TRANSLATING EVALUATION CONCEPTS USED BY MULTIMEDIA DEVELOPERS THROUGH THEIR EVALUATION APPROACHES FOR INSTRUCTIONAL CD-ROMs

---

A Dissertation
Presented to
The Faculty of the Curry School of Education
University of Virginia

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

---

by

Mark Daniel Bardini, B.A., M.B.A.

August, 1997

Abstract

The purpose of this study was to describe the evaluation concepts used by six multimedia developers for instructional CD-ROMs. This was achieved through interviews with the six developers using a naturalistic, case study methodology. Transcripts of the interviews were content analyzed and presented in individual case study format. The discussion section comprises a cross-case analysis of all six conversations.

Two major categories emerged from the content analysis: types of evaluation performed and overall views and uses of evaluations. The first primary category includes two sub-categories: formative and summative evaluations, each of which detailed various evaluation methods used by the developers during and after the development process. The second primary category described the developers' views of evaluation and how they believed that evaluations were helpful for improving their products.

The interviews revealed that developers are performing evaluations of CD-ROMs, but the evaluation language that they use is frequently different than the language used by program evaluators. The responses from the developers communicated their view of evaluation as an integral component of the development process and their belief that evaluation is critical for product improvement.

The evaluation approaches used by the six developers offer practical methods by which to evaluate instructional media. Moreover, many of these approaches are especially useful for increasing the utility of evaluations and also are effective strategies to improve the utilization of scarce resources of money and time that are constraints to development and evaluation.

This study concluded that CD-ROMs can be properly developed and evaluated

through the establishment of new partnerships between evaluators and developers in which they agree to share their evaluation methods and have them published as well. Currently, the disparity in the evaluation language used by evaluators and developers inhibits these partnerships. The researcher sought to translate the methods and terms used by developers to provide the impetus by which evaluators and developers might better communicate and establish mutually beneficial working partnerships. This study will provide those professionals who undertake evaluations of CD-ROMs and other multimedia with a valuable source to consider in their endeavors.

Department of Leadership, Foundations and Policy

Curry School of Education

University of Virginia

Charlottesville, Virginia

APPROVAL OF THE DISSERTATION

This dissertation, Translating Evaluation Concepts Used By Multimedia Developers Through Their Evaluation Approaches for Instructional CD-ROMs, has been approved by the Graduate Faculty of the Curry School of Education in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Robert W. Covert, Advisor

Michael S. Caldwell, Committee Member

Patricia M. Lampkin, Committee Member

Jerry G. Short, Committee Member

June 10, 1997 Date

Acknowledgments and Dedication

I am forever indebted to a number of people who helped me reach this final stage of my studies--so many, in fact, that I cannot mention everyone here. To my committee: Bob, thank you for being my advisor and my friend. You gave me the chance to succeed and stuck with me while others might have given up. I will always appreciate your help; Mike, I enjoyed your sense of humor and learned a great deal in your classes that I know will always be useful in my career; Jerry, you gave me the impetus for this study by letting me manage the evaluation of *Cases in Cardiology*. Your help, friendship, and encouragement meant a great deal to me; Pat, you also were very helpful. Your encouragement and enduring support will always be very special to me. Thank you all.

To my parents, family, and relatives who were there for me in both good and bad times. They all inspired me and encouraged me to succeed. I will always appreciate and love you all.

Finally, to my fiancee Roberta who means everything to me. If it were not for you and your love and encouragement, I am not sure that I would have ever made it through my doctoral studies. You put up with all of my mood swings throughout the challenges of this study and I love you. I cannot wait until we are married. I dedicate this dissertation to you because you will always be the most special person in the world to me.

To all of my friends at UVA who were there for support and help throughout my studies. My wish for you is that you are all successful in your careers in education--you truly are an inspiring, talented, and dedicated group of students.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Though it is just one step toward educational improvement, evaluation holds greater promise than any other approach in providing educators with information they need to help improve educational practices. Blaine Worthen and James Sanders, 1987.

The implementation of multimedia capabilities in computers is just the latest episode in a long series: cave painting, hand-crafted manuscripts, the printing press, radio and television...These advances reflect the innate desire of man to create outlets for creative expression, to use technology and imagination to gain empowerment and freedom for ideas. Glenn Ochsenreiter, Managing Director, MPC Marketing Council, 1992.

Statement of the Problem

The field of evaluation has ventured into new territories and fields in the last several years, particularly with the advent of innovative technologies and media. As a result, evaluation has had to adapt to these technological advances to maintain its efficacy as a profession. With the onslaught of emerging and expansive advances in the fields of science and technology, this is not an easy task but rather a new challenge. To maintain their efficacy, evaluators will have to meet these challenges and conform with new rules, understandings, and methods in approaching them.

One such emerging technology confronting the field of evaluation is new instructional media such as CD-ROMs (Compact Disc-Read Only Memory). As the popularity of CD-ROMs continues to grow, the challenge facing the field, then, is very real: how will professionals who undertake evaluations of these discs conduct them? This is a challenge that has already been undertaken in software publishing companies that produce CD-ROMs, as will be explained in this study.

CD-ROMs are somewhat new to most of today's society. Many individuals know audio CDs (compact discs) from the discs that are placed into their stereos to listen to as an alternative to an album that was once "pressed" onto a vinyl record. However, the same technology that was devised in the 1980s to create audio compact discs for stereos was expanded to allow their use with computers. The benefits of such an approach were clear: the CD-ROM could store more information than a standard high density floppy disk, and because it has such a large storage capacity, it could be used to include many of the technological advances made in multimedia all on one disc rather than using several hundred floppy disks. Most of these advances in multimedia (including the Internet and World Wide Web, video, sound, animation, and digital technology) were too immense for a single standard floppy disk to handle. The CD-ROM, however, with its readable memory and up to 650MB storage capacity, could store as much information as about 451 high density floppy disks. Moreover, CD-ROMs are a new interactive medium and have introduced an innovative method of teaching using computers. With their appealing, easy-to-use, and graphical interfaces, CD-ROMs have become a boon to educators, students, consumers, and corporations because they offer such a friendly and exciting learning environment. Many current instructional CD-ROMs created by software companies have incorporated these multimedia advances into their products.

The Use of CD-ROMs For Instructional Purposes

The popularity of CD-ROMs for instructional purposes continues to grow. Paralleling this growth is an increase in the popularity of computer-assisted instruction

(CAI) for educators, students, consumers, and corporations alike. The most prevalent use of CAI has included multimedia, such as CD-ROMs. In educational settings, CD-ROMs are being used by students for self-directed, interactive learning both inside the classroom and at home. In corporate settings, CD-ROMs are being used to train employees. Today, many consulting firms and corporations are using CD-ROM-based multimedia programs to better train and inform their employees. This has opened a new and rapidly expanding field for evaluators to pursue.

The true advent of the computer age began when IBM introduced the first personal computer in 1981. Since that time, many have used computers in one way or another to assist with learning or instruction. A massive influx of computers into schools also has taken place since then. From virtually no computers in the late 1970s, school purchases of computers had grown to approximately one million units by the mid-1980s (Becker, 1987). Some estimate that this number has grown to over three million machines, with tens of thousands of teachers enrolling in computer literacy courses, and educational software developers offering twenty thousand items for sale in a recent year (Papert, 1992). It appears, then, that school expenditures (in addition to personal consumer and corporate expenditures) for computer hardware, software and teacher training has steadily increased, will likely increase even more, and has come to command an ever-growing portion of school resources.

During much of the 1980s, most instructional software was distributed on floppy disks. By the late 1980s and early 1990s, floppy disks began to be replaced by an improved form of instructional media--multimedia, which included CD-ROMs. Currently,

CD-ROMs are the standard and the most popular media by which instructional software is used (Nicholls, 1997).

Purpose and Importance of the Study

To gain a better understanding of how multimedia developers evaluate their CD-ROMs, this study described the evaluation processes used by six educational multimedia producers who participate in the production, development, and evaluation processes of their instructional CD-ROMs. By exploring this source of information both program evaluators and software developers will learn from each other how they can improve their evaluation approaches. This goal could be established through an agreement or a partnership between evaluators and developers themselves, once they consented to disseminate and share their evaluation approaches with each other. Ultimately, this would be a mutually beneficial learning process for both individuals. Moreover, by examining these evaluation approaches, both evaluators and developers can determine whether they are using the same evaluation concepts and then make modifications to their approaches if necessary. This translation and exchange of evaluation concepts would also be helpful in establishing a strong partnership between evaluators and developers.

As both the number of instructional CD-ROMs produced and subsequently used by consumers, students, and teachers increases (and the competition among software companies increases commensurately as well), the implications for consumers are apparent: there exists a fundamental concern over which CD-ROMs will be best-suited to

meet the needs of consumers. How will these software publishing companies that produce and market CD-ROMs be better informed to improve their products and make them more useful for potential purchasers? The answer to this question lies in the actual evaluation methods that these companies use to improve their products.

Software publishers and software developers, however, do not often reveal or explain their evaluation processes, nor do they publish them. This may stem from a fear of revealing trade secrets to their competitors which could result in a loss of their market share and also offers a possible explanation as to why there is a shortage of available literature on CD-ROM evaluations. It is more plausible that these publishers fear being evaluated and subjected to public scrutiny by revealing their evaluation approaches. There may be other undisclosed or unforeseen reasons for this anomaly which remain unexplained. Therefore, it is the intention of this study to identify and describe the evaluation concepts and methods that software developers use for their CD-ROMs. By doing so, this study will provide a valuable resource for evaluators, instructional designers, and other professionals who evaluate CD-ROMs.

The researcher's primary purpose in this study is to describe and summarize the evaluation processes and methods used by software developers and compare them with existing educational evaluation theories and practice. This will allow those professionals undertaking CD-ROM evaluations to choose between several evaluation designs and decide for themselves which method might work best for their own purposes.

This study will be a useful resource for the field of evaluation since the field has neither embraced nor endorsed a standard, practical model of evaluation for CD-ROMs

and other multimedia. Within the fields of evaluation and instructional design lies another problem to be explored in Chapter 2 of this study, which Cambre (1978) also searched for: Why are evaluations of instructional media not consistently performed, even though they have been long known to improve both the design and instructional processes of these media?

Methodology

By using the naturalistic inquiry model of qualitative research, this study examined the evaluation methods used by six software developers who produce instructional CD-ROMs. The data was collected through interviews with the developers who are involved in the evaluation, development, and production processes.

Patton (1990) defines naturalistic inquiry as studying real-world situations as they unfold naturally in an unobtrusive, non-manipulative, and non-controlling way. The point of using this qualitative data, he contends, is to "understand naturally occurring phenomena in their naturally occurring states" (p. 41). A detailed explanation of these naturalistic inquiry methods and a description of how they will be employed in this study are presented in Chapter 3.

Organization of the Study

In Chapter 2, a review of the literature related to this study is presented. Chapter 3

provides an explanation of the qualitative methodology employed including the limitations of naturalistic inquiry. Through write-ups of individual case studies, Chapter 4 presents the results of the interviews conducted with software developers as case studies and reveals the categories which emerged from the interviews. Chapter 5 discusses the importance of the interviews presented in Chapter 4 through a cross-case analysis, describes some limitations of this study, explores how program evaluators and software developers can learn from each other how to increase the utility of their evaluations, makes recommendations for designing and improving evaluations of CD-ROMs, and offers suggestions for further research. Appendix B provides several definitions of product testing and multimedia terms which will assist readers in understanding the terms used by multimedia developers in this study.

CHAPTER 2

REVIEW OF THE LITERATURE

Introduction

This chapter reviews literature related to the evaluation of instructional media, such as CD-ROMs, and explores a rationale for the importance of formative and summative evaluations in this field. It also documents the enormous growth rates in CD-ROM use and production.

To provide readers of this work with a historical foundation of evaluations of instructional media, how they were started, and how important they are to the field of evaluation, the first section of this chapter reviews the history and development of formative evaluation procedures for instructional film, television, and radio. These evaluations were enormously influential for instructional media of the past and will certainly provide the impetus and standard by which all multimedia can be evaluated in the future. The next section of this chapter examines how formative evaluation was defined, why there is a need to evaluate CD-ROMs and how formative evaluation can meet this need, and also explore the potential benefits of formative evaluations. Another section reviews the available literature that is in print on actual CD-ROM evaluations.

This chapter also explores the growing popularity of CD-ROMs by examining three educational settings that are currently using CD-ROMs for instructional purposes: medical education, student use in the classroom and at home, and corporate training programs, and will also discuss the implications for each of these consumers as they

consider evaluating and using CD-ROMs. Two other areas of literature which also detail the increasing popularity of CD-ROMs are reviewed in this chapter: those that document the increased production of CD-ROMs in the marketplace with statistics and figures, and those that describe the increased use of CD-ROMs by educators, consumers, and companies supported by statistics and figures, as well.

A Historical Overview of the Development of Formative Evaluation Procedures for Instructional Film, Television, and Radio

To better understand the history and role of evaluation for instructional media, the definitions of *formative evaluation* and *summative evaluation* and their relationship to the development phases of instructional media must be clarified. In relation to educational technologies, Flagg (1990) defines *formative evaluation* as:

> The process of trying out unfinished materials and acquiring information to improve curriculum design and production. It helps the designer of a product, during the early development stages, to increase the likelihood that the final product will achieve its stated goals (p. 3).

On the other hand, *summative evaluation* is defined as:

> The assessment of the intended and unintended impact of the program as the final program is implemented in its usage environment. It is usually performed independently of the project, and is not intended for the developers, but instead for the program consumers, purchasers, funders, and so on. *Summative evaluation* may yield formative information, but that is not its goal (Flagg, 1990, p. 6).

Flagg (1990) explores what the terms research and development mean in the fields of evaluation and instructional media by establishing a common vocabulary and by equating four phases of development in instructional media with four phases of evaluation

to better understand each of their objectives:

> In practice, evaluators often use the tools and methods of research. To the extent allowed by practicality, utility, and cost, evaluation meets the research criteria of objectivity, reliability, and validity. Thus, many prefer the compromising name, evaluation research.
>
> I assume four phases of program development and four parallel phases of program evaluation...to establish a common vocabulary. The names for these phases in practice sometimes vary depending on the scale and kind of program, the developing organization, and the training of the program team members. The names listed are relatively widespread or generic in the development of electronic learning materials:

Phases of Development	**Phases of Evaluation**
Phase 1: Planning	Needs assessment
Phase 2: Design	Pre-production formative evaluation
Phase 3: Production	Production formative evaluation
Phase 4: Implementation	Implementation formative evaluation *(field testing)* Summative evaluation

(Source: Flagg (1990). *Formative evaluation for educational technologies*, p. 4)

These four stages of program development and four parallel phases of program evaluation are notable and serve to define and contextualize the terms that are referenced throughout this study. Additional terms related to multimedia and to product testing are defined in Appendix B.

Cambre (1978, 1981) and Flagg (1990) argue that evaluations of instructional media, particularly for electronic technologies, are not a new phenomenon. In fact, formative evaluation of instructional media have been conducted since the 1920s, but little evidence or literature has been made available which details the procedures and developments of formative evaluations during the last seventy years (Cambre, 1978). Just

as a few, scattered articles were found in print on CD-ROM evaluations for this study, this has also been the case with evaluations of instructional media such as television, film, and radio--little has been published or been made available even though formative evaluations of these media have long been conducted.

To provide readers with a historical foundation by which they might better understand both the importance and role of formative evaluations now and in the past, this section will discuss formative evaluations that have been conducted as early as the 1920s and how they developed into the standards by which professionals in the field of evaluation can now use to evaluate new multimedia, such as instructional CD-ROMs. Moreover, a historical analysis of the development of formative evaluation techniques for instructional film, television, and radio will provide a convenient framework in which to study the field of instructional media evaluation for possible answers to the questions raised in this study--especially of why formative evaluations of instructional media are still largely ignored or not produced in the body of available literature.

Lastly, a review of the literature of the history of formative evaluations in instructional film, television, and radio will reveal the influences of these three media on the development of formative evaluation techniques and illustrate how different techniques emerged as a function of the development of these instructional media.

The Origins of Formative Evaluation in Instructional Media

In two studies, Cambre (1978, 1981) conducted what are perhaps the most complete sources detailing the history of formative evaluations in instructional media.

Both studies argue that the practice of trying out unfinished material and acquiring information to improve curriculum design and production for electronic technologies is not new. Cambre (1978, 1981) strengthens this argument by tracing such activities in the development of film, television, and radio in the United States back to the 1920s and 1930s. However, as curriculum development and curriculum evaluation have evolved and matured into defined possessions over the last two decades, certain names such as *research*, *feasibility*, *effectiveness* and *value*, were adopted for the various phases of development and evaluation (Cambre, 1978, 1981). The use of these names as phases of evaluation further confused the field of researchers, evaluators, and instructional developers who were attempting to refine what the actual definition, role, and importance of evaluation signified. In fact, the term *formative evaluation*, as understood and used today, was not officially defined until 1967 by Michael Scriven, as this chapter will attest. Dick and Carey (1996) best summarize this dilemma:

> The problem of untested materials was magnified in the 1960s with the advent of large curriculum development projects. At that time the concept of "evaluation" tended to be defined as the determination of the effectiveness of an innovation as compared with other existing products. When such studies were carried out, researchers often found a relatively low level of student achievement with the new curriculum materials. In reviewing this situation, Cronbach and Scriven concluded that we must expand our concept of evaluation. They proposed that developers conduct what has come to be called *formative evaluation*--the collection of data and information during the development of instruction which can be used to improve the effectiveness of the instruction (pp. 233-234).

The origins of formative evaluation for instructional media can be traced back to the introduction of modern audiovisual materials into the instructional process in the early 1900s. This presented educators with new choices of instructional materials, hence with decision points requiring frequent value judgments (Cambre, 1978). It was the necessity

for assessing the value (i.e., evaluation) of these "revolutionary" materials that gave rise to the need for evaluations of instructional media. Cambre (1978) defined this need as a call for "quality control":

> The need resulted from a characteristic of many technological innovations: initial energies are devoted to the development of hardware, often to the neglect of correlative development of the materials to be conveyed by the hardware. The imperative for developing suitable materials seems especially urgent in educational (instructional) technologies, where the design of the products often must be based upon more than creative intuition to achieve educational effectiveness. The result of a failure to develop appropriate materials along with the hardware is often a call for quality control from early adopters, which, if not heeded, may threaten the life of the medium itself. Such a situation existed in the early days of instructional film (p. 23).

As far back as 1920, the first documented evidence of frustration in evaluating instructional media was evident, as the demand for educational films far surpassed the supply. In 1920, Thurstone elaborated on the problems of product development as it is understood even today:

> It is obvious that the development of motion pictures in education will depend on the combination of the talents of the expert in the subject matter to be taught, the expert photographer with the experience in motion pictures, the scenario writer, the film laboratory and others. Possibly a professional pedagogue should be added to the staff to pass on the purely educational features of presentation with or without portfolio in subject matter, as the case may be. The first problem in the development of acceptable motion pictures for educational purposes is one of organization (p. 24, as cited in Cambre, 1978, p. 25).

Overall, Cambre (1978) asserts that several events occurred in the early 1920s that set the stage for the use of formative evaluation procedures in film. Schools were beginning to acquire motion picture projection equipment, only to find a shortage of acceptable films to use in their classes. The "halo effect" of the new medium soon wore off, and educators began to call for quality in educational films. Cambre (1978) states that:

> Empirical evaluation was recommended to improve product effectiveness. Some people suggested that educators be represented in the planning and production process of motion pictures for school use. The attainment of educational objectives was understood as a criterion by which educational media effectiveness was to be judged, And, finally, experimental research and evaluation on single educational films began to influence the growth and direction of the young industry (p. 52).

As the next decade of the 1930s emerged, radio and television began to assume a critical importance in both the instruction of students and in the creation of formative evaluations. At that time, subject matter experts, teachers, and educational "specialists" were frequently involved in instructional film development and production. Review by expert judgment became an established preproduction evaluation technique. The term "evaluation" assumed two meanings in the 1930s, the most popular of which was the critical review of or appraisal by the user of the educational film (Cambre, 1978). Checklists of evaluative criteria were designed to serve both producers and user in judging the worth of educational motion pictures. The second and more significant meaning of the word "evaluation" for formative evaluation purposes was that of a "continuous process" accompanying the development of an educational product. In this context, radio ushered in the technique of try-out and revision of an instructional product with groups of the intended targeted audience, which was perhaps the most important advance in the evolution of formative evaluation for educational media. Applied psychologists also began showing an interest in improving educational radio effectiveness (Cambre, 1978).

Early Motion Picture Evaluations

In the early 1920s, educational motion pictures attracted the most attention and

seemed most in need of evaluation; however, one of the first large-scale instructional film evaluations that was accomplished did not come from within the schools, but instead by Lashley and Watson, who in 1921 and 1922 conducted what was described as "practically an initial attempt at measuring and evaluating on a large scale the effects of any [motion] picture" (p. 186, as cited in Cambre, 1978, pp. 27-30). Their study was commissioned and funded by the U.S. Interdepartmental Hygiene Board in 1919 and conducted at the Psychological Laboratory of the Johns Hopkins University under a $6,000 grant. According to Lashley and Watson in 1921, they investigated "the informational and educative effect upon the public of certain motion-picture films used in various campaigns for the control, repression, and elimination of venereal diseases" (p. 181, as cited in Cambre, 1978, pp. 27-30). Specifically, they investigated the civilian version of a sex-hygiene film, "Fit to Fight," a silent film made for army use in the U.S. and France during World War I. This film, however, received a fury of opposition which caused the government to withdraw its research funding, but did not prevent the completion of most of the film.

In 1921, Lashley and Watson took this opportunity to develop and test a methodology for evaluating educational films, expressing their beliefs that their findings would "be of value in both planning future productions and in expert judgment and testing of product samples on the target market" (p. 212, as cited in Cambre, 1978, p. 28). Although the authors distinguished between the terms "measurement" and "evaluation," they did not define either, and they seemed to use the terms interchangeably. They did, however, get reactions to their film by field-testing it with seven different groups

(male/female, black/white audiences) and showing it commercially in two towns.

