ITEM	DEFICIENCIES	RECOMMENDATIONS	CHANGES MADE
PT 6B	Didn't define an information processing system (N.B. this term was defined in prototype.).	None given.	1. Definition of an information processing system defined in script. 2. Graphic used for definition.
Q 8	Terminology - some undefined concepts.	None given.	1. All concepts defined in script. 2. Graphics used for all definitions. 3. "Keys" used for all terms.

ITEM	DEFICIENCIES	RECOMMENDATIONS	CHANGES MADE
PT 2A	1. P spoke too fast. (7/7) 2. P had poor screen presence. 3. P smacked his lips.	1. Speak slower. 2. More professional P needed.	Different P used who had acting experience.
Q 10	1. P read from cue cards.	Rehearsals.	Used rehearsals.
	2. P not relaxed or natural.	1. Look at camera. 2. Put microphone down or make more hidden.	Used cue cards held under camera by cue card holders. Used a lavaliere microphone.
Q 10	3. P didn't have much credibility.	1. Need more footage of P. 2. Need better balance between P and graphics.	1. Used more "shots" of P. 2. Switched camera shots between P and graphics more to get better balance. 3. Required P to wear jacket and tie.
Q 10	4. P was too hurried-had poor enunciation and pronunciation.	None given.	Used a more professional P.

Figure 3.3
SG Debriefing Data and Changes Incorporated into the SG Revision, Relating to Presenter's Performance.

ITEM	DEFICIENCIES	RECOMMENDATIONS	CHANGES MADE
PT 2A	1. Quality of the audio and video was terrible.	More professional production.	1. Used facilities of WKAR-TV. 2. Used a director and crew who were telecommunication students at Michigan State University (MSU). 3. Used portable television equipment from MSU College of Education's Media Laboratory and Computer Access Center. 4. Used professional graphics supplies.
Q 8	1. Distortions (glitches). 2. Transitions were noisy.	None given.	1. Used professional editing equipment. 2. The director did the editing. He had much experience in use of the equipment.
Q 8	3. Couldn't read the graphics.	1. Use larger lettering (7/7) 2. Show only small portions of graphics at one time. 3. Use arrow indicators at each step of a chart.	1. Used professional graphics supplies. 2. Used "key wipes" on special effects generator during TV production.

Figure 3.4
SG Debriefing Data and Changes Incorporated into the SG Revision, Relating to the Television Production.

The implementation of the general recommendations into changes in the instructional system will not be displayed in charts. This information will be provided in standard type listings, which will be presented after the figures that follow. The debriefing data indicates that there was general consensus about the general recommendations.

The general recommendations from the small group tryout and their implementation into changes in the instructional system are as follows:

1. Better (professional) production.

This general recommendation was implemented through the following changes:

1.1 Different P with acting experience used who was dressed appropriately and who had read script in advance.

1.2 Professional television studio used for all production work, except for one remote scene for which professional-type portable equipment was used.

1.3 Rehearsals used to help insure quality performance.

1.4 A Michigan State University senior in telecommunications was used as the director.

1.5 Professional graphics supplies used, which were prepared under the director's supervision.

1.6 Michigan State University telecommunication students used as crew.

2. Better script.

This general recommendation was implemented through the following changes:

2.1 Different script co-authored by the director and the producer was used.

2.2 The script was the subject of several meetings and was revised several times. Published sources and expert opinion was consulted during the script's preparation.

3. Longer script-divided into specific parts.

This general recommendation was implemented through the following changes:

3.1 The program was almost twice as long as the prototype.

3.2 The script used a series of three humorous skits to separate the four segments of the program in a manner somewhat similar to the recommendation made. Although the recommendation for 4 separate programs was an impractical suggestion, it was felt that a single program, divided into four segments, might be a suitable alternative.

3.3 The segments were similar in nature to the separate programs suggested in the tryout data. That data specified that one program be dedicated to the overall idea and the other three programs be dedicated to each of the three information processing systems discussed in the program.

4. Better instructional design.

This general recommendation was implemented through the following changes:

4.1 More emphasis is placed on terminology during the early part of the program. This terminology is repeated in the later stages of the program by pointing out how each of the three basic information processing systems that have evolved through the course of history has been an improvement in the way in which the basic functions of information processing systems are carried out. Thus, the terminology is given first in general terms at the beginning of the program and then repreated in more specific terms throughout the program.

4.2 By doing so, the demonstrations of the three information processing systems are made more meaningful to the overall program than in the case of the prototype.

4.3 Motivation is provided at the beginning of the program by asking interesting questions about information processing systems.

4.4 A mental set is provided for by giving the purpose of the program and an advance organizer in the form of an overview.

4.5 At the end of the program the three information processing systems are summarized in terms of the functions of information processing systems and the advantages of the digital computer are given.

Computer conferencing debriefing data

The computer conferencing debriefing data centered around five major themes, as follows:

1. Coverage of terminology.
2. The presenter's performance.
3. The production of the television program.
4. The presentation of content.
5. Use of media and techniques.

The latter two points are unique to the computer conferencing debriefing. This difference will be discussed later in this section of the chapter.

In the remainder of this subsection of this chapter, the deficiencies identified by the learners and their recommendations for eliminating these deficiencies which were made during the discussion of specific posttest and questionnaire items will be presented in the form of charts. Their general recommendations, referred to as "priority items for improvement" at the end of the discussion of Questionnaire item 9 (See Appendix F), will be discussed through standard type listings, rather than by use of a chart.

The chronological order of the debriefing was as follows:

1. Posttest item 6B.
2. Questionnaire item 10.
3. Posttest item 5B.
4. Questionnaire item 9.
5. Priority items for improvement.

The figures that follow indicate the content of the discussion of specific posttest and questionnaire items.

ITEM	DEFICIENCIES	RECOMMENDATIONS	CHANGES MADE
PT 5B		A comment (quasi-recommendation) was made regarding the importance of vocabulary in learning a concept, especially in a TV program where you can't "go back to something" said earlier. (3/6) (Although this was not a recommendation, directly, it seemed to suggest their interest in improving information given on terminology.)	1. Greater emphasis placed on the terminology through use of graphics a the beginning of the program. 2. Use of keying of words over visuals. 3. More emphasis placed on terminology in script.
PT 5B	Programming function not explained well.	None.	Same as above.
PT 6B	Control function not adequately defined. (2/6)	None given.	Same as above.
PT 6B	Information given about programming function determining "control."	None given.	Same as above.
PT 6B	"Sequence" might imply programming function.	Implies faulty wording of posttest item.	None. Sequencing implies what the control function does.

Figure 3.5
CC Debriefing Data and Changes Incorporated into
CC Revision, Relating to Coverage of Terminology.

ITEM	DEFICIENCIES	RECOMMENDATIONS	CHANGES MADE
6 B	P was poor (4/6)	None	Better P used.
Q 9	Monotone P (2/6)	None	Same as above.
Q 10	P stuttered (1/6)	None	Same as above.
Q 10	Oral presentation hindered me from learning. (1/6)	P should be more professional.	Same as above.
Q 10	Hesitations made him seem uncertain of content and nervous. (1/6)	Better prepared P needed. (1/6)	Same as above.
Q 10	His grammer bothered me. (1/6)	None.	Same as above.
	He was obviously reading and sometimes faltered. (1/6)	More professional P needed. (4/6)	Same as above.
		Try not to read presentation. (1/6)	Same as above.

Figure 3.6
CC Debriefing Data and Changes Incorporated into
CC Revision, Relating to Presenter's (P) Performance

ITEM	DEFICIENCIES	RECOMMENDATIONS	CHANGES MADE
6 B	Printing should have been more legible. (2/6)	Deficiency implies that graphics should be improved.	Better graphics used.
Q 9		Audio quality should be improved. (6/6)	1. Telecommunication major with special interest in radio production used as audio person. 2. Professional audio equipment at WKAR-TV, E. Lansing, Michigan used for production.
Q 9, 10	The video cassette "skipped" (editing error).	Deficiency implies that video editing be improved.	Video editing carried out by the director, who was experienced in video editing and a telecommunications major.

Figure 3.7
CC Debriefing Data and Changes
Incorporated into CC Revision, Relating
to Media Production

ITEM	DEFICIENCIES	RECOMMENDATIONS	CHANGES MADE
PT 5B	Not enough time spent on computer functions. (1/6)	None.	Greater emphasis placed on computer in script.
PT 6B	Too much time spent on the abacus, compared to time spent on the computer. (1/6)	None.	Same as above.
PT 6B	Insufficient information given about digital computer. (1/6)	None.	Same as above.
PT 6B	Content not treated adequately for this posttest item. (4/6)	None.	1. Greater emphasis placed on terminology in script. 2. Graphics used to improve information given on control function.
PT 6B	Computer example (demonstration) was hard to follow. (1/6)	None.	Changes made in recording of scene. More emphasis placed on computer in script.
PT 6B		Need to present information slower (1/6)	1. More professional P used. 2. Effort made in script writing to present information at a slower rate than in prototype.
Q 9		Entire presentation needs more pizzazz.	1. Skits placed between segments. 2. Taxpayer scene added.

Figure 3.8
CC Debriefing Data and Changes Incorporated into
CC Revision, Relating to Presentation of Content

ITEM	DEFICIENCIES	RECOMMENDATIONS	CHANGES MADE
5 B		P should have selected media that were related to goal of explaining computer operations. We're confused about content, purpose and reason. (1/6)	None. Implications of recommendations for study not clear.
6 B	This should not have been the only teaching technique. (1/6)	Many other learning and teaching techniques should have been used. (1/6)	Taxpayer scene added.
		Tape recording and picture book would have served just as well. (1/6)	None. Recommendation conflicted with available resources.
		It would have been helpful if a handout were available for reference along with the video tape. (5/6)	Viewing Guide added. (See Appendix G)
		On-site visit should have been included.	None. Recommendation too difficult to implement.

Figure 3.9
CC Debriefing Data and Changes Incorporated into
CC Revision, Relating to Use of Media and Techniques

The general recommendations (priority items for improvement) and their implementation are shown below. Again the degree of consensus obtained is indicated by arithmetic fractions.

1. Use a better presenter. (6/6)

This recommendation was implemented through use of the same presenter used for the SG revision. Again, this person had previous acting experience.

2. Use clearer visuals. (2/6)

This recommendation was implemented through the use of professional type graphics supplies and their preparation by students who were telecommunication majors. As in the case of the SG revision, some graphics were transferred to photographic slides, so as to facilitate their use during the television production.

3. Spend more time on computers. (1/6)

This recommendation was only to some degree implemented. An effort was made to explain each of the information processing systems discussed in the television program as well as possible, with respect to the 5 information processing functions which were also discussed in the program. However, no new content was added, except to the extent that what was said in the prototype regarding computers was explained more thoroughly, which meant more time spent on computers.

4. Use of different types of models. (1/6)

Although only one learner recommended this type of improvement, it was perceived as an important one. On the basis of this recommendation in combination with the suggestion for a presentation with "more pizzazz" see Figure 3.8), the use of "other learning and teaching techniques" (see Figure 3.10), and general recommendations 10 and 11, shown below, called "Use a more interesting script," and "use additional visual aids," a scene was added depicting a taxpayer filling out income tax forms.

As the relationship between this solution and the recommendations just referred to may not be clear, the rationale for this decision will be given in the next few paragraphs.

It was not feasible to make use of additional visual aids other than graphics, within the context of the television program. That is, the use of other non-print media external to the television program was not feasible, as part of this study. However, it was felt that an alternative approach to the subject matter within the television program, could satisfy all of the above-mentioned recommendations.

The type solution that was sought was one in which the concept of an information processing system could be explained by showing an everyday-type application of information systems, which could be explained in terms of the five information processing functions. Thus, the teaching model that was used was the use of an intuitive approach augmented with rational explanation.

Thus, the scene that was included in the program was one in which a taxpayer is depicted using various files, an income tax form, an instruction booklet from the Internal Revenue Service and scratch paper showing computations. The presenter explains, off-stage, the materials shown in the scene and the tasks performed by the taxpayer, in terms of the 5 information processing functions.

This scene comes immediately after the information processing functions have been named. Following the "taxpayer" scene the information processing functions are formally defined. Thus, the scene is a prelude to the formal definition of several concepts. In doing so it can, perhaps, be thought of as a different visual aid, as a different teaching model, as a different learning and teaching technique, as a means toward adding "more pizzazz," and as part of a more interesting script.

5. Provide more review of material. (1/6)

This recommendation was implemented in several ways. First, a supplementary handout was used, which will be discussed in Recommendation 6, below. Second, the demonstrations given of the three different information processing systems included information in each case about the five information processing functions, as these applied to the type of information processing system being discussed. Third, the names of the information processing functions were repeated immediately prior to answering questions in the handout, as described below, and at the end of the program.

6. Use supplementary handouts. (1/6)

A viewing guide was developed in response to this recommendation (see Appendix G). The use of a handout was also recommended during the discussion of Posttest Item 6B (See Figure 3.10), which had a consensus rating of (5/6).

The viewing guide included an advance organizer, definitions of terms, a guide to viewing the taxpayer scene, review questions, and a table comparing the three information processing systems discussed in the program. After the information processing system functions are defined by the presenter, he asks the learner to stop the tape and answer the review questions in the viewing guide. Answers to the questions are also provided in the viewing guide. The handout (viewing guide) was distributed to the learners immediately prior to viewing the program.

7. Guidance from computer oriented personnel. (1/6)

This recommendation was not implemented, as such, as its meaning was not clear to the developer. One of the co-instructor's for the course which uses the prototype was consulted during the preparation of the two revisions, with respect to some of the terminology used in the program. To this extent, then, the recommended guidance was obtained. However, no effort was made to include computer personnel in the television program as presenters of information.

8. Provide a better visual quality. (1/6)

This recommendation was implemented through the use of professional television facilities and equipment and students who were telecommunication majors as the director and for the various studio positions (crew) during the television production. The facilities and equipment of WKAR-TV, East Lansing, Michigan were used for most of the production. The exceptions were two "on location" scenes. For these scenes, portable equipment was borrowed from the Michigan State University, College of Education's Instructional Resource Center.

9. Provide a better explanation of the computer example (demonstration). (1/6)

As indicated in the discussion of Recommendation 3, above, more time was spent on the demonstrations of each of the three types of information processing systems. Thus, this recommendation was implemented in a more general way than specified.

10. Use a more interesting script. (1/6)

This recommendation was implemented in various ways. Two of the major ways was through use of the "taxpayer" scene and the humorous skits. The latter were placed between the major segments of the program and indicated the subject-matter of the segment that was about to begin. Thus, they served as "visual labels," for the CC revision. The "taxpayer" scene, was described in the discussion of general recommendation 4, above. Other aspects of making the script more interesting included the use of more interesting graphics, and asking interesting questions and providing an overview at the beginning of the program.

11. Use additional visual aids. (1/6)

This was also provided for, in a sense, through use of the taxpayer scene, which was described in the discussion of general recommendation 4. Even though its in the same television format, this scene makes use of this visual medium in a way that is in addition to the way it was used in the prototype. Other additional aids used in this revision are the graphics used providing definitions of key terms, and the use of the "key-wipe" effect on the special effects generator, which enabled the ability to place key words in context with specific audio and visual stimuli.

The process for utilizing the debriefing data

Although the first priority in utilizing the debriefing data was the implementation of the general recommendations, all the data was examined and utilized during the revision process, whenever possible. The charts shown earlier in this section of the chapter indicate if and how each deficiency identified and recommendation made during the discussion of specific posttest and questionnaire items was incorporated into the revisions or the revision process. In the final analysis, almost all the data was utilized except where practical limitations or lack of specifity in the data precluded such utilization.

Although consensus was a concern in evaluating the importance of a recommendation or of an identified deficiency, the lack of a general consensus did not preclude the incorporation of the idea into the revision, unless that was impractical to do or the idea was not specific enough to guide its utilization in the revision process. In all cases, reasons are indicated in the charts or in the discussion in this section of this chapter, in cases where an idea was not incorporated into one of the revisions. The issue of consensus will be included in the next subsection of this chapter.

Major differences between the two debriefings and the two revisions

The two debriefings differed from each other with respect to the breadth of the respective perspectives represented in the data resulting from these debriefings. In the case of the small group debriefing, the data reflected a discussion centered on the presenter, the coverage of the terminology and the production of the television program. The data from the computer conferencing debriefing, on the other hand, reflected these three discussion topics plus the use of media and the presentation of context. For example the learners in this group made suggestions for a handout, and other visual aids, and they questioned the amount of time spent on the demonstration of the computer. Among their priority items for revision were to provide a supplementary handout, and the use of other models.

By comparison, then, the computer conferencing debriefing group seemed to be open to a broader range of revision possibilities than the small group debriefing group, as the latter group did not make suggestions that inferred such fundamental changes in the instructional system. By and large, the small group debriefing group accepted the videocassette, used without outside instructional materials in approximately the same manner as used in the prototype, as a given. The changes the latter group suggested, represented improvements within that specific context.

Another difference between these two groups is the greater amount of consensus found in the small group debriefing data than in the computer conferencing debriefing data. It is, perhaps, because of this that all the debriefing data for both debriefings was examined, rather than examine only that data for which there was a high level of consensus. Otherwise, the computer conferencing debriefing data would have been, in the main, of little use for revision purposes.

By and large, the two revisions incorporated most of the same changes, though sometimes for different reasons. An example of the latter is the use of the humorous skits. In the small group revision, these skits were used for the purpose of separating the program into four separate segments. In the computer conferencing revision these skits were used as part of the effort to "add pizzazz," and to "use a more interesting script," since humor usually provides interest and pizzazz. Thus, in spite of basic differences in these two debriefing groups, some of their recommendations resulted in the same changes in the revisions.

The two debriefing groups also identified many of the same deficiencies and recommended many of the same revision strategies, which resulted in the same improvement being incorporated in the two revisions. In view of this similarity and the similarity in some changes resulting from different recommendations from the two debriefing groups, it is important to identify and discuss the differences between the two revisions.

The only major differences in the two revisions is that the computer conferencing revision incorporates two instructional design elements that are not found in the small group revision. There are no instructional design elements or any other kind of major elements that are incorporated in the small group revision which are not also found in the computer conferencing revision.