In all, 4,800 people viewed the film; 1,200 responded to questionnaires; 100 personal interviews were conducted to serve as a validity check on the questionnaires; and 73 voluntary field workers gathered unobtrusive measures of specific behavioral effects of the film up to six months after showing. All in all, the researchers collected an enormous amount of data on informational, emotional, and behavioral effects which they hoped would be used to improve other sex hygiene films in the future. By analyzing this data the authors discovered that there were several places in which the film failed to convey accurate information. This evidence offered a convincing argument for them to improve their next film by using this type of data during the formative stages, rather than after its release.

In 1928, the Eastman Kodak corporation, sensing the enormous market for educational films, formed Eastman Teaching Pictures, Inc. It was this company which directed the first large-scale attempt to systematize the use of expert judges before and during the production of educational films. Kodak's version of expert judgment procedures in using classroom teachers as scenario writers (these were silent films) and educational and technical experts as the judges of the material caught on very quickly and remains, to this day, one of the most consistently used formative evaluation techniques. This procedure was modified and extended in the 1930s by ERPI (Electronic Research Products, Inc.) Films, Inc., one of the first and the largest sound-film production companies in the 1930s (Cambre, 1978).

ERPI developed the "Conference Method" of film production, involving the

cooperative efforts of an educational research specialist, a subject matter expert, and a technical film expert. The technique was described as "group judgment which weighs the teaching value of each individual element" (Devereux, 1933, p. 6). Their process began with a comprehensive survey of a field to determine needs, topics, and general educational objectives for a proposed film series, and ended with a final review of the completed picture by the production and research specialists and sponsors of the new film. In between was a task analysis and a production and revision schedule.

Perhaps the most significant evaluation instrument to evolve from ERPI's "Conference Method" was that of a film appraisal checklist. These checklists, sometimes called evaluation forms, are still used today in appraising and selecting educational films. The ERPI checklist, called "Checklist for Evaluating Talking Pictures," was suggested for both production and selection use. It involved a five-point rating scale from "excellent" to "objectionable" and consisted of six major categories: Objectives of the Picture, Content of the Picture, Development of Content, Technical Audio-Visual Elements, Contributions to other Curriculum Materials, and Overview of General Effectiveness (Devereux, 1933, pp. 204-210).

In the ensuing years, other checklists continued to appear on the market with similar dual purposes, i.e., formative and summative evaluation needs. Doane (1936) reviewed some of the literature of motion picture research in education and created a "Check List" divided into only three major categories: Subject Matter, Method of Presentation, and Technical Make-up. He also provided a prediction chart to assist producers in estimating probable demand for first or trial bookings.

Goodman (1941) determined that the checklists to that date were not sufficient to meet the needs of either the school people or the film producers. He believed that a series of checklists for evaluating media should be developed and utilized using the nationwide opinions and reactions of everyone concerned--producers, distributors, administrators, teachers, pupils, etc. He also suggested that a national committee representing both consumers and producers of educational media be formed to over see a series of evaluation projects for each type of media. The end result would be a series of standards "basic to all further developments in the field of visual education...a tool for choosing the good from the bad, the useful from the harmful in the ever-growing sea of visual aids being produced..." (p. 362).

The Importance of Early Radio Evaluations

The second and more significant meaning of the word "evaluation" for formative evaluation purposes was that of a "continuing process" accompanying the development of an educational product. In this context, radio ushered in the technique of try-out and revision of an instructional product with groups of the intended target audience, perhaps the most important advance in the evaluation of formative evaluation for educational media (Cambre, 1978).

With the increased popularity of radio in the 1930s emerged a commensurate increase in the sources of data collection for formative evaluation purposes to the intended target audience. Marketing researchers for some time had been measuring the effectiveness of their products by crude methods of audience analyses before the program

reached the public. Radio was particularly suited to audience analysis, as "transcriptions" could be made of programs rather inexpensively and taken to groups for pretesting. Motion picture equipment was much less portable, and film much more costly (Cambre, 1978).

Coutant (1939) first explained the "feature analysis of radio programs." It involved gathering sample audiences together at "parties" put on by radio program sponsors. The programs were presented by means of a "playback machine," and the party guests were asked to rate various features. The result was an "appeal profile" which suggested changes to be made in the program before it was aired. Cambre (1978) suggests that Coutant's techniques served as a forerunner to the Program Analyzer and audience appeal profiles developed with scientific precision during the 1940s and 1950s for radio and film evaluations.

The use of mechanized audience polls was an important development in the evaluation of radio programs. The Program Analyzer developed in 1940 by Paul Lazarsfeld of Columbia University and Frank Stanton of the Columbia Broadcasting System for evaluating radio programs was the first major mechanical device to influence educational media evaluation. Basically, a polygraph machine that recorded audience responses at the touch of a button, the Program Analyzer and its later versions allowed for simultaneous and continuous data collection of from 10 to several hundred respondents during the course of one program. When used in conjunction with questionnaires and interviews, the device allowed producers to analyze audience reactions to their products on a second-to-second basis and investigate characteristics as they might bear upon those

reactions. These responses were then generated in the form of a rating profile--a graphic representation of the continuous reactions of the audience to the program as it proceeded. Profiles would be based on a statement of like/dislike, learning/not learning, or some variation thereof (Cambre, 1978, 1981).

As far back as 1939, WBOE, a Cleveland radio station owned and operated by the Cleveland Board of Education, began using audience pretesting techniques for their programs. The producers of this station viewed evaluation as a "continuous and democratic" process (Report of Radio Activities: WBOE, p. 78). They requested "evaluations" from a variety of sources, including technicians, students, and teachers. Their "Pre-Broadcast Evaluation" report read as:

> It is a mistake to assume that evaluation cannot take place until the radio lesson has been put "on the air." In fact, that assumption would imply a waste of much time and effort for, if the radio material is to be of increasing value, the process of evaluation and revision must commence with the writing of the script and it must be a continuous process. Unless the proposed material is "pre-judged," a waste of the listener's time is a probable result (p. 76).

The WBOE personnel regarded evaluation as a continuous process. They introduced the term "pre-audit" to describe their preliminary evaluation procedures:

> It is reasonable to assume that the most effective form of preliminary evaluation should take place in a situation similar to that in which the proposed material will be received later. Thus, it has been the practice in the Cleveland Schools to "pre-audit" the rough draft in a typical classroom which corresponds in grade level, ability, size, equipment, and so forth to the usual listener situation for that particular subject. With the radio teacher, and often the supervisor, thus noting listener reactions to the new material, the conference which follows may result in important revisions. It is sometimes found, for example, that too much has been attempted in one lesson, or that the pauses for listener activity may not be properly timed, the vocabulary may be too difficult, the questions may not be clear, the directions confusing, and so forth. Certainly, it is less wasteful to experiment thus with one class than to broadcast doubtful material city-wide (p. 76).

In addition, as early as 1939, the Cleveland producers reported that "evaluation techniques entailing the use of radio and control groups" were applied to lessons presented by radio (p. 74). Interestingly enough, however, most of the extensive measurement instrumentation (checklists, rating scales, pre-post testing, and other tests and scales) were used to evaluate the effects of the programs in a summative manner, while it appears that precise objective measurement was ignored during production. Instead, evaluation during production seems to have happened through interview or through observation techniques, and was very qualitative in nature. This may be due to the fact that time and subject availability contingencies dictated what types of data would be used to contribute to product improvement (Cambre, 1978).

WBOE's involvement with the "Evaluation of School Broadcasts" project conducted by the Ohio State University Bureau of Educational Research, Radio Division from 1936-1942, apparently influenced their decision for program improvement through evaluation techniques (Cambre, 1978). This project was sponsored by the Federal Radio Education Committee of the Federal Communications Commission, with grants from the General Education Board and the Rockefeller Foundation.

The "Evaluation of School Broadcasts" was described as a "research and service project" with a three-fold purpose: to analyze the effects of programs and their value as instruments of education, to be of practical service to broadcasters in their planning and evaluation of programs, and to assist educators and teachers in their use of their radio and sound recordings (Reid & Woelfel, 1941).

Reid and Woelfel (1941), along with their colleague I. Keith Tyler, introduced the

set of criteria for radio broadcast based on what was found to be effective in facilitating the achievement of educational objectives:

> Teachers and broadcasters alike find frequent need for techniques by which they may quickly judge the educational effectiveness of programs intended for use in schools. While ultimate judgments about programs can best be made in terms of the effects which the programs have in producing growth toward desirable educational objectives, evaluating this growth is a complex and difficult process. It requires considerable time...Out of such long-term evaluations, however, there have been derived some standards and criteria regarding the content and form of programs which can in turn be applied to new programs with a fair certainty that judgments based upon these criteria will be reasonably fair and accurate. These criteria have been derived from evaluation studies, from conscientious critical listening to and appraisal of classroom situations in which programs are being used. The validity of these criteria is therefore based upon the experience of qualified persons whose work over a period of years has made them competent experts (p. 4).

World War II and Beyond

The 1940s were arguably the most critical period in the history of formative evaluation for instructional film and television. Although the concept of evaluation was over twenty years old in audio-visual education, no attempt had been made as yet to legitimize it as a respectable activity. Evaluation studies were not included in early compendia of film research, leaving room for speculation that either evaluation was not being done to any great extent or the results of its application were not considered worthy of publication. Overall, there seemed to be little theoretical base in an empirical sense for evaluation activities (Cambre, 1978).

Outside the field of audiovisual education, three academic movements were occurring that were to have an impact on the theoretical development of audiovisual education in this period: the psychometrics movement in psychology and education, the

growth of psychologists' interest in evaluation activities as a form of applied research, and the emergence of the field of mass communications. Two of these areas, psychometrics and mass communications, did develop through the traditional theory-building stages and for that reason offer legitimate and diverse theoretical frameworks for evaluation of audiovisual products in the 1940s.

Three men, in particular, had a profound influence on the audiovisual field at this time. Charles Hoban, Jr., C.R. Carpenter, and A.A. Lumsdaine all shared an interest in the development of quality instructional products and a concern that valid research and evaluation methodologies be developed to promote progress toward that end (Cambre, 1978). Each was responsible for shaping the future of formative evaluation of instructional film and technology during and after World War II.

The Influence of Psychometrics, Applied Research, and Mass Communications in Educational Evaluation

Traditionally, evaluation has been associated with the measurement activities of psychologists and educators. Worthen and Sanders (1973), for example, contend that DuBois' *History of Psychological Testing* (1970) is also the history of formal evaluation. Hoban (1942) strongly believed, however, in the subjective approach to audiovisual research and evaluation. He acknowledged the newness of evaluation in education "at least in terminology," (p. 127) and asserted that evaluation is too often confused with measurement:

> Evaluation is a step beyond measurement, and measurement is so relatively new in education that it is difficult, particularly for those most active in the measurement

> movement, to acknowledge the more encompassing function of evaluation...evaluation is much more than measurement. Measurement may be necessary in obtaining data upon which to make an evaluation, but measurement is not evaluation. To measure a piece of string is not to evaluate a piece of string; to measure behavior is not to evaluate behavior.
>
> Evaluation is a subjective, not an objective process. It rests on empirical data and is derived from them, but the empirical data upon which evaluation rests are not merely quantitative, nor do they contribute to evaluation merely because they are quantitative. The evaluation of motion pictures in education, then, does not consist merely, exclusively, or necessarily in "measuring" the results of film showings on behavior through the use of tests that can be reduced to a mathematical integer (pp. 127-128).

The one major issue that emerged at this time, then, was the subjective-objective nature of evaluation. Psychometricians strongly believed in the objective nature of data collection for evaluation purposes, principally through objective testing procedures. Those who favored the mass communications orientation accepted more subjective, qualitative techniques. Psychologists placed evaluation in the applied research context and allowed for a combination of objective and subjective data collection methods (Cambre, 1978).

It was these three theoretical orientations, but particularly that of mass communications, which shifted the emphasis from the product itself to the effects of the product on the intended audience, i.e. from intrinsic to payoff evaluation, in Scriven's terms.

Hoban, Lumsdaine and Carpenter emerged as proponents of formative evaluation in audiovisual education, yet each had a distinct belief about the proper manner of conducting this type of evaluation. Hoban (1942) fought against the intrusion of the psychometrics movement, feeling it was limiting to the needs and purposes of evaluation. Specifically, he believed that evaluation was more encompassing than measurement and

must always involve "subjectivity."

Lumsdaine (1947) placed evaluation in the applied research context, putting somewhat equal value emphasis upon "objective" and "subjective" components in evaluation procedures. He described the problem of improving the effectiveness of instructional films as a dual challenge--improving the film production on the one hand, and improving film utilization on the other. He defined "experimental research" as "measurement of the effects on pupils produced by film instruction" (p. 254), and suggested that the appropriate measuring instruments for determining whether a teaching film is successful in accomplishing its intended purposes are factual test, attitude scales, and/or indices to measure interest, incorporated into a pre- and post-test experimental design.

Carpenter (1948) promoted data collection methodologies form the mass communications field, thereby broadening the theoretical base for formative evaluation even further. He was the Director of the Instructional Film Research Project sponsored principally by the U.S. Naval Training Device Center, and conducted by the Division of Academic Research and Services at Pennsylvania State University form 1947-1955. The objective of the program was:

> ...to discover and derive principles which should govern the scientific development and effective use of sound motion pictures and other such related media for achieving the most rapid and complete learning by individuals in groups (Greenhill, 1957, p. 1, as cited in Cambre, 1978, p. 104).

Carpenter (1948) and his researchers maintained that the extensive efforts of wartime production units were "pitifully lacking in terms of evidence of suitability and effectiveness

of their products." (p. 119).

One of the invaluable contributions of this Penn State study to the formative evaluation movement was the development and extensive use of an instrument for film analysis called the "Film Analyzer," a version of the Program Analyzer device developed by Lazarsfeld and Stanton in 1940. It suggested seven formative evaluation uses for the Film Analyzer:

1. To assist in the evaluation of scripts.

2. To be used with a "story board" treatment or filmstrip presentation of the pictorial content of the film, with a recorded narration.

3. To be used appropriately by specialists to evaluate film structure. To analyze and to study motion pictures in the "rough cut" editing stages. Valuable data may be collected for improving the final editing or production.

4. For use in preview and final decision conferences which are held to determine the acceptability of the film, in terms of technical accuracy, treatment, and overall quality.

5. For use with all of the above purposes with selected samples of people who represent the target audience for which the film was produced.

6. For use in evaluating existing films.

7. For use with other kinds of programs such as lectures, demonstrations, dramatic performances, radio, and television programs (Carpenter, Eggleton, John, and Cannon, 1950, pp. 16-17).

The Film Analyzer was designed to be used in conjunction with the "Classroom Communicator," an elaborate device that allowed for registering and tallying instant responses on test questions, or as a feedback technique in a large presentation situation.

By the end of the 1940s, and the end of World War II, it was clear that evaluation was emerging as a distinct entity in the development of educational materials.

Twyford (1951) attempted to establish the reliability and validity of the rating profiles used in conjunction with the Film Analyzer. He identified 15 different audience response devices developed during the early 1950s. It was his contention that, if he could develop an audience rating profile that correlated significantly with objective test results, the Film Analyzer Technique could be used to serve the needs of instructional developers as well as it had served the entertainment industry.

Twyford's work is one of the more significant contributions to the development of formative evaluation procedures for instructional films in the applied research tradition. He established with a respectable degree of rigor, experience, and common sense that:

> 1. Learning may be measured by means of a subjective rating profile with a degree of accuracy comparable to that obtained by testing.
> 2. A target audience of from 10 to 40 people can provide the necessary data to evaluate a film effectively.
> 3. Profiles obtained from the exact audience the film is designed to teach will have the most value for producers.
> 4. Repeated use of a Programer Analyzer technique during the production of a film results in a more effective instructional product.
> 5. If the rating profile is interpreted in conjunction with respondent interviews it is possible to determine why certain segments succeeded or failed more than others, and revise accordingly (Cambre, 1978).

In addition, Twyford (1954) compared the effectiveness of six different rating scales for measuring learning, the most effective one being the subjective indication of learning--"I am learning." The subjective indication of learning was based on a five point rating scale. In 1954, he reported an evaluation study using the "I am learning" scale on a "Nerve Gas" kinescope during three phases of production. The first profile was made while the script was read and simple drawings and actual objects were demonstrated. The second profile was made while the program was presented over closed circuit television

with appropriate visuals. A different group of 10 persons rated this program, and obtained essentially the same results. A third group of 10 persons rated the final kinescope film recording as it was projected. Twyford (1954) indicated that the similarity among the three profiles suggested that this was a useful device for pretesting film and television products. His emphasis on the importance of pre-production testing of an instructional film was solidified in this evaluation. Moreover, Twyford both refined and validated the formative evaluation technique of rating scales.

Evaluation Studies in the Air Force

The Air Force was another contributing factor to the evaluation of instructional film. Working under the direction of A.A. Lumsdaine at the Human Resources Research Laboratory (HRRL) at the Air Force, Zuckerman (1951, 1954) conducted a validation of the storyboard as a predictor of film effects in conjunction with a large-scale evaluation study of an Air Force training film "Flight Capacities of the F86A." Zuckerman's study was not only an "evaluation" procedure; it was "an attempt to use experimental methods to assess the extent to which a film strip made from storyboard art work can serve as a device for predicting how many people would learn from a completed factual motion picture" (Zuckerman, 1954, p. 50).

The study involved comparing the instructional effects of a 16mm black-and-white storyboard filmstrip with narration, with the completed color sound film. Zuckerman tested the two versions on two groups of the target population randomly assigned to one of the treatments. The results were based on correct responses on a pre- and post-test

questionnaire on information presented in the productions. From his analyses (Chi-square, Pearson Product-moment coefficients), he determined that the level of learning for items varied greatly, though the two treatments were generally well enough correlated to yield a useful prediction of the relative effectiveness of the film in conveying specific items of information. In terms of formative evaluation, the study showed that, in the case of this training film, the data resulting from field-testing a storyboard prototype of the film could be used profitably to identify strengths and weaknesses and to suggest which parts of the film treatment should be revised.

Saettler (1968) reviewed several Air Force studies and commented on a paradox that plagued formative evaluation both then and now:

> ...it is clear from the available studies that they [Air Force program research studies] hold significant implication for the use and design of instructional films and related media. However, we are simultaneously confronted with the puzzling fact that this extensive Air Force program failed to influence the film production techniques employed while these studies were in progress. Moreover, it is even more baffling when we find that none of the results of the hundred or more studies following from both the Army-Navy Pennsylvania State University program and the Air Force program have yet to be implemented in the instructional films produced by the Armed Forces. One explanation for this paradox may lie in the almost complete isolation of researchers from film production personnel which occurred in the Army-Navy studies as well as the Air Force studies...It is clear from the history of both military studies...that there is a need for a close working relationship between research and film production and that some effective methods of dissemination and innovation must be devised whereby the results of media research can be readily put into effect (p. 336).

As this section suggests, methodologies for evaluating instructional film were given a solid boost as a result of America's involvement in World War II. Audio-visual education specialists emerged to play a prominent role in the development of instructional products. A systematic approach to product development coupled with a strong need to

communicate clearly to a specified audience resulted in production planning systems that set the stage for those in use today. The military recommendations for film production promoted the clear and specific definition of educational objectives, knowledge of the intended audience, specification of a definite teaching situation, expert review during production, and a change of attitude with regard to the role of films within a given curriculum. For the first time, films began to be used as basic rather than as supplementary teaching devices (Cambre, 1978).

Most of the research studies conducted during and after the war were performed by social scientists from the applied psychology tradition. Testing procedures for evaluation purposes were elucidated and redefined. Storyboards were refined and validated. Curiously enough, however, much of the work that was done at this time still failed to make a significant impact on actual practice because of its low visibility, the "distance" between researchers and producers, and pragmatic constraints of time and effort (Cambre, 1978). Interestingly enough, the same dilemmas that faced instructional film, television, and radio producers in the past, even to this day, still plague the evaluation of instructional media (and even multimedia) and appear to diminish its importance.

Problems Inherited From the Past: A Lack of Continuity

As mentioned previously, even though a sizable amount of activity in the area of evaluation was conducted in the past decades, it is evident that, with a few exceptions such as scant and relatively obscure technical reports, most of these studies were held in the hands of a few individuals who often did not reveal their procedures or results. As a

result, the flow of theoretical information and technical information to persons involved in producing educational materials was alarmingly small and almost nonexistent. This lack of continuity was explicit in the 1950s and very much remains a problem even today (Cambre, 1978). While research was being performed, Finn (1953) acknowledged, much of it was scattered "throughout the literature of the social sciences and needs a staff of detectives to trace it down" (p. 15):

> The post-war years have brought an increase in research activities in audio-visual education and related areas. Much of this research is government sponsored and financed, but it is being published in pamphlet form, in psychological journals, or in other places more or less inaccessible to the practicing worker in the audio-visual field...The audio-visual field is in the peculiar position of having much of its research carried on by workers in other disciplines using hypotheses unknown to many audio-visual workers, and reporting results in journals that audio-visual people do not read, and at meetings that audio-visual people do not attend. While the research is expanding the intellectual background of the profession it seems to be having little effect (pp. 15-16).

An attempt to address this lack of continuity was made through the publication of *Audio Visual Communication Review* in 1953, the first professional journal for the audio-visual field. The issue of evaluation was given theoretical consideration in this journal by Hoban (1956) several years later. He listed six reasons for the failure of evaluation research to take hold of the field of audio-visual education which, ironically, are still an impasse in today's multimedia world: (a) its function is not thoroughly understood, or, if understood, is not accepted as necessary, important, or desirable, (b) it is expensive and time consuming, (c) it is restrictive on the researcher, (d) it carries few if any professional rewards, (e) it requires skills in human relations and group processes not taught in the graduate training program, and (f) its results frequently include negative findings which are

understandably unwelcome, actually threatening (p. 15).

To be maximally effective without threatening those who produce whose products are being evaluated, Hoban (1956) suggested three necessary conditions: (a) pre-release evaluation, (b) cooperation between producers and evaluators in the planning and participation in the evaluation study, and (c) communication of results on a need-to-know basis. Of the first he stated:

> Pre-release evaluation permits correction of error prior to public distribution of the product being evaluated. It has the disadvantage of delaying the completion of the product with accompanying increase in overhead cost, plus the possibility of additional time and cost of reshooting, re-editing, or renarrating to bring the material up to the required standard of acceptance and effectiveness. If no revision is necessary, this fact alone is generally worth the time and cost it takes to find it out, and if revision is necessary, pre-release evaluation makes it possible to do it (pp. 17-18).