The two instructional design elements in question are as follows:

1. The use of a viewing guide, together with the request by the presenter that the learner stop the tape during the program and answer the questions found in the viewing guide. The questions are all constructed response type questions which require the learner to define one of the functions of an information processing system.

2. As described previously, the "taxpayer" scene depicts a taxpayer using various files and materials for preparing his income tax forms. The presenter, who is off-stage, describes the various

components of the scene in terms of the functions of an information processing system. The presenter had named those functions immediately before the "taxpayer" scene. Immediately after this scene, the presenter provides formal definitions of thse functions. The purpose of the scene was to provide an intuitive notion and visual representation of each of the information processing system functions immediately prior to a more formal presentation of definitions of these functions.

Immediately after the formal definitions are stated, the presenter asks the learner to stop the tape and answer the questions in the viewing guide. Thus, these three events of intuitive description of functions, their formal definition, and answering review questions about them form a sequence of events.

There is a fourth factor which may be related to the learning of these functions, which was completely unplanned. During the television production, different "crew" members served as cue card holders for prompting the presenter. These "crew" members worked at different levels of efficiency in putting one card down and picking up the next. As a consequence, unintended pauses were imbedded into the computer conferencing revision between the definitions of the information processing system functions. It is possible that such pauses contributed to the learning of those definitions to a greater degree than in the case of the small group revision, which did not have those pauses.

These factors, all relate to the questions on the posttest that related to Instructional Objective 2, which is, "To identify the functions of an information processing system." During the data analysis, special attention was given to examining the possibility as to whether the computer conferencing revision led to greater learning than the small group revision with respect to Instructional Objective 2, in view of the above mentioned differences in the two revisions.

Revision Process

Each of the two revisions of the television program were based on the debriefing data resulting from the debriefing session for one of the two alternate models of tryout-revision discussed previously. In the case of the small group debriefing, the data was derived from the cassette audiotape recording of the session and the sheets of drawing paper used during the session for posting the major comments made during the session.

The data for the computer conferencing session was derived from computer printouts of the conference items which were agenda items during the debriefing. The general recommendations were printed out as part of the conference item 42, since that was the item dedicated to the discussion of questionnaire item 9 which came immediatley before the debriefer changed the topic of the discussion to that of the group's general recommendations. The latter discussion could, of course, have been made the topic of a seperate conference item had the writer chose to plan for it in that way.

The writer served as the instructional developer for both revised versions of the television program. Consideration was given to the possiblity of using two instructional developers who were at the same level of skill. However, some more experienced instructional developers suggested that a more objective strategy would be for the writer to develop both revisions and keep control over the objectivity of his behavior.

In view of the fact that the services of a television director was obtained, with the writer serving as executive producer, co-producer with the director, and as co-script writer with the director, it was necessary for him to exercise his various prerogatives to control the production in such a manner that revisions were based upon debriefing data and not on television production considerations.

Along these same lines, it was necessary for the writer to maintain objectively in the development of the two revised versions, so that both versions received approximately equal attention. There were many overlaps between the sets of debriefing data, with respect to deficiencies in the prototype and recommendations for eliminating such deficiencies. Hence, time and effort spend on the development of one revised television program often resulted in the development of the other. Thus, it cannot be said that equal time was spent in the development of the two revised versions. All that can be truthfully said is that equal attention, psychologically speaking, was given to the needs relating to the development of the two video programs, regardless of the time required to attend to such needs.

In the case of the small group debriefing session, the drawing sheets were first reviewed in order for the developer to get a general orientation as to what transacted during the debriefing. Then, the cassette recording of the debriefing was listened to by the developer and notes were made from it. Due to the fact that the tape was difficult to hear easily at certain points, the tape was also listened to together by both the developer and the debriefer on the day after the debriefing. The debriefer was able to interpret portions of the tape that had been unclear to the developer.

After developing some general ideas concerning the two revisions of the video program, the developer met with the director in May, 1981. The director viewed the prototype and they then looked at some of the debriefing data from the two debriefing groups. They then agreed on some general procedures. The developer agreed to write the storyboards (that is, sets of index cards describing video sequences and related audio content) and the scripts since he was familiar with the program content. They planned to meet again after the story boards were ready.

The storyboards were developed through the following process. The debriefing data for the small group revision was reviewed many times and notes were made for handling each of the deficiencies and revision recommendations. In addition, computer literature and experts were consulted for definitions of terms relating to concepts covered in the prototype. Eventually, "shot cards" for video sequences were developed for alternative storyboards, which were ultimately reduced to a single storyboard. The same process was used for the development of a storyboard for the computer conferencing revision.

They met in June, 1981 to discuss the completed storyboards. In general, they were in agreement as to what needed to be done. The director took charge of finding a television crew and the talent (that is, those persons who would appear in the video program) and all arrangements for equipment, television studio facilities, and related television production matters.

After the meeting on the storyboards with the director, the developer reviewed the meeting notes and made some changes in the story boards. He then began the script writing process, for the television program to be based on the small group debriefing data. The script for the computer conferencing revision was not written until after the script for the small group revision had been completely written.

In early July, 1981, the scripts were ready and the director and the developer met to develop a "shooting script" (that is a script that included all camera directions, camera shots, camera angles, audio content, and other matters relating to the production). Again, there was general agreement on these matters, with the developer keeping in mind the debriefing data while letting the director plan the production, without going either beyond the implications of the debriefing data or ignoring same. A good mutual understanding on duties, responsibilities, dissertation needs, and production needs was established, which lasted throughout the production process.

It should be noted that these revisions were planned as complete television productions and not as a series of changes to be inserted into the prototype. This was due to the nature and extent of the changes implied in the debriefing data for both revisions.

Phase II Data Collection -- Tryout of the Revisions

Having produced the two revisions it was now possible to try them out with learners. As mentioned previously, arrangements had been made for such tryouts with the three Summer term classes of the three instructional media classes in the College of Education, Michigan State University.

In each class, the following procedures were followed:

1. A short introduction was given by the writer about the purpose of the research and the activities in which they would be involved. Their cooperation was solicited, and questions were answered.
2. Students who had already seen the video program were told that their participation could not be used and were excused from participation in the study.
3. Immediately prior to distribution of the pretest, the learners were told that those receiving a pretest marked CC in the upper right corner of the cover sheet would be going to another classroom for the viewing of the television program and would then return to their regular classroom to answer a posttest and a questionnaire. Therefore, the students were asked to make a mental note as to which form of the pretest they completed.
4. The alternatively coded pretests were distributed to the students in a random manner.
5. Due to the fact that the first two classes had an uneven number of students, the distribution of pretests in the third class, was adjusted so that there would be an equal number of learners assigned to the two experimental groups.
6. The learners were asked to begin the pretest as soon as they

received it, and were asked to complete it as quickly as possible.

7. After all the completed pretests had been collected, the learners assigned to the computer conferencing revision were taken to another room to view the television program. This version of the television program also involved the use of the viewing guide, mentioned earlier in this chapter. The viewing guide included questions to be answered while the playback monitor was turned "off;" they were also told to check their answers on another page in the viewing guide. This meant that the learners assigned to the computer conferencing revision would take longer for the tryout.
8. The learners for the small group tryout were left on their own to view the television program. A volunteer was selected for turning the television monitor "off" after the viewing.
9. At the conclusion of the viewing of the television program, the writer brought the learners assigned to the computer conferencing revision back to their regular classroom.
10. The posttest was distributed to all the learners and were told to raise a hand when ready to complete the questionnaire.
11. The questionnaire was distributed in response to the raised hands.
12. When a learner had completed both forms he or she was thanked for participating in the study and asked to leave the room quietly, so as not to disturb any other learners completing the forms, when appropriate.

13. The total data collection process for each class lasted between one hour and one hour and fifteen minutes.

Summary

In this chapter, the idea of the comparison of two new models of tryout-revision, directly or indirectly based on Abedor's small group model (MK II), was introduced again, with a flowchart for these models given. The models are called the Computer Conferencing Model of Tryout-Revision and the Small Group Model of Tryout-Revision, and both of them rely solely on group debriefing data as the tryout data. The only difference between these models is that the former uses computer conferencing as the communications channel during debriefing, while the latter uses face to face communication.

For purposes of this study, the computer conferencing model uses computer conferencing in a synchronous mode, so that it may match the natural synchroneity of face to face communication in the small group model.

The development of the computer conferencing model prior to its use in the pilot study was described, including the many arrangements that were necessary. This is followed by discussions of how the instructional prototype was selected, the sample and the development of the instruments.

Then, the research study was described, which presupposed the existence of these models of tryout-revision. This discussion included the research design, the research questions and hypotheses, the variables, and the rationale for using the statistical procedures used in the data analysis.

At this point a pilot study was described. This was used for obtaining preliminary insights on how the tryout and debriefing sessions for the instructional prototype for the two models would operate, as well as data on the workability of the instruments and outcome data, based on use of the prototype. This data was then evaluated, and modifications based on this evaluation data were described.

The actual experiment then commenced. First, the collection of tryout and debriefing data, based on use of the prototype was discussed. The debriefing data was then analyzed and reviewed, which led to further development of the computer conferencing model, due to problems encountered during the debriefing for this model. A second tryout and debriefing for the computer conferencing model only was then described, which is followed by a limited comparison of the outcome data for the two models.

Two revisions were then developed, based on the debriefing data from each of the two debriefings, respectively. The processes involved during the development of the revisions were discussed. Also described were the changes incorporated in the revisions, based on the debriefing data.

Finally, having produced two revisions, tryout data was collected, based on their use.

Having discussed the procedures required for collecting the research data, the research findings will be discussed in the next chapter.

CHAPTER 4

ANALYSIS AND DISCUSSION OF RESEARCH

Introduction

In this chapter, the research results are reported. These results are followed by a summary of the research findings and a discussion of some of these research findings. Then, some factors that might explain some of the research findings are discussed.

The following abbreviations are used in this chapter:

CC revision = the revised version of the instructional system, based on the data from the debriefing session for the computer conferencing model of tryout-revision.

SG revision = the revised version of the instructional system, based on the data from the debriefing session for the small group model of tryout-revision.

Research Results

The research results are reported in two subsections. The first of these give the results based on the posttest scores. The second subsection provides the results based on the questionnaire.

Posttest data research

In this subsection, the hypotheses refer to a treatment group consisting of learners who tried out a CC revision and learners who tried out an SG revision. As explained in Chapter 1, this procedure was necessitated by the limit in the number of planned comparisons that can be made in analysis of variance. This limit is based on the number of degrees of freedom in the data, which is equal to the number of treatment groups minus one. Thus, in this study, which had 3 treatment groups, the number of degrees of freedom and, hence, the number of planned comparisons that are possible is equal to 2.

However, in order to provide for analysis based on each of these two groups of learners as separate treatement groups, post hoc analysis was carried out. This analysis will be presented after the hypotheses have been tested with data based on the combined treatment group.

Hypotheses

The null and research hypotheses used in analyzing the posttest data are as follows:

a. Hypotheses for comparing the prototype to the two revisions:

H_0: The mean posttest score, after adjusting for pretest score, for learners who try out either a CC revision or an SG revision of an instructional system will be less than or equal to the mean posttest score, after adjusting for pretest score, for learners who try out the prototype version of that instructional system.

H_1: The mean posttest score, after adjusting for pretest score, for learners who try out either a CC revision or an SG revision of an instructional system will be greater than the mean posttest score, after adjusting for pretest score, for learners who try out the prototype version of that instructional system.

b. The hypotheses for comparing the two revision against each other:

H_0: The mean posttest score, after adjusting for pretest score, for learners who try out a CC revision of an instructional system will be less than or equal to the mean posttest score, after adjusting for pretest score, for learners who try out an SG revision of that instructional system.

H_2: The mean posttest score, after adjusting for pretest score, for learners who try out a CC revision of an instructional system will be greater than the mean posttest score, after adjusting for pretest score, for learners who try out an SG revision of that instructional system.

Hypothesis testing

To test these hypotheses, an analysis of covariance of the posttest scores was carried out. In order to take into account any differences in the three groups on the basis of pretest scores, pretest score was used as a covariate. The analysis is shown in Table 4.1, which follows. The probability level used for rejecting the null hypotheses was .05.

Table 4.1

Analysis of Variance of Posttest Scores,
Using Pretest Score as a Covariate

Source of Variation:	(a) DF	(b) Mean Square	(c) F	(d) Significance of F
Within Cells	45	8.088		
Regression	1	155.422	15.219	.000
Prototype vs. Revisions	1	2.517	.311	.580
CC Revision vs. SG Revision	1	60.406	7.468	.009*

*Significant at the .05 level

Table 4.1 shows that the difference in posttest scores between the learners who tried out the prototype and those who tried out either of the revised versions is not statistically significant, as indicated by the value, p=.580, shown in column (d). This table does, however, show a significant difference in posttest scores between learners who tried out the two revisions, as shown by the value, p=.009, in column (d). This difference was in favor of the learners who tried out the CC revision. However, in view of the non-significance of the difference between learners who tried out the prototype and learners who tried out the two revisions, the difference between the two revisions is of limited importance. Hence, the null hypothesis corresponding to research hypothesis 1 was not rejected but the null hypothesis corresponding to research hypothesis 2 was rejected.

<u>Post hoc analysis for hypothesis testing</u>

Since it was the SG and CC revisions, and ultimately the models used for developing them, that was of main interest in this study, post hoc analysis was carried out in order to compare each of the revisions, respectively, with the prototype. In view of the fact that post hoc analysis is not as strong as planned comparisons, a .025 significance level was used.

Table 4.2

Analysis of Variance of Posttest Scores, Using Pretest Score as a Covariate, for Comparing the SG Revision to the Prototype

Source of Variation:	(a) DF	(b) Mean Square	(c) F	(d) Significance of F
Within Cells	28	9.713		
Regression	1	97.645	10.053	.004
Prototype vs. SG Revision	1	4.594	.473	.497

Table 4.2 shows that the difference in mean posttest scores between the learners who tried out the prototype and those who tried out the SG revision is not statistically significant, as indicated by the value, p=.497, shown in column (d).

Furthermore, Table 4.3 shows that the difference in mean posttest scores between the learners who tried out the prototype and those who tried out the CC revision is not statistically significant, as indicated by the value, p=.120, shown in column (d).

Table 4.3

Analysis of Variance of Posttest Scores, Using Pretest Score as a Covariate for Comparing the CC Revision to the Prototype

Source of Variation:	(a) DF	(b) Mean Square	(c) F	(d) Significance of F
Within Cells	28	9.898	6.531	.016
Regression	1	64.640	6.531	.016
Prototype vs. CC Revision	1	25.440	2.570	.120

<u>Further analysis for combined treatment group</u>

In view of these findings, it was decided that a multivariate analysis of posttest scores by instructional objectives, after adjusting for pretest score, might indicate that for one or more instructional objectives the data may lead to the rejection of the null hypothesis corresponding to research hypothesis 1.

This analysis is shown in Table 4.4, which follows. In this table, a linear combination is shown, accounting for a significant difference between learners who tried out the prototype and learners who tried out one of the revised versions, as indicated by the value, p=.022, shown in column (b). Moreover, another linear combination is shown in Table 4.4, which accounts for a significant difference between learners who tried out the CC revision and learners who tried out the SG revision, as indicated by the value, p=.012, shown in column (b).

In order to determine which instructional objective(s) within the linear combinations has/have the strongest influence, the univariate F values and correlations between the dependent variable and each of the canonical variables respectively, representing the linear combinations just mentioned were examined.

For the first planned comparison (Prototype vs. Revisions), it is indicated in Table 4.4 that Objective 2 makes the largest contribution to the first linear combination noted above, as indicated by the value, .572, shown in column (e) for this objective. Instructional Objective 2 is: "To identify each of the functions of an information process system." A large contribution is also shown for Objective 4, as indicated by the value, -.527, in column (e). However, the difference is in the opposite direction, as indicated by its negative correlation with the canonical variable, and therefore is not important to this analysis. These are the only two objectives which show a significant difference between the prototype and the revisions, as can be seen in column (d).

Table 4.4

Analysis of Variance for Posttest Scores by Instructional Objectives, after Adjusting for Pretest Score, for Comparing the Prototype to the Revisions and for Comparing the CC Revision to the SG Revision

Effect:	(a) Approx. Multivariate F	(b) Significance of Multivariate F	(c) Univariate F	(d) Significance of Univariate F	(e) Correlation w/Canonical Variable
Within Cells					
Regression	2.064	.001			
Prototype vs					
Revisions	2.859	.022*			
Objective 1			.001	.971	.008
Objective 2			6.411	.015*	.572
Objective 3			1.398	.244	-.267
Objective 4			5.443	.025*	-.527
Objective 5			.433	.514	-.149
Objective 6			.416	.522	.146
CC Revision vs					
SG Revision	3.261	.012*			
Objective 1			.216	.645	.098
Objective 2			11.779	.001**	-.726
Objective 3			3.353	.075	-.387
Objective 4			.076	.785	-.058
Objective 5			.051	.823	.047
Objective 6			2.276	.139	.319

*Significant at the .05 level.
**Significant at the .001 level.

For the second planned comparison (CC Revision vs SG Revision), it is indicated in Table 4.4 that, again, Objective 2 makes the largest contribution to the second linear combination noted above, as indicated by the value, -.726, in column (e) for this objective. The mean score for the learners who tried out the CC revision was significantly higher than that of the learners who tried out the SG revision. This difference is indicated by the univariate F value in column (c), which was significant at the .001 level, as shown in column (d).

Further post hoc analysis

Again, it was of interest to carry out post hoc analysis in order to compare each of the two revisions, respectively, to the prototype. This data is shown in Tables 4.5 and 4.6, which follow.

Table 4.5 shows no linear combination accounting for a significant difference between learners who tried out the prototype and learners who tried out the SG revision, as indicated by the value, p=.144, shown in column (b) for SG Revision vs Prototype. Thus, the null hypothesis corresponding to Research Hypothesis 1 could not be rejected, with respect to any of the instructional objectives.

Table 4.6, on the other hand, shows a linear combination accounting for a significant difference between learners who tried out the prototype and learners who tried out the CC revision, as indicated by the value, p=.008, shown in column (b).

In order to determine which instructional objective(s) within this linear combination has/have the strongest influence, the univariate F values and correlations between the dependent variable and the canonical variable were examined. Table 4.6 shows that Objective 2 makes the second largest contribution to this liner combination, as indicated by the value, .702, shown in column (e) for this objective. Objective 1 makes the largest contribution to this linear combination, as indicated by the value, .765, shown in column (e) for this objective. However, Objective 2 is the only instructional objective which shows a statistically significant difference between the prototype and the CC revision, as indicated by the value, .003 shown in column (d), based on a significance level of .025. Thus, for the CC revision, the null hypothesis was rejected with respect to Objective 2, even though the null hypothesis, as formally stated, could not be rejected.

Table 4.5

Analysis of Variance, for Posttest Scores by Instructional Objectives, after Adjusting for Pretest Score, for Comparing the SG Revision to the Prototype

Effect:	(a) Approx. Multivariate F	(b) Significance of Multivariate F	(c) Univariate F	(d) Significance of Univariate F	(e) Correlation w/Canonical Variable
Within Cells Regression	1.391	.103			
SG Revision vs Prototype	1.860	.144			
Objective 1			.099	.756	.799
Objective 2			.246	.625	.676
Objective 3			2.241	.148	.503
Objective 4			4.134	.054	.266
Objective 5			1.008	.328	.199
Objective 6			.409	.529	.074

Table 4.6

Analysis of Variance for Posttest Scores by Instructional Objectives, after Adjusting for Pretest Score, for Comparing CC Revision to the Prototype

Effect:	(a) Approx. Multivariate F	(b) Significance of Multivariate F	(c) Univariate F	(d) Significance of Univariate F	(e) Correlation w/Canonical Variable
Within Cells Regression	1.508	.058			
CC Revision vs Prototype	6.609	.008*			
Objective 1			.633	.435	.765
Objective 2			18.485	.003*	.702
Objective 3			.008	.926	.607
Objective 4			4.653	.042	.494
Objective 5			.063	.805	.162
Objective 6			.035	.853	.106

*Significant at the .   level.

Questionnaire data research

As was discussed in Chapter 3, the questionnaire used in this study measured attitude toward the instructional system that was tried out. This attitude was measured by ten 5-point Likert-type scales, with values "1" to "5" assigned to the points on the scales, ranging from "strongly disagree" ("1") to "strongly agree." ("5") (cf. Appendix D for a copy of the questionnaire).

Hypotheses

Thus, it was possible to compute total questionnaire ratings for each learner and to derive "mean ratings," which are referred to in the hypothesis shown below. In the discussion of this research, terms like "favorably" and "more favorable" are used. Such terms are based on the numerical values used with these scales.

The hypotheses used for analyzing the questionnaire data are as follows:

> H_0: The mean rating for attitude toward the instructional system for learners who try out either a CC revision or an SG revision of an instructional system will be less than or equal to the mean rating for that attitude for learners who try out the prototype of that instructional system.
>
> H_3: The mean rating for attitude toward the instructional system for learners who try out either a CC revision or an SG revision of an instructional system will be greater than the mean rating for that attitude for learners who try out the prototype of that

instructional system.

H_0: The mean rating for attitude toward the instructional system for learners who try out a CC revision of an instructional system will be less than or equal to the mean rating for that attitude for learners who try out an SG revision of that instructional system.

H_4: The mean rating for aatitude toward the instructional system for learners who try out a CC revision of an instructional system will be greater than the mean rating for that attitude for learners who try out an SG revision of that instructional system.

Hypothesis testing

To test these hypotheses, a multivariate analysis of variance for this attitude rating was carried out. As in the case of the posttest data research, the learners who tried out the two revisions have been combined into one treatment group, due to the restriction in the number of planned comparisons which are possible. This analysis is shown in Table 4.7, which follows. The probability level used for rejecting the null hypotheses was .05.

Table 4.7

Multivariate Analysis of Variance for Questionnaire Ratings of Attitude Toward Instructional System

Effect	(a) Approximate F	(b) Significance of F
Prototype vs. Revisions	1.209	.317
SG Revision vs. CC Revision	.547	.845

In Table 4.7 it can be seen that a multivariate analysis of the questionnaire ratings indicates by the value, p=.317, shown in column (b) that no significant difference was found between learners who tried out the prototype and learners who tried one of the two revisions. This research result refers to ratings for individual questionnaire items, and all the possible combinations of such ratings. Likewise, no significant differences were found between learners who tried out the SG revision and learners who tried out the CC revision, as indicated by the value, p=.845, shown in column (b).

Hence, the null hypotheses corresponding to research hypotheses 3 and 4 could not be rejected.

Post hoc analysis

Post hoc analysis was also carried out for the questionnaire data in order to compare each of the two revisions, respectively, to the prototype. This analysis is shown in Table 4.8, which follows:

Table 4.8

Multivariate Analysis of Variance for Questionnaire Ratings of Attitude Toward the Instructional System, for Comparing the two Revisions to the Prototype

Effect	(a) Approximate F	(b) Significance of F
Prototype vs. SG Revision	.544	.839
Prototype vs. CC Revision	.544	.839

In Table 4.8, it can be seen that a multivariate analysis of the questionnaire ratings indicates by the value, p=.839, for both comparisons shown in column (b) that no significant difference was found between learners who tried out the prototype and learners who tried out either the SG revision or the CC revision. Again, this research result refers to both individual questionnaire items and all the possible combinations of these items.

Summary of the Findings

Please note that two different levels of statistical significance were used in this study, which are shown in the findings listed below. The reason for doing so was that Findings 2, 3, and 5 relate to post hoc analysis, whereas the other findings relate to planned comparisons. Since planned comparisons are more stringent statistical tests than post hoc analysis, it was recommended to the writer by a statistical consultant that the lower level of statistical significance be used with the post hoc analysis. However, it was found that other statistical experts don't agree with this approach.

The research findings for this study can be summarized as follows:

1. The difference in posttest scores between learners who tried out the prototype and learners who tried out either the SG or CC revisions was not statistically significant at the .05 level.

2. The difference in posttest scores between learners who tried out the prototype and learners who tried out the SG revision was not statistically significant at the .025 level.

3. The difference in posttest scores between learners who tried out the prototype and learners who tried out the CC revision was not statistically significant at the .025 level, except as noted in Finding 5.

4. The difference in posttest scores between learners who tried out the CC revision and learners who tried out the SG revision was statistically significant at the .05 level, in favor of the learners who tried out the CC revision. The actual significance level for this finding was $p = .009$. However, in view of Findings 1, 2, and 3, this finding must be interpreted with caution.

5. The difference in posttest scores, with respect to Instructional Objective 2, between learners who tried out the CC revision and learners who tried out the prototype was statistically significant at the .025 level in favor of the learners who tried out the CC revision. The actual significance level for this finding was $p = .003$. This instructional objective is: "To identify each of the functions of an information processing system."

6. There was no linear combination, with respect to the instructional objectives used in the study, accounting for a statistically significant difference between learners who tried out the prototype and learners who tried out the SG revision, at the .025 or .05 levels of significance, for any of the instructional objectives used in the study.

7. The difference in posttest scores, with respect to Instructional Objective 2, between learners who tried out the CC revision and learners who tried out the SG revision was statistically significant at the .05 level. The significance level for this finding was at the $p = .001$ level, in favor of the learners who tried out the CC revision.

8. No significant difference in attitude toward the instructional system at the .05 level of significance was found between learners who tried out either of the revisions and learners who tried out the prototype.

9. No significant difference in attitude toward the instructional system at the .05 level of significance was found between learners who tried out the SG revision and learners who tried out the CC revision.

<u>Discussion</u>

On the basis of these research findings it might appear that either the use of tryout-revision, in general, does not lead to the improvement of an instructional system or that the use of the models used in this study are incapable of leading to such improvement and/or the manner in which this study was carried out prevented such improvement from being realized.

Due to the limitations of this study, there is no way to determine, without the benefit of further research, whether the models used in this study are effective improving an instructional system. For this reason, the writer does not wish to contend that the research supporting the general utility of tryout-revision, despite the limitations of that research, as well as the successful application of tryout-revision by instructional developers is ill-founded. He would rather assume that the lack of success met in producing revisions that led to greater cognitive learning and a more positive attitude toward the instructional system than did the prototype was due to the models and/or the procedures used in this study.

Summary

In this chapter, four hypotheses were tested in order to evaluate the effectiveness of two new models of tyout-revision; namely a Computer Conferencing Model of Tryout-Revision (CCTR) and a Small Group Model of Tryout-Revision (SGTR). Two types of research data were used; namely, posttest data and questionnaire data. Analysis of variance was carried out for analyzing the data. In the case of both types of data, one hypothesis was tested to compare one group of learners who tried out an instructional prototype with a combined group of the learners who tried out two revisions of the prototype. Each of these revisions was based on tryout data resulting from the use of each of the two new models of tryout-revision. A second hypothesis was tested to compare the learners who tried out each of the two revisions to each other.

Learners were combined in the case of the first of these types of hypotheses in order to limit the number of planned comparisons to two.

This was done because the number of planned comparisons that can be carried out in analysis of variance is one less than the number of treatment groups. Since this study made use of three treatment groups, only two planned comparisons could be made.

Thus, by combining the two treatment groups that tried out the revisions, it made possible the comparison of the two models of tryout-revision to each other, and the comparison of the prototype to what might be called "the average of the two revisions." However, the primary interest in this study was to compare the treatment group that tried out the Computer Conferencing revision with the treatment group that tried out the prototype, in order to determine whether computer conferencing could be effectively used in tryout-revision. To accomplish that purpose, post hoc analysis was carried out, which made possible the comparison of each of the treatment groups that tried out one of the revisions with the treatment group that tired out the prototype.

The revision based on the tryout data obtained through computer conferencing was shown to be more effective than the revision based on the use of a small group model of tryout-revision. However, no significant differences were found in the comparison of each of the treatment groups that tried out one of the revisions to the treatment group that tried out the prototype. Therefore, further analysis was carried out for the purpose of determining whether the hypotheses relating to prototype - revision comparisons might hold for one or more of the instructional objectives upon which the test items were based.

This, further analysis showed that the mean posttest scores of the learners who tried out the revision based on the use of computer conferencing were significantly greater than those of the learners who tried out the prototype, with respect to one of the main instructional objectives.

The results were then briefly discussed, with respect to the fact that neither of the tryout-revision models led to general improvement in cognitive and affective outcomes for the instructional system, with no improvement shown in attitude toward the instructional system. It was suggested that these outcomes were the result of the specific operationalization of the models and procedures used in the study rather than being due to the inability of the tryout-revision process, in general, to lead to the improvement of instructional systems.

CHAPTER 5

SUMMARY, CONCLUSIONS, DISCUSSION, AND RECOMMENDATIONS

Summary

This study was instigated by the concern of some instructional developers that prototype revision should be based on the application of prescriptions rather than on the use of a trial and error process. Gropper (1975) suggested many categories of learner and program failures that he feels could be the basis for specifying revision strategies. Scriven (1967) said the same thing when he coined the term, formative evaluation. Speaking of tryout data he said, "Percentages are not very important. It is the nature of the mistakes that is important in evaluating the curriculum, and in rewriting it (p. 61-62)."

However, collecting data on learner mistakes has neither been easy to do, nor has it been systematically carried out. Since the beginning of instructional development the tutorial model of tryout-revision has been used in a limited way for collecting data on learner errors as they are made. The literature indicates that the time consuming and intensive nature of the tutorial process discourages the extensive use of this model. Abedor (1971, 1972) developed a small-group-type model which included a group debriefing process for collecting data on learner mistakes after the fact. However, no further use of this model is indicated in the literature, though Abedor (1981) and others have used it since then. One of the problems with the group debriefing process is that the group size is limiting. Abedor (1981) suggests 6-8 learners as the optimal size.

Not suprisingly, then, the literature indicates a preference among instructional developers for the collection of test data, either by itself or preceded by the use of the tutorial model and, the small group model. Thus, test data seems to be the alternate basis for determining how well a prototype performs. However, if the prototype does not perform well, the causes of the mistakes are not easy to determine.

In view of this history of tryout-revision efforts the recent development of computer conferencing suggests the possibility of a new solution to the old problem of determining the nature of mistakes and other qualitative data relating to prototype tryout. Computer conferencing accomodates both 2-person communication and communication among larger groups. Palme (1982) suggests 30-40 as an optimal size for the latter. Since computer conferencing consists of the exchange of messages in whatever form is specified, it appears that this communication mode could be used for collecting all the types of tryout data that have been collected through the use of other tryout-revision models. Moreover, in view of the fact that message exchange easily lends itself to the communication of personal thoughts and opinion, it would appear that detailed explanations of mistakes during prototype tryout and attitude toward the prototype could easily be made the topics of discussion in a computer conference.

However, one question that comes to mind in considering the use of computer conferencing in group debriefing is whether the remote nature of this communication mode would significantly inhibit communication about mistakes made during prototype tryout and the expression of negative attitudes toward the prototype. That is, some individuals may need the encouragement of sympathetic looks and words to admit to mistakes among a group of peers and to give detailed explanations of such mistakes. The same may hold true for expressing negative opinions about the prototype. On the other hand, the remote nature of this communication mode might encourage some learners to express such thoughts, since there is no immediate reaction from peers. The same type of questions could be asked of the use of computer conferencing in a tutorial-type interaction.

The focus of this study, then, was the question, "is it feasible to use computer conferencing to collect tryout data relating to mistakes made in prototype tryout and attitude toward the prototype?"

Since the literature indicates that previous work has not been carried out in this area, various starting points were possible. However, it seemed that group debriefing would be a better starting point than the tutorial process since more data can be collected in a single implementation of this process than in the tutorial process. Although the use of Gropper's (1975) categories of learner and program failures was attractive, it was decided not to include such considerations, in order to keep the study simpler.

Since Abedor's (1971, 1972) MK II small group model involved a group debriefing, this model appeared to be a logical starting point for the application of computer conferencing to tryout-revision. However, the model involved the use of not only group debriefing data but also test and questionnaire data and, to some degree, tutorial data as the basis of prototype revision. In view of the interest in this study of the debriefing process and the effect of the remote aspect of computer conferencing on group debriefing, it was decided to modify the MK II model, so that only debriefing data be used as the basis of prototype revision. Such a modification, in essence, created another model of tryout-revision, since it couldn't be assumed that its modifed form would perform just as well as its unmodified form.

This modified form of the MK II model, then, served as the basis for what was called in this study the Small Group Model of Tryout-Revision (SGTR). By substituting a computer conferencing debriefing of learners for a face-to-face debriefing of learners a second model was developed; namely the Computer Conferencing Model of Tryout-Revision (CCTR). This latter model is, of course, related to a theoretical Computer Conferencing Model of Tryout-Revision, described in Chapter 1. In that description, various types of uses of computer conferencing and the use of categories of learner and program failues are included. Thus, the computer conferencing model operationalized in this study can be thought of as a prototype version of a more complex computer conferencing model that could be developed in future research and development.

These two models, then, were the central interest of the study in examining the basic question, mentioned earlier in this summary; namely

"Is it feasible to use computer conferencing to collect tryout data relating to mistakes made in prototype tryout and attitude toward the prototype?" Two types of comparison were involved, as follows:

1. To what extent does either the modified MK II small group model or the computer conferencing model of tryout revision (SGTR and CCTR, respectively) lead to the development of a revision which brings about better cognitive and affective learning outcomes than does the prototype, if any?
2. To what extent does CCTR lead to the development of a revision that brings about better learning outcomes than does SGTR, if any?

To obtain answers to these questions, an instructional television program about information processing systems, developed at Michigan State University (MSU), was used as the prototype for this study. The learning outcomes investigated were based on answers to a posttest and responses to a questionnaire designed to measure attitude toward the television program. The sample was drawn from instructional media classes in the MSU College of Education during the 1981 Spring and Summer terms. The prototype was tried out during Spring term, with learners randomly assigned to either a small group debriefing or a computer conferencing debriefing. Each set of debriefing data served as the basis for the development of a revision. These were called the small group revision (SG revision) and the computer conferencing revision (CC revision). CONFER II, originally developed at the University of Michigan was used as the computer conferencing system. The version of CONFER II residing on the University of Michigan mainframe computer was used for this study.

The test items on the pretest and posttest were based on six instructional objectives. Therefore, it was possible to use both univariate and multivariate analysis of variance. For the latter, a score for each instructional objective was derived for each subject. The posttest scores were adjusted for pretest score to eliminate variance due to differences on the pretest for the three treatment groups. The three treatment groups were as follows: 1) the learners who tried out the prototype; 2) the learners who tried out the SG revision; and 3) the learners who tried out the CC revision.

The general findings for the study were as follows:

1. Cognitive learning, as measured by mean posttest score, was not significantly greater ($p = .05$) for learners who tried out the CC revision or the SG revision than that of the learners who tried out the prototype, except as noted in "3", below.
2. Cognitive learning, was as measured by mean posttest score, significantly greater ($p = .009$) for learners who tried out the CC revision than that of those who tried out the SG revision. However, in view of the lack of a significant difference in learning noted in "1", above, the difference in learning noted here concerns the relative effectiveness of the two models of tryout revision. The computer conferencing model was shown to be more effective than the small group model for improving an instructional system, with respect to cognitive learning. However, both tryout-revision models were generally ineffective, or improving the prototype.

Conc

conc

to t

tryo

appl

invo

prog

medi

3. Attitude toward the instructional system was not significantly more positive (p = .05) for learners who tried out one of the revisions than that of learners who tried out the prototype.

4. There was no significant difference (p = .05) in attitude towards the instructional system between learners who tried out the SG revision and learners who tried out the CC revision.

The sub-findings of the study were as follows:

1. With respect to Instructional Objective 2, cognitive learning was significantly greater (p = .001) for learners who tried out the CC revision and for learners who tried out the SG revision.

2. With respect to Instructional Objective 2, cognitive learning was significantly greater (p = .003) for learners who tried out the CC revision than for learners who tried out the prototype.

Conclusions

This section will be divided into two subsections; namely, conclusions relating to the research findings and conclusions relating to the workability of the computer conferencing model of tryout-revision. These conclusions are limited to the use of a single application of the two models being compared. This single application involved a single prototype, which was a self-instructional televison program and a sample selected from graduate courses in instructional media at Michigan State University.