He goes on to suggest that in terms of the cooperation between producers and evaluators, evaluation should not be something one group does to the product of another group, but rather it should be a team activity "in which the evaluator works as a member of a team to improve the performance of other team members and the effectiveness of their product" (p. 18).

For fulfilling the "need to know" or feedback function, Hoban (1956) recommended face-to-face conferences to minimize the danger of punishment for negative results. In effect he discouraged "rushing into print" with evaluation results without careful consideration of alternative ways of communicating the information.

This article was important for several reasons: it gave wide exposure to the theoretical issues and problems surrounding the evaluation of audiovisual products and it

also reflected the cumulative knowledge about the attitudes toward the concept of evaluation up to that time, endowing it with something akin to a theoretical tradition (Cambre, 1978).

The Arrival of New Methodologies

As instructional film and television productions increased in popularity in the 1960s and 1970s, new methodologies to evaluate them, based upon previous testing and experience in the field, began to emerge.

Levonian (1960, 1962, 1963) developed an interesting approach to target audience analysis before film production. His "Audience-Tailored Film" technique involved soliciting opinions via an "opinion film questionnaire," consisting of a film scene, a narrated question appearing on the screen, and an abbreviated repetition of the narrated question on a duplicated form in the audience's hand. There were 69 questions in the questionnaire, measuring knowledge and opinions about India from a sample of the intended audience. A factor analysis of the responses to these items determined the development of a film about India which, it was hoped, would mediate opinion change in a desired direction. Results from the opinion film questionnaire were used to determine the content areas to be included in the film, the order of film sequences, and, in a less significant way, the type of appeal to be employed.

The results of this study were "The New India," an Audience-Tailored Film developed by means of a pre-production film questionnaire, a multi-variate analysis of the responses, and the utilization of the quantitative results in determining film content and

sequences. It may be argued that content was determined before the film questionnaire was developed, implied in the decision of what to include and leave of the pretest--a common weakness to most needs analysis techniques. Nevertheless, Levonian's method worked in the sense that it produced a film that was highly successful in changing opinion in the desired direction (Levonian, 1963).

Gliessman and Williams (1966) reported a method for collecting problem situations to be included in a teacher-training film series. Sixty brief descriptions of different teaching and classroom problems were developed by the film team and presented to undergraduate students, professors, and practicing teachers to determine interest, usefulness, and realism. The problem situations were ranked, and treatments developed for the 24 highest ranking problems. Classroom evaluations were then conducted at treatment, script, and final product stages.

Programed Instruction

The notion of feedback so prominent in the communication models of the 1940s and 1950s became a central component of the "programmed" technologies. Nowhere was it more visible than in the programed instruction movement (Cambre, 1978). Many of the same ideas that were developed during this time remain an integral part of the programed instruction of today's multimedia.

The programed instruction movement was closely tied to learning theory, unlike the haphazard, atheoretical development of other audiovisual materials and equipment (Cambre, 1978, 1981). Lumsdaine (1964) contrasted the systematic, empirically based

efforts of early programed material writers to provide rules for constructing their products with the lack of such efforts on the part of the instructional film and television developers:

> ...it seems that empirically verifiable programing rules will increasingly be developed in the future for film and TV instruction, and that provision will be made for greater use of test data for improving program effectiveness. This could help greatly to increase the potential of these media for group instruction...(p. 380).

Markle (1967) exposed the fact that the consumers of educational products had exerted no pressure for empirical validations. She claimed, "one might hypothesize that, since the apathy toward quality control is so strong among educators, the idea may disappear altogether" (p. 107). She defined "quality control" as "precise measurement of the performance characteristics of the product" (p. 107), and suggested that a technology of quality control developed in education when three emphases converged: (a) measurable objectives, (b) explicit, observable, and preplanned reproducible teaching methods, and (c) the feedback loop from learner to instructor (p. 109). While feedback and revision were not new concepts in education, Markle (1967) maintained what the programed instruction movement did, however, add to the evaluation and revision process was:

> The close observation of the individual learner in the early stages of program design...Whereas, the curriculum study groups take chunks of finished material into classrooms, the programer typically tests small parts of his unfinished designs for material with a single student at a time (pp. 115-116).

Even in the multimedia computer society of today, the three phases of product testing, the laboratory phase (development testing, i.e., formative evaluation), the demonstration phase (validation testing), and the utilization phase (field testing) that Markle (1967) described remain the same phases which today's instructional designers

still follow.

Instances of the use of formative evaluation procedures in the development of programed instructional materials were almost commonplace by the late 1960s (Cambre, 1978). Improved programed materials through formative evaluation procedures were reported, for example in diverse sources such as the Ellson programed tutoring program in basic reading (Ellson, Harris, and Barber, 1968), the American Institute for Research industrial training projects (Markle, 1967), and a National Science Foundation mathematic program (Allendoerfer, 1969).

The Role of Instructional Television

By 1967, Gropper (1967) acknowledged the increased instructional effectiveness from Instructional Television (ITV) lessons prepared with programed instruction procedures, and stated that, "the need for empirical tryout and subsequent revision is, it would seem, well understood today (p. 9). Coincidentally, this same assertion came in the same year that Scriven (1967) baptized the concept with the name "formative evaluation." It serves as impressive testimony to the fact that the idea was around long before the name (Cambre, 1978).

The state of instructional television evaluation in the 1950s and 1960 was very reminiscent of the early days of film. Orr (1966) reported that, by the mid 1960s, about 80 cities in the United States had Educational Television stations, most of which were devoting a substantial portion of their broadcast time to in-school programing. Rutting (1967) found that the quality of instructional television had not reached beyond a low level

of effectiveness because, among other things, committees for ITV curriculum planning and development lacked specialists in research and evaluation and in educational psychology on the staff.

A review of the literature that might be labeled "evaluation studies" reveals at this time most of the effort expanded in terms of "evaluating" instructional television products centered on demonstrating the relativeness effectiveness of television in communicating an instructional message. This resulted in a rather large body of media comparison studies (TV vs. face-to face) that had little significance beyond showing that, all things being equal, television was at least effective as conventional teaching (Cambre, 1978).

In 1962, Egon Guba addressed the problem of evaluating instructional media in conjunction with the Midwest Program of Airborne Television Instruction (MPATI). He saw evaluation as "a continuously ongoing process, with the feedback from evaluation at any instant being used to further refine, extend, or redirect practice in fruitful ways" (p. 8, as cited in Cambre, 1978, p. 147). As perplexing new media variables were introduced in major projects, the focus of evaluation quickly and necessarily became the medium itself rather than the material being taught (Cambre, 1978). It was only after researchers had determined that their media comparison studies had run their course and found little or no significant differences, that evaluators could finally begin to apply the formative evaluation process with a satisfactory degree of skill and rigor. By the late 1960s, the inception of the Children's Television Workshop (CTW) model was evidence of a trend in a new direction--one which would continue well into the 1970s, and still remains influential today.

Since its inception in 1968, formative evaluation has been an integral part of CTW. The CTW Model for formative evaluation includes a few key elements--adequate funding, adequate time for prebroadcast planning and research, and a recruiting and organizational policy that ensures optimal understanding and working between the research and production staffs. This notable model, which was their destiny toward success, then proceeds though six stages:

1. Establish behavioral goals
2. Determine existing competence of target audience
3. Determine appeal of existing material
4. Test prototype materials
5. Test progress during first broadcast season
6. Conduct summative evaluation

Significantly enough, the Workshop also identifies its evaluation staff as "formative researchers"--one of the first such instances of that title in its field. "Research" in the sense of trying to identify the relative effectiveness of program variables plays a large part in the CTW Model; the aim is to generate principles of presentational learning which will have potential generalizability to other television programs, and even to other media. The model for research on presentational learning serves three functions in the formative evaluation process: (1) to act as a convenient checklist for both producers and researchers, suggesting the program attributes they need to take into account in creating new segments or designing new formative evaluation field studies; (2) to provide an organizational structure for attributes categories and for the field research methods appropriate for

measuring outcomes in each of the categories; and (3) to bring together in a convenient number of categories the great number of hypothesized program-design principles resulting from the formative research (Palmer, 1974).

Certainly over the years, many have heard of the results of the CTW research program, mainly through the visibility of its results. Numerous publications have resulted from the years of formative and summative evaluations of "Sesame Street" and "The Electric Company," many of which are available in varying sources of literature today. The enormous popularity and endurance of these shows are testaments to the strength of the instructional evaluation model that CTW uses with great success.

CTW is not the only major instructional television production company conducting formative evaluations for their programs. At the Agency for Instructional Television (AIT), evaluation is a significant feature of their funding requests. They incorporate a policy of "decision-oriented research" in every series undertaken (Rockman and Auh, 1976). Feedback and revision are major activities in the AIT process, from needs analysis through final product. "Preformative" and formative evaluation activities include the collection of data on appeal, attention, recall, comprehension, and cognitive gains.

New Directions: The 1980s and the Present

As new technologies began to make significant advances into the educational arena, the focus on formative evaluations switched from television and film to computers, even though some start-up products originated in the television industry. When the first personal computer was marketed by IBM in the early 1980s, educational producers

realized the enormous opportunities they had to create exciting and new products for instruction. This was the advent of instructional computer software and, consequently, the arrival of interactive CD-ROMs and other interactive forms of multimedia.

In 1984, Ekulona and Strohmer (Ekulona, 1985) of Maryland Instructional Television (MITV) produced an interactive disc which was a forerunner to the CD-ROM. They used pre-production formative evaluation methods to assist with their design. The two researchers had received a commission to design an innovative educational computer-controlled interactive videodisc, which they titled "The Business Disc: How To Start and Run a Small Business." It was a:

> simulation which leads the user through the steps of creating a successful small business by requiring him or her to make planning decisions, then experience twelve months of running a business based on those and subsequent decisions (Ekulona, 1985, p. 35).

One of the decisions facing this small development team was how to assure the user friendliness of the videodisc without waiting until the mastered disc and computer software were completed. At that late date, budget and time restrictions would limit significant changes. To obtain feedback early enough in the development process to make an economically feasible impact on the videodisc, formative evaluation activities occurred during the design and production stages. During the late design phase, the evaluation method of paper proofs was used frequently and informally to test the effectiveness of the disc design. Adults who had experience with their own small businesses and those without any experience played the disc--on paper. The paper proof method involved showing individual participants paper sketches of the video or computer screens while reading the

narrative storyline to them--a similar method which had been used by many instructional film producers in the past. After respondents made each business decision, they experienced the next branch of paper screens dependent upon that decision. After a few hours of work with a participant, enough information was collected to guide changes in the design (Flagg, 1990).

The revised paper design was then evaluated to check whether those changes were improvements or not. In the continuous testing and revision process, respondents participated repeatedly until fewer and fewer problems turned up. This technique provided exploration of a number of evaluation questions about user friendliness at a time when changes were easily made (Flagg, 1990).

Not all evaluations during this time were conducted in the design and production stages. Formative evaluation also played a role in the implementation phase of a product, as it did with the development of "Puppet Theater," an interactive computer software program that allows children to design puppet characters on a computer screen and lets them choose names and voices for them as well. This program was designed by Steven Ocko, a developer for Microworld Learning, and described by Flagg (1990), who worked in cooperation with the designer. Ocko and Flagg worked together to define the research questions underlying his decision whether "Puppet World" needed to be revised to use with fourth graders. The evaluators applied a range of methods to examine: (a) appeal--how much the users liked different parts of the program; (b) responsiveness--how well the program responded to users' input and wishes; and (c) flexibility--how flexibly the program could adapt to users' needs.

Ocko found the research data of his out-of-house formative evaluators to be constructive in reconfiguring "Puppet Theater" for implementation with older elementary school users. In response to feedback from user observation and interviews, the designer planned additional tools to increase the program's appeal, responsiveness, and flexibility for these users (Flagg, 1990).

One prime example of an evaluation conducted by the Children's Television Workshop occurred between 1983 and 1984. Faced with the death of one of the main characters of "Sesame Street," Mr. Hooper, the production staff met to decide how to handle this death in a way that would not negatively affect the distant preschool viewers. To help guide their decisions, the producers turned to their formative evaluation staff. What the evaluators envisioned to obtain feedback was to show a prebroadcast program portraying the death of Mr. Hooper, while communicating to children that he would not be coming back and that he would be missed (Lovelace, Schwager, & Saltzman, 1984; Sesame Street Research, 1983). Three and five year olds viewed segments about Mr. Hooper's death in small groups and were individually interviewed to appraise their understanding of the main messages.

To estimate longer term effects, the evaluators asked 20 pairs of parents and preschoolers to watch a short segment on Mr. Hooper's death, at pickup time at their day-care center. About 10 days later, parents were questioned by phone as to any effects they observed after the viewing. No negative behaviors were reported, and those children who initiated conversations discussed the facts with no emotional overtones (Flagg, 1990). Overall, this was a fine example of a pre-production formative evaluation which worked

very well.

Conclusions on Evaluations for Instructional Media

In conclusion, the usefulness of evaluation has been demonstrated in instructional technology and media repeatedly over the past decades (Cambre, 1981). In television and film, as the design and production process became increasingly more sophisticated, the evaluation techniques did also. The next section of this chapter will further explore the utility of evaluations for instructional technology and multimedia, particularly for CD-ROMs.

Why There is a Need to Evaluate CD-ROMs: The Role of Formative Evaluation

As both the number of CD-ROMs and the number of multimedia users spiral, the need for evaluation of these products becomes more apparent. Since many software companies view their market audience as a vital source of feedback in the design and development of their products, the market research divisions of these companies may have to work more closely with the evaluation teams and project managers of their products in order to improve them, thus creating new synergies in software companies. As the potential threat of fierce competition among CD-ROM companies grows, companies may react by employing more evaluation teams to assist with improving their product features, thereby offering better opportunities to increase their market share and surpass their competitors. This growth in the number of CD-ROMs has already transpired and currently

shows no signs of diminishing. Paralleling this growth is the actual targeted market (users) for CD-ROMs, who will be relied upon more substantially to assist both evaluators and multimedia manufacturers with the requisite feedback on how CD-ROMs can be improved. Clearly, a stronger link between the evaluation teams and the project teams themselves at software companies could help these organizations to improve their products.

Strommen and Revelle (1990) assert that basic research alone cannot be relied on to ensure product effectiveness of CD-ROMs and other interactive media. The authors contend that it is impossible to create optimal interactive materials without doing a careful analysis (evaluation) of how a particular product performs with actual users in real-life contexts. Flagg (1990) has also argued that instructional materials developed solely on designer intuitions are consistently inferior to those developed with feedback from actual users during product development. The methods used for obtaining this type of feedback are known as formative evaluations. The term *formative* refers to the fact that the product is still being developed, or formed, while the studies are being conducted (Flagg, 1990).

Formative evaluation was introduced by Scriven (1967), originally referring to "outcome evaluation of an intermediate in the development of the teaching instrument" (p. 51). However, over 20 years of use has broadened the application of this term to cover any kind of feedback from target students or professional experts that is intended to improve the product during design, production, and initial implementation (Flagg, 1990). Cronbach (1975) authored one of the original articles on the need for formative evaluations of instructional materials. In that article, he broadened Scriven's theory to

include the concept of instructional materials, defining formative evaluation as "the collection of data and information during the development of instruction which can be used to improve the effectiveness of the instruction" (pp. 243-244). As described by Flagg, formative research involves the pretesting of proposed educational or other materials to ensure that they can achieve their desired goals, while formative evaluation helps the designer of a product, during the early development stages, to increase the likelihood that the final product will achieve its stated goals.

One of the best-known works in the field of instructional design is *The Systematic Design of Instruction* by Dick and Carey (1996) who explain that:

> Formative evaluation of instructional materials is conducted to determine the effectiveness of the materials and to revise them in areas where they are ineffective. Formative evaluations should be conducted on newly developed materials as well as existing materials that are selected based on the instructional strategy...The evaluations should be designed to produce data to pinpoint specific areas where the instruction is faulty (pp. 285-286).

In examining the importance of CD-ROM evaluations, a fundamental question arises of how and why these evaluations should be approached differently than other kinds of product evaluation. For example, textbook evaluations are similar to evaluations of CD-ROMs in that they use both buyers (students and teachers--the target user audience) in their evaluations, but one way in which they differ is that they do not offer immediate, active feedback as CD-ROMs do with user tryouts. As a form of multimedia, CD-ROMs are a much more interactive medium than videos (and many other forms of programmed instruction) and textbooks. As a result, these discs give immediate evaluative feedback to the development and production team on how they can be improved--an important

evaluation process. Moreover, user tryouts of CD-ROMs will tell you whether the discs will run or not, thereby providing more instantaneous feedback of their effectiveness and utility than other forms of instructional media. It can be argued, then, that no textbook in use now has this type of evaluation data.

Why Evaluations of Instructional Media Are Not Consistently Performed

In exploring the historical development of formative evaluations in instructional media, Cambre (1978) contends that despite a consensus for decades among audiovisual educators that evaluation during the development process results in more effective instructional products, and in spite of the existence of models and suitable techniques, the process of formative evaluation is not regularly or consistently applied in the development of many instructional products. Moreover, there are few explanations for this anomaly, other than the constraints of time and money, or the minimal payoff it involves. Cambre (1978) argues that pretesting films or film prototypes with members of the targeted audience for purposes of improvement seemed to be regarded for the most part to be either a desirable luxury beyond the time constraints of film producers, or less often, a needless academic exercise that would not contribute to the quality of the films being produced. This also suggests that the relatively long history of product improvement through evaluation techniques has often been overlooked by contemporary commentators who tend to view formative evaluation as a relatively new phenomenon, even to this day (Cambre, 1978). Dick and Carey (1996) further this argument:

> Recent studies have shown that thousands of the instructional products sold in the United States each year have not been evaluated with students and revised prior to

> distribution. Other studies have demonstrated that simply trying out materials with a single-learner and revising the materials on the basis of that data can make a significant difference in the effectiveness of materials...[Evaluation] emphasizes the necessity of gathering data from members of the target population about the effectiveness of materials and using that information to make the materials even more effective (p. 234).

Komoski (1974) found that, of the approximately 300,000 commercial instructional materials available even at that time, only about 3,000, or 1 percent, "demonstrated one or more of the attributes of empirically developed and improved material" (p. 365). The record for the improvement of instructional films was even lower, with fewer than 1 percent being revised on the basis of empirical data, either during or after production. To extend Cambre's argument with relation to multimedia, Flagg (1990) argues that pretesting instructional media (multimedia, or multimedia prototypes) with members of the actual target audience for purposes of improvement continues to be viewed by the industry as a desirable luxury beyond the time and financial constraints of multimedia producers, and even as a needless academic exercise that does not contribute to the quality of multimedia being produced. This is one feasible explanation of why formative evaluations have not been conducted and why, if they are actually conducted, the targeted audiences are largely forgotten or ignored during the evaluation process.

Reiser and Kegelmann (1994) also maintain that all organizations which evaluate software should incorporate actual users (such as students) in the evaluation process. They suggest that targeted audience members of the CD-ROM or software being designed should be observed as they use the program and conclusions about the quality of that program should be based, in part, upon what the evaluators observe. Evaluators should also ask students and users to share their opinions of each of the software programs they

work through. More importantly, an examination of how the software program affects learning should be reported and analyzed within the evaluation. Through these processes, organizations conducting evaluations will assist educators and users alike in identifying software that will enhance learning and knowledge.

The most important feedback from an evaluation will likely come from the actual targeted audience of the product. Flagg (1990) maintains that evaluators should measure the effects of a program on the targeted population. She believes that it is a misuse of experts to expect them to estimate the materials' effect on learners. Experts are not always able to predict the effectiveness of programs (Rosen, 1968; Rothkopf, 1963).

The Benefits of Formative and Summative Evaluations For Instructional Software

According to Strommen and Revelle (1990), there are many benefits to evaluations of interactive technologies both before (formative) and after (summative) production:

> Formative testing can provide crucial feedback on the strengths and weaknesses of particular materials and also can allow for the modification of problematic features prior to release of the final product. It also provides a corpus of documented in-house knowledge based on past experience. As studies of previously developed products accumulate, they form an indispensable resource for anticipating problems with new products still in the conceptual stages of development...Firsthand knowledge of what works and what doesn't means not having to 'reinvent the wheel' with each new product, and it ultimately leads to a streamlined design process and better products (p. 72).

Only the actual testing of a product can reveal whether it works as expected, and whether users enjoy it and find it beneficial.

Helgerson (1992) contends that the evaluation of a CD-ROM can be used as both a promotional tool and a source of improvement:

> Evaluated strengths and benefits become promotional reasons for adopting the product, using it more often, or using it more productively. Evaluated weaknesses and costs become the basis for identifying improvements that can be promised for subsequent editions...Evaluation of the CD-ROM title by your users will be essential as you plan revisions, corrections, and updates. The same evaluation reports can be used to promote knowledge and acceptance of the disc among current and prospective users (pp. 43-44).

A summative evaluation conducted after the CD-ROM is released may also be beneficial for software companies. As the current trend toward the increased use of instructional software on CD-ROMs continues, it becomes more difficult and overwhelming for teachers, administrators, and other multimedia users to review a significant percentage of software on the market. Thus it has become increasingly important that educators and users be able to select software that is instructionally effective (Reiser and Kegelmann, 1994). This portends that educators and users alike will increasingly rely on evaluation organizations and trained evaluators that can provide them with the information they need to select instructionally and functionally effective software. This selection process can be facilitated through a summative evaluation because the shared results of both the review and assessment of the quality of instructional software with educators and other users will likely make them better informed to make wiser decisions in their selection.

Finally, the main reason for performing formative evaluations of CD-ROMs is clear: to inform the decision-making process during the design, production, and implementation stage of an educational program with the purpose of improving the program (Flagg, 1990).