Conclusions Relating to the Research Findings

The conclusions discussed in this subsection are based on the findings from this study, listed at the end of the summary, provided in the previous section.

Conclusion 1: The use of the Computer Conferencing Model of Tryout-Revision, as operationalized in this study, and the Small Group Model of Tryout-Revision, as operationalized in this study, do not appear to lead to the general improvement of an instructional system, with respect to cognitive learning.

Discussion of Conclusion 1

This conclusion, based on Finding 1, suggests one of four possiblities. One possibility is that previous research on tryout-revision, showing the general usefulness of learner tryouts for the improvement of instructional prototypes, is ill-founded. The second possibility is that the models used in this study, as operationalized, are not capable of producing a set of tryout data that is sufficient for making changes in the prototype which can lead to significant improvement in the cognitive learning outcomes of learners. The third possibility is that the developer for this study did not make changes in the prototype, based on the tryout data, that were sufficient for bringing about significant improvement in the cognitive learning outcomes of the learners. The fourth possibility is that the manner in which the study was carried out had the overall effect of lessening the cognitive learning outcomes of the learners to such a degree that less than significant improvement in such learning outcomes resulted.

In view of the limitations of this study and the problems met in carrying it out, the writer does not wish to contend that previous research in this area is ill-founded. In Chapter 3, the section, "The Debriefing Data" is provided so that the reader may judge whether the developer made proper use of the tryout data.

In the next section of this chapter, "Further Discussion of Some of the Findings of No Significant Difference" the reader is provided with numerous suggestions as to how the manner in which the study was carried out may have had a negative influence on the level of improvement in learning outcomes resulting from the tryout of the revisions.

This leaves the matter of the effectiveness of the models, to the way in which they were operationalized in this study, as the remaining possibility. Conclusion 2, which follows, deals with that possibility.

Conclusion 2: Conclusions cannot be drawn about the effectiveness of tryout-revision models of tryout-revision based on the use of debriefing data only.

Discussion of Conclusion 2

The small group model and the computer conferencing model used in this study were based on debriefing data only. Since these models were shown to be generally ineffective for improving an instructional system, as indicated in Finding 1, and as discussed in Conclusion 1, it is possible that debriefing data alone is not sufficient as a basis for revising an instructional system. However, since these models were not compared to models which used more extensive sets of tryout data, no general conclusions can be drawn about the relationship between types of tryout data and the effectiveness of tryout-revision models based on such tryout data.

Conclusion 3: The use of the Computer. Conferencing Model of Tryout-Revision, as operationalized in this study, appears to be capable of leading to the improvement of the content of an instructional system that is related to the changes incorporated into a revision based on the application of this model.

Discussion of Conclusion 3

This conclusion is based on the fact that learners who tried out the CC revision had significantly higher posttest scores than did learners who tried out the prototype, with respect to Instructional Objective 2, as stated in Sub-finding 1. Instructional Objective 2 was as follows: "To identify the functions of an information processing system."

As was indicated in Chapter 3, in the section, "The Debriefing Data," the CC revision incorporated three significant changes in the prototype that relate to the content covered by Instructional Objective 2. Those three changes occur in three sequential segments of the instructional system. In addition, less significant changes were incorporated into the CC revision, relating to Instructional Objective 2. As was indicated in Chapter 3, these 3 significant changes are as follows:

1. The "taxpayer" scene. In this scene, the viewer is provided with an intuitive notion of the functions of an information processing system. The scene shows a man preparing his income

tax forms. The presenter, who is "off-stage", explains the scene in terms of information processing system functions.

2. The presenter formally defines the functions of an information processing system. The change represented in this revision is that these definitions are simultaneously shown on graphics while the presenter is stating them. The graphics were prepared through use of professional-type graphics supplies. They were prepared by the program's director, who was a senior at MSU majoring in telecommunicaton and another MSU senior also majoring in telecommunication.

3. The presenter tells the learner to "stop" the tape and answer the questions in the viewing guide. These questions are all constructed-response-type questions, requiring the learner to define a particular information processing system function. The learner is referred to the correct answers by a note at the bottom of the page listing the questions.

Other less significant changes in this revision relating to information processing system functions are as follows:

1. The viewing guide provides definitions of these terms, which the learners are free to read since the viewing guides were distributed to them immediately before the television program was played back.

2. The names of the information processing system functions are shown in later segments of the program through the use of the "key wipe" feature of the special effects generator during the television program. This technique made it possible to

associate portions of the demonstrations of information processing systems with the names of these functions.

Thus, a possible reason for the scores of learners who tried out the CC revision being significantly higher than those of learners who tried out the prototype, with respect to Instructional Objective 2, is the combined effect of the changes incorporated into the CC revision. Since these changes were based on the debriefing data from the CC debriefing, it suggests that the Computer Conferencing Model of Tryout-Revision, as operationalized in this study, is capable of leading to the improvement of the portions of an instructional system that are related to the changes incorporated into a revision based on the application of this model.

Hence, in spite of the fact that when the prototype was compared to the CC revision there were no significant differences in cognitive learning outcomes on the basis of differences in the means of the total posttest scores, as pointed out in Conclusion 1, does not necessarily mean that some differences in cognitive learning cannot occur as the result of trying out a different version of the instructional system.

This conclusion points to three different types of phenomena. First, in the writer's opinion learning is not a uniform monolithic structure. Learning comes about as the result of specific components of the instructional process, differences in the learner's capability to learn from different stimuli, and from the interrelationships between these two factors. Therefore, the implications of this type of phenomena for instructional development is that specific aspects of learner behavior and instructional stimuli need to be kept in mind as ir

ne

de

or

a

of

le

sp

ov

ca

st

mod

Con

Ess

the

char

instructional systems are being developed.

Second, in the process of measuring instructional outcomes it is necessary to keep in mind that different results can be obtained depending upon what is measured and how it is measured. Hence, if criteria and/or instructional objectives are used in the development of a test instrument, it may be wise to define learning outcomes in terms of such specific components of the total learning outcome for each learner. Moreover, it may be best to write hypotheses based on specific instructional outcomes, rather than on total test scores or overall attitude measures. In this way the more specific predictions can be made prior to the data collection. This would strengthen the study in cases where a null hypothesis is rejected.

Conclusion 4: The Computer Conferencing Model of Tryout-Revision, as operationalized in this study, is more effective for improving an instructional system with respect to cognitive learning than the Small Group Model of Tryout-Revision, as operationalized in this study.

Discussion of Conclusion 4

This conclusion is drawin on the basis of Finding 2. Since both models were shown to be ineffective on the basis of Finding1, Conclusion 4 needs to be interpreted with extreme caution. Essentially, it indicates that, although neither model can improve upon the prototype, the computer conferencing model is relatively better than the small group model.

Perhaps this relatively better performance of the computer conferencing model is primarily due to the effect of Instructional Objective 2. That instructional objective accounts for 8 of the 21 test items on the posttest or 38% of the total posttest score. No other instructional objective accounts for such a large proportion of the total posttest score. The instructional objective that accounts for the next highest percent of the test items is Instructional Objective 4, which accounts for only 28% of the test items.

Further evidence for this contention is provided by Finding 3. Here, the data analysis indicated that when the two models are compared, with respect to Instructional Objective 2, application of the computer conferencing model led to learning that was significantly greater than the learning resulting from the application of the small group model.

As was indicated in Chapter 3 in the section, "The Debriefing Data," the computer conferencing revision incorporated instructional design elements that are not found in the small group revision; namely, the "taxpayer" scene and the use of the viewing guide, including the request by the presenter that the learner stop the tape and answer the questions in the viewing guide. Both of these elements pertain to the learning of the content related to the Instructional Objective 2. On the other hand, there are no instructional design elements, or other elements related to learning, contained in the small group revision that are not found in the computer conferencing revision.

Thus, it can be concluded that the computer conferencing model provides extra instruction with respect to Instructional Objective 2. It seems reasonable to conclude, then, that when the two revisions were compared with respect to Instructional Objective 2, the significantly greater learning resulting from the computer conferencing revision was primarily due to the extra instruction provided in that revision that related to Instructional Objective 2. Hence, this research finding seems to be the resultant of the good matching that was done between the revision effort, on the one hand, and the number of points on the posttest relating to the content in the prototype for which the major portion of the revision effort was carried out, on the other hand. But, it cannot be concluded that this revision effort which had a high "payoff" was necessarily attributable to the superiority of the tryout-revision model that was used for that revision.

Furthermore, no conclusion can be made about the general superiority of the computer conferencing model over the small group model, since it was tried out with only one prototype and especially because that prototype was tried out with only one posttest. Therefore, the results are inconclusive as to which of the two tryout-revision models is the better model. Only through further research based on these models with other instructional systems will conclusive results be obtained.

Conclusion 5: Neither the use of the Computer Conferencing Model of Tryout-Revison, as operationalized in this study, nor the use of the Small group Model of Tryout-Revision, as operationalized in this study, led to improvement in attitude toward the instructional system.

Discussion of Conclusion 5

This conclusion, which is based on Finding 3, suggests that perhaps the changes incorporated into the two revisions on the basis of the tryout data were not sufficient for bringing about sufficient improvement in attitude toward the instructional system. The attitude scales dealt mostly with the presentation aspects of the television program's audio and video stimuli. The fact taht the experimental subjects were drawn from instructional media courses, may mean that they were an unusually critical sample of experimental subjects, with respect to media considerations.

It should be pointed out that the revisions were produced on a very low budget and though the use of a volunteer director, a volunteer production crew, and volunteer talent. Given that the learners were students of instructional media development it might have been extremely difficult to satisfy their expectations through the sue of such resources. This is especially true in view of the fact that changes in the prototype were limited to ideas expressed during the debriefing, based on the prototype. Thus, other revision possibilities could not be considered. It remains to be seen how more typical learners would react to the revisions, and how a more professional production of the revisions might result in different findings with respect to these attitude measures.

Conclusion 6: The use of the Computer Conferencing Model of Tryout-Revision, as operationalized in this study, did not lead to more improvement in attitude toward the instructional system than did the use of the Small Group Model as operationalized in this study.

Discussion of Conclusion 6

This conclusion, which is based on Finding 4, suggests the possibility that the computer conferencing model is not as effective for handling questionnaire items as it is in handling posttest items. This interpretation is based on the fact that Conclusion 4 referred to the greater effectiveness of the computer model in comparison with the small group model, with respect to cognitive learning.

However, this interpretation does not appear to be completely well founded, since the process for discussing both types of items in the computer conference is the same. Yet, it may be possible that it is easier to give feedback on posttest items than on questionnaire items, since the former are more concrete than the latter. In addition, as in the case of Conclusion 5, the problem of improving a television program to meet the standards of instructional media students may have played a role.

Conclusions relating to the workability of the Computer Conferencing Model of Tryout-Revision

The conclusions discussed in this section all relate to the development of the computer conferencing model.

Conclusion 7: It appears that computer conferencing is operationally feasible as a means of debriefing learners in tryout-revision.

Discussion of Conclusion 7

This conclusion is, of course, based on very limited data; that is, it is based on the use of computer conferencing, where all the participants were "on-line" at the same time in the debriefing of one sample of learners who tried out one instructional system. The criteria or deciding whether the use of computer conferencing was operationally feasible were as follows:

1. Were the learners able to participate in a computer conference dedicated to the collectionof learner debriefing data?

2. Were the learners willing to participate in such a computer conference?

In the case of both criteria the answer was a definite yes. In the case of the first criterion, only one person out of the six participants in the second tryout for the experiment has serious difficulites. However, she did contribute, substantively to the debriefing data. The other five participants performed with only minor problems. It should be pointed out that two technical assistants were on hand to facilitate the process.

With respect to the second criterion, none of the six participants expressed serious negative feelings about the experience, and most of them expressed appreciation for being given the opportunity to participate.

Conclusion 8: It appears that learners can be trained in the use of computer conferencing software and computer terminals within a reasonably short training period.

Discusison of Conclusion 8

It was found that, on the basi of a 15-20 minute individual tutoring training session plus the use of a training manual that was distributed to the learners prior to the training sessions, it was possible to train learners for participation in a computer conference.

This included training them to "log-on" the computer and make use of a few commands in the CONFER computer conferencing system.

Conclusion 9: Consensus among the learners was difficult to reach.

Discussion of Conclusion 9:

It was difficult to arrive at consensus among the learners due to various problems. First, they did not seem to be aware that the debriefer was awaiting responses from them for follow-up questions asked in response to learner disscussion votes. This may have been due to the manner in which the debriefing was conducted. The learners and the debriefer were required to continuously request displays of the votes from the computer and to provide additional discussion votes, in response to the other discussion votes.

However, there were time delays in obtaining the displays of the votes. This was primarily due to the "problem" in the CONFER II software in being able to respond properly to the CONFER command modifer, "votes new." After a few uses of this modifer, and receiving the response, "NO ITEM SATISFIES THIS REQUEST," the learners were instructed to ask for all the votes, through use of the command modifier, "votes." This caused the computer to display all the votes for a conference item, which caused annoyance and boredom for the learners, since they were already familiar with the older votes.

As a result, it appeared to the writer, acting as a technical assistant during the conferencing that they were inclined not to read all the votes. They simply waited for all the votes to be printed at the terminal, and then entered a discussion vote. Thus, if the debriefer had asked a follow-up question, the learners would not necessarily have been aware of that, since, in many instances, they were busy mentally composing their next discussion vote or jsut waiting for the prompt to enter another vote. Hence, the problem with the "votes new" modifer had a serious effect on the quality of the debriefing, since it meant that the whole process was slowed down and that the learners were overloaded with information they already had and were not able to easily concentrate on the information.

Second, the debriefer did not always seem to be following up with additional questions, based on responses from the learners. The debriefer indicated after the debriefing session that she "kept waiting" for responses from all the learners for a question before entering her next question. This, of course, caused delay, since the learners were not always answering questions.

In fairness to the designer of CONFER II, Robert Parnes, it should be pointed out that the "votes new" modifier had not originally been designed to be used more than once for a conference item during a single CONFER session. Normally, a participant in a CONFER conference has no need to check the new votes for an item more than once during a CONFER session. In thus study, an attempt was made to use the CONFER software, for a synchronous conference, even though the software had had been primarily designed for asynchronous conferences. However, the CONFER II software has been modified since the time of the computer conference used in this study, so that it responds as it should whenever the "votes new" command modifier is used.

In general, the debriefing aspect of the model needs to be seriously examined, so that a regular procedure is developed and taught to the learners prior to the debriefing session. Another solution would be the use of "ALL CAPS" typing for the debriefer only, so that communications from the debriefer stands out from the other communications. Further comments will be made about the debriefing process in the section, "Recommendations for the Future Development of the Computer Conferencing Model of Tryout-Revision".

Further Discussion of Some of the Findings of No Significant Difference

The following factors may be related to the inability to reject the null hypotheses concerning the comparison of the revisions to the prototype:

1. The subject matter of the instructional prototype was very abstract and not necessarily suited to a television program alone. The subject matter dealt with definitions of and the names of the functions of an information processing system. Such instructional content is difficult to learn in a television program through the use of a presenter and definitions printed on graphics cards. Thus, it might have been a case of not using a more appropriate instructional strategy for such content. If the learners who tried out the prototype had greater skill in handling verbal information provided through a visual medium than the learners who tried out the revisions, this factor could have been a countervailing factor contributing to the finding of no significant difference. No attempt was made to control for this factor in selecting experimental subjects and assigning them to the treatments.

2. The learners were all graduate students, and therefore experienced in studying, learning how to learn, and taking tests. Thus, perhaps an instructional medium at any quality level might have provided the information needed for such learners to answer the posttest, especially since a pretest was used. The literature reviewed for this study did not reveal any studies on the effectiveness of tryout-revision that involved the use of graduate students.

3. The learners did not volunteer to participate in any real sense. Their instructors gave permission to the writer to make use of regular class time for the experiment and their cooperation was sought during the writer's introduction prior to administering the pretest. Therefore, there could have been resentment due to the use of class time for participating in the study. This situation could have caused poor learning, regardless of the quality of the instructional system. This factor is especially significant in view of "4," below.

4. The learning material was not related to the subject matter being studied in the classes used in the study. Thus, the subject matter was unexpected, albeit planned in advance, during the class time and perhaps perceived as a poor substitute for the regular learning that would occur during that class period. The literature indicates that most studies in tryout-revision have involved the use of instruction that fitted into the regular learning for the learners used in the study. Thus, it may be that such correspondence between normal learning and the subject content of materials used in tryout-revision studies is a necessary condition for the effectiveness of tryout-revision.

5. The classes used for the study, including the pilot study were instructional media classes. Since such classes place considerable emphasis on the production aspects of such media, they may have focused more on how the program was produced than on the content. If so, they may have "missed" the learning material being taught.

6. On the other hand, there was a very significant difference between the two debriefing groups. As was pointed out in Chapter 3, the experimental subjects assigned to the prototype were randomly assigned to the two debriefing groups. All these students were from the class in Photography in Instruction. However, due to insufficient debriefing data from the computer conferencing tryout, a second tryout was carried out for the latter model. This second tryout would have included tryouts for both models had there been a sufficient number of learners available. However, when it became apparent that only one class in "Effective Use of Instructional Media" with 10 students was available for a second tryout, it was decided to retain the data for the small group model and carry out a second tryout for the computer conferencing model only.

As a result of this eventuality, the learners in the two debriefing groups were not randomly assigned from a common pool of subjects. Rather, two academic classes were assigned, but not randomly, to the two treatments.

To make things even worse, though it wasn't considered a serious problem at the time it occurred, the instructors and the course content for these two instructional media classes represented two very different approaches to the development of instructional media. In the case of the "Photography in Instruction" class, both the instructor and the course are strongly oriented to the production aspects of instructional photography. However in the case of the "Effective Use of Instructional Media" class both the instructor and the course are strongly oriented to a very theoretical and conceptually global approach to the subject even though media production is involved. The instructor is known for his questioning attitude, particularly the questioning of underlying assumptions relating to instructional development.

Hence, it can be argued that the learners for the two tryout-revision models had different mental sets in their approach to the tryout and debriefing process. In the case of the photography students they could have been looking strictly at the visual aspects of the television program. In the case of the other group of students they could have been concerned with more fundamental considerations, such as "Why use television for this instruction?" and, "What other media might be combined with the television program?" Thus, it is felt that, although in many respects equivalent groups of learners were used for the two debriefing groups, in many ways they were not equivalent.

7. In carrying out the study, it's quite possible that the learners who tried out the prototype were more motivated than those who tried out the revisions. This could have resulted from the fact that prior to the prototype tryout the learners had been told that they would be participating in a debriefing; however, the learners assigned to the other treatmetns were not told they would be involved in a debriefing. Therefore, if the expectation of the learners assigned to

the prototype were more attentive during the tryout, in order to be prepared for the debriefing, than the learners assigned to the revisions, some inequality in the learning could have occurred.

8. The sample was small and not randomly selected and only to a degree randomly assigned. Only 13 learners tried out the prototype, and 18 tried out each of the two revisions, for a total of 49 learners. As explained in Chapter 3, the learners for the two revisions came from three different Summer Term classes and were randomly assigned to the two revisions. The learners for the prototype came from two Spring Term classes. The phrase "came from" is used here instead of "selected from," in view of the fact that all the students in these classes participated except those who were ineligible due to prior viewing of the instructional system used in this study.

9. The posttest had only 22 items of which only 21 items were usable, for data analysis purposes. One test item was eliminated during data analysis, due to contradictory information given in the prototype which was covered by one test item. The split-half realibility coefficient was .777, which is reasonably high. However, the test was, first of all, based on instructional objectives. Then, for each of an original set of 11 multiple choice questions, a true/false question was written. This procedure suggested the use of split-half reliability as an appropriate measure of reliability.

However, the test was divided into two parts, labeled, "Part A" and "Part B." The Alpha reliability coefficients for these parts were only .459 and .612, respectively. Moreover the correlation between the two parts was only .657.

These latter reliability coefficients are quite low and indicate that the test items were not measuring the same underlying construct. It appears, then, that there was consistency among the test items for each instructional objective, as indicated by the split-half reliability coefficient, but that the test as a whole was not reliable, as indicated by the Alpha coefficients.

10. The debriefing process in the two debriefing sessions were not as well carried out as was anticipated, as mentioned in the last section of this Chapter. Consequently lower levels of consensus then anticipated were reached especially in the computer conferencing debriefing.

In the small group debriefing, a consensus, of sorts, was easily obtained with respect to the matter under discussion by asking for a "showing of hands." Even though consensus generally implies total agreement after group discussion of all the variations of opinion, this more limited type of consensus indicated by the outcome of a vote is still useful.

The logical equivalent to this in the computer conference was attempted by asking the learners to indicate their agreement to a particular point, by entering a discussion vote on the matter. In a computer conference where all the participants are on-line at the same time, (synchronous mode), this type of voting seems to have certain weaknesses.

First, there is a time lag between the asking of a question and the receipt of the answers to the question by the question asker.

This is due both to the nature of computer conferencing in general, and to the way computer conferencing was operationalized in this study. In this computer conference, all the discussion was carried on through discussion voting, as was explained in Chapter 3 and in the last section of this Chapter. Thus, the learners were required to continually request the new votes from the computer and then make additional comments and respond to whatever questions might have been asked in those new votes, especially those asked by the debriefer. For example, the debriefer could ask how the rest of the group feels about a comment just made by one of the learners. To respond to that question, the learners would enter a new vote.

This is, by necessity a much slower process than the equivalent process found in face-to-face debriefing, where a debriefer leads the discussion and can elicit responses immediately, which can be posted on a drawing pad for all participants to see. The debriefer can then ask for a "showing of hands," with respect to agreement on any of the ideas that were posted.

Another aspect of this factor was that the "votes new" command modifier in CONFER did not function as it was expected, as was explained in Chapter 3. Thus, the learners would try to obtain the new votes only through use of this command modifier. If they had no luck with it, they were instructed to use the "votes" modifier. However, this slowed down the process even more, because it meant that a learner would have to wait for the computer to print out all the votes for a particular conference item, in order for the learner to see a relatively few new votes, and to enter a new vote. This also caused some negative feelings toward the conferencing system.

A second factor is what might be called "the cocktail party factor."

In the case of both cocktail parties and synchronous computer conferences, one may feel a need to respond to several comments at the same time, but only one response can be made at a time. As in the case of cocktail party behavior, one must choose whom to respond to. In this computer conference, such a response was accomplished by entering one new discussion vote. (N.B. this is not a constraint of the conferencing system, but, rather, an element of the debriefing process used in this study.) Thus, it is easy to see how some questions asked by the debriefer were never answered by some learners, due to the simultaneous multiplicity of input, to put it in formal language.

Thirdly, in debriefings of this type, the amount of time that a debriefer can spend on a given issue is dependent upon the total number of items on the debriefing agenda, and the total amount of time available.

In the case of this study there was only about an hour and a half available until the end of the regular class period, at which point the debriefing had to end. Furthermore, many test items and questionnaire items met the 30% criterion discussed in Chapter 3 (i.e., 30% or more of the learners answering incorrectly to a test item or responding unfavorably to a questionnaire item).

Thus, it was a situation in which the debriefers were under presure to keep the discussion "moving" in order to cover as many agenda items as possible in the limited time available. In addition, it should be added that Abedor (1981) claims that interest in debriefing begins to lag after about an hour. Thus, the time available was not inadequate with respect to what was required, on the basis of Abedor's experience with face-to-face debriefing since 1971.

The debriefers were instructed to place more emphasis on covering individual agenda items adequately rather than covering as many agenda items as possible, if a choice had to be made. This, it was felt, relieved some of the pressure, but the debriefers were still conscious of the time factor, according to post-debriefing debriefings that the writer held with them.

Hence, with respect to the fact that some questions of the computer conferencing debriefer was not answered by all the learners, there are two aspects of the time limitation factor. On the one hand, the computer conferencing process is slower. With the limitation of total time available, the computer conferencing model was at a disadvantage in view of its natural slowness. But on the other hand, the computer conferencing debriefing could have made use of its time more efficiently by following a more structured process than it did, so as to compensate for the slower process.

This problem was not anticipated and planned for during the training of the computer conferencing debriefing. One possibility would have been for the debriefer to try to discourage the numerous extraneous remarks made in the conference that had nothing to do with the debriefing process. A second procedure would have been for the debriefer to enter more votes summarizing the remarks made up to that point with respect to a given debriefing item and ask for new ideas to add to the list. This was done only to a small degree. Then, this list could have been voted on in terms of priority rankings for the individual items on the list, and then voted on as a whole, as the group's consensus of opinion.

By contrast, this problem of accomplishing the task in the time required didn't seem to arise in the small group debriefing. But it seems that the nature of such debriefings lends itself to more efficient and cooperative activity than does computer conferencing, as carried out in this study. The debriefer can efficiently post the deficiencies in the prototype and recommendations on a drawing pad, as the learners make their comments, one after the other. Thus, there can be more concentrated effort. This seemed to be the case in this study.

Thus, the time lag in a computer conferencing debriefing can act as an inhibitor of human cooperation, when the total time available is limited, as when all the participants are on-line at the same time. This is not as great a problem when the conference is spread out over several weeks or months, when various other features of computer conferencing can be used that were not used in this study. But, as previously indicated, various means for structuring the communication process are necessaary in order to overcome the naturally slower communciation process of synchronous computer conferencing. Such structuring of the communication process would make such computer conferences more like face-to-face debriefing with respect to structured communication.

Fourthly, there is a limit in the amount of energy that learners are willing to expend and the amount of interest that learners can be expected to have in a computer conferencing debriefing. Except for one learner who had experience with word processing none of the learners participating in the computer conference had previous experience in the use of computer terminals or microcomputers.

On the basis of experiential knowledge regarding the use of a computer terminal for the first time and learning how to use a few computer commands, it seems that the initial experience should be quite short. In addition, limitations in typing skill can be very demanding on one's energy when one is expected to participate in a process that is dependent upon the use of typing skills in order to communicate.

In this study, the computer conference was held from 8:30 to 10:00 in the evening. Five of the six learners held full time jobs and the sixth person had responsibilities as a single mother. Thus, the timing of the conference was not highly suited to their energy and interest levels, to say the least. Two of the learners had serious limitations in typing skill. Thus, in combination with the fact that computer conferencing, and computer use in general, was a unique experience for five of the six participants in the computer conference and that the conference lasted for one and a half hours, it appears that limitations in energy and interest may have been a limiting factor in arriving at consensus in the computer conferencing debriefing.

In summary, with respect to the limited amount of consensus reached in the computer conferencing debriefing, it seems that 1) the slowness of computer conferencing when carried out with all participants on line at the same time; 2) the limitation of structure in the conferencing process; and 3) various limitations in energy and interest all played a role in limiting the amount of consensus that was reached in the computer conferencing debriefing.

In summary, then, all these factors taken together might account for, at least to some extent, the inability to reject the null hypotheses relating to differences between the prototype and the revisions; that is, H_1 and H_3. In other words, the insignificant differences in learning found in this study may have been due, at least partially, to the experimental conditions rather than to the experimental treatments.

Recommendations for the Future Development of the Computer Conferencing Model of Tryout-Revision

The following suggestions seem to be among the most critical:

1. The model needs to be operationalized in the asynchronous mode

   The asynchronous mode makes it possible for all parties concerned to fit computer conferencing into their schedules when it's most convenient and when they have the needed amount of interest, motivation and creativity to participate most effectively. Since learners are tied to very specific schedules, in the case of full time students, it is important to avoid making any specific demands of them as to when they should participate. This is not only a practical necessity, in view of class schedules, but it also encourages their participation if they are allowed the freedom to choose the most convenient time for participation. However, there is also the problem of insuring that they participate on a regular basis.

2. <u>Consensus needs to be improved</u>

Consensus could be improved through several strategies. The following are key possibilities:

a. <u>Use the Delphi technique</u>, whereby the participants are provided with group data on the level of consensus from the previous "round" of votes, as well as the opportunity to request clarification on the meaning of any of the comments made by other participants. Delphi could be used as a means of establishing the agenda for the debriefing rather than depending upon the analysis of outcome data. By doing so, the "sense of the group" would be the basis for discussion rather than test and questionnaire items, which may not reveal the basic deficiencies in the prototype.

b. <u>Summarize</u> the comments made in the conference on a fairly frequent basis and then ask for <u>prioritization</u> of the comments in terms of their importance to revision strategies. Then ask for <u>implementation strategies</u> for the high priority deficiencies. This would keep the group working on the most serious matters and provide for <u>revision data</u>, rather than just prototype deficiencies. Revision ideas could also provide the basis for the establishment of "working groups" which could interact as separate small groups and then recommend revision strategies to the larger group.

c. Ask for clarifications and/or alternative opinions when an idea is suggested that is difficult to understand or sounds like an ill-advised suggestion.

3. Operationalize on-line tutorials and individual debriefings

a. On-line tutorials will need to provide the impetus for the learner to feel inclined to contact the tutor when help is needed. Therefore, special strategies for developing trust and personal warmth in the absence of the usual non-verbal communication used in face-to-face and telephone communications will need to be developed if successful on-line tutorials are to be realized.

b. Individual debriefings need to be developed as a follow up to the receipt of an individual's test results in order to determine the sources of his problems. Here is where categories of learner and program failures, as suggested by Gropper (1975) are needed in order to collect process data in an orderly, scientific manner. However, debriefers would need to be mindful of the need for new categories to handle problems that don't fit the existing categories. Training in the use of a set of categories would also be an important requirement.

4. Compare use of conference "items", and their equivalent in other computer conferencing systems with CONFER-type discussion voting to determine which type communication is most

feasible or whether both are needed.

5. Compare EIES and CONFER for their relative advantages and disadvantages for debriefing and on-line tutorials.

Recommendations for Future Research

On the basis of this research the following recommendations are made for future research with the Computer Conferencing Model of Tryout-Revision.

1. Improve the model on the basis of the Recommendations for the Future Development of the Computer Conferencing Model of Tryout-Revision, mentioned in the last section.
2. Test the model with a variety of types of instructional formats, types of learners, and types of content, making use of categories of learner and program faults such as those suggested by Gropper (1975).
3. Carry out what was indicated in "2," above but also collect tryout data on program design elements and media characteristics that faciliate learning. This type of research would be in line with the work of Palmer (1974) at the Children's Television Workshop (CTW), Salomon (1974), Bruner (1966) and Suchman (1967). Palmer and others at CTW are concerned with formative research in television program development, whereby they identify and refine program design features which are reliable predictors of learning outcomes. They are interested in the development of principles that relate such predictors and outcomes which are potentially generalizable

and thereby contribute to "the science and techology of learning (Palmer, 1967, p. 329)." He perceives such research as being a step between basic research and educational practice, as well as providing for better practice.

Salomon's interest is in the identification of the inherent characteristics of different media, particularly their symbol systems, which may be related to the learning of specific types of learning tasks for particular types of learners and, thus, the attainment of types of instructional objectives.

Suchman (1967a) and others are concerned with a science of evaluation research, within which cause-effect hypotheses could be developed and thereby offer "a bridge between 'pure' and 'applied' research"(p. 350).

Bruner (1966, pp. 40-42) in commenting upon a prescriptive theory of instruction says that a theory of instruction should specify experiences that provide predispositions toward learning; it should also specify how knowledge should be structured so as to facilitate learning; it should specify effective sequences for presenting learning outcomes; and it should specify rewards and punishments that lead to learning.

These, then, are some research ideas that could be pursued through the use of a model of tryout-revision which collects greater amounts of process data through the application of computer conferencing.

Computer Conferencing, Cost-Effectiveness, and the Future of Instructional Development

One of the questions that many people raise in discussing the appropriateness of computer conferencing is its cost-effectiveness. As in the case of other technological innovations, the cost-effectiveness of computer conferencing is, by no means, easy to determine. This is primarily due to the fact that no one knows all the types of uses to which this medium might be put and the resulting benefits. If, for example, computer conferencing were used for instructional development, in general, then its application to tryout-revision would be a relatively minor application, since the bulk of instructional development work is concerned with other activities. In many ways, several other aspects of instructional development are more suited to computer conferencing than tryout-revision, since more professional time and effort could be saved through such uses.

In tryout-revision, the involvement of tryout learners at several institutions could come about as the result of inter-institutional cooperation in the development of an instructional system, in which case the cost of the computer conference could be shared among the cooperating institutions. In such a case, the cost-effectiveness of computer conferencing could be quite high. If this medium were used by a single institution for the development of an instructional system for local use only, the cost-effectiveness of computer conferencing could be quite low.

However, the use of computer conferencing and the involvement of learners at several instiutions could lead to the involvement of some or all of the other institutions in the development of that instructional system or perhaps a future instructional development project. In such an eventuality, computer conferencing serves as a tool for the solicitation, intentionally or otherwise, of co-sponsors for the project. In these days of limited development funds, computer conferencing could serve as the harbinger mechanism for bringing about a major shift in instructional development from a "local" to a "network" scale of operations. Under such circumstances, central agencies might be needed to serve as "match makers" for development projects.

Moreover, this type of inter-institutional cooperation could cut across industry, universities, government, public school systems, consulting firms and international non-governmental organizations. Another type of inter-institutional cooperation could be cooperation across national boundaries.

The question could be raised, of course, "Why do these types of cooperation need to depend upon the development of computer conferencing?" One answer to that question is that, by and large, cooperation usually involves a considerable number of meetings, and the reading of memoranda and other reports, and a considerable amount of travel.

Computer conferencing allows for the attendance at meetings by staying at the office or at home or wherever one's computer terminal is located. The computer conference contains all the materials to be read, questions to be responded to and/or voted on, and messages to answer. Moreover, the participant can schedule his/her participation according to individual needs and not be required to make schedule changes or travel away from home at inconvenient times. In addition, the individuals is not absent from regular activities for a day or more while attending an out-of-town meeting.

Thus, computer conferencing provides a means for inter-institutional cooperation in a manner that is very cost-effective, time-efficient, and personal-energy-efficient. From a cost standpoint alone, the time of professional and managerial personnel is usually the most expensive element in a cost estimate. For an administrator earning $30,000 per year to be absent from "the office" all day to attend a two-hour meeting out-of-town, it would cost approximately $14.50/hour x 8 hours = $116.00, not including travel costs.

By comparison computer conferencing is cheap. To use the Wayne State University version of CONFER II, where non-University of Michigan CONFER conferences are being organized, costs about $2.00 per hour for use of the computer. In addition, it costs about $6.00 per hour to use the telecommunication utility, Telenet, which would be needed to connect with the Wayne State computer from a distant city. The cost of using the CONFER II software is $8.00 per hour. If a person is located in a city where there is no Telenet port, there would be long distance charges to dial such a port. If it were assumed that the total cost per hour for using CONFER were $18.00, two hours of conferencing would cost $36.00. Therefore, by comparison to traveling to an out-of-town meeting, using CONFER is very cheap, less time consuming, and less wearing on the meeting participant. In addition, he or she does not have to be absent from regular office and personal activities.

Ultimately, the question is reduced to one of cost-accounting and who pays for what costs. With respect to tryout-revision, carrying out local tryouts is very cheap. Learners aren't usually paid for their participation and instructional developers are paid regardless of their specific duties on a given day. Therefore, by comparison, computer conferencing is very expensive for a single institution developing materials for local use only.

But, is this the way to develop instructional materials? Should local resources be the sole basis for instruction provided by an institution? Should local learners be the sole source of tryout data, even for local resources? If knowledge is created through the combination of ideas generated by many individuals at many locations, should not instruction also be created through the combination of ideas of many individuals at many locations? These questions cannot, of course, be answered here, but they need answers.

Turoff and Hiltz (1981) have predicted a network society. If instructional development is to be part of this network society, it seems important to begin planning for that eventuality now. One way to begin the process would be through the development of small-scale computer conferences. Tryout-revision is one area in which such conferences could be begun with only minimal resources.

Some Final Thoughts

It seems to the writer that computer conferencing is an idea well worth pursuing in the field of instructional development. The tryout-revision area seems like a good place for that process to begin, s

a

a

d

b

c

d

c

i

t

r

b

f

h

c

tr

al

t

f

o

po

o

since this area can be examined in a research and development context without involving an actual instructional development project.

At the beginning of this research project, it was not known whether computer conferencing would "work". In spite of various difficulties, it worked, at least operationally. But it could work better. For that reason, it is important that the idea of a computer conferencing model of tryout-revision be pursued through further development and research.