As the use and production of CD-ROMs and other forms of multimedia computer

use increase for instructional purposes, similar growth in the evaluation of these computer materials will surely follow. The next section will discuss one such growth area by documenting actual evaluations of CD-ROMs that are in print, bringing this discussion up to date in the 1990s.

CD-ROM Evaluations in Print

A review of the available literature on evaluations of CD-ROMs reinforced the researcher's hypothesis that there is some literature which exists on this topic, but it is limited. This was an influencing factor in the researcher's reasoning to conduct this study. A primary goal was to organize and condense what little information was available in the literature on evaluation of CD-ROMs and then explain how software companies actually perform these evaluations of their products so that a body of work on this process would exist for evaluators and others to utilize and reference as they prepare to conduct evaluations of CD-ROMs. Richards and Robinson (1993) confirmed the researcher's belief stating that "very little literature has been written about how to evaluate CD-ROM software" (p. 92). Flagg (1990) contends that current instructional design texts and evaluation texts treat the topic of formative evaluation briefly, if at all:

> Since the early 1980's when computers were first produced, the field of formative evaluation has grown informally through project-specific conference papers, limited dissemination of unpublished in-house reports, and isolated articles and chapters in media research and instructional design journals and book. Researchers are still "nibbling at related problems," but widely available collective knowledge about this field is scarce. Moreover, published discussion focused on formative evaluation of computer-based materials is almost nonexistent (p. 2).

This scarcity of published research material also supported the researcher's notion that the evaluation of CD-ROMs is a very real challenge for current evaluators to undertake. What few articles and books that were found on this topic contained a paucity of information on the procedures, processes, and explanations of how the evaluations were designed or conducted. In fact, most of the literature on evaluating CD-ROMs merely details the conclusions or results of the evaluation, while ignoring the actual design and context of the evaluation. Zink (1991) supports this theory confirming that while CD-ROMs have been evaluated since their introduction, "reviews of CD-ROM products in the professional literature have largely been uncritical in their discussion of the user interface or have been limited to subjective general comments about difficulty of use" (p. 16). This can, he suggests, be attributed to the fact that existing guidelines for the evaluation process are essentially lists of features to look for.

Just as Flagg (1990) argued that scant literature is available on evaluations of instructional media and has been too few and far-between, that theory is also confirmed by Cambre (1978) who asserts that:

> The [preceding] discussion [dissertation] testifies to the fact that a respectable amount of activity in the area of evaluation during development of instructional products had been going on since as early as the 1920s. It is also evident that, with a number of notable but isolated exceptions, most of the work was concentrated in two or three major projects run by a handful of interested individuals. The results were made public for the most part in relatively obscure technical reports. Consequently, the flow of theoretical and technical information to persons involved in the work of producing educational materials in the field was alarmingly small, almost nonexistent. It is not surprising that, even today, most people involved in evaluation are unaware of this historical perspective and consequently look upon themselves as tillers of new ground (p. 121).

Just as this researcher had originally hypothesized that CD-ROM evaluations were being

conducted, even though they were unavailable or undocumented in the literature, Cambre (1978) and Flagg (1990) both confirmed this belief with evidence that there actually has been little written or documented in the arena of formative evaluations for instructional media, even though these evaluations have been conducted for decades.

In the literature reviewed for this study, most journal or magazine articles addressed the evaluation as being conducted by the authors themselves, rather than by potential users or the targeted audience for the CD-ROMs. Nicholls, Han, Stafford, and Whitridge (1990) assert that the most important part of a CD-ROM is the user interface and evaluation of the interface should concentrate on its user-friendliness. LaGuardia (1992) also argues that the best means of evaluating a CD-ROM is by trying it out.

Most of the articles reviewed for this study on CD-ROM evaluations read like a review or critique which consumers read in *Consumer Reports*, detailing only what the best and worst features of the product were, recommending a purchase or not, while completely ignoring an explanation to the reader of how the actual product evaluation was designed and implemented. In short, most of the available literature on the evaluation of CD-ROMs are product evaluations, lacking detailed information on the evaluation process itself. With this in mind, the researcher approached this study differently. He wanted to reveal how software companies were conducting evaluations of their CD-ROM products, especially since they did not often make their evaluation processes, designs, and strategies available to the general public. Moreover, the researcher hoped that this study would also challenge both designers and evaluators to change their existing guidelines for the evaluation process to something other than just lists of features to look for on CD-ROMs.

Through a literature search for possible articles that exist on actual CD-ROM evaluations, the researcher's efforts were often fruitless. Many of the articles that were uncovered merely focused on the use of CD-ROMs for library databases, evaluating only how efficient these discs were as retrieval instruments. This is due to the fact that CD-ROMs are the current industry standard and the most common retrieval media for literature searches in today's libraries (Nicholls, 1997). These "evaluations" of CD-ROMs reviewed in the literature appear to constitute a majority of the available literature found on the use of CD-ROMs in education--there is a dearth of literature available on the evaluation of educational CD-ROMs, and what little exists fails to detail the evaluation process carried out in the study.

Even fewer articles exist which expressly detail the evaluation of one entire CD-ROM--either one that was in the process of being produced or one that was already created. Most articles simply offer guidelines to follow in evaluating CD-ROMs or state that some evaluation information was collected, but offer little else.

One author, Hagenbruch (1994), mentions the need for an evaluation of "American Memory," a multimedia computer system which provides access to archival materials on American History at the Library of Congress, but only describes that she collected user surveys, comments from tryouts of the system and results from interviews with users. She neither details nor explains how the evaluation process was designed nor shares its results.

The UNESCO (United Nations Educational, Scientific, and Cultural Organization) in Paris (1992) reports the result of a study carried out to evaluate the installation and use of the UNESCO CD-ROM prototype, but neglects to explain for what purpose the CD-

ROM is being used other than stating it is a database for "information retrieval" (p. 36). There are examples of questionnaires, surveys and qualitative data on the evaluation, but little information on how these instruments were created and how the data was collected and analyzed. A low response rate of 36% may be one reason why such scant evaluation data are reported in this study, even though UNESCO characterizes this rate as "satisfactory" (p. 5). One example of how incomplete the report is in its declaration that the questionnaire was sent out to 267 institutions around the world, but does not state the rationale for how these institutions were chosen. Moreover, measures to insure instrument reliability and validity are not mentioned.

Another study examines issues in the context of the development of an interactive CD-ROM, but fails to offer both the qualitative and quantitative results of the data collected during the evaluation. Hedberg (1994) evaluates the learning outcomes from the use of an interactive CD-ROM entitled "Investigating Lake Luka," an ecology simulation CD-ROM used to instruct students in the physical components of an ecosystem. The evaluation involved three main approaches: expert review of the package, one-on-one testing of the prototype materials via video observation and interviews, and case studies. While an analysis of the collected data was conducted, the author offers no quantitative results and scant qualitative information from the conducted interviews and from the evaluation itself. Without this information the study appears incomplete, as does the evaluation of learning itself. The one important result from this evaluation which the author does detail is that the CD-ROM did, in fact, require users to take control of their learning, and that the cognitive supporting features helped with problem solving. Although

this article was far more comprehensive than others in its description of the evaluation design and efforts, it does not, however, offer any suggestions on how the evaluation design could be improved nor does it conclude that, overall, the evaluation itself was useful. This suggests that the evaluation was not an integral feature in the development of this product.

The most complete discussion of an actual formative evaluation of a CD-ROM was described by Wilson and Tally (1990), who conducted a series of formative evaluations for the "Palenque" project, an interactive optical disk prototype for children, allowing them to explore a Maya ruin in the Yucatan called Palenque. This optical disc prototype was based on themes, locations, and characters from the "Second Voyage of the Mimi" television series produced at Bank Street College in New York. The study is comprehensive in its explanation of the evaluation design and processes (which included observation of child users during the development stage, including an examination of ease of interaction, appeal, comprehensibility, and accessibility--i.e., beta testing), but lacks information on how the data was analyzed. Moreover it does not contain a reference to any evaluation instruments that may have been created. Overall, however, this was the most complete description of the design of a CD-ROM evaluation found in the literature. The study also suggested that the evaluation was extremely beneficial:

> By creating and evaluating a series of prototype discs and software, we were able to use the reactions of the child users to catch design problems early, before they became major problems. Thus, the ongoing activities of the formative evaluators served to inform the efforts of the designers, producers, and programmers. By observing children representative of the targeted audience actually using the prototype, the formative researchers were able to bring reactions from "real world" child informants to the design and development process. In this way, the creative hunches and intuitions of the designers,

> producers, graphic artist, and programmers were confirmed, modified, or unconfirmed relatively early in the development process, as an ever more appealing and comprehensible product for the target audience emerged (pp. 97-98).

Three Educational Settings Currently Using CD-ROMs for Instructional Purposes

To illustrate the growing use of CD-ROMs in actual educational settings, this section will discuss how three important fields are currently using CD-ROMs for instructional purposes: medical education, student use in classrooms and at home, and corporate training programs.

The Need For New Instructional Technology in Medical Education

In medical education, the role of CD-ROMs and other instructional technologies appears to be expanding. As technological advances continue many medical educators perceive both the apparent benefits and additional implications in using modern technology to teach physicians, as will be explained in this section of the literature review. A related factor to consider along with these implications is that as CD-ROMs continue to be used in medical education, how will they be evaluated and who will evaluate them?

Bewley (1992) contends that the education of doctors has not kept up with changes in medicine and medical technology. He maintains that training at undergraduate and postgraduate levels can be rigid and at times inappropriate. To keep up with all these changes, continuing medical education needs expansion. Junior hospital doctors need to have more opportunities to learn at their own convenience, especially when not making

their rounds. With this free time medical students, residents, and fellows could use computers and interactive media to refine and increase their medical knowledge and skills. Access to information could be improved through computer linkage with CD-ROM databases, which would allow the physician to directly review an abstract relating to a particular clinical problem and offer on-the-spot answers to specific patient-care questions (Mayne, 1994). Moreover, computers can deliver individualized interactive learning packages using simulated patients.

Many others medical educators insist that the traditional approach to continuing medical education (CME) will be inadequate to prepare practitioners for the twenty-first century (Conn, 1992). In the past, physicians could earn CME credits by attending meetings or conferences. But while these "seminars at regional and national meetings are useful to provide updates or to fill in more detailed information on the basic knowledge that all physicians must acquire during their training, a different, a more imaginative approach is needed to acquire the necessary knowledge and skills to utilize newly developed technologies" (p. 602). In that regard, self-directed learning will continue to be an essential approach to CME, and the availability of computer programs, including CD-ROMs and videodiscs, will be increasingly important. *Cases in Cardiology*, a CD-ROM created by two cardiologists at the University of Virginia, is one such example of how technology in the field of medical education is expanding and presenting physicians with greater alternatives and with more creative and viable means to augment their medical knowledge and skills. This CD-ROM was distributed to primary care physicians and allows them the opportunity to earn CME credits by using it. Moreover, this new type of

learning and education is self-directed because physicians can conveniently use CD-ROMs at their own leisure.

There are other positive and negative factors to consider when using CD-ROMs for medical education. Learning by a computer medium is a radical departure from the traditional, Socratic classroom style of teaching for physicians, yet it offers exciting new learning opportunities. Physicians could use instructional CD-ROMs at any place, time, or manner that they chose without having to sit in a classroom and listen to a lecture. With rapid changes in the field of medical diseases, many physicians who are in practice are not skilled in using some of the new approaches to treating these conditions. But there are many other advantages to using a CD-ROM. As an instructional tool, the CD-ROM is an especially appropriate medium for medical training about the heart, since it is able to combine information such as EKGs (electrocardiograms), heart sounds, x-rays, and echocardiograms in one convenient locale. Often, in training and hospital settings, these sources of information are located in different areas. In addition, most CD-ROM developers hope that their products can make learning easier, more fun and interesting for physicians.

Conn (1992) and Jennings (1994) argue that the area in which traditional approaches to CME fail most spectacularly is the introduction of the practicing physician to new technology and science. While medical fields, such as Pathology and Gynecology, are being revolutionized by advances made in technology, the traditional classroom lecture lags in providing the knowledge and skills necessary to apply new science and technology to everyday patient care. In that sense, new approaches must be found and the

introduction of computers in medical education can help in utilizing newly developed techniques such as flow cytometry and image analysis. Apparently as the technology for information management becomes increasingly important in all areas of medicine, computer literacy will be a requirement for the practicing physician. The personal computer, then, can be invaluable for CME activities and it also makes the vast resources of the literature immediately available on CD-ROM.

One question which arises immediately is how, if at all, are CD-ROMs and other technologies replacing the traditional didactic form of instruction? How effective are these new technologies in medical education and what are its advantages over the tractional methods of instruction? Mayne (1994) suggests that CME is successful when it results in improved outcomes for patients, but there may not be much connection between traditional didactic instruction and improvement in clinical practice. To remedy this, he suggests that greater emphasis could be placed on self-directed learning. By using computer technology, doctors can audit their own practices to detect specific deficiencies and reveal individual educational needs. Through this process, learning and assessment can then be linked with improvements in practice.

Folberg, Dickinson, Christiansen, Huntley, and Lind (1993) indicate in their study that interactive technologies could substitute for lectures; however, the issue is not only whether new instructional technologies are superior to traditional methods, but also whether multimedia platforms such as CD-ROMs offer opportunities for learning that are not otherwise available. The goal of all media projects should be to find a balance of form in which the value of traditional teaching methods can be preserved with the maximum

opportunity for innovation (Calhoun & Fishman, 1994). Medical education will have to further explore these issues and test them to discover the most appropriate combinations to teach its students.

The benefits of using new interactive technologies for medical education were evident in a continuing study (Folberg et al., 1993). The authors designed and implemented a complete curriculum in ophthalmic pathology using IBM- and Macintosh-based interactive videodisc technology. They also redesigned a portion of the curriculum for a new television-based platform using interactive compact discs. The results were:

> 1) Rapid access to thousands of high-quality illustrations; 2) the ability to view enlargements of photographs; 3) an online glossary to view definition of terms coupled with high-quality photographs; 4) a dynamic introduction to pathophysiology using interactive animation sequences (pp. 849-850).

The authors also concluded that while computer-based multimedia workstations are relatively expensive for personal use, they may be useful if the equipment can be shared in a learning center or library. "Compared with other interactive computer-based solutions...the interactive compact disc is a relatively inexpensive vehicle for providing medical education programs intended for use in the individual practitioners' office or home" (p. 850).

New Technologies and New Implications for Medical Education

New technologies are accompanied by new implications for the medical field. With the impending use of new technologies in medical education, a basic question immediately arises--are physicians ready for them? While it is evident that computers are poised to become key players in the delivery of health care, how useful will they be if practitioners

do not know how to use them properly? A recent conference on medical communication in the electronic era examined the potential of computers to assist in diagnosis, provide continuing medical education, disseminate evidence and research findings, and simplify practice management. The report concluded that the potential use of technologies in medical education are enormous, but in order for them to gain wider acceptance for medical purposes, physicians need to be made more comfortable with their use at an early stage of training--particularly in medical school and during their residencies and early practice (Lowry, 1995).

There are additional, imminent challenges imposed on CME programs that implement new technologies. The current pressure for public accountability of medical practitioners clearly indicates that physicians must accept the reality that CME will not be recognized unless it is provided by an accredited organization and attendance is documented (Conn, 1992). CME institutions should expect their programs to be recertified if required procedures such as peer review accompany documentable evidence that their programs are meeting national accreditation standards. This suggests that CME will be based on needs assessment, educational objectives, more effective formats, and an evaluation of whether CME and their use of new technologies and programs, in fact, improved the physician's effectiveness in practice. In this regard, evaluations of new technologies and programs in CME will surely become more critical in the future.

Instructional CD-ROM Use By Students in the Classroom and at Home

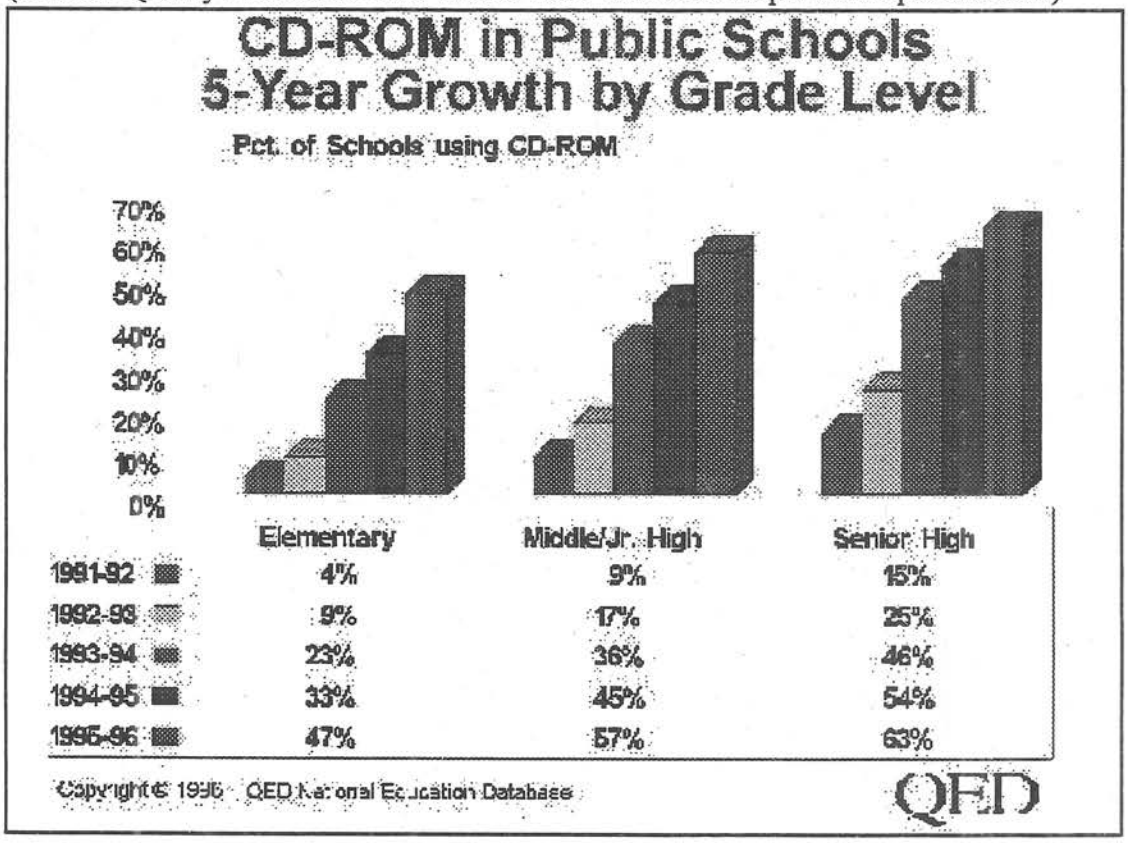
Instructional or educational CD-ROMs, sometimes referred to in the category of

"edutainment" (combining both educational and entertainment features), are rapidly growing as the medium of instruction for students both inside the classroom and at home. Home learning CD-ROMs titles for children of school age constitute one of the two primary market driving forces. A larger installed base of computers in homes with children has resulted in higher growth rates for educational title sales than for other categories: 386% during fiscal year 1994, as compared to 267% for entertainment games (Gussin, 1996). These growth rates are expected to continue. In addition, home learning CD-ROM titles are also envisioned to be the bridge to formal education, where the market for electronic materials is entering a period of considerable growth (Gussin, 1996).

There are other statistics from schools which also detail the increased use of computers and multimedia in the classroom. According to World Wide Web Page of Quality Education Data (http://www.qeddata.com/sttech.html), an education research firm, a survey of 14,201 public school districts and 84,851 public schools found the United States national average for the ratio of students to multimedia computers is 35 to 1. The same survey showed that the number of these schools which have CD-ROMs is 43,449, over 50%. CCA Consulting, Inc. reports that the number of personal computers in K-12 schools has now reached approximately 6.5 million (*CD-ROM Professional*, 9(12), December, 1996, p. 24).

Three figures from Quality Education Data are also of interest for this study. Figure 1 details the percentage of public schools using CD-ROMs through a five-year growth analysis.

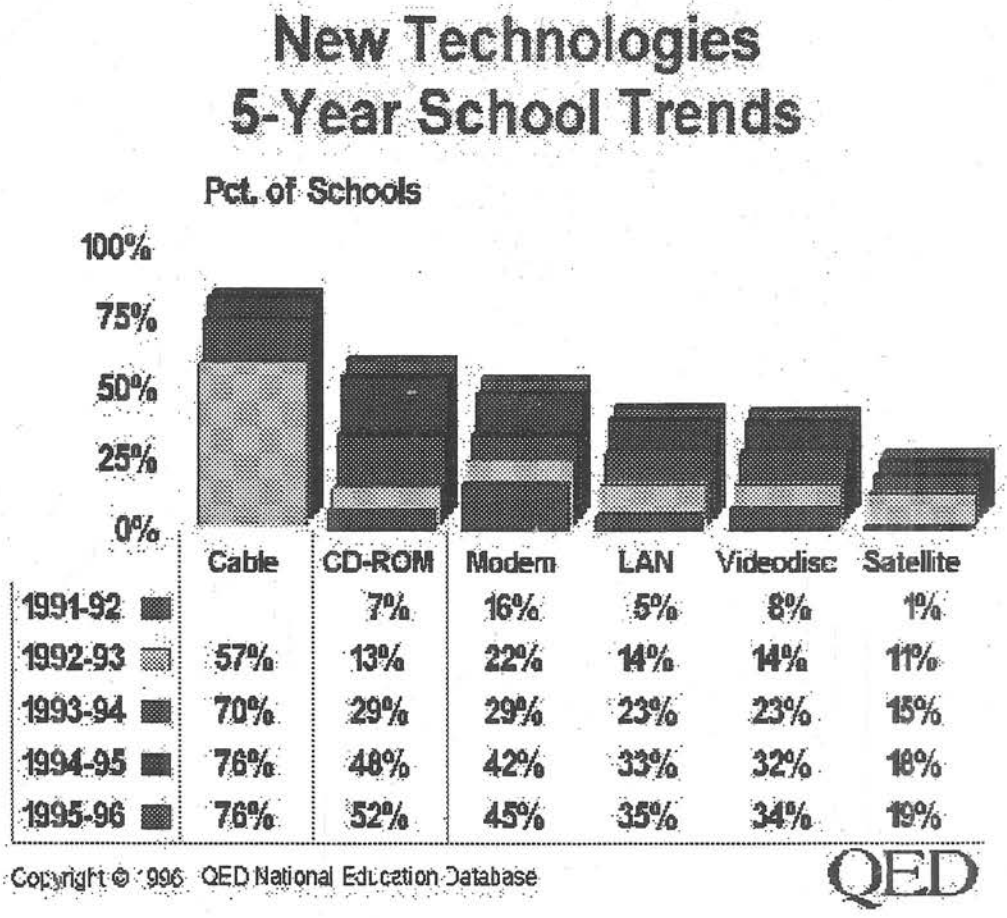
Figure 1: Percentage of Schools Using CD-ROMs: 5-Year Growth Analysis
(Source: Quality Education Data World Wide Web Site: http://www.qeddata.com)



This figure shows that the use of CD-ROMs in public schools has grown enormously over the past five years at each scholastic level.

Figure 2 details the new technologies and the percentage of schools that are using them, over a five-year trend.

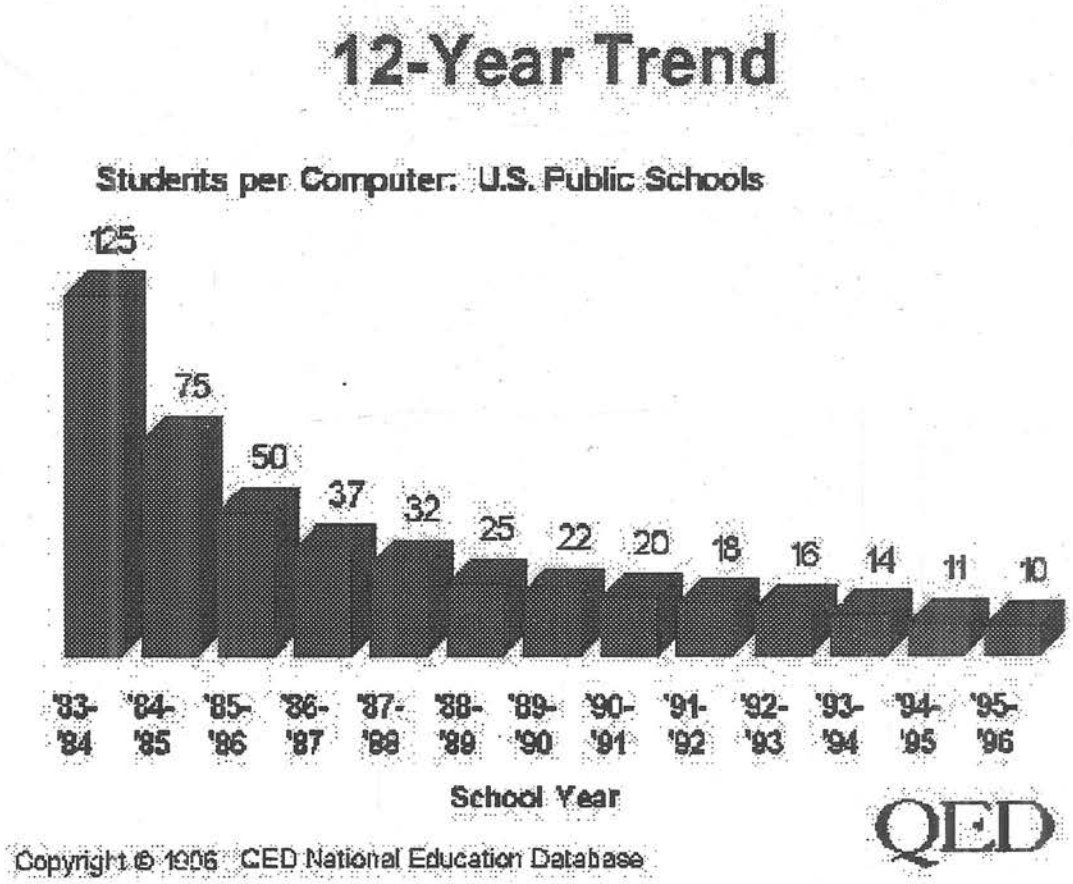
Figure 2: Percentage of Schools Using New Technologies: 5-Year School Trend
(Source: Quality Education Data Web Site: http://www.qeddata.com)



Other than cable, the figure shows that the most popular new technology is the CD-ROM, which, since 1991-1992, has grown from 7% use in schools to 52%, outpacing the growth of every other new technology over the past five years.

Figure 3 shows the ratio of students to computers in U.S. public schools over a twelve-year trend.

Figure 3: Students Per Computer: 12-Year Trend
(Source: Quality Education Data Web Site: http://www.qeddata.com)



This figure shows that over a 12 year trend the number of computers in K-12 schools continues to grow. In fact, in 1995-1996, there were only 10 students per computer.

As school districts become increasingly computer-oriented, the allocation of funds for computers in their classrooms and the total expenditures on technology continue to increase. Another survey reported in the November, 1996 issue of *CD-ROM Professional*

reported that the average United States school district is expected to spend $92.70 per student on technology in 1996-1997, up from $90.17 in 1995-1996. Total expenditures for the 1996-1997 school year are expected to reach $4.1 billion, up from both $3.9 billion in 1995-1996, and from $3.6 billion in 1994-1995 (p. 18).

The Federal Government also intends to spend more money to improve technology in the schools. The Congress of the United States recently approved a $600 billion bi-partisan omnibus appropriation measure that provided significant increases for the Department of Education *(USA Today*, September 30, 1996, p. A1). In addition, with its newly approved $26.3 billion budget for the 1997 fiscal year, the Department of Education will release funds to school districts on July 1, 1997, for the 1997-1998 school year. Some of these funds include:

1) Goals 2000.The Goals 2000 budget is $491 million for fiscal 1997, an increase of 40% from the previous year. Goals 2000 supports a spectrum of educational reform efforts, including planning for technology integration.

2) Title I Concentration Grant Programs. While the Title I Basic Grant Program increased only 2% for fiscal year 1997, Title I Concentration Grant programs which help support instructional technology will increase significantly by 46% for fiscal 1997 (*USA Today*, September 30, 1996, p. 1).

In households, the number of personal computers continues to multiply. By the end of 1995, 38.5% of all households owned a computer, an increase of 5% over the previous year (*CD-ROM Professional*, 9(12), December, 1996, p. 24). The popularity of CD-ROM use even appears to be outpacing that of the Internet. Odyssey reported that 20% of all

United States households have a CD-ROM drive, while only 14% have on-line access (*CD-ROM Professional*, 9(12), December, 1996, p. 24).

The Use of CD-ROMs for Corporate Training and Education Programs (Computer-Based Training, or CBT)

CD-ROMs are also emerging as an important instructional medium by which companies can train or better educate their employees. Many corporations and smaller companies are now using CD-ROMs and other multimedia to train their employees, and this is expected to increase at enormous rates in the near future. As the costs of bricks-and-mortar-based learning experiences skyrocket, companies increasingly are turning to technology to deliver training and education. According to an article in *Information Week*, Quality Dynamics Inc. predicts that by the year 2000, half of all corporate training will be delivered via technology. A separate study in the same magazine by the Gartner Group projects the demand for technology-based training rising 10% a year for the next two years, to $12 billion. The study states that corporate America spends $50 billion a year on continuing education to improve their employees' skill sets and retrain them to deal with the rapid pace of change in the workplace. More and more of that funding will be going into learning (*Information Week*, November 4, 1996, p. 32).

An increasing number of corporations are rushing to put multimedia-based instruction onto CD-ROM. This migration of company training video and textual information onto disc has created an abundance of work for service bureaus that specialize in company training titles. In addition to this, many new software developers are

producing authoring software especially designed for computer-based training production (Kroeker, 1996).

Businesses continue to spend enormous amounts of money to train their employees through multimedia applications. Multimedia training sales are predicted to reach into the billions in the near future, according to Datamonitor USA, a research firm. This firm also reports that multimedia for training sales will approach $8 billion in the business sector and $2 billion in home markets by 2005 (*EMedia Professional*, 10(1), January, 1997, p. 21). These figures will continue to multiply as the popularity of CD-ROMs for instruction and training grows within industry settings.

Allen (1996) argues that while traditional, stand-up classroom training will not disappear, alone it simply cannot provide adequate solutions to the problems facing those companies seeking to develop computer-based training programs. Many corporations, he states, are now jumping aboard what they see as a "new" CBT bandwagon that has actually been lumbering along successfully for nearly 20 years (p. 36). Furthermore, multimedia training is excellent for teaching basic information (such as names, dates, facts) and is effective in testing and certification and in providing much meaningful practice. Allen suggested that this type of training can also be crafted to teach critical thinking skills and to model and modify attitudes. Employed appropriately, then, multimedia training in businesses can yield significant positive results: less time and cost, better learning and performance, improved competitive advantage, and allow businesses to deliver training in a more consistent way (Allen, 1996; Schroeder, 1996).

Table 1 documents the amount of money a hypothetical business (an average sized

company with 500 employees) can save by using CBT--approximately 20% in the first year of implementation.

Table 1:
Classroom Training vs. Multimedia Training: CBT Costs With Custom Training

	Classroom Training	Multimedia Training
Wages of Trainees ($20/hr)	$400,000	$240,000
Opportunity Cost ($400/day)	$1,000,000	$600,000
Travel Costs (50% of people traveling)	$250,000	-----
Trainer Wages	$47,500	$11,400
Trainer Travel	$20,000	-----
Development Costs (custom training)	$160,000	$600,000
Delivery Systems (first year amortization)	----	$35,000
TOTALS	**$1,877,500**	**$1,486,400**

(Source: Allen (1996). The return on investment of computer-based training. *CD-ROM Professional*, 9(10), p. 44).

As this table demonstrates, given the appropriate use of multimedia training, the cost savings associated should provide businesses and institutions with at least a 30 percent return on investment within one to two years (Allen, 1996).

Table 2 takes the figures from Table 1 further by comparing classroom training with multimedia training by using off-the shelf training.

Table 2:
Classroom Training vs. Multimedia Training: CBT Costs With Off-The-Shelf Training

	Classroom Training	Multimedia Training
Wages of Trainees ($20/hr)	$400,000	$240,000
Opportunity Cost ($400/day)	$1,000,000	$600,000
Travel Costs (50% of people traveling)	$250,000	-----
Trainer Wages	----	$11,400
Trainer Travel	----	-----
Purchase/License Costs (off-the-shelf)	$150,000	$100,000
Delivery Systems (first year amortization)	----	$35,000
TOTALS	**$1,800,000**	**$986,400**

(Source: Allen (1996). The return on investment of computer-based training. *CD-ROM Professional*, 9(10), p. 44).

Table 2 shows a 45% savings to the same business if it uses off-the-shelf training (which is basically using computer software that is available at consumer stores, i.e., "off-the-shelves," rather than software self-customized for a company), as opposed to custom training. No matter the number of employees, off-the-shelf multimedia can be immediately cost-effective, despite costs related to hardware systems (Allen, 1996).

Through the documentation of current statistics and future predictions, the next section of this chapter will attest to the enduring popularity of CD-ROMs.

The Increased Production and Use of CD-ROMs:

Current Statistics and Future Predictions

To gain a better understanding of instructional (educational) CD-ROMs, this section begins by detailing the current, numerous categories into which these discs fall. Table 3 explains the top ten CD-ROM education top sellers ranked by category according to unit sales.

Table 3:
Top Ten Educational CD-ROM Categories

CATEGORY	PERCENTAGE
Reading/Storybooks	16.5
Math	14.9
Early Learning	9.6
Science	8.4
Creativity	7.8
Multiple Educational	6.7
All Other Educational	5.5
Languages	5.0
Geography	3.5
Standardized Tests	3.4

(Source: *CD-ROM Professional*, 9(6), June, 1996, p. 28)

The data above reveals that reading and storybooks for children and mathematical CD-ROMs continue to be best sellers for educational publishers.

As CD-ROMs become the medium of choice for multimedia instruction, the

number of CD-ROM titles continues to grow. The number of CD-ROMs in print grew to 3,919 in 1995, a 93% increase over 1994. Nicholls (1997) maintains that consumer CD-ROM titles in print will top 13,000 worldwide in 1997. Nearly 80% of total worldwide CD-ROM title revenues are derived from institutional markets such as the professional and educational arenas, rather than consumer market sales. (Block, 1996). During the first half of 1995 alone, CD-ROM revenue jumped 208% to $529 million (Feibus and Silverthorne, 1995).

The revenues for multimedia publishers from CD-ROMs also continues to grow, as these discs become the medium of choice for these companies. Multimedia title publishers generated 84% of their 1995 revenues from CD-ROM, as opposed to cartridge, floppy disk, or online media, according to a recent survey (*CD-ROM Professional*, 9(11), November, 1996, p. 18).

The use of CD-ROM technology is rapidly expanding. According to a report in *CD-ROM Professional* conducted by IDC/Link Resources, 80-90% of new computers are being shipped with CD-ROM drives, and currently 13 percent of households had CD-ROM drives in 1996, compared to 9 percent in January 1995 (Block, 1996). Currently, personal computers are in 38 percent of American homes, and CD-ROM drives are in 20 percent of them (Nicholls, 1997).

Table 4 details the growing popularity and return rates of CD-ROMs for 1995.

Table 4: **CD-ROM Statistics for 1995**		
Description	**Statistic**	**% Change from '94**
Household Penetration of CD-ROM Drives	13%	+4%
Titles in Print	3,919	+93%
Average Title Sales	**20,000 units**	n/a
Average Return Rate	10-30%	n/a

(Source: *CD-ROM Professional*, 9(6), June 1996, p. 64)

This table demonstrates that household penetration of CD-ROM drives and the number of CD-ROM titles in print have increased dramatically from 1993-1994.

Yetton (1996) claims that nearly 25 million CD-ROM drives were shipped in 1995, with that number expected to increase to over 50 million units by the end of 1998, and upwards of 70 million by the year 2000; however Nicholls (1997) contends most recently that the worldwide installed base of consumer CD-ROM drives is on target to almost double to 80 million in 1997. Another recent report discloses that CD-ROM drive shipments were anticipated to reach an estimated 54.5 million drives by the end of 1996, an increase of over 2000% from 2.5 million shipments in 1992 (*CD-ROM Professional*, 9(11), November 1996, p. 18).

CD-ROMs are now the most popular media used to distribute software. Due to its large capacity to hold up to 650MB of information (which is the equivalent of 130K pages of text), durability, low production costs, and ease of packaging and distribution, virtually all software titles are produced using CD-ROMs (Yetton, 1996). Also, these discs are

easily transported with no fear of data due to magnetic or other interference (which can harm a standard floppy disk by erasing all or most of its data) and can also be stamped out by the thousands for far less than the same information in any other format (Herther, 1996). The justification, then, for software developers to invest in this technology are obvious: the extremely low cost of data storage and delivery using CDs is just one major economic benefit, as explained in Table 5.

Table 5:
Costs Per Megabyte of Project Delivery on Various Media

Media	Capacity	Cost per Megabyte
Online service	$6/hour for download	More than $8 at 2400 baud
Hard Disk	100MB	About $7
Paper	2K per page	About $5
Magnetic Tape	60MB	Less than $1
Floppy disk	1.44MB	Less than $.50
CD-ROM	650MB	About $.01

(Source: Vaughan, T. (1994). *Multimedia: Making it work.* Berkeley, CA: Osborne McGraw-Hill, p. 461).

The table shows that CD-ROMs are one of the least costly mediums for a project when compared to other forms of media.

CD-ROMs are also emerging as a popular networking vehicle for companies. As more and more software applications are distributed and deployed on CD-ROM, companies are beginning to look at new ways to share this information resource within

corporate networking environments. One solution to this idea is CD-ROM networking--a networked CD-ROM subsystem incorporates three distinct technologies: CD-ROM drive devices, network connectivity technology, and data storage management products. New developments in CD-ROM technology will offer storage capacity of up to two full-length feature films or 18GB. Network CD-ROM servers attached to towers and jukeboxes will be the enabling technology central to this new concept of network storage management (Yetton, 1996).

As this section details, the popularity and use of CD-ROMs continues to grow at phenomenal rates and shows little evidence of slowing down. They have truly become the medium of choice for instructional learning.

Summary of the Literature

In reviewing literature for this study, this chapter detailed several important conclusions related to the growing importance and need to evaluate instructional CD-ROMs:

1. Both formative and summative evaluations can be beneficial and useful in the overall improvement of instructional CD-ROMs.

2. Formative evaluations of instructional media have long been conducted, but were often overlooked, unknown, treated as unimportant or extravagant, hidden away in obscure technical reports or conference papers, or disguised in terminology rather different from the evaluation and research terminology used today. This dilemma was

heightened due to widespread confusion and misunderstanding regarding the role, importance, and definition of formative evaluation since, until 1967, professionals lacked a formal definition and understanding of what that term represented.

3. Even though evaluations of instructional media have long been conducted, there is little literature to document them. This may lead to a false assumption that evaluations of instructional media are not conducted or are regarded as unimportant. It also may lead researchers interested in this field to falsely conclude that they are "tillers of new ground," as Cambre (1978) stated.

4. There is a dearth of literature available on the evaluation of educational CD-ROMs, and what little exists fails to detail the evaluation processes that were carried out in the study.

5. The use and production of CD-ROMs continues to expand at enormous rates, as does their popularity for instructional purposes.

6. As the popularity and use of instructional CD-ROMs continues to grow, evaluation of these discs will become more vital to both consumers and to the software companies that produce them--simultaneously elevating the need for trained evaluators in this field.

The next chapter describes the methodology used in researching this study. These methods will be useful in exploring the evaluation techniques that software publishers use in evaluating their CD-ROMs.

CHAPTER 3

METHODOLOGY

Introduction

This chapter begins with a description of naturalistic inquiry as described by Lincoln and Guba (1985) and Patton (1990). The second section of this chapter concentrates on the research procedures employed in this study. A third section explains the techniques used to determine the trustworthiness and rigor of the research. Information on the naturalistic researcher, which serves as the research instrument in naturalistic inquiry, is presented in the final section, along with the researcher's qualifications in conducting this study.

Naturalistic Inquiry

The focus of this section is on describing the philosophical background of naturalistic inquiry.

Naturalistic inquiry has been so named because it is *not* the aim of the naturalistic researcher to manipulate the research environment (Patton, 1990). Naturalistic inquiry is supported by the axioms of the naturalist (or constructivist) paradigm rather than the positivist paradigm, upon which quantitative research methods are based.

The Naturalistic Paradigm

Lincoln and Guba (1985) stated that the naturalistic paradigm has five basic beliefs or axioms:

1. Nature of reality. Naturalistic researchers believe that reality is constructed by the individual, has multiple perspectives, and should be studied from a holistic point of view.

2. Relationship of knower to the known. In the naturalistic view, the researcher (knower) and the object of the research (known) interact and influence each other--they are not independent and separated.

3. Possibility of generalization. The intention of the naturalistic researcher is to develop a unique body of knowledge in the form of a "working hypothesis" that describes an individual case. The naturalistic researcher seeks to present this working hypothesis in such a way (described in the thick description) that a reader may decide if outcomes from one context might hold in another context, or what is the basis for transferability from one context to another.

4. Possibility of causal linkages. According to Lincoln and Guba (1985), naturalistic researchers believe that all events or actions are mutually interactive and shape each other. Because naturalistic inquiry investigates the constructed reality of humans this makes it difficult, if not impossible, to attribute an action to a specific cause.

5. Role of values. The naturalistic paradigm proposes that all inquiry is value-bound because of several influences. These influences include the researcher's choice of the problem, the research methodology, as well as the choice of the theoretical basis for guiding the data collection and analysis (Lincoln and Guba, 1985). In addition, naturalistic

researchers believe that the values of the environment where the inquiry was conducted act as an influence on the results.

Characteristics of a Naturalistic Inquirer

Lincoln and Guba (1985) describe several characteristics for the employment of naturalistic inquiry that are grounded in the five assumptions of the naturalistic paradigm. Ten characteristics that are most relevant to this study are described below.

1. Conduct research in the natural setting. Naturalistic research is conducted in the natural setting. For example, the software developers involved in this study were interviewed while they were in their job settings.

2. Employ humans as the research instrument. In a naturalistic inquiry the researcher uses himself or herself as the primary data gathering instrument. A naturalistic inquirer may also engage other individuals as data gathering instruments.

3. Use primarily qualitative data. Although not exclusively used, qualitative data is the principal source of data for the naturalistic researcher. Quantitative data may also be used as secondary data for triangulation purposes.

4. Apply purposeful sampling methods. The naturalistic inquirer is likely to apply purposeful, or theoretical, sampling rather than the traditional random and representative sampling methods used in quantitative research. Purposeful sampling allows the researcher to specifically choose the sample based on a specific purpose, so as to maximize the scope of the data revealed.

5. Conduct inductive data analysis. Naturalistic researchers prefer inductive data

analysis because it is more likely to provide the rich background necessary for determinations of transferability. In addition, inductive analysis is more likely to disclose the "multiple realities" in the data gathered in a naturalistic inquiry because the researcher does not approach the study with a predetermined set of categories.

6. Provide for grounded theory. In naturalistic inquiry, the guiding theory emerges from the data. The naturalistic researcher approaches the study from an open-minded perspective (being as neutral as possible) to allow for the description of and understanding of the "multiple realities" of the respondents. This approach is an attempt to devise theory from the context of the study rather that from the values of the researcher (Lincoln and Guba, 1985).

7. Provide for emergent design. The design of the study emerges as the naturalistic inquiry advances. This does not mean that the researcher begins the study without any expectations. For example, the naturalistic inquirer may begin the interview process with several questions. As the interview process continues, however, other questions emerge during the conversation with the individual. These additional questions that emerge during the interview process could also be included in the interview of a subsequent respondent and might turn out to be key questions of the inquiry. Other design decisions, such as which respondents to include in the sample, are also determined during the naturalistic inquiry.

8. Negotiate outcomes with participants. In a naturalistic inquiry, the researcher is attempting to reconstruct the participants' realities. This makes it imperative that the naturalistic researcher obtain verification and confirmation that the information presented

in the case reports and the interpretations are accurate.