As the cost of using computers is reduced, and as the use of computers increases, computer conferencing should be playing an important role in instructional development. But, it seems important to prepare the way for that role now, so that the profession will be ready to use this technology when the idea of computer conferencing becomes more popular.

Computer conferencing in instructional development could be a way for institutions to share their resources better, especially their human resources. This would be due to savings in travel time, travel cost, and the cost of professional staff time that is made necessary by travel. Though computer conferencing will never be a substitute for all other types of communication, it could serve as a substitute for those types of communication which are not necessarily improved by face-to-face or telephone communication.

The important matter, however, is to concentrate now on the operationalization of a model of tryout-revision that makes the best possible use of computer conferencing. Once a good model is developed, its diffusion could take place without much effort, due more to its intrinsic worth than to selling the idea. If this comes about, instructional development could enter a very vital phase in its development.

APPENDICES

APPENDIX A

A GUIDE FOR THE PARTICIPANTS IN PAUL TONER'S RESEARCH PROJECT INVOLVED WITH COMPUTER CONFERENCE

PAUL TONER

PHONE: 3-2047 (leave message)

1. What is this research project about?

The purpose of this research project is to compare the small group model of tryout-revision (formative evaluation) with a new model that makes use of computer conferencing as an alternative to the small group debriefing process. Essentially, the study is attempting to determine whether this new model is just as effective for the revision of prototype instructional systems as the small group model. If it is, then it may mean that larger, more heterogeneous, and more geographically diversified samples of learners can be used in formative evaluation, especially in large development projects, for obtaining both outcome and process data.

The study will make use of an early prototype of this new model, with future research providing the refinements not possible with the existing constraints on available resources.

Another group of students have already tried out the prototype material and were debriefed by use of the small group debriefing technique.

The results from this study will provide the basis for a doctoral dissertation in educational systems development.

2. What is computer conferencing?

Briefly, computer conferencing is a type of communication system made possible by the use of computer terminals interconnected to a common computer by the telephone system. The participants of a computer conference share access to a particular computer file and thereby are able to communicate with each other. This particular computer conference will make use of a very limited number of features of a computer conferencing system, called CONFER, which resides in the University of Michigan mainframe computer.

3. <u>What will we be doing in this conference?</u>

The purpose of this computer conference is to provide feedback on a videocasssette used in the ED882 Seminar on computers in Education. We want to know what is wrong with this videocassette and what should be done to correct its deficiencies. Therefore, you are encouraged to be as honest and straightforward as you possibly can in your comments about the videocassette, and to be as specific as possible so that your comments can be translated into specific revision procedures. Since the conference has been designed to approximate as closely as possible a face-to-face debriefing group, you should try to communicate in a style that is common to such groups; namely, be informal, don't be too concerned with typing errors or the use of non-standard English, and don't be reluctant to react to the comments entered by the other participants.

4. <u>What is the agenda for the tryout-debriefing session?</u>

The tryout-debriefing session will consist of two stages of activities, as follows:

a. <u>Tryout of the prototype material in your regular classroom (approximately 45 minutes):</u>

(1) Take a pretest.

(2) View the videocassette.

(3) Take a posttest.

(4) Answer a questionnaire.

b. Debriefing at Lyman Briggs College

This will take about two hours. We will try to finish by 10:00, but in no event will it extend beyond 10:15.

5. How will the debriefing be structured?

The debriefing will be structured on the basis of a quick analysis of the posttest and questionnaire results. The posttest items will be ranked according to the number of wrong answers given. The questionnaire items (Likert-type scales) will be ranked according to the number of neutral and negative ratings given. The debriefer will begin with the highest ranking questionnaire item. The debriefer will continue alternating between posttest and questionnaire items and decending to the next lowest ranks until we have either covered all the items or until we have about only 10 minutes remaining for the debriefing session. The last matter that you will be debriefed on will be your overall prioritized recommendations for the revision of the videocassette. Each posttest and questionnaire item has been made the subject of a computer conference item. The debriefer will lead you through the process, by telling you which conference item to view and discuss as the debriefing procedes and specifying the kind of feedback is needed (e.g. deficiencies in the videocassette, recommendations for correcting deficiencies, clarifications of individual comments made, requests for consensus).

6. What will happen prior to the tryout-debriefing session?

   a. Your name will be entered into the conference, so that your conference comments will be identifiable during the conference.

   b. If you can spare the time (10-15 minutes) you could practice viewing and entering comments into the conference (called discussion votes). Item 11 was entered into the conference as a model item to nelp you understand the process that will be used. Item 44 was entered into the conference as a practice item. It asks you to enter a comment about one of your courses.

   c. I would be happy to meet with you at your convenience for a brief tutorial session in order to facilitate your practicing of viewing and entering conference items. There are computer terminals in Erickson Hall, the MSU Library and the Computer Center, where we could meet. (See my phone number on the cover page of this Guide.) Also the computer consultants in the Educational Media lab have been given some training with regard to the procedures to be used in this computer conference, so you would check with the consultant on duty when you want to practice. The advantage to practicing prior to the tryout-debriefing session is that it would reduce the amount of frustration that you might encounter during the conference. The pilot study participants seemed to enjoy themselves after they learned how to make proper use of the computer terminal and the conference procedures. So, a few minutes of practice could make the debriefing session a much more enjoyable experience for you as well as helping you learn how computer conferencing is done.

7. <u>Procedure for signing-on the computer</u>
The procedure for signing-on the computer (located at the University of Michigan) and joining the conference is shown in the dialog found below. You will please note that all your responses are shown in the left hand column. Also, remember that the (RETURN) key, found on the right side of the keyboard, must be pressed after typing each response.

YOU	COMPUTER
1) Turn on the terminal's power (See on/off switch on left side of terminal, next to keyboard.)	
2) Turn on the power for the modem (a box-like unit in which you will place the telephone's handset).	
3) dial 3-3500 on the telephone.	
4) When you hear a high-pitched tone place the handset into the modem, with the wire-end of the telephone lined up with the wire-end of the modem.	
5) Press the (RETURN) key	(Junk) % Terminal
6) PRESS THE (RETURN) key	WHICH HOST?
7) UM (RETURN)	(Junk) #
8) SIGNON_______(RETURN) (leave space)	ENTER YOUR PASSWORD
9) WORLDNET (RETURN)	(Junk) #
10) SOURCE SP1U:CTR (RETURN) (That's the number 1)	(Junk) No new messages

8. Procedure for viewing and entering practice discussion votes prior to the tryout-debriefing session

In order to practice the process we will be using during the tryout-debriefing session, please use the procedure shown below.

YOU	COMPUTER
FIRST PRACTICE SESSION: (5 minutes)	
1) I 11 (RETURN)	(Displays item 11 - a model item) VOTE, FORGET, OR PASS:
2) P (RETURN)	DO NEXT?
3) I 44 (RETURN)	(Displays item 44 - a practice item) VOTE, FORGET, OR PASS:
4) V (RETURN)	GIVE YOUR DISCUSSION VOTE:
5) (Enter your comment. REMEMBER to press (RETURN) at the end of each line. For your last line press (RETURN) only; i.e. a blank line, which tells the computer you have completed your discussion vote.)	DO NEXT?
6) (See sign-on procedure, Section 10)	

YOU	COMPUTER
SECOND PRACTICE SESSION (5-10 minutes)	
1) V NEW 44 (RETURN)	(Displays the votes entered since your first practice session.) VOTE, FORGET, OR PASS: (N.B. If no discussion votes have been entered since your first practice session, it will display: NO ITEM SATISFIES THIS REQUEST. DO NEXT?
2) (a) To enter another discussion vote, type: V (RETURN)	(a) GIVE YOUR DISCUSSION VOTE:
(b) Otherwise, type: P (for PASS) (RETURN) (N.B. NEVER type FORGET or F, as this tells the computer to never displays the item to you again.)	(b) DO NEXT?
3) (a) (Enter your discussion vote.)	(a) DO NEXT?
(b) (See sign-off procedure, Section 10)	
4) (See sign-off procedure, Section 10)	

9. <u>Procedure for the beginning of the debriefing session</u>

The general procedure will be as follows:

- a. Have a short practice session:
    - (1) View item 11 (a model item).
    - (2) View item 44 (a practice item).
    - (3) Enter a discussion vote on item 44.
    - (4) View new votes on item 44.
- b. Begin the debriefing
    - (1) The debriefer will enter a discussion vote on item 44, informing the learners of the first item to view for the debriefing session.
    - (2) From this point on, you will be viewing and entering discussion votes on items, as directed by the debriefer.

The procedure that begins on the next page is an attempt to provide you with a primer for helping you to proceed through the early part of the debriefing session, including the practice session. If you find the details confusing you might want to check the general procedure, as outlined above, to orient you to the general process you will be participating in.

Detailed procedure for the beginning of the debriefing session

YOU	COMPUTER
1)	DO NEXT?
2) I 11 (RETURN)	(Displays item 11 - a model item) VOTE, FORGET, OR PASS:
3) P (RETURN)	DO NEXT?
4) I 44 (RETURN)	(Displays item 44 - a practice item) VOTE, FORGET, OR PASS:
5) V (RETURN)	GIVE YOUR DISCUSSION VOTE
6) (Enter a general comment about the videocassette you viewed during the tryout. REMEMBER to press (RETURN) at the end of each line. Then, enter a blank line by pressing (RETURN) only.)	DO NEXT?
7) V NEW 44 (RETURN)	(a) (Displays new votes you have not viewed yet.) VOTE, FORGET, OR PASS: (b) (If there are no new votes you have not

YOU	COMPUTER
8)(a)(i) If the debriefer has entered a discussion vote, asking you to type: P (RETURN) and to request a certain item in resposne to the next DO NEXT?, follow that procedure.	(a)(i) DO NEXT? (Displays item requested.) VOTE, FORGET, OR PASS:
(a)(ii) If the debriefer has NOT entered a discussion vote yet, type: P (RETURN)	(a)(ii) DO NEXT?
(b) V NEW 44 (RETURN) (in response to 7(b), where the computer displays message that there are no new votes)	(b)(i) (Displays new votes) VOTE, FORGET, OR PASS: OR (b) (ii) NO ITEM SATISFIES THIS REQUEST DO NEXT?
9)(a)(i) V (RETURN) if an item was displayed in 7(a)(i)	(a)(i) GIVE YOUR DISCUSSION VOTE
(a)(ii) V NEW 44 if DO NEXT? was displayed in 7(a)(ii)	(a)(ii)(a)(Displays new votes.) VOTE, FORGET, OR PASS OR (a)(ii)(b) NO ITEM SATISFIES THIS REQUEST DO NEXT?
(b)(i) P (RETURN) if new votes were displayed in 8(b)(i)	(b)(i) DO NEXT?
(b)(ii) Repeat step 8(a)(ii) if DO NEXT? was displayed in 7(b)(ii)	(see (a)(ii)(a) and (b)
10)(a)(i) (Give your comment) (response 9(a)(i))	(a)(i) DO NEXT?
(a)(ii)(a) P (RETURN) (response to 9(a)(ii)(a))	(a)(ii)(a) DO NEXT?
(a)(ii)(b) V NEW 44 (response to 8(a)(ii)(b))	(a)(ii)(b) (Displays votes) VOTE, FORGET, OR PASS:
(b)(i) By this time you should have received directions form the debriefer, which you should follow. (Response to 10(b)(i))	

11) (a)(i) V NEW (item specified by debriefer in 7(a)(i) response to 10(a)(i))	(a)(i) (Displays new votes) VOTE, FORGET, OR PASS:
(a)(ii)(a) By this time you should have received directions from the debriefer, as to which item to view. (Response to 9(a)(ii)(a))	
(a)(ii)(b) (RETURN) (response to 10(a)(ii)(b))	(a)(ii)(b) DO NEXT?
12) (a) V (RETURN) (reponse to 11(a)(i))	(a) GIVE YOUR DISCUSSION VOTE
(a) By this time you should have received directions from the debriefer, as to which item to view. (Response to 11 (a)(ii)(b))	
13) (Enter your comment) (response to 12(a))	

10. Procedure of signing-off the computer

To sign off the computer, follow the following procedure:

	YOU	COMPUTER
1)		DO NEXT?
2)	STOP (RETURN)	(Junk) #
3)	SIG $ (RETURN)	Displays account balance and cost of your computer session.

(If there are any problems in signing-off, be sure to turn off the terminal, which will automatically sign you off in a less than graceful manner.)

11. Special function keys on the Decwriter terminal

RETURN	Must be pressed at the end of every response and line of a vote.
CTRL H	Used to backspace one character at a time. First press CTRL and then press H as many times as needed.
CTRL X	Used to delete an entire line, prior to pressing (RETURN). Hold CTRL down and press X .
SHIFT	Used to type a capital letter. Hold down, while typing the letter.
LINE FEED	Used to advance the paper one line. Useful for separating output for different items.
	The spacebar, found at the bottom of the keyboard is used to insert a space, as you are typing.
ESC (SEL)	Used to interrupt the computer from giving you unwanted output. For example, if you requested the wrong item, you can stop the printing of the item by pressing this key. Unfortunately, the computer sometimes interprets this command to mean that you want to leave the conference. This will be the case if it prints a # at the extreme left of the paper. To return to the conference <u>type</u>: $RES.

____
(Student Number)

APPENDIX B

PRETEST

Please write your student number on the line shown in the upper left hand corner of this page, so that your pretest score may be matched with your posttest score. No other use will be made of your student number, and the results of the pretest and posttest will be used for research purposes only. The pretest should not take more than 5 minutes. As soon as you finish, please raise your hand and someone will collect your test paper.

For the following questions, plese check the most appropriate alternative.

1. The five basic functions in an information processing system are:

____a. storage, retrieval, progrmming, processing, and control
____b. retrieval, input/output, programming, processing, and control.
____c. storage, processing, input/output, control, and programming.
____d. storage, processing, input, retrieval, and control.

2. The computer can handle one more function than does the calculator; this function is

____a. memory.
____b. input/output.
____c. programming.
____d. control.

3. In executing a set of instructions in information processing systems, the sequence for carrying them out is determined by

____a. the control function.
____b. the processing function
____c. the programming function.
____c. the input/output function.

4. When a person presses keys on a calculator, he/she is carrying out

____a. a part of the input/output function.
____b. the control function.
____c. the programming function.
____d. the processing function.

5. Programming is

____a. a set of instructions designed to tell the machine what to do and the sequence for executing the instructions.
____b. the data that is entered into the machine.
____c. the combination of "a" and "b", above.
____d. none of the above.

6. The abacus is capable of

____a. storage only.
____b. storage and processing.
____c. processing only.
____d. storage, processing, and control.

7. Computing machines have been part of human culture

____a. only since the late 1950s.
____b. only since the late 1940s.
____c. only since the early 1950.
____d. since at least 3,000 B.C.

8. The processing function in an information processing system is

____a. the storage and retrieval of information.
____b. the retrieval of information.
____c. the input and output of information to and from the system's memory.
____d. the ability to alter the contents of a register in a prescribed way.

9. Output from the abacus is indicated by the position of the beads at the end of a mathematical operation. The statement is

____a. never true.
____b. uusually true.
____c. always true.
____d. sometimes true.

10. The function that the calculator can perform which the abacus cannot peform is

____a. input.
____b. input/output.
____c. retrieval.
____d. processing.

11. The control function in an information processing system refers to

____a. its ability to control the storage of data.
____b. its ability to control the alteration of the contents of its registers.
____c. its ability to control the sequence for executing the instructions.
____d. its ability to control the input and output of data.

Answer the following questions by checking whether they are "true" or "false."

12. When a person presses keys on a calculator, he/she is carrying out the processing function.

____ TRUE ____ FALSE

13. In executing a set of instructions in information processing systems, the sequence for carrying them out is determined by the programming function.

____ TRUE ____ FALSE

14. Programming is a combination of the instructions that are designed to tell the machine what to do and the sequence for executing the instructions, on the one hand, and the data that is entered into the machine.

____ TRUE ____ FALSE

15. The computer can handle one more function than does the calculator; this function is input/output.

____ TRUE ____ FALSE

16. The five basic fucntions in an information processing system are: storage, retrieval, programming, processing, and control.

____ TRUE ____ FALSE

17. Computing machines have been part of human culture only since the early 1950s.

____ TRUE ____ FALSE

18. The control function in an information processing system refers to its ability to control the alteration of the contents of its registers.

____ TRUE ____ FALSE

19. The function that the calculator can perform which the abacus cannot peform is processing.

____ TRUE ____ FALSE

20. Output from the abacus is indicated by the position of the beads at the end of a mathematical operation.

____ TRUE ____ FALSE

21. The abacus is capable of storage only.

____ TRUE ____ FALSE

22. The processing function in an information processing system is the input and output of information to and from the system's memory.

____ TRUE ____ FALSE

____

(Student Number)

APPENDIX C

Posttest

There are two parts to the posttest, both of which are attached. The posttest should not take you more than 10 minutes. As soon as you finish the test, please raise your hand and your test paper will be collected from you. That will mark the end of your participation in the experiment. Your participation has been appreciated.

PART A.

For the following questions, please check the most appropriate alternative.

1A. The five basic functions in an information processing system are:

___ a. storage, processing, input/output, control and programming.
___ b. storage, processing, input, retrieval, and control.
___ c. storage, retrieval, programming, processing and control.
___ d. retrieval, input/output, programming, processing, and control.

2A. The processing function in an information processing system is:

___ a. the input and output of informatin to and from the system's memory.
___ b. the ability to alter the contents of a register in a prescribed way.
___ c. the storage and retrieval of information.
___ d. the retrieval of information.

3A. Computing machines have been part of human culture:

___ a. since at least 3,000 B.C.
___ b. only since the early 1950s.
___ c. only since the late 1940s.
___ d. only since the late 1950s.

4A. Output from the abacus is indicated by the position of the beads at the end of a mathematical operation. This statement is:

___ a. sometimes true.
___ b. never true.
___ c. usually true.
___ d. always true.

5A. The computer can handle one more function than does the calculator; this function is:

___ a. control
___ b. programming
___ c. input/output.
___ d. memory.

For the following questions, answer "true" or "false".

6A. The abacus is capable of storage and processing.