9. Employ the case study reporting mode. Presenting the data in case studies provides for the thick description necessary for transferability determination. In addition, case studies "provide more valid portrayals, better bases for personal understanding of what is going on, and solid grounds for considering action" (Stake, 1981, p. 32).

10. Apply special criteria to determine trustworthiness of the research. This final characteristic of naturalistic inquirers presents the criteria that are specific for affirmation of trustworthiness. The traditional means of determining trustworthiness employed with quantitative research of validity, reliability, and objectivity, according to Lincoln and Guba (1985), are inconsistent with the five axioms of the naturalistic paradigm. Instead, credibility, confirmability, and transferability are sought to confirm the trustworthiness of naturalistic research.

The next section of this chapter will describe the research procedures and sampling techniques used in this study.

Research Procedures and Sampling Techniques

This section is intended to clarify the naturalistic research procedures and sampling methods employed in this study. These procedures include: the process of purposeful sample selection, data collection strategies, the guidelines for the case report development, and case analysis strategies.

The six software developers identified for this study were selected because they

were involved in the evaluation process in a major capacity. All of them had a broad range of experiences and roles in evaluating instructional CD-ROMs at their companies--either managing the evaluation, or having some form of vital participation in that process.

Purposeful Sampling

For naturalistic inquirers, the process of purposeful sampling is employed, rather than the representative, or random method utilized by quantitative researchers (Lincoln and Guba, 1985; Patton, 1990). Purposeful sampling enables the researcher to choose a sample with the intent of extracting the most useful information for the purposes of inquiry. The six individuals interviewed in this study were selected through purposeful sampling. Specifically, the researcher identified three respondents through contacts that he had established with friends in the field of instructional design. One individual also volunteered for this study after the researcher posted a request for participants in an electronic newsgroup for evaluators. The remaining participants were individuals whom the researcher already knew and had agreed to be interviewed for the study. As these individuals were contacted, they were screened by asking two primary questions: one, did the companies they worked with produce instructional CD-ROMs and two, did the individual have a participatory role in the evaluation process of those discs. Once this was ascertained, the researcher asked them if they would take place in this study.

Together, these six individuals represent a group who have a wide range of experiences in evaluating CD-ROMs for their publishers. The researcher's primary rationale for choosing them was that they were involved in the evaluation processes of

their CD-ROMs, and would consent to being interviewed about that process. The group which was interviewed turned out to be diverse--some from small software publishing companies, others from large to very large companies.

Qualitative Data Collection

Primary and secondary data were collected for this study. Interviews with those who had worked with software publishing companies served as the primary data source and were developed into case reports. Secondary data was also obtained from these individuals.

Primary sources of data. The primary source of data for this study was collected in interviews with each of the six participants. The researcher conducted these six interviews over a period of several months.

Two different interview techniques, as described by Patton (1990), were used for this study: an open-ended interview and an informal conversational interview. There were a few guiding questions developed for each of the interviews in reaction to responses received through prior interviews and based on the interview transcripts. The first question the researcher asked of the participants in this study was an open-ended one: "Tell me about the evaluation methods you use for your CD-ROMs." Appendix A includes a sample of the interview questions used in this study.

Secondary sources of data. This researcher obtained additional data from the participants in this study. These included feedback forms, questionnaires, and surveys used in conducting their evaluations. These also help support the credibility of statements made

by the interviewees about their evaluation processes.

Case Analyses

Each interview provided in this study was organized into a separate case study. Cases were analyzed using inductive data analysis, specifically content analysis. The experiences of the participants interviewed as described in the cases were analyzed and coded into categories that are representative of their responses from the interviews.

Overall, there were several steps the researcher took in performing a content analysis of the interviews. Specifically, in order, they were: (a) identified the universe of content--the responses from the six developers which were obtained through interviews over a period of several months; (b) obtained examples of the content to be analyzed--the transcripts of the interviews; (c) read through all the transcripts, identifying the coding unit as themes (categories); (d) read through the transcripts again, specifying the category system (of which two major categories and several sub-categories were identified); (e) applied the selected categories (themes) to the individual coding units; (f) revised the categories after re-reading through the transcripts and their analyses; (g) created a cross-case analysis matrix to better analyze all the categories; (h) reviewed all the categories with the peer debriefer and external auditor to verify them; (i) began to write case studies using quotes and categories from the analyzed transcripts.

Through a cross-case analysis of each case, Chapter 5 explores the categories that were mentioned by all six developers as well as the categories not mentioned by all six developers. The categories were then examined by the naturalistic researcher for their

completeness. Guba (in Lincoln and Guba, 1985) lists several guidelines that should be applied to determine completeness. One of the guidelines stated that the categories, when viewed externally, should constitute the whole picture; and when viewed internally, the categories should appear consistent. A second guideline stated that there should be enough categories to include all the relevant and important facets of the study. Guba also suggested that an external auditor should be able to determine the completeness of categories be replicating the inclusion of units into the same categories and should be able to verify that the categories are appropriate for the data.

Cases

The primary data that were collected in the interviews were organized into cases. The interviews with the six software developers provide a complete description of the context of the inquiry, and are presented as case studies, each providing a comprehensive overview of the interviews. Every attempt was made for the case to stand alone and allow, "the researcher to understand the case as a unique, holistic entity" as suggested by Patton (1990, p. 387).

Since the cases will be analyzed in greater depth in Chapter 5 through a cross-case analysis, Chapter 4 presents the essential categories that emerged from these interviews after they were transcribed and analyzed by the researcher. To give examples of how these themes arose, actual quotes from the interviews are used. Chapter 4 will also include a brief description of the setting, including background information on the participants themselves to give the reader a better understanding of the companies they work for and

their job responsibilities. Through a cross-case analysis, Chapter 5 will examine all the categories that emerged in Chapter 4 and compare the categories that were mentioned by all six developers as well as those that were not mentioned by all six developers. A cross-case analysis matrix will be used to examine this.

Lincoln and Guba (1985) listed several guidelines to be followed when preparing cases. They recommend that cases be written in a non-interpretive fashion and in a manner that allows the respondents to express their constructs in their own language and with enough detail to give the reader of the case the feeling that he or she was "suddenly transported to the site" (Lincoln and Guba, 1985, p. 365). With this in mind, the cases in Chapter 4 are written in informal, active voice with little narrational interruptions. Cross-case analyses and interpretations are included in Chapter 5. The cases are written so as to honor the agreements to maintain the confidentiality of the respondents.

These cases provide a thick description of the individuals as they detailed their roles in the evaluation processes at their companies. The thick description provided in these cases includes a rich description of the location and circumstances surrounding the interview as well as frames of references supplied by the researcher.

Trustworthiness

In conventional quantitative research, a researcher seeks to determine the levels of validity, reliability and objectivity in a study. According to Lincoln and Guba (1985), it is inappropriate to apply these same criteria in qualitative research as they are in conflict with

the axioms of the naturalistic paradigm. Instead, they proposed that investigators, using qualitative research strategies, seek credibility, dependability, transferability, and confirmability. Lincoln and Guba (1985) stated that these criteria, "together with corresponding empirical procedures...adequately (if not absolutely) affirm the trustworthiness of the naturalistic approaches" (p. 43).

For each of these four elements of trustworthiness, Lincoln and Guba (1985) clarified the definition of each element by listing a question qualitative researchers should ask themselves:

> 1. Credibility or Truth Value. How can one establish confidence in the "truth" of the findings of a particular inquiry for the [respondents] with which and the context in which the inquiry was carried out?
>
> 2. Transferability or Applicability. How can one determine the extent to which the findings of a particular inquiry have applicability in other contexts or with other [respondents]?
>
> 3. Dependability or Consistency. How can one determine whether the findings of an inquiry can be repeated if the inquiry were replicated with the same (or similar) [respondents] in the same (or similar) context?
>
> 4. Confirmability or Neutrality. How can one establish the degree to which the findings of an inquiry are determined by the [respondents] and conditions of the inquiry and not by the biases, motivations, interests, or perspectives of the inquirer? (Lincoln and Guba, 1985, p. 290).

Several techniques were employed in this study to address these questions. Three qualitative techniques: triangulation, member checking, and peer debriefing were employed to determine credibility. For transferability, the researcher provided thick descriptions in the cases. The three strategies used to determine credibility were also used to establish dependability. An external audit was conducted to insure dependability and

confirmability. Each of these strategies is described below as it was employed in the context of this research project:

Credibility

Triangulation. One way to reinforce the design of the study is to utilize triangulation strategies. Triangulation is a technique that is used to check to see if what the person interviewed said is credible. A primary triangulation technique that the researcher used for this study was to discuss the evaluation approaches and methods revealed to me by the six individuals with other members of their organization to verify that they had actually been conducted.

Data triangulation, or comparison of data collected by applying qualitative methods, was also employed with this study. To verify the evaluation practices of the participants interviewed, various instruments such as feedback forms, questionnaires, and surveys used by software developers in their evaluations were collected and reviewed by the researcher.

Member Checking. In order to address the credibility of the research, the interviewees were asked to confirm the information presented in the interviews. In addition to the researcher asking for clarification of information during the interview sessions, the interviewees reviewed the drafts (transcripts) of the cases. The respondents were also given the opportunity to review the final cases as written in Chapter 4, including the case analyses, and interpretations. Each respondent was asked to determine the accuracy of the interpretations and had the right to amend his or her statements or add

additional statements, clarifications, or explanations. All six individuals interviewed for this study made some minor corrections to the final cases, which the researcher went back and corrected.

Peer debriefing. A peer reviewer was designated for the study. According to Lincoln and Guba (1985), "the debriefer must be someone who is, in every sense the inquirer's peer, someone who knows a great deal about both the substantive area of the inquiry and the methodological issues" (pp. 308-309). To meet this criteria, the researcher chose a graduate education student who had conducted several naturalistic studies. She was chosen because she held other qualifications similar to those of the researcher, including a formal background in evaluation and instructional technology. In addition, she had training in qualitative methodologies and had employed them in other research projects.

The peer reviewer examined the raw data, case studies, case analyses, and interpretations as they were completed. She provided feedback and recommendations to the researcher during the data collection and analyses processes.

Transferability

Thick description. In order to determine if the information provided in this study is transferable to other situations, the researcher provided *thick description*, or an extensive presentation of the salient contextual features of the inquiry (Hipps, 1992). This thick description should provide everything that a reader needs to understand the interpretations made by the naturalistic researcher. The description must also provide the reader with

enough information so that they can determine the transferability of the study to another situation. According to Lincoln and Guba (1985): "The naturalist cannot specify the external validity of an inquiry, he or she can provide only the thick description necessary to enable someone interested in making a transfer to reach a conclusion about whether transfer can be contemplated as a possibility" (p. 316).

Dependability and Confirmability

Near the completion of the data collection and data analysis phase, an external auditor was asked to complete an inquiry audit. The external auditor selected for this study had recently completed his own naturalistic inquiry study using similar methods to determine trustworthiness. He had been identified by other qualitative researchers as having competence in naturalistic methods of inquiry. He was provided with all pieces of the audit trail necessary for him to complete his audit which included: the cases, interview transcripts (with the categories and the researcher's analysis on them), cross-case analysis, audiotapes, methodological log, and the questionnaires, surveys and feedback forms supplied by several of the interviewees.

The naturalistic inquirer and the external auditor determined the goals of the audit. The external auditor was given the interpretations and analyses of the study and he affirmed the emerging categories described by the researcher.

After the auditor reviewed the components of the audit trail he completed an assessment of confirmability and dependability. This assessment included an investigation into whether the inferences and category structure based on the data were logical--he

confirmed that they were. In his audit, he verified that findings of the study were well-grounded in the collected data and that they could be replicated.

In addition, the researcher maintained a methodological log as part of the audit trail created in the study. It included the dates of the interviews, decisions on the use of particular methodologies, as well as critical information on how and when necessary changes and modifications were made in the study. This was also reviewed by the auditor, who again expressed confidence in the inquiry methods of the researcher and confirmed its dependability.

Researcher as Instrument

In qualitative studies the researcher serves as the instrument; therefore it is necessary to thoroughly describe his or her background and research qualifications. The following section highlights the researcher's background as a naturalistic inquirer and as a student who has completed a number of graduate level courses in both evaluation and instructional technology.

The Naturalistic Researcher

The researcher came into this study with a background in educational program evaluation and instructional technology. During his doctoral studies in evaluation, he managed an evaluation project involving an instructional cardiology CD-ROM *(Cases in Cardiology*) created for physicians to study heart diseases in a multimedia environment.

This project was mostly qualitative in nature, using interviews, videotaped user tryouts sessions, and questionnaires with physicians as the primary sources of data. It was during this project that the premise for this study was largely conceived. The researcher had worked with several developers during the project and perceived that interviews with software developers describing their experiences and designs of evaluation processes for CD-ROMs would be an interesting study. Moreover, he concluded that, as a member of the management team and lead evaluator of the *Cases in Cardiology* CD-ROM, that a common vocabulary needed to be established between evaluators and developers. That project experience, along with the graduate courses completed in evaluation, qualitative research, and instructional technology, led him to undertake a study of this nature, and also qualifies him a naturalistic researcher.

Other naturalistic studies, qualitative research projects, and evaluation projects were conducted by this researcher during his doctoral studies which also lend credibility to his background in these areas. He also completed several evaluation case projects as a doctoral student in which content analysis and qualitative research were used as the primary methods of data analysis and research. All of this background provides the context for his interpretation of this study and also in providing him with a framework to conduct a study of this kind.

Limitations of Naturalistic Inquiry

This study shares the experiences of six software publishers who evaluate

instructional CD-ROMs. While this may not allow a great deal of generalization, it does, however, provide an in-depth look at their experiences. Patton (1990) acknowledges that "qualitative methods typically produce a wealth of information about a much smaller number of people and cases. This increases understanding of the cases and situations studied but reduces generalizability" (p. 14). The results of naturalistic inquiry are not meant to be generalized to any population as the sample is not representative. In fact as Hylton (1988) explained,

> Naturalistic inquiry rejects the notion that generalizability in the 'scientific' sense is either possible or useful, particularly in the areas of human behavior and individuals' perceptions, and accepts that what is found in a particular context has idiographic meaning for that context at that time" (p. 41).

The naturalistic researcher must provide case studies that contain enough information, or thick description, to enable the reader to determine the transferability. According to Lincoln and Guba (1985), "It is...not the [researcher's] task to provide an index of transferability; it is his or her responsibility to provide the data base that makes transferability judgments possible on the part of the potential applier of the research" (p. 316). The data base is the thick description provided in the case studies in Chapter 4.

Summary of Research Methods

By using the naturalistic paradigm of qualitative research, the experiences of six software developers who evaluate instructional CD-ROMs were investigated. Interviews with the six developers were the primary source of data and case studies were developed from these interviews. The cases were analyzed using inductive analysis techniques,

specifically content analysis. Credibility of the research was ascertained through strategies such as triangulation, member checking, and peer debriefing. An external audit was conducted to determine the confirmability and dependability of the research findings. The thick description provided in the cases in Chapter 4 should be sufficient for other researchers to determine the transferability of the research results.

Table 6 offers an overview of the six developers and information about their interviews.

Table 6:
Overview of the Six Developers and Their Interviews

	Jackie	Bob	John	Gwen	Jack	Mary
Job Title	VP of Marketing & Project Manager	President	Instruc-tional Designer	Project Manager	Professor & Developer	President
CD Topic	Algebra, Trigo-nometry	Account-ing	Anatomy	Computer Net-working	Marketing	Breast Cancer
Size of Company or Publisher	<10 F/T employees	<7 F/T employees	>300 F/T employees	>1000 F/T employees	>1000 F/T employees	<10 F/T employees
Selection Criteria	Produces and evaluates instruc-tional CD-ROMs	Produces and evaluates instruc-tional CD-ROMs	Produces and evaluates instruc-tional CD-ROMs	Produces and evaluates instruc-tional CD-ROMs	Produces and evaluates instruc-tional CD-ROMs	Produces and evaluates instruc-tional CD-ROMs

Table 6:
Overview of the Six Developers and Their Interviews

	Jackie	Bob	John	Gwen	Jack	Mary
Method of Contact	Through friend in software industry	Through friend in software industry	Through friend in education (also knew him)	Through posting to electronic news-group	Previously worked with him	Through contacts in education
Interview Method	Taped in person at her office	Taped in person at his office	Taped in person at his office	Taped by telephone while in her office	Taped in person at his office	Taped by telephone while in her home office
Length of Interview	1.5 hours	1.25 hours	1.5 hours	1.5 hours	1.5 hours	1.75 hours

This table shows the diverse job titles of the six developers. It also displays the topics of the CD-ROMs discussed in the interviews and reveals the sizes of the companies for which the six developers work. A detailed background of the six developers (including job titles, education and experience, as well as the companies they work for and the products they produce), how the researcher chose them, the setting of the interviews, and the researcher's method of contacting and interviewing the developers are explored in greater depth at the beginning of each case in Chapter 4.

The next chapter details the evaluation approaches used by six software developers through individual write-ups of the interviews, which are presented as case studies.

CHAPTER 4

THE CASES

Introduction

As Lincoln and Guba (1985) state, one method of seeking answers in a naturalistic inquiry is through the development of case studies. This chapter is comprised of the write-ups of each participant's interview. The interviews serve as the primary data for this inquiry and are presented individually as six case studies, supported by actual quotes from the participants. Each interview was conducted while the six developers were in their natural job settings. The researcher analyzed each interview and divided them into primary categories, within which are sub-categories, that emerged. Background information concerning the six individuals and the companies they work with are offered so that the reader will have a better understanding of the context and setting of the interviews. At the end of each case, a brief analysis of the case study is provided. The order of the case studies is the same as the order in which the individuals were interviewed.

The next section of this chapter will detail the two primary categories that emerged from a content analysis of the six interviews and explain their importance to the study.

The Two Primary Categories and Their Sub-Categories

Overall, two primary categories emerged from a content analysis of the six interviews: 1) types of evaluation performed, and 2) overall views and uses for

evaluations. Each category has its own sub-categories. In the first primary category, one sub-category was preordinate--product testing. The researcher entered this study anticipating that all developers used product testing of some nature to evaluate their CD-ROMs. That belief was confirmed by the six developers in this study. The categories and sub-categories were:

Category One: Types of Evaluation Performed

1) Formative Evaluation (Methods Used):

- Needs Assessment
- Product Testing (Beta, Alpha, Platform and Cross-Platform)
- Partnerships with Educational Institutions
- Target Audience (User) Tryouts
- Content Expert Review
- Focus Groups
- Conferences and Seminars
- Surveys, Questionnaires

2) Summative Evaluation (Methods Used):

- Market Feedback (Interviews, Surveys, Post-Release Testing, Feedback at Conferences and Seminars)

Category Two: Overall Views and Uses of Evaluations

- Financial Constraints for Evaluations
- Evaluation Results Used to Make Changes on CD-ROM
- New Synergies Created Between Divisions of Company for Evaluation
- How the Development and Evaluation Processes Differ Between CD-ROMs and Textbooks
- Overall, Evaluations Helpful for Product
- Suggestions for Improving Evaluations

Category One: Types of Evaluations Performed

In analyzing the six interviews, the first major category which emerged was that of the types of evaluation performed by the software developers. These included the two major types of evaluations: formative and summative, each of which has examples of the

types of method used. Within formative evaluation, there were eight sub-categories: needs assessment; product testing; partnerships with educational institutions; target audience (user) tryouts; content expert review; focus groups; conferences and seminars; and surveys and questionnaires. Summative evaluation had only one sub-category: market feedback. Since formative evaluations appeared to be the more dominant category, there were many more types of formative evaluation methods used than summative.

Overall, this category helps to explain the two evaluation processes that these developers use in evaluating their CD-ROMs and also provides an in-depth look at the methods used within each category. To provide the reader of this study with insights into how each method is employed, actual quotes from the interviews are provided within the cases. These quotes reveal that software developers are conducting effective evaluations of CD-ROMs.

Category Two: Overall Views and Uses of Evaluations

Another major category that emerged from the interviews with the six participants was their overall views and uses of evaluations. It was through this category that a stronger sense of whether or not evaluations were helpful for improving their CD-ROMs became evident. In addition, some constraints that the developers faced in performing evaluations of their products were revealed.

A total of six sub-categories emerged from the six participants for their overall views and uses of evaluations: financial constraints for evaluations; evaluations results used to make changes on the CD-ROMs; new synergies created between divisions of

company for evaluation; how the development and evaluation processes differ between CD-ROMs and textbooks; overall, the view that evaluations were helpful for the product; and finally, some participants, in reviewing their evaluation processes, offered some suggestions for improving evaluations.

The next section of this chapter provides an in-depth look at the participants interviewed for this study through brief background information on each participant, as well as through write-ups of individual case studies and individual analyses of each case study.

The Cases

Case Study 1: Jackie

The Setting

Jackie has diverse roles as Director of Marketing, Project Manager, and Evaluator for a small educational software publishing company in Virginia. The company has less than ten full-time employees, is almost three years old, and produces two mathematical CD-ROMs on algebra and trigonometry which are used primarily by high school and college students. As described in one of their pamphlets, the company's mission is to "improve the learning process by creating innovative educational software programs." Jackie has been with the company since its inception and has witnessed its increasing growth. Her MBA degree in Marketing provided the necessary skills to lead several projects. Jackie was the Project Manager for both mathematical CD-ROMs and is currently working toward producing new versions of these discs. During our interview in her office, we discussed the development and evaluation of the only two CD-ROMs which the company produces.

The researcher first contacted Jackie by phone a week prior to our interview to arrange a time and place to meet. At that time, we also discussed her background and spoke a little about her company. On the day of the interview, we exchanged stories about graduating from business school and talked about our careers in business. After this introduction, the interview began.

Jackie had mentioned a few ideas about her company in our phone conversation that the researcher wanted to ask her more about. She had spoken of how the concepts of the two mathematical CD-ROMs produced and evaluated by her company started, even though a needs assessment was not actually performed, so the researcher asked her to clarify this:

> MB: Just as we spoke on the phone last week, tell me how your company goes about evaluating its CD-ROMs.
> JK: We were talking about what we did with our *Trigonometry* program--to evaluate it. The idea for the product of why people started developing it in the first place is basically that a programmer wanted to replace his linear algebra teacher with a computer program by taking advantage of all the things that technology can do within a software program that was really easy to use and could actually teach mathematical concepts. So that development pretty much started based on that idea. It wasn't really that we were doing any type of markets or needs assessment for the type of product that we were producing before...Basically, our product was trying to add value from an educational standpoint to teach mathematical concepts through the computer medium and they're designed to be a self-lead tutorial for students to sit on their own and basically just go through a lesson and learn the basics of whatever concepts that we were trying to teach.

During our phone conversation, Jackie had spoken of a concept that her company used called "conceptual learning." To better describe the idea of using a computer medium to interactively teach a concept, Jackie explained that her company uses this term as a means of validating the need for their products and adding value to them as well:

> MB: Can you tell me more about the concept of "conceptual learning" that your company uses.
> JK: It's really basically taking advantage of the computer medium. That's where we're adding the most value. It's more than just a textbook on a computer, it's really the type of things that you can have in terms of showing a sine wave or a sine function, and circular motion, and what a radian actually represents graphically through animated pictures of the math stuff involved. One of the other things that we do in our programs is we call "user manipulation" where students can actually reach in and interact with objects on the screen--they can drag rays to form angles, to understand what they represent. There's also one example of

> a...when we're teaching the Pythagorean Theorem we've got a right triangle and what you can do is basically click on either of the two acute angles and resize the triangle and when you do that you can actually see the different side lengths of C, A and B and then you also see that C^2 always equals $A^2 + B^2$. That's an example of what we're trying to encourage tactile and visual learning of mathematical concepts. The way math is traditionally taught is for an auditory person that can memorize a way to do something and just basically not...it doesn't matter if they understand what it represents if they can memorize it and they just regurgitate it out to do the same type of problem. Where it's just a teacher writing something on the chalkboard and then taking notes--a lot of people don't get it that way. And so that's what our programs are trying to do--to encourage students to take an active role in the learning process and to play around with those things so they really understand what they're doing as well as provide practice and testing which I think is definitely important to be able to validate that learning that has taken place--that they can answer questions.

The two major categories that emerged from this interview are: 1) types of evaluation performed (formative and summative) as well as the methods that were used for each type of evaluation; and 2) overall views and uses of evaluations. Each major category has sub-categories that will be discussed as well.

Types of Evaluation Performed

Formative Evaluation Methods Used

1. Product Testing

According to Jackie, a vital evaluation process used by her company was product (field) testing--specifically, alpha, beta, platform and cross-platform testing. In the earliest stages of their CD-ROMs, testing was a vital component of the evaluation process:

> JK: Basically what our testing process for *Trig* really was what our own people did internally and bringing students in and getting feedback on design and interface. But what we did for *Trigonometry* was that once we had a full product, we sent it out. And actually once we had a full product and sent it out all that we had done internally in terms of upgrades had been bug fixes.

To clarify her description of bug fixes, the researcher asked her to elaborate on that term. In doing so, she explained several other testing methods that she equates with evaluation during the development stages of the CD-ROMs, such as market feedback (including "table of content" feedback and prototype feedback), class testing, field testing, and bug testing:

> JK: We have called it [bug fixes] beta testing as far as when we send out a product that we know has bugs in it and we want people to knock it around to find out what is actually wrong--are there typos, a math statement that's wrong, if something doesn't work. That's what we called beta testing for our *Trig* program. With our *Algebra* program we are developing--'cos it hasn't been officially released yet--and what we have done to date in terms of market feedback have been, besides internal knocking around and testing, the "table of content" feedback and the prototype feedback.

Currently, Jackie is arranging for other schools to test her CD-ROMs so she can gain more feedback on her products. She calls this "class testing":

> JK: What we're setting up now is class testing for our *Introductory Algebra* program. We actually have four very large community college systems that have agreed to this. And the way we got that in the first place is because they were involved earlier on and saw what we were doing. And actually we're also just selling that at this time since we don't have a product yet--we're asking people...And that pretty much validates the need. Does anyone want to test our *Algebra* program once it's ready or test it out it to see how it works with your students in a math lab setting--to have little feedback forms for students to fill out. And, of course, the institution likes it because they're getting it for free.

She continued this response with a description of field testing and bug testing for the CD-ROMs:

> JK: But as far as the class testing and field testing, setting that up is just really...having..they're getting it for free for a while, to be able to test it and give us feedback on it.
>
> As far as bug testing, basically what we're doing with CD-ROMs is getting feedback on a lot of that stuff, basically...when we're comfortable internally and whomever we've let install and test our software, we basically press small quantities and send it out a bunch of press people or educators and say, "Take a

look at this new product--it's a demo. Knock it around and give us e-mail feedback...."
MB: Students and teachers.
JK: Yes.

MB: So you send out a prototype and try to solicit feedback.
JK: Also from an internal point of view we have Windows NT here, so we try to go find Windows 3.1 or '95 and go into a UVA lab and install it [laughs]. There's so many different systems out there--we need to test it out. We also went to Office Depot and installed it on all of the computers there! [laughs]
MB: Did they know you were doing it?
JK: No, we're just looking at the computers, pressing install and just seeing what happens. "Ok, it's working on all of the products--out the door." Circuit City had us a little worried because the sales people...
MB: They're very aggressive...
JK: Yeah [laughs], they're always around! Testing. We actually had a pretty common installation problem with our program--in order to install it, this one file has to be out of memory. So we said that some start-up program is using this program--this file--if it's in use then you can't install it. Stuff like that it only helps by getting more and more people using it and looking at it. The more people that do, especially if it's a huge bug and people paid for it, then we'd give them a free upgrade, and we turn it around.

From her responses above, it is apparent that Jackie uses many different types of testing for her CD-ROMs.

To verify that Jackie performs testing of her CD-ROMs, the researcher obtained three evaluation instruments from her--a feedback form and two surveys. She first provided an eleven page instrument titled "Introductory *Algebra* Feedback Form." Specific questions on the form pertaining to product testing included: "What is your opinion on the topics covered in the *Algebra* CD-ROM? Does it cover what would be taught in your introductory algebra course? What topics are missing, if any? What topics would you add? What is your opinion of the suggested order of topics?" Another five page survey titled "Survey for Task Force on *Algebra"* also asked for feedback on the

table of contents such as: "What other models do you suggest that we use in *Algebra* to teach any of the concepts listed in our table of contents? (e.g., spreadsheets, etc.)." These questions are good examples of how Jackie specifically obtained "table of contents" feedback. An example of a product testing question during beta testing also included on the feedback form was: "What is your opinion of the Lesson 4 prototype? Please be specific." The two forms were given to the instructors who had agreed to test the product and offer their feedback on the disc.

To corroborate cross-platform preference and testing issues, a one-page technical follow-up survey created by Jackie called "Survey" was also obtained. It asked teachers: "Given your background and knowledge of the hardware available in your schools, which of the following operating systems would you recommend that *Algebra* be available: Windows 3.1, Windows 95, Macintosh (68000 or PowerPC Processor), OS/2?" This survey was also given to instructors who had agreed to test the CD-ROM and was returned to Jackie once they had completed their testing procedures.

2. Partnerships Created With Educational Institutions

In analyzing the many formative evaluations procedures used by Jackie and her company, perhaps the most beneficial one was the establishment of partnerships with educational institutions throughout Virginia, particularly with instructors and students from these schools. These collaborative partnerships allowed Jackie to gather critical product feedback while simultaneously arousing user interest in the company's products:

> JK: What we actually did with our *Trig* program was we brought in some friends and home schoolers in the area, math teachers as well as trig students from both

> a local high school and a K-12 school and anybody that we pretty much knew who we could get to come in for a few three or four hour sessions in the evening. Basically what were trying to get from them was really an idea of how long it would take to get through an entire lesson, as well as user interface, primarily--that type of design of "do you know what do when you get through the program?" "What do these buttons represent to you--do you know what you're doing when you're going through it?" As well as some testing and evaluating of whether or not learning was taking place--of whether or not we were...Basically, our product was trying to add value from an educational standpoint to teach mathematical concepts through the computer medium and they're designed to be a self-lead tutorial for students to sit on their own and basically just go through a lesson and learn the basics of whatever concepts that we were trying to teach.
>
> Where we find that we're providing the most value to a lot of our customers is by really good, guided instruction that takes advantage of the computer medium. That really is a turning point for those instructors--they could just say, "Oh, buy this program." Or, "Sit in front of this program in the math lab." But we need an instructor buy-in to be able to show that learning has taken place. And because they are the one's who are involved in the purchasing process, from a marketing standpoint, from schools.

As a means of gathering feedback on the organization and content of the disc, Jackie sent out a table of contents for the Algebra CD-ROM to community college professors, thereby establishing partnerships with them:

> JK: What we're doing for our new *Algebra* program is we sent out tables of content to probably about 10-15 math instructors at the community college level. That was our target market that we were trying to definitely meet that need. And that we were trying to line up with whatever Introductory algebra course--Elementary Algebra course, Algebra One course--whatever they happened to call it in their school system. And so we were definitely trying to get their feedback--"When we're developing products, what should we put in it?" Then, the next stage after we got through that from the table of contents, was to send out a prototype which was more... which was not completely functional program wise, but gave them an idea of...as well as our product design and our product goals--there was a statement of goals in there and the survey that I gave you was what we wanted feedback from--from the teacher/instructor points-of-view. It wasn't necessarily the end user at that stage, when we were sending out the prototype, since it didn't have functionality it really was just looking at ideas for the user interface. There's a lot of...when I talk about user manipulation and interactivity that we're building in --there's so many different ways that anybody can go about designing it in the computer. So what we're really trying to do is mock something up quickly in a

> Director prototype and be able to send it off to people. What we also found was that some people--some of our instructors when they were testing the product--felt that they were testing it for, "Oh this doesn't work, this doesn't work. When you press this I put the answer as 4 and it said I was right but the answer was supposed to be 5." And we weren't looking for that type of feedback at that stage when we sent it out. We were looking at that stage for more overall pedagogical, "What do you think if this were more developed, would you be interested in purchasing this product?"

By getting professors to try out the CD-ROM, Jackie used this as a marketing ploy both to establishing closer partnerships with educational institutions and also to sell her product:

> MB: So you were looking more at the instructional design at this stage...
> JK: Right, exactly-- "Did it meet the goal that we were trying to do. How would you use it? What other software programs are you using?" It's not just...it's basically a marketing survey to find out validation for the need as well as getting people...what we really do is use this as a marketing ploy because if people are involved in giving you feedback at an early stage, they're the people that are going to be talking about it, they're the people that are most likely to buy it.

As can be seen from her quotes, it was through the partnerships with schools that Jackie strategically made her products more visible to consumers.

At the end of one feedback form Jackie asked the instructors: "Are you interested in continuing to review and to beta test our product?" to gain a better understanding of their level of commitment to the development and evaluation process. One other survey obtained by the researcher asked teachers: "Are you interested in continuing an advisory relationship with our company?" These serve as further evidence that she was very interested in continuing the partnerships created with educational institutions and their instructors.

3. Use of Target Audience (User) Tryouts

The review of literature section in this study discussed that some evaluations of CD-ROMs in print appeared to ignore their intended users (target audience) or not mention them in the evaluation process. This, however, was not the case with Jackie. She utilized both teachers and students in the evaluation process:

> JK: As far what we're really trying to do is make sure that the market is involved in our development process. And it really is pedagogically are we doing the right thing? Are we lining up with the table of contents for an algebra program--for an algebra course. Are there any more subjects that we need to add? What other features do they want? Do they need book markings, do they need a summary and a glossary. Do they need unlimited numbers of practice problems and testing problems? There's a lot of features that when we're just sitting in this building just trying to develop products we can all come up with different opinions of what's cool, and what's not cool, and what's needed, but really what we do for a product validation is really get our product in front of the market in various stages and listen to what they have to say in terms of what they would like to see in the type of product that we're producing.
>
> MB: Mostly you send out your prototypes to teachers who will be your "lead-ins" to students, then?
> JK: Yes. As far as after that once the product is completed and done for a version upgrade or for just establishing a presence, we set up field testing and getting a lot of students feedback as well as teacher feedback. And that ultimately is what is important. We're not really trying to say, ok this student did learn or didn't learn. We're trying to say how would a student use our product and what do they think about it.

4. Content Expert Review

Another essential product feature which was used in the development and evaluation of the CD-ROMs produced by Jackie's company was review of the disc's content by subject-matter (content) experts, notably math teachers (including retired high school teachers and professors) who are employees of the company. Jackie speaks of the content writers on the staff:

MB: Tell me more about the focus groups you conducted and the teachers you used.
JK: We also had several meetings with algebra teachers at a local high school as well as some people in the Math Education Department at the Curry School of Education. We had Sunday [review] sessions and they we're coming up with...say you need to do anything with a computer, like helping our content writers--we do have full-time content writers on staff (content coordinators), people that are math teachers that are writing this stuff and are responsible for writing. Basically, anything in the computer...what are effective ways to teach algebra to kids 'cos these people are in front of students all the time and they can think of ways to teach kids and articulate them. It's not so much what do you think of our program, it's paying for ideas of what to put in the first iteration of the program. And we definitely are involving people with that as well. Basically, it's just asking them the concept of how you would teach this subject on a computer--you have to be in a different state of mind--it's different than teaching it on a blackboard.

5. Focus Groups

The use of focus groups for evaluation purposes is not a new one. Jackie's company found the use of focus groups as an important and informal source for feedback on how their CD-ROMs could be improved. She found focus groups especially useful by utilizing teachers and by videotaping interested users at conferences as they explored the CD-ROMs:

JK: It's not just within community colleges, but locally with a high school and a military academy in Virginia--the trig teachers there have been involved in a lot of our focus groups. That's been a good way for us to be able get the type of feedback that we want in a way that's not as hard as filling out a 22 page survey (which we did get feedback was a little onerous).

MB: How many people were in these focus groups?
JK: We had a focus group at NCTM (National Council of Teachers of Mathematics) in April and it was 10 people and we paid them. This is what we paid them $100 each for participating--both in terms of doing a pre- and post-survey as well as talking to us for about 3 hours. We have it on videotape. We showed them a prototype, got their feedback. I think what we were doing by going to NCTM...and we hired a consultant to get us people that were big math educators...we ended up getting a lot of people that really wanted to change the way that math was being taught and not necessarily the people that would be

> interested in buying a tutorial program, because these are the one's who love being in front of a classroom and love being the one's that are actually teaching their students and coming up with ideas and activities, and things like that and the thought of how we would create a computer program that could take away a lot of that was a little bit threatening--but that was good to know from our point-of-view, in terms of any type of feedback that we got. I actually might be able to drag out a few of those survey's if you'd like.

Usually, the most helpful aspect of focus groups was their non-threatening, informal environment. The use of this approach often works well because users feel most comfortable, rather than constrained, and therefore are likely to give more valuable feedback. Jackie and her company make sound use of informal meetings and casual conversations with their users:

> JK: When we're comfortable internally and whomever we've let install and test our software, we basically press small quantities and send it out a bunch of press people or educators and say, "Take a look at this new product--it's a demo. Knock it around and give us e-mail feedback." We got a lot of general comments. It's always good--we're never going to stop saying, "What do you think about the program?" That always helps the sales call. To have people say, "Well, this, this, this didn't work" and hearing more about it other than just "I didn't like it"--asking them in a <u>non-leading</u> way to try to figure out.

In one follow-up survey from the focus groups, Jackie asked users: "What type of information would you liked tracked in the program?" This is a good example of the type of open-ended questions that were representative of the focus groups as a whole.

<u>6. Conferences and Seminars</u>

As explained in the previous example, Jackie's company uses conferences and seminars as a forum for gathering feedback from their target users. She states:

> JK: ...But, yeah, there are several meetings of...especially when Susan [President

> of the company] and I came back from our big conferences of showing our products to instructors in a booth, where we had traffic in and out the whole time, as well as whenever I go and talk to superintendents or people on visits. It's basically just an analysis of, "Yes, this is really important" based on ten or more people saying the same thing to you and it's relatively easy to do, then it's a no-brainer. Then you can actually go thorough it and make the changes. Actually, the hardest thing that we've been able to change just internally--I don't know if this is a factor of the people that are involved, or a lot of the "look and feel" and graphical elements--we had a big problem with the button design and whether or not anybody understood what they were and being able to put a translation on there and our Director was, "Oh, we need words on top of that--it's messing up everything." That's more of an art versus market feedback.

Jackie elaborated on how she attends conferences to better position her company, as well as to gain feedback and increase the product's visibility:

> JK: There's a lot of stuff going on at the academic level. We go to math conferences and seminars all the time. We look how to position ourselves--we're even positioning ourselves in Canada. As a young company you've got to be doing all of that--you've got to be training people, you've got to go out there and getting feedback. Even if it's just people looking at it. That's how you ultimately get sales. And we're definitely willing to be spending money to do that, 'cos that's the only way we're able to get our name out there.
>
> MB: That's the best form of advertising--word-of-mouth...
> JK: Right, definitely, definitely. And we definitely found that out when we went back to the mathematics and teacher conferences this year. We just came back from them in November and we saw people that saw the product when it was released last year and had been testing it giving us feedback telling us, "I can't wait for *Algebra*. When is it going to be done? I just can't wait for it!"

7. Use of Questionnaires and Surveys

As previously mentioned, Jackie and her company developed several evaluation instruments, three of which she provided to the researcher upon request. The instruments were used for product testing, "table of content" feedback, and for platform and cross-platform testings.

As with the use of focus groups for evaluations, using questionnaires and surveys

with their target audience was another constant for software publishing companies for their evaluations. Jackie explains:

> JK: We often ask our teachers, "Does anyone want to test our *Algebra* program once it's ready or test it out it to see how it works with your students in a math lab setting?"--to have little feedback forms for students to fill out.
>
> Once we get all these surveys, what we do is have our office manager type in everybody's responses into one word document so that we could see...depending upon when they come in, you kind of toss them around..."oh, this person hated it, this person loved it, this person was too hung up on the fact that it was a Director prototype." Once you see them all together you're able to say, "Oh, these are the competitors' products that they're using; this is what they really liked about it; this is what our strengths are." And this really helps the selling process.

Learning from past mistakes is commonplace in many evaluations. Jackie learned from some users that a 22-page survey she created was too long. This also helped her realize that focus groups might work better than long surveys:

> JK:...That's [focus groups] been a good way for us to be able get the type of feedback that we want in a way that's not as hard as filling out a 22 page survey (which we did get feedback was a little onerous).

Moreover, the use of too many surveys often complicates time and financial resources even more:

> JK: So, where we've been limiting the number of surveys we send out--actually even with unlimited money we wouldn't want to be managing...going through 500 customer surveys.

One example of a user interface question asked by Jackie on the *Algebra* feedback form during prototype testing was: "What, if anything, did you find confusing about the prototype? Did you find it easy or difficult to get around in the prototype?"

Summative Evaluation Methods Used

1. Market Feedback

Jackie and her company view summative evaluation as a vital means of obtaining market feedback. This sort of feedback is mostly used to make revisions in the next product upgrades (versions) of the same product. For instance, Jackie's company had released three versions of their *Trig* CD-ROM, each of which had changes made on it from post-release feedback:

> JK: Basically we've only done 3 iterations of *Trig* (versions 1.0, 1.05, 1.1). There's also some dissention as to what constitutes an entire digit upgrade. What we did from 1.0 to 1.1 was that we added some student user reactions--some things to make it easier for instructors. The stuff that we're doing is what we're able to teach, otherwise if you don't have any instructors guiding you, then you can just have it up on the screen and leave.

Perhaps the most vital summative feedback that Jackie received from the teachers who used the *Trig* and *Algebra* CD-ROMs was that the discs allowed a form of individualized instruction that was previously unheard of in the classroom. This allowed students and teachers to interact even more and offered additional, individualized feedback on math areas that were problematic with students:

> MB: I've read that CD-ROMs can enhance instruction, but that we should never replace the student-teacher interaction--never lose the human touch of teaching. Does that also come out in your evaluations?
> JK: Oh, definitely, definitely. The whole goal might have been, when the product first comes out, to replace the teacher, but you really do find the limitations of what you're able to do well. A lot of what's going on in mathematical thinking now, in terms of a lot standards, is to teach algebraic thinking and problem-solving and I think that's impossible to do in a computer program, especially without student-teacher interaction. Basically there's a lot of collaborative problems where students are thinking how they would approach a certain problem using mathematics even if they might not know what the mathematics are--proportional reasoning or something like that. Those type of big, complicated problems are really hard to do on a computer. Also, it's really hard to do on the computer where the student is really lost and they're forcing him to figure it out--and they're saying, "I don't even know where to start!" What you ultimately want the student

> to be able to do is to know where to start, even if they don't know when you've given them the problem, to come up with a bunch of different strategies or solutions on their own based on what they know, and not on what computers tell you to do. So, there's definitely limitations in terms of being able to satisfy a lot of the different standards.

Jackie explained how one of her competitors tried to advertise that their products could replace teachers. She mentioned how this aggravated many consumers and teachers. Also, she illustrated how her company's products were different by engaging the user more often:

> JK: Academic Systems, which is one of our biggest competitors, started out their whole marketing campaign saying that "we replace teachers" and their sell to administrators was "you don't have to pay an instructor--that everything can be done here."And that totally pissed everyone off.
>
> You can't do that. What we claim with our products is that we definitely enhance the instructor, we're not trying replace the interaction. But what we're trying to do is to have the student engaged and sitting at a computer and learning 80% of the material on their own. And for the more difficult 20% where you definitely need a human instructor...that the instructor is able to, in an ideal world, work with our product, be able to give individualized feedback to every student...and the teacher as more of the coach to be able to help the students with the harder stuff instead of the person conveying information more in the traditional classroom setup. That they're telling students what they know and students are writing it down and learning it passively. We're really trying to change the model with our program by having students to have an active role in participating in the learning process--by taking on responsibility for going through a tutorial that is supposed to teach stuff themselves, and not a lecture. And with some of the features that our program will be able to do in terms of having it be the basis of the course is to be able to give progress on how the student is going through the program back to the instructor. So the instructor can pull up you record and say, "Mark, you're having problems with multiplying negative numbers together. You don't get it, because you're getting all the questions wrong."