___ TRUE ___ FALSE

7A. The control function in an information processing system refers to its ability to control the input and output of data.

___ TRUE ___ FALSE

8A. The function that the calculator can perform which the abacus cannot perform is input/output.

___ TRUE ___ FALSE

9A. When a person presses keys on a calculator, he/she is carrying out the programming function.

___ TRUE ___ FALSE

10A. Programming is a set of instructions designed to tell the macnine what to do and the sequence for executing the instructions.

___ TRUE ___ FALSE

11A. In executing a set of instructions in information proccessing systems, the sequence for carrying them out is determined by the control function.

___ TRUE ___ FALSE

PART B

For the following questions, please check the most appropriate alternative.

1B. the abacus is capable of
___ a. storage only.
___ b. processing only.
___ c. storage and processing.
___ d. storage, processing, and control.

2B. The control function in an information processing system refers to
___ a. its ability to control the storage of data.
___ b. its ability to control the input and output of data.
___ c. its ability to control the alteration of the contents of its registers.
___ d. its ability to control the sequence for executing the instructions.

3B. The function that the calculator can perform which the abacus cannot perform is
___ a. processing.
___ b. input.
___ c. input/output.
___ d. retrieval.

4B. When a person presses keys on a calculator, he/she is carrying out
___ a. part of the input/output function.
___ b. the processing function.
___ c. the control function.
___ d. the programming function.

5B. Programming is
___ a. the data that is entered into the machine.
___ b. a set of instructions designed to tell the machine what to do and the sequence for executing the instructions.
___ c. the combination of "a" and "b", above.
___ d. none of the above.

6B. In executing a set of instructions in information processing systems, the sequence for carrying them out is determined by
___ a. the processing function.
___ b. the control function.
___ c. the input/output function.

___ d. the programming function.

7B. The five basic functions in an information processing system are retrieval, input/output, programming, processing, and control.
___ TRUE ___ FALSE

8B. The processing function in an information processing system is the ability to alter the contents of a register in a prescribed way.
___ TRUE ___ FALSE

9B. Computing machines have been part of human culture only since the late 1940s.
___ TRUE ___ FALSE

10B. Output from the abacus is indicated by the positon of the beads at the end of a mathematical operation.
___ TRUE ___ FALSE

11B. The computer can handle one more functions than does the calculator; this fucntion is programming.
___ TRUE ___ FALSE

APPENDIX D

LEARNER QUESTIONNAIRE

____
(Student number)

Hi

I'd really appreciate it if you could spend some time and effort in answering the attached questionnaire. There are two main reasons why we are interested in your honest opinion about the instructional materials you have just used. These materials will be revised on the basis of the feedback you give us on what is wrong with them and how they should be improved. Furthrmore, since these instructional materials are for a newly developed course in the College of Education at MSU, the kind of feedback you provide might be applied, in general, for the revision of the other instructional materials used in the course.

Second, your opinions will be part of the data I am collecting for my Ph.D. dissertation. With your help, I will be able to obtain my degree and then help teachers and administrators improve their courses and curricula, so that students will learn more and learn more easily.

We would also appreciate getting your reaction to this questionnaire. Did we ask the right questions, were the questions confusing, poorly worded, not specific enough, etc. Can you think of any other questions that should be included? Please feel free to write your comments about these things on the questionnaire.

THANKS A LOT FOR YOUR HELP!!!!!!!

Paul Toner
Doctoral Student
Educational Systems Development
College of Education
Michigan State University

LEARNER QUESTIONNAIRE

For the statements listed below, please check the most appropriate reaction shown to the right of the statement. The following legend gives the word or phrase represented by the initials used for these reactions:

SA - STRONGLY AGREE
A - AGREE
N - NEUTRAL
D - DISAGREE
SD - STRONGLY DISAGREE

1. The questions on the posttest covered what I consider to be the most important information in the presentation. ( ) SA ( ) A ( ) N ( ) D ( ) SD

2. I found the presentation interesting. ( ) SA ( ) A ( ) N ( ) D ( ) SD

3. The order in which the information was presented helped make the presentation easy for me to follow. ( ) SA ( ) A ( ) N ( ) D ( ) SD

4. The presentation's content was clear to me. ( ) SA ( ) A ( ) N ( ) D ( ) SD

5. The equipment demonstrations used to illustrate the subject matter helped me learn. ( ) SA ( ) A ( ) N ( ) D ( ) SD

6. When the presentation began, I already had the necessary background knowledge to learn from the presentation effectively. ( ) SA ( ) A ( ) N ( ) D ( ) SD

7. The visual displays (graphics) used in the presentation helped me learn. ( ) SA ( ) A ( ) N ( ) D ( ) SD

8. The video quality of the presentation helped me learn. ( ) SA ( ) A ( ) N ( ) D ( ) SD

9. The audio quality of the presentation helped me learn. ( ) SA ( ) A ( ) N ( ) D ( ) SD

10. The quality of the oral delivery of the presenter(s) helped me learn.

( )	( )	( )	( )	( )
SA	A	N	D	SD

APPENDIX E

COMPARISON OF THE OUTCOME AND DEBRIEFING DATA FOR THE TWO MODELS OF TRYOUT-REVISION

In the case of the small group model, the posttest and questionnaire items for which the number of incorrect responses reached the 30% criterion level are shown in Table 3.2, which follows. The criterion level for 7 learners equals 3 responses, rounding to the next highest integer (7 learners X .3 2.1=3). The table also shows the number of such responses for each of these items as well as the items placed on the debriefing agenda.

Table 3.1

Posttest and Questionnaire Items with Responses Reaching the 30% Criterion Level in the Small Group Tryout.

Item Numbers	No. of Incorrect or Unfavorable Responses
Posttest items:	
2A*	5
5A	3
7A	5
9A	4
11A	3
2B	5
4B*	6
5B	5
6B	6
11B	3
Questionnaire Items:	
2	4
3	4
4	5
5	4
6	4
7	5
8*	6
9	5
10*	7

* = item placed on debriefing agenda

As indicated by the asterisks in Table 3.2, the items discussed during the debriefing were posttest items 2A and 4B and questionnaire items 8 and 10. The reader may recall that it was left to the debriefers to decide which items to place on the debriefing agenda in the case of ties. Therefore, as shown in Table 3.2, posttest items 7A or 2B could have been selected for the debriefing agenda.

Table 3.3, below, shows this same type data for the computer conferencing model. The criterion level for 6 learners equals 2 responses (6 learners x .3 = 1.8 =2).

As can be see from Tables 3.2 and 3.3, the outcome data for the two tryouts are fairly consistant. In each table, only one posttest items is listed which is not also listed in the other table; namely, item 4B in Table 3.2, and item 8B in Table 3.3. However, for the small group tryout, that item is one of the items for which the responses reach the criterion level. For the questionnaire data, all the items on Table 3.3 (computer conferencing tryout) are listed on Table 3.2. But, two of the items on Table 3.2 (small group tryout) do not appear on Table 3.3; namely, items 3 and 4.

The biggest differences between the two tryout groups are the items placed on the debriefing agenda. Questionnaire items 10 is the only item placed on the agenda for both groups. This is primarily due to the effect of ties among the items plus an error in agenda setting on the part of the small group debriefer.

Table 3.4, below, shows the amount of discussion for the items debriefed in the computer conferencing group during the second debriefing session.

Table 3.2

Posttest and Questionnaire Items with Responses Reaching the 30% Criterion Level in the Computer Conferencing Tryout.

Item Numbers	No. of Incorrect or Unfavorable Responses
Posttest items:	
2A	3
5A	2
7A	2
9A	2
11A	3
2B	3
5B*	3
6B*	6
8B	2
11B	2
Questionnaire Items:	
2	3
5	4
6	4
7	4
8	4
9*	4
10*	4

* = item placed on debriefing agenda

Table 3.3

Statistics on substantive Discussion Voting in the Second Computer Conferencing Debriefing

Item No.	No. of Votes	No. of Voters	Average No. of Votes
Posttest Items:			
5B	10	6	1.7
6B	22	6	3.7
Questionnaire Items:			
9	7	5	1.4
10	20	6	3.3
Average Item	14.8	5.8	2.5
General Recommendations	7	6	1.2

Although Table 3.4 shows considerable improvement in the average numer of substantive discussion votes (14.8 versus 11) per item between the first and second computer conferencing debriefings, the average number of votes per active participant per item shows an improvement of only .3 votes (2.5 versus 2.2). (See the previous section, "Brief Analysis and Review of the Debriefing Data") The biggest improvement was in the overall tone of the voting in the second debriefing.

However, in regard to the non-substantive votes, it should be borne in mind that they were at the computer terminals for 1 hour and 40 minutes, including the time during which the connection with the computer had been lost. Such a time period is a long session at a computer terminal, even for seasoned computer conferencers. None of the learners had had much previous computer experience except for one learner who used word processing equipment on a regular basis.

No similar attempt to analyze the small group debriefing session in terms of frequency of comments was made. The only bearing such statistics has on this study is its relationship to the improvement of the model through its development during the life of this study through the elimination of the various problems that have been alluded to. No research interest in the relative differences in debriefing group behavior across models of tryout-revision is being pursued in this study, as interesting as those dimensions of debriefing may be.

APPENDIX F

PROCEEDINGS OF COMPUTER CONFERENCE

This appendix consists of a transcription of the computer conference carried out during this study. The discussion votes of the conference participants were transcribed exactly as typed during the conference.

Item 28 21:09 Apr 19/81

LOU ANN

POSTTEST ITEM 6B

In executing a set of instructions in information processing systems, the sequence for carrying them out is determined by

a. the processing function.

b. the control function.

c. the input/out function.

d. the programming function.

("b" is correct)

Why do you think some of the learners answered this question incorrectly? To give your comments (votes):

1. TYPE: v after computer displays "VOTE, FORGET, OR PASS:"

2. Type your comments.

3. Press (RETURN) key only for your last line.

4. TYPE: v new 28 to see new comments, after computer responds "DO NEXT?", and be sure to consider voting again.

5. Repeat step 4 until you see Lou Ann's comment telling you to proceed.

37 Discussion votes

Betty Walker: what is thhe control fucntion and who performs the same? This question was not answered during the explanation.

Chuck : not enough was explained about it v new 28

Patricia Reynolds: It would be helpful if a handout were also available for reference along with the video tape. (The speakers words were also slurred or else the quality of the tape was poor.

Linda: The videotape gave me the impression that the programming function gave the computer instructions for carrying items out in sequence.

Chuck: i agree with linda

LOU ANN: do you all agree that this content was not treated adequately in the tape?

Susan: I agree with betty coment. This question was not answered to my satisfaction in the presentation.

Betty: The presentation did not meet the expected mgoal of giving infor mation about the computer. It gave information about the abacus and the calculator which wer much more detailed andunderstandanble.

Carmen: sequence might suggest programming

Susan: I have no additional comment

Linda: I'm afraid I was distracted by the speaker's grammar on the videotape At one point he said, "has began." I found my self listening to his grammer instead of to what he was saying.

Patricia: I would like to see the tape again.

LOU ANN: do any of you agree with Patricia that a handout to use along with the tape would hve been helpful

Linda: I think I would have found a handout helpful.

Susan: a handout would have been asadditional help Linda sounds likean english teacher. is she?

Chuck: yes

Betty: Yes, I agree that a handout would have been helpful. Also, many other learning and teaching techniques should be used to make the presentation acceptable for the purpose intentedxxx and tested.

Patricia: I would also like to make a comment about the general quality of the presentation. It was difficult to see the visuals that were used such the writing on the blackboard which gave the five functions, i.e., programming, input/output etc. The the five functions, i.e., programming, input/output etc. The examples for use of the abacus were also hard to see. I think I would have retained more information if the words which were printed on the screen were ligible. Did anyone else have difficulty dicerphering?

Linda: I think I'd be a grammar fanatic even if I were't an English teacher.

Carmen: yes

Chuck: why cant we go home

Susan: I had trouble following the material because I have little experience with this field. I need to move through things more slowly

LOU ANN: to summarize, it sounds as if yo all agree that this content was not treated adequately in the tape, and perhaps the presenter may have distracted you from getting what little was explained. do you agree? please vote "yes" or "no"

Chuck: yes

Betty: What is the goal of this experiment? It seems lide a waste of time and money.

Patricia: wwwe cam(t gp hpome yet -- I just figured this ti thing out!

Susan: yes

Linda: I agree with Pat on the quality of the visuals. Also, too much time was spent on the example of the abacus compared to the amount of time spent on the computer example. I found the computer example almost im possible to follow, even if the picture would have been clearer.

Chuck: you stay and i will go

Carmen: I found the whole presrntation quite interesting actually because I know very little about computers

Linda: I don't want to leave now. I'm just starting to get the hang of this. It's kind of fun, sort of like talking to a computer.

Chuck: who is buying

LOU ANN: as soon as you vote on my last question (answering "yes" or "no" please go on to view item 53. type "P" (return) and then type "Item 53" in response to "do next?"

Linda: I though Chuch was buying.

Betty: Yoe most certainly eight Lou Ann. The content was poorly treated. xxxxxxxxxxxxxxxxxxxxx There is veery little possibility of saving this presentation. Pleases start the wohole thing over.

Patricia: yes

Carmen: what is everybody else saying

VOTE, FORGET, OR PASS:

Item 53   20:27 May11/81   13 lines

LOU ANN

QUESTIONNAIRE ITEM 10

Item: The quality of the oral delivery of the presenter(s) helped me learn.

Why do you think some of the learners did not rate this questionnaire item positively? To give your comments (votes):

1) TYPE: v   after the computer displays "VOTE, FORGET, OR PASS:"
2) Type your comments.
3) PRESS: (RETURN) key only for your last line.
4) TYPE: v new 53 to see new comments, after computer displays "DO NEXT?", and be sure to consider voting again.
5) Repeat step 4 until you see Lou Ann"s comment telling you to proceed.

42 Discussion votes

Chuck: stuttered too much

Susan: The quality of the oral presenter could not possibly help one to learn. If anything, it hindered me form learning

Betty: Ha ! Pat and Linda have alreadyxx discussed this probm

Linda: His hesitations made him sound uncertain of the content or maybe he was just ne nervous. Plus, as I said earlier, his grammar bothered me.

Chuck: linda is buying

Susan: The quality of the machine was also poor

LOU ANN: i think most of you have already said that the quality of the oral presenter was not good. If you don't agree please say so. If you do agree, please give us some suggestions for improving his delivery. Paul is going to redo this tape, and he needs your suggestions.

Carmen: He ws obviously reading and sometimes he faltered

Betty: I: m so-o-osleepy and this impairs my ability to think and respond. Watch out for the next vote.

Chuck: the tape kept skipping

Susan: Perhaps you could fine someoneto present the presentationwowho was able to sound more prepared

Betty: He should walk tall and carry a big stick and maybe we would pay more attention

LOU ANN: chuck, are you saying that the audio quality on the tape wasa problem too?

Chuck: try not read the presentation

Patricia: It seemed a shame to take a tape which will show moving pictures and use it to show static photos. A tape recording and a picture book would have served as well. Snyoneagree? Anyone agree?

Linda: I'm finding it hard to keep up with this as though it were a regular conversation. My comments don't necessarily end up directed toward the persom to whom I was talking. There!! Does that sound like an English teacher?

Susan: It would help if the visuals were more clearand visible

LOU ANN: Pat, that's a really interesting comment. What do the rest of you have to say about that? Was this really appropriate content for a videotape?

Chuck: audio was a problem because of reading

Carmen: I do not wish to continue this discussion

Linda: I really think that more of a computer background would have helped me in understanding the whole point of the presentation.

Susan: This should not have been the only teaching techniquee used. However I don't feel that it should not be used on a videotape at all.

Betty: NO, it wasa presentation that should hved beenmadewith theassis ance of an on-site visit.

LOU ANN: Please vote "yes" or "no". would a more professional presenter have made it easier for you to learn from this tape?

Susan: yes

Patricia: yes

Betty: YES !!!!!!!!!!!!!!!!!

Linda: YES!!

Chuck: no i will be different

Betty: Chuck, are you sleepy too?

Susan: Linda I hope I didn't offend you by pointing out that you myst be an eenglish teacher. We all need better grammar.

Chuck: i never done did had no english

LOU ANN: after you finish votin on this item please type "Item 27" in response to "do Next?" We'll resume this discussion in the votes on Item 27.

Patricia: Lou ann, are there more questions you would like to ask about the tape? I would like to know why we are using this for (this machine, I mean) for discussion since it really would have been more profitable to Paul if we gave him feedback on his tape face to face instead of struggling with this computer -- is there a hidden agenda?

Susan: I know you have to have a thick crust to teach

Linda: No problem, Sue. I love being an English teacher. / AND it's really hard to offend me.

Carmen: would someone tell me where we go from here

Betty: Linda-- would you pleas provide us with some instruction in the English languageas soon as possible? This computerxxxxxxxx sure isless useful than the English lessons.Let! get with it.

Susan: Carmen , If I knew where we went from her, I would go there.

Linda: I agree with Pat. This seems like an awful waste of money.

Carmen: I think I will write a letter to a friend

Chuck: its 9:56

VOTE, FORGET, OR PASS:

Item 27   21:05 Apr19/81   16 lines

LOU ANN

POSTTEST ITEM 5B

Programming is

a. the data that is entered into the machine

b. a set of instructions designed to tell the machine what to do and the sequence for executing each instruction.

c. the combination of "a" and "b", above.

d. none of the above.

("b" is correct)

Why do you think some of the learners answered this question incorrectly? To give your comments (votes):

1) TYPE: v   after computer displays "VOTE, FORGET, OR PASS:

2) Type your comments.

3) Press (RETURN) key only for your last line.

4) TYPE: v new 27 to see new comments, after computer responds "DO NEXT?", and be sure to consider voting again.

5) Repeat step 4 until you see Lou Ann's comment telling you to proceed.

16 Discussion votes

LOU ANN: do any of you have any different comments for this item? Or was the problem here the same as for item 28?

Chuck: not enough time spent on the computer functions and i am getting tired of this

Linda: I didn't have any problem answering the question on programming. I thought the videotape made it quite clear.

Chuck: same as 28

LOU ANN: chuck, while we are waiting for the others to respond , why don't you think about what three things would be most important to improve on this tape? Okay?

Betty: A listing of programming type instructions wasa not shown or very clearlyoutlined in the presentation.

LOU ANN: who agrees with linda?

Patricia: I have a comment. (of course!) I have found that learning is largely a matter of semantics. If I understand the vocabulary, I can understand the concept. If there are too many words in one sentence of paragraph with which I am unfamiliar, it takes me longer to absorb the information. On a video tape there is no going backwards to catch what I missed or to clarify a point.

Susan: I don't

Patricia: I can't really remember anymorey how good the tape was. This experience on the computer (has overshadowed my normal ability to criticise effectively.

Linda: Why is Chuch always so far ahead of the rest of us? I thought he was the one who couldn't type?

Chuck: im still thinking

LOU ANN: after you finish this vote, please go on to item 42. type "Item 42" in response to "do Next?"

Carmen: I agree with Patricia If too much is unfamiliar there is nothing to hang the new knowledge onto

Susan: good point Pat. The hour also adds to the fog

Betty: Think! Think! The speaker should hve selected media that was related to the goal of explaining computer operations. We are now confused about content, purpose and reason.

VOTE, FORGET, OR PASS: p

Item 42   23:16 Apr19/81   11 lines

LOU ANN

QUESTIONNAIRE ITEM 9

Item: The audio quality of the presentation helped me me learn.

Why do you think some of the learners did not rate this questionnaire item positively? To give your comments (votes):

1) TYPE: v   after the computer displays 'VOTE, FORGET, OR PASS:"
2) Type your comments.
3) PRESS: (RETURN) key only for your last line.
4) TYPE: v new 42 to see new comments, after computer responds "DO NEXT?", and be sure to consider voting again.
5) Repeat step 5 until you see Lou Ann's comment telling you to proceed.

32 Discussion votes

LOU ANN: if you think the audio quality should have been improved, please vote "yes"

Chuck: monotone speakers are boringzzzzzzzzzzzzzzzzz

LOU ANN: we're almost finished; please stay with us a little longer.

Carmen: yes

Patricia: Yes the audio quality should have been improved. There was one point where the tape skipped erasing a lot of words (or a few words - there's no way of telling).

Susan: yes

Chuck: yes

Chuck: i am sinking fast

LOU ANN: aside from the speaker himself, should the sound quality be improved?

Betty: YES !!!!!!!!!! AND THAT:S NOT ALL.

Linda: What happened to my last disscussion vote? It just disappeared.

Chuck: no

Susan: Linda, I had that same problem. Betty, I'm interested in knowing how you would complete your last statement

Patricia: Sure the sound quality could be improved -- and yes the presentor could have been more professional but I really felt that the entire presentation could have had more pazzzzzzzz. After all, a script obviously was written for this.

Carmen: Lou Ann may we have a new question?

Chuck: yes

LOU ANN: okay, this is the last discussion item. What 3 things would you re commend as being priority items for improvement in this tape? (chuch, i told you to be thinking.)

Chuck: new speaker more time spent on computers and use different types of models

Betty: Ixxxxxxxxxxxx I smell food and it is thundering outside. Improvementsto the audio quality would hve been made through a more attractive and confident speaker.

Susan: 1. clearer visuals 2. a more professional speaker 3. more review of material.

LOU ANN: carmen, the new question is: what 3 things would be most important to change in this tape? How would you change them?

Patricia: supplementary handouts, more interesting copy (not the topic), and better announciation (how doe you spell that?)

Chuck: annunciation

LOU ANN: are you all still there? We appreciate the help you have given us tonight - and the patience!

Carmen: This machine is malfunctioning i t does not print all the questions or comments

Betty: He also neede additional visual aids, more command of the English language and some guidance form computer oriented personnel.

Linda: First, I think the speaker should be rehearsed or a more professional speaker used. Second, the video quality seemed a little fuzzy. Finally, I think that iin the presentation itself, more explanation is necessary on the computer example The point being made about the abacus seemed labored. I think the point was made and then some.

LOU ANN: look at the three items the other paticipants have mentioned. if after reading their responses dyou want to change any of yours, please do by entering another vote. If you still have the same opinion, vote on anything you want and sign off. Thanks again for all your help!!!!!!!!!!!!!!!!!!!!!

Chuck: anything

Betty: YOU:RE MOST WELCOME) I REALLY ENJOYED PARTICIPATING.

Linda: This was kind of fun, but too long a session. Thanks for letting me get the experience on a computer,though.

Carmen: Firstly a more accomplished narrator Secondly bettervisibility of details like th e writing Thirdly cannot think of anything else

VOTE, FORGET, OR PASS

APPENDIX G

Viewing Guide for LIM 2 Television Program: "History of Computing Machines and Basic Concepts"

Written by:
Paul D. Toner
July 5, 1981

Table of Contents

A. Advance Organizer

The following is an outline of the ideas that will be presented in this lesson.

1. System
2. Subsystem
3. Information processing system
4. Approaches to examining systems
    a. Functional approach
    b. Historical approach
5. Functions of information processing systems
    a. Storage
    b. Processing
    c. Programming
    d. Control
    e. Input/Output
6. Information processing systems to be examined
    a. The abacus
    b. The calculator
    c. The digital computer

B. Definitions

1. System

"A system is the collection of integrated entities which have arbitrarily been designed as of central interest, such as a school. (Thomas Harries, 1971)

2. Subsystem

A subsystem is a smaller collection of entities which comprise a portion of the system of central interest; i.e., teachers, students, physical facilities, etc. (Thomas Harries, 1971)

3. Information Processing System

An information processing system is a collection of devices and people designed to process symbols.

4. Storage

Storage is the ability to keep information, either temporarily or permanently, in a location from which it can be retrieved.

5. Processing

Processing is the ability to alter the contents of registers.

C. Analysis of Scene Showing Man Completing Tax Form

The purpose of this scene is to suggest to you that information processing systems can be thought of as an extension of the way humans handle information without the benefit of machines.

Storage function: File folders, check file, and check registers.

Programming function: Tax form and instruction booklet.

Processing function: Making calculations on scratchpad.

Control function: Entering information on tax form according to sequence specified by the form and instruction booklet.

Input/Output:

Input: Entering information on tax form with pen.

Output: Completed tax form.

D. Review Questions

After you have stopped the tape temporarily, when asked to do so by the television program presenter, please complete the statements shown below. All the statements refer to information processing system functions.

1. Processing is ____

____

2. Programming is ____

____

3. Control is ____

____

4. Storage is ____

____

5. Input/Output is ____

____

Input is ____

____

Output is ____

____

PLEASE TURN TO NEXT PAGE.

Answers to Questions on Previous Page

1. Processing is the ability to alter the contents of registers.
2. Programming is the ability to specify the steps to be performed and the sequence for performing them.
3. Control is the ability to determine which instruction to execute next.
4. Storage is the ability to keep information, either temporarily or permanently, in the location from which it can be retrieved.
5. Input/Output is the exchange of information between the machine and its user. Input is the information fed to the machine by the user; Output is the informatin fed or displayed to the user by the machine.

E. Classification Table Showing Each Function According to Whether its Carried Out by Humans or tne Machine, For Each Information Processing System.

	SYSTEM		
FUNCTION	Abacus	Calculator	Digital Computer
Storage	M	M	M
Processing	H	M	M
Programming	H	H	H
Control	H	H	M
Input/Output	H	H	H/M

Legend:

H = Function carried out by humans

M = Function carried out by the machine

N.B. In the case of the digital computer, the human carries out input and the machine carries out output.

APPENDIX H

Posttest Items Arranged by Instructional Objectives

	Instructional Objectives	Items
1.	To identify the names of the basic functions of an information processing system.	1A, 7B
2.	To identify each of the functions of an information processing system.	2A, 7A, 10A, 11A 2B, 5B, 6B, 8B
3.	To identify what takes place when a person presses keys on a calculator.	9A, 4B
4.	To identify the information processing functions the machine carries out in information processing systems.	5A, 6A, 8A 1B, 3B, 11B
5.	To identify the period of time during which computing machines have been part of human culture.	9B
6.	To identify how output is indicated on an abacus.	4A, 10B

BIBLIOGRAPHY

✓ Abedor, Allan J. "Development and Validation of a Model Explicating the Formative Evaluation Process for Multi-Media Self-Instructional Learning Systems." Ph.D. dissertation, Michigan State University, 1971.

✓ Abedor, Allan J. Personal communication. April, 1980.

✓ Abedor, Allan J. "Second Draft Technology; Development and Field Test of a Model for Formative Evaluation of Self-Instructional Multi-Media Learning Systems." Viewpoints 48 (July 1972): 45-78.

Alkin, Marvin C. "Evaluation Theory Development." In Evaluating Action Programs, pp. 105-117. Edited by Carol H. Weiss. Boston: Allyn and Bacon, 1972.

✓ Anderson, Richard C. "The Comparative Field Experiment: An Illustration from High School Biology." Proceedings of the 1967 Invitational Conference on Testing Problems, New York, pp. 3-27. Princeton, N.J.: Educational Testing Service, 1968.

Baker, Eva L. "Formative Evaluation of Instruction." In Evaluation in Education; Current Applications. Edited by W. James Popham. Berkeley, CA: McCutchan Publishing Corp., 1974.

Baker, Eva L. "Generalization of Rules for Empirical Revision." AV Communication Review 18 (Fall 1970): 300-305.

Baker, Eva L. and Alkin, Marvin C. "Formative Evaluation of Instructional Development." AV Communications Review 21 (Winter 1973): 389-418.

Berk, Ronald A. "Introduction." In Educational Evaluation Methodology: The State of the Art, pp. 1-10. Edited by Ronald A. Berk. Baltimore: John Hopkins University Press, 1981.

✓ Briggs, Leslie J. and Wager, Walter W. "Formative Evaluation" (Chapter 11) Handbook of Procedures for the Design of Instruction. 2nd Ed. Englewood Cliffs, N.J.: Educational Technology Publications, 1981.

Bruner, Jerome S. Toward A Theory of Instruction New York: Norton, 1966.

Campbell, Donald T. and Stanley, Julian C. Experimental and Quasi-Experimental Designs for Research. Chicago: Rand McNally, 1963.

✓ Cronbach, Lee J. "Course Improvement Through Evaluation." Teachers College Evaluation 64 (May 1963): 672-683.

✓ Dick, Walter. "Formative Evaluation." In Instructional Design: Principles and Applications, pp. 311-333. Edited by Leslie J. Briggs. Englewood Cliffs, N.J.: Educational Technology Publications, 1977.

✓ Dick, Walter. "Formative Evaluation in Instructional Development." Journal of Instructional Development 3 (Spring 1980): 3-6.

✓ Fleming, Malcomb. "What is a Good Film?" Journal of the University Film Productions Association, 15 (1963) 8-9, 18.

Gropper, George L. Diagnosis and Revision in the Development of Instructional Materials. Englewood Cliffs, N.J.: Educational Technology Publications, 1975.

✓ Gropper, George L. "Does 'Programmed' Television Need Active Responding?" AV Communications Review 15 (Spring 1967): 5-14.

✓ Gropper, George L. and Lumsdaine, Arthur A. "The Use of Student Response to Improved Television Instruction: An Overview" (Report No. 7). In Studies in Televised Instruction. Pittsburgh: American Institutes for Research, 1961.

Hiltz, Starr Roxanne, et al. The Effects of Human Leadership and Decision Support Feedback on Group Problem Solving in Computerized Conferences: A Controlled Experiment. Newark, N.J.: Computerized Conferencing & Communications Center, New Jersey Institute of Technology, 1981.

Hiltz, Starr Roxanne, et al. Face-to-Face vs. Computerized Conferences: A Controlled Experiment. Volume I: Findings. Newark, N.J.: Computerized Conferencing and Communications Center, New Jersey Institute of Technology, 1980.

Hiltz, Starr Roxanne et al. Replicating Bales Problem Solving Experiments on a Computerized Conference: A Pilot Study (Research Report No. 8. Newark, N.J.: Computerized Conferencing & Communications Center, New Jersey Institute of Technology, 1978.

Hiltz, Starr Roxanne and Turoff, Murray. The Network Nation. Reading, MA: Addison-Wesley, 1978.

Jorgensen, Erling S. Personal communication. March, 1982.

✓ Kandaswamy, Subramaniam, et al. "Learner Verification and Revision: An Experimental Comparison of Two Methods." AV Communications Review

24 (Fall 1976): 316-328.

Light, Judy A. and Reynolds, Larry J. "Debugging Product and Testing Errors." Viewpoints 48 (July 1972): 45-78.

Linstone, Harold and Turoff, Murray. The Delphi Method: Techniques and Applications. Reading, MA: Addison-Wesley, 1975.

Markle, David G. The Development of the Bell System First Aid and Personal Safety Course: An Exercise in the Application of Empirical Methods to Instructional Systems Design. Final Report. Palo Alto, CA: American Institutes for Research in Behavioral Sciences, 1967; Bethesda, MD: ERIC Document Reproduction Service, ED 026 871.

✓ Markle, Susan M. "Empirical Testing of Programs" (Chapter V). In Programmed 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.

Merrill, M. David. Review of Diagnosis and Revision in the Development of Instructional Materials, by George L. Gropper. AV Communications Review 23 (Winter 1975): 446-449.

Merton, Robert K. Social Theory and Social Structure. 1968 enlarged ed. Chapters IV and V. New York: The Free Press, 1968.

✓ Nathenson, Michael B. and Henderson, Euan S. Using Student Feedback to Improve Learning Materials. London: Croon helm Ltd., 1980.

Palme, Jacob. "Computer Conferencing Is More than Electronic Mail." Paper presented at the EDUCOM Annual 1982 Conference, Baltimore, November, 1982.

Palmer, Edward L. "Formative Research in the Production of Television for Children" (Chapter XII), pp. 303-329. In Media and Symbols: The Forms of Expression, Communication, and Education. (The Seventy-third Yearbook of the National Society for the Study of Education. Part I. Edited by David R. Olson. Chicago, National Society for the Study of Education, 1974.

Parnes, Robert "Learning to CONFER: The Interplay of Theory and Practice in Computer Conferencing." Ph.D. Dissertation, University of Michigan, 1981.

Parnes, Robert "Learning to CONFER: The Interplay of Theory and Practice in Computer Conferencing." Dissertation Abstracts International 42 (August 1981): 672A.

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

Education, 1969.

Reigeluth, Charles M., et al. "Science Approach to Instructional Development." Performance and Instruction Journal 20 (September 1981): 19-22.

Reigeluth, Charles M., et al. "What Is the Design Science of Instruction?" Journal of Instructional Development 1 (Spring 1978): 11-16.

Robeck, M.J. "A Study of the Revision Process in Programmed Instruction." M.A. thesis, University of California, Los Angeles, 1965.

Robinson, Tobie Joyce. "Replicable Training in Revision Techniques." Dissertation Abstracts International 33 (September 1972): 1573-A.

Robinson, Tobie Joyce. "Replicable Training in Revision Techniques." Ph.D. dissertation, University of California, Los Angeles, 1972.

Rosen, Marvin J. An Experimental Design for Comparing the Effects of Instructional Media Programming Procedures: Subjective vs. Objective Revision Procedures. Final Report. Palo Alto, CA: American Institutes for Research in Behavioral Sciences, 1968; Bethesda, MD: ERIC Document Reproduction Service, ED 025 056.

Salomon, Gavriel. "What Is Learned and How it Is Taught: The Interaction between Media, Message, Task, and Learner" (Chapter XV)., pp. 383-406. In Media and Symbols: The Forms of Expression, Communication, and Education. (The Seventy-third Yearbook of the National Society for the Study of Education. Part I. Edited by David R. Olson. Chicago, National Soceity for the Study of Education, 1974.

Scott, Dow and Deadrick, Diana. "The Nominal Group Technique: Applications for Training Needs Assessment." Training and Development Journal 36 (June 1982): 26-33.

Scriven, Michael. "The Methodology of Evaluation." In Perspectives of Curriculum Evaluation (AERA Monograph Series on Curriculum Evaluation, No. 1), pp. 39-83. Edited by Ralph W. Tyler. Chicago: Rand McNally, 1967.

✓ Short, J.G. et al. Strategies of Training Development (Report No. AIR-E97-2/68). Palo Alto, CA: American Institutes for Research, 1968.

✓ Silberman, H., et al. Use of Exploratory Research and Individual Tutoring Techniques for the Development of Programming Methods and Theory (TM 895/200/00). Santa Monica, CA: System Development Corporation, 1964.

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

Suchman, Edward A. Evaluative Research. New York: Russell Sage Foundation, 1967.

Suchman, Edward A. "Principles and Practice of Evaluative Research." In An Introduction to Social Research, 2nd ed., pp. 327-351. Edited by John T. Doby. New York: Appleton-Century-Crofts, 1967.

Sulzen, Robert H. "The Effects of Empirical Program Revision and the Presentation of Objectives on Student Performance." Paper presented at the American Educational Research Association Annual Meeting, New Orleans, February 27, 1973; Bethesda, MD: ERIC Document Reproduction Service, ED 010 851.

Thiagarajan, Sivasilam. "Instructional Product Verification and Revision: 20 Questions and 200 Speculations." ECTJ 26 (Summer 1978): 133-141.

Turoff, Murray and Hiltz, Starr Roxanne. "Exploring the Future of Human Communication Via Computer." Technology and Society 1 (Fall 1981): 5-10.

VanderMeer, Arthur W. An Investigation of the Improvement of Informational Filmstrips and the Derivation of Principles Relating to the Effectiveness of these Media. Study Number I. University Park, PA: College of Education, the Pennsylvania State University, 1964.

VanderMeer, Arthur W. An Investigation of the Improvement of Informational Filmstrips and the Derivation of Principles Related to the Improvement of these Media: Study II. University Park, PA: College of Education, The Pennsylvania State University, 1964.

VanderMeer, Arthur W. and Montgomery, Robert E. An Investigation of the Improvement of Informational Filmstrips and the Derivation of Principles Related to the Improvement of these Media: Study Number III. University Park, PA: College of Education, The Pennsylvania State University, 1964.

Weiss, Carol H. "Utilization of Evaluation: Toward Comparative Study." In Evaluating Action Programs, pp. 318-326. Edited by Carol H. Weiss. Boston: Allyn and Bacon, 1972.