Jackie also used the market feedback and tracking which she received from post-release testing to customize her product for the user. This also demonstrates her use of summative evaluation procedures:

> JK: As far as feedback through the program and how an instructor will be using it

> and be able to individualize their instruction and coaching them whatever they say to a particular student by knowing that information which is what a computer program can track. It can track how long it took you to do this, what your scores on your test were, and compare it to everyone else's who used the program. And that level of sophistication and tracking is what we see our ideal product being able to do. Actually to date, what we've been able to develop being a small company is really the full tutorial aspect of it--the interactive, user-driven tutorial.

In the future, Jackie and her company envision a more comprehensive summative evaluation, once the company earns greater profits. This will include more partnerships with colleges and professors:

> MB: So what kinds of costs are involved for your evaluations? Is there some kind of pay involved for teachers?
> JK: Yes, basically but what we've budgeted in for when we get more finances is to pay for honorariums for professors. I know some professors at UMass that you may be able to talk to from an academic standpoint of evaluation. They're going to bring in graduate students to try out our software and give us feedback.

One other means of conducting summative evaluations for Jackie was through the assessment of whether the company was improving both internally and externally. Again, partnerships with educational institutions and getting public opinion on her company's side were helpful with this objective:

> JK: Basically, we do a lot of things internally about whether or not we're improving. Ultimately, there's a lot of people in academic school systems telling us that their pass rates and grades have improved this much by using this product. We're not really trying to do that. And I don't know even if we had a lot of money that we'd try to do that or sell that. If we were talking with somebody just like the UMass students who are willing to test the effectiveness of it and who are willing to write a dissertation about the effectiveness of it, then we wouldn't turn it down, but it's not something we try to pursue to say that your students will do better by using our products because the amount of value...we're trying to get public opinion on our side. To get that you have to get the teacher to say that this is educational value.

Jackie also evaluates customer satisfaction with the product once it is sold as part

of the summative evaluation process, but often discovers that not all the ideas she receives can be used:

> MB: What other aspects do you evaluate for?
> JK: Customer satisfaction--yes, although I don't think we have enough customers to say anything right now; so the people that say I love this...we have a little bulletin board of customer comments and e-mail, and all that great stuff and that's good being in a small business. Yeah, customer satisfaction is important, but I think that where we're falling short is not being able to do everything the customers ask.

Overall Views and Uses of Evaluations

There were six sub-categories of overall uses and views of evaluations which Jackie mentioned in her interview: financial constraints for evaluations; evaluation results used to make changes on CD-ROMs; how the development and evaluation processes differ between CD-ROMs and textbooks; new synergies created between divisions of company for evaluations; overall, that evaluations were helpful for the product; and made suggestions for improving evaluations.

1. Financial Constraints for Evaluations

Working for a small company has its drawbacks for Jackie, especially in her attempts to expand its evaluations. Smaller companies often encounter the financial constraints inherent with production, development, and evaluation. One money-saving idea that Jackie used as an incentive to get feedback from instructors was to offer them a free copy of the CD-ROM as a reward for their feedback on the disc. That responsiveness to their customers is one positive that helps separate her company from the larger ones:

JK: What we did find even in our process with getting table of contents feedback and prototype feedback is that a lot of people were requesting honorariums of $50-100. But we, being a small company, even for that level of feedback of only sending it to about 25-50 instructors, that we couldn't afford to do that at the stage where we were. So basically, we sent it out not saying we'll give you a $100 if you do this. We basically sent it out to 50 people that were people that we had good contact with that had liked our *Trig* program, or had liked what they had seen, or had mentioned to us somehow that they were interested in testing our *Algebra* program. We probably got about 10 out of the 50 call us back and say, "How much money would you give us for this?" And then people knowing that they weren't going to get paid for it, saying "Well I'll just kind of give your a brief overview of it, go through it, not spend too much time on it." Or we'd work something out with them to get a free copy of it when it's released. Being a small company we're unable to afford doing that.

But as far as the class testing and field testing, setting that up is just really...having..they're getting it for free for a while, to be able to test it and give us feedback on it. And what we've been able to show...I think a good advantage of how we stand out from larger companies is that we listen to that and are very responsive to what customers want and need. We might not actually have the resources to implement all their suggestions, but that's what they're getting--they're getting involved in creating a new multimedia program that teaches algebra in a way that would suit their needs. That, as well as it's free for however long it takes for them to get it on their network to get their student looking at it.

The decision whether or not to make changes to the CD-ROM from user feedback is often a compromise between financial resources and the strength of the feedback, according to Jackie. But, again, there are financial obstacles:

MB: What kind of implications are involved with your evaluations?
JK: As far as what implications are...if we had more resources we would definitely spend more money on getting market feedback and validation as well as...it's kind of interesting--because the more you do that, the more buzz you create about your own product, but the more demand you have that if you have the money to do that that you've got to be doing something with it--you can't just ignore all the feedback. Or if you ignore it, you have to have better reasons than you didn't have the resources to do it. And you definitely get a lot of feedback--some of strongly saying it's great, some saying it's not great. And really that type of analysis is...you can only meet a certain percentage of the market with one product anyway. And that's really the make-or-break feature that would make them not buy the product --there's plenty of other people that it doesn't matter, or that they're actually willing to buy the product anyway. People are willing to buy the product anyway--

> people are willing to buy the product from a little paragraph in *Newsweek*. The school sale is much longer. It involves a lot more...you have to do the field testing to get the sale. So, the more that we can afford to be able to do that type of field testing and setting it up, the more potential for sales and the easier marketing is for us.

Another way that Jackie and her company try to save money is by providing value through a product that is customized for each educational institution which uses it. This requires a great deal of field testing and teacher training which is often very costly, but essential for increasing the value of the CD-ROM:

> JK: In order to do the basics of a course, and in order to make money as a business trying to sell to that, it costs a lot of up front money. Field testing is almost required in order for that to happen. The sales cycle is probably two years from when they first see the product to when they would actually adopt it. And as far as the features of the demand, tons of customization, customization varies from institution to institution. So your development has to go around and change this, change that, just to get it to sell. It requires teacher training which obviously we're willing to do to show people how to use our products, but ultimately what we're trying to do is to provide a turn-key supplement or solution where we say put your student in front of it. We don't want to show you that is your lesson plan for the day--that these are the homework problems we assigned. The actual features and customization that's involved in order to be the basis of a course in an educational marketplace for schools, high schools (you need to be on a state list to begin with for state high schools)--that's just a huge debt. Academic Systems have been funded by Microsoft and TCI--they're probably spend about $4-7 million right now and basically they have a bunch of beta test sites. Yes, we can spend that much money, and the bet is obviously on computer classrooms, and everything is technology education, Clinton Goals 2000, whatever. That's going to happen, but from a small company standpoint it's a lot. We're trying to come up with solutions of yes, we're providing value here, how can we make money on it. And we're not ignoring the school market...it gives the consumer validity to know that this is being used at this community college; this is being used at the high school where I went.

2. Evaluation Results Used to Make Changes on CD-ROM

One practical approach that the researcher used to determine how valuable

evaluations were for software developers was by asking them to provide examples of changes they made on their CDs from the evaluations. Jackie gave several examples of this:

> MB: Yeah, I think it's important to take the time and label everything...it can get confusing...
> JK: Yeah, I agree with that. In an earlier version of our *Trig* program when you clicked into the introduction it said "Introduction" it was a blank screen. And one of the first things you had to do was to press the continue button, but if you didn't see anything how are you supposed to know you're supposed to know to hit the continue button? So, there's a lot of things like that that had to be adjusted which we learned from our evaluations, but rightly so. What we end up having was a pretty linear back-and-forth lesson, and you can go forward and backward. But it's not a lot of jumping around. It's not like...that's what we have in our *Trig* program. As well as our designed *Algebra* program--you're not jumping around from one place to the other.
>
> MB: And you have a navigator at the top...
> JK: Yes, just to continue and go backwards. But one of the things that we found out people like about our programs is that you can go to the main menu at any time and you can exit at any time. Anytime you go into something you don't want to go, you can't go down these major mazes and have to hyperlink on top of hyperlink in terms of "where am I in the program? I just want to get back to see what section I want to go into now." As far as user interfaces, that's been something that we don't want people to go back more than two steps away from where they know they can get back.

3. New Synergies Created Between Divisions of the Company for Evaluation

Jackie's company offered a good example of how different employees from various divisions of the company were utilized to assist with the evaluation process. Field test coordinators, programmers, office managers, and both marketing and development worked staff members worked together to evaluate her company's CD-ROMs, as she explains:

> MB: Which employees or divisions of your company conduct evaluations of your products?

JK: As far as what divisions of our company--we're a small company, so it's really a combination of management, marketing, and development that are leading these efforts--our field test coordinators getting feedback to see what's out there. It's really to ensure that our developers are getting all the feedback.

Once we get all these surveys, what we do is have our office manager type in everybody's responses into one word documents so that we could see...depending upon when they come in, you kind of toss them around..."oh, this person hated it, this person loved it, this person was too hung up on the fact that it was a Director prototype." Once you see them all together you're able to say, "Oh, these are the competitors' products that they're using; this is what they really liked about it; this is what our strengths are." And this really helps the selling process--to be able to say with conviction that our user interface is the best and simplest to use and people can sit down with the program and they know what to do with it--which is really hard--and which is not what other products can say. We know what our strengths are. To say, "I really like the graphics"--that's one of our big strengths.

MB: As far as development, you're the Director of Marketing...
JK: I'm the Director of Marketing currently. Our Director of Development right now is the person coordinating it...is also president of the company, who has the ten years of experience in the field. I also served as the Director of Development for *Trig*. So, as far as...right now the way the company is structured, both big people in management have a really solid understanding of what it takes both to market the product and to develop it.

MB: Would you have periodic meetings, you know, in terms of testing--with the President, the programmers, you? How do the programmers get the feedback from you to improve your products? Where there summary reports, for instance--how do you go about getting all that information to the programmers?
JK: Yes, what we did was we put together all of the surveys and basically it's a meeting where everybody takes a look at all of this stuff and gets to see it on their own. But that's definitely handed out--every single survey is handed out to every employee involved in development. And then it's pretty much management's decision.

Since Jackie's company had only a few employees, it is a testament to the company's strength and structure that it efficiently utilized each employee to assist with the evaluation of its CD-ROMs. Even though the divisions of marketing, upper management, programming, development, field testing, and others may have only consisted of one employee each, clearly they all worked together to develop and evaluate its products,

creating new synergies throughout the company.

4. How the Development and Evaluation Processes Differ Between CD-ROMs and Textbooks

In Chapter 2 of this study, the question of how evaluations of CD-ROMs differ from other forms of instructional media was discussed. Jackie realized the difference between CD-ROMs and textbooks as instructional tools. While most textbook publishers are venturing into electronic multimedia publishing with great aspirations, they often mistakenly treat the development and evaluation of these products like textbooks:

> JK: Basically, it's just asking them [teachers] the concept of how you would teach this subject on a computer--you have to be in a different state of mind--it's different than teaching it on a blackboard. I think that the best textbook authors obviously make the worst multimedia authors.
>
> MB: I've seen some textbooks that weren't as good as a CD-ROMs.
> JK: Yes, Harper Collins has a program called *Summit* which is an Algebra tutorial program and they... that's basically their textbook on the subject--it's the same thing and it's pretty clear that they don't know how to write multimedia content. They think, "Oh, this is something big that we should enter this market program."

5. Overall, Evaluations Helpful for Product

Another way of determining whether developers believed evaluations were useful for improving their CD-ROMs was by asking them this question outright. Jackie strongly believed that evaluations were helpful for her products, even though she hoped her company could do more. She still maintains that evaluation is an efficient way of establishing word-of-mouth advertising. In that belief, she equates market feedback with evaluation in two ways: one, as a determination of whether the product is meeting market

needs; and two, for establishing a market in the first place:

> MB: Overall, then, would you say that evaluations of your products have helped?
> JK: I definitely think that they help. I definitely think that p.r. in a small company or word-of-mouth is the only way to established yourself. As well as knowing who the right people are to evaluate your products. The head of Math Education--IBM Education--loves our program. People are coming to him about the IBM Algebra program that it's old--look at this one! Evaluation is extremely important both from, is our product meeting market needs as well as establishing a market presence in the first place. Everything that we're doing we couldn't have been making tremendous amounts of money without doing that first, especially that we're not a big company. Evaluation, in general, and validating what we do is very important for us--even paying others to look at it, but our funds are limited.

6. Made Suggestions for Improving Evaluations

By working for a small company, Jackie realized that the most beneficial way to improve her evaluations would be through obtaining more market and development feedback. One way this can be facilitated is through phone calls with customers:

> MB: There are a lot of issues involved on the marketing end--customers, feedback....
> JK: When we do hire a big marketing staff we can get a customer on the phone and it's easiest to start with, "What do you think about the program?" That's definite information that can help development--ok, we need this and that--we need Windows 3.1, they're not interested in Internet products. Just to get that information about things we really don't know --we don't have a good sense of that sometimes--of what can actually help the development process. The manager of the Marketing Department is just me, right now and I can think about what people need and formalize it. There are a lot of tele-marketers or sales force people--if you're actually talking to people as much as I can, I want to know about that--about what type of interaction you had.

Analysis of Case Study

From reading Jackie's interview it is evident that, as an individual trained in marketing, she views it as one of the most important aspects of evaluation and development; therefore the primary source evaluation for her is market feedback--

responses from actual targeted users such as instructors and students. Jackie utilized many forms of testing to gather market feedback. They included beta, alpha, platform and cross-platform testings (bug testings), prototype feedback, "table of content" feedback, and class testing. Using these various forms of testing suggests how important they are to Jackie for improving her products.

Most importantly, Jackie's marketing background supports her view that evaluation and development are contiguous processes, critical to the improvement of her products. She contends:

> JK: As far as what we're really trying to do is make sure that the market is involved in our development process. And it really is pedagogically are we doing the right thing?...But really what we do for a product validation is really get our product in front of the market in various stages and listen to what they have to say in terms of what they would like to see in the type of product that we're producing.

Jackie's responses fit into every category except for one: needs assessment. Working for a small company, that was surprising because they usually have to conduct a needs assessment to learn whether there is a market for their products or not. That is an integral means of saving money in smaller companies. However, the high demand for any math product that could help students having problems with math is always present. This validates the need for their products:

> JK: ...Actually what we're also providing is people that are fed up with 50% of their students failing Introductory Algebra at the college level. It's ridiculous that a successful department has a 50% failure rate--that students are just fed up and looking for other things to do. So, if you can put them in front of a math product and provide that, then that's our sale.

It is this high demand for math products which made a needs assessment unnecessary,

although it can be contended that Jackie and the president of the company must have conducted a needs assessment or market research of some nature to determine that the need for their products was already established. However, the researcher honored Jackie's original statements that she did not conduct a needs assessment--that is the primary reason why she was not placed in the needs assessment category in this study.

One instance in which Jackie's company might consider a needs assessment, however, would be if they ventured into new CD-ROM titles, other than math:

> JK: Efficacy of a market for our product--that is not really...except for the fact that we have only been able to generate little sales at this point. That market need [for math] is pretty much established and has been validated a lot by somebody in educational publishing who has been in the industry for 10 years [the President] so she knows the market well.
>
> MB: So you really don't do a needs assessment?
> JK: What we really would do with math--it's very straightforward that math education is important. Where we will do a lot of market research is deciding once you fill out the arithmetic curriculum--right now we have *Trig* and *Algebra*--once we do that we would definitely be doing a needs assessment for what product we might produce next. But that's based on what the market wants...it could be business or accounting...it depends what the competitive environment is there, where do we have synergy based on what we've already done and established--of math and physics, is it math and English as a Second Language 'cos those are the kind of customers at community colleges--people that take algebra may not know English well. There's just a lot of different things that we could go by to evaluate, but that's really where a lot of needs assessment would come in from our customers--trying to figure out what the next product line will be. It could be nursing or medicine. Wherever it makes the most sense for us to go.

It is also apparent that one of the most important aspects of Jackie's evaluations were partnerships with educational institutions, which she mentioned numerous times throughout the interview. Clearly, that was one way her company saved considerable money and also gained exposure for their products through word-of-mouth advertising.

The partnerships that she created with high schools, colleges, instructors, parents, and students were clearly beneficial. The company also plans to venture into the Internet to entice parents to buy their products:

> MB: The academic level, with students and teachers, is certainly what you're focusing on first...
> JK: Yes. What we're also trying to do is with our Java-based stuff is getting into parents places and sites, education sites where...a lot of that is growing on AOL, we're trying to work at it from both ends. Also, ultimately getting a high school teacher using it or saying, "It's a good program, and I evaluated it"--that's for parents of kids who don't do well in math--to go to the math instructor and say "What else can I buy for my child to help them?" That word-of-mouth from teachers to parents is what we're going for as well.
>
> MB: What about the Internet?
> JK: We're also considering Java based programs that we can put up over the Internet--that's where the potential is, especially for a small publisher like us--to make us compete in the marketplace. As far as getting rid of the middleman, as far being able to immediately access anyone (consumer). "My son is having problems with functions. Oh, there's a function lesson right there that I can buy for five dollars, right now"--Boom! Credit card transaction and the parent is helping their child with homework.
> MB: It's real-time.
> JK: Yep, it's real-time. The implications for that from a marketing standpoint are huge!

It became obvious from her responses that Jackie wished she could have completed more comprehensive evaluations of her products, but the financial limitations of working for a small company often hindered her efforts. She mentioned financial constraints many times throughout the interview, stating that evaluations were important, but costly:

> JK: Evaluation, in general, and validating what we do is very important for us--even paying others to look at it, but our funds are limited.

Currently, her company even lacks the money to update its web site, which could be

helpful for marketing and evaluation purposes, even though it still maintains a toll-free customer number:

> MB: Do you have an 800 number to get feedback?
> JK: Yes. And we have a Web Site, although it's pretty dated. When we have more money, we'll support it.

Case Study 2: Bob

The Setting

Bob is President of a small educational software publishing company in Virginia, which he helped to find nearly ten years ago. The company has less than seven full-time employees and has several hundred thousand dollars in sales. He has also been a visiting associate professor of business at a prominent graduate school in Virginia for over ten years. Bob's company produces seven software titles, all of which are introductory financial and accounting products primarily used by college students.

The researcher first contacted Bob by phone several weeks before our interview. Since it was near the end of the year and the beginning of the holiday season, he was very busy and had limited time to talk. We arranged to meet at the beginning of the next year for our interview.

Before the actual interview began, Bob and the researcher discussed several students whom we both knew--some of them were hired to be part-time programmers at his company. We also spoke about the business school for which he worked, the researcher explaining to him that he had completed an MBA degree several years ago. As the interview began, Bob first explained how his background as an accounting professor was instrumental in translating his expertise to develop courses in accounting onto computer software, using it as a responsive process:

> BH: This comes from years of teaching--I don't believe you learn anything unless its painful. You have to sweat it out and agonize a little, and when you emerge from that process, you'll have learned. Computer-based training is great, but I don't want to give the user a whole lot of input--I want them to figure a problem

out first, then I'll give you some feedback. It a responsive process.

Moreover, he elaborated on the fact that his company did not need to hire external individuals to review the content of the software produced, since he is a subject matter expert in the area of accounting. This saved them quite a bit of money each year. Although the company had traditionally produced their software onto floppy disks, it recognized the needs of its customers to produce new forms of multimedia, such as CD-ROMs. During the interview in his office, we discussed the development and evaluation of their most recent product in accounting, which is produced onto a CD-ROM. Bob explains his need to react to the market place for CD-ROMs and its inherent risks:

> MB: So you produce financial and accounting software?
> BH: Most of our stuff is sold to the education market--2 year and 4 year colleges. This type of software lends itself very well to CD-ROM--crunching numbers kind of thing. We sell to the educational market...We are totally reacting to the market place. Fortunately, the market place in educational publishing is way behind the real world market place. Students are up to the times, but most faculty are not. Originally we introduced our software for Windows and DOS, but we know that there's going to be a demand from a few customers for CD-ROM, so that's why we're venturing into this market now. We're trying to respond to that and put ourselves ahead of the market place. It takes a lot of development and a lot of money. CD-ROM is a big industry and it's incredibly risky because you spend a lot of money on development.

The two major categories that emerged from this interview are: 1) types of evaluation performed (formative and summative) as well as the methods that were used for each type of evaluation; and 2) overall views and uses of evaluations. Each major category has sub-categories that will be discussed as well.

Types of Evaluation Performed

Formative Evaluation Methods Used

1. Product Testing

The need for product testing, especially beta testing, was one which Bob knew was essential for product improvement:

> BH: In terms of what we do [for evaluations], we go though it ourselves--beta testing--checking for mistakes, glitches, spelling errors. Our material and text hasn't changed much from the Windows-based product. We went through the product very, very thoroughly ourselves--we passed it around and made the necessary changes.

Another important testing procedure for Bob was platform testing:

> BH: We also had platform testing across different types of machines, such as pentiums and other computers to see if it worked well. Most of it was done in-house here.

Often the small size of Bob's company limited his ability to perform comprehensive product testings, which will be explored more in the next category (overall views and uses of evaluations). As a result, they attempted to do most of it internally, but sometimes saw the need to test it with students:

> MB: Tell me more about your production evaluation methods (before your product was officially released).
> BH: Once we knew our authoring system worked, and we were satisfied with it, we would play around to see if the disc would work. We didn't change the design drastically from the Windows one. We knew the system worked, so we placed full faith in the authoring system.
>
> We let our programmers design a disk and then we would go through it and see if it would crash, if it was OK, if it could go back and forth, etc. If it did, then they had to make the necessary changes. Time and money were our main constraints--we couldn't go the full evaluation route in that sense, although we would have liked to. We also had platform testing across different types of machines, such as pentiums and other computers to see if it worked well. Most of it was done in-house here.
>
> We have another product for writing that we allowed graduate business

> students to go through it to test it--to find errors. It doesn't cost us anything--we just give it out and solicit feedback from the students--we don't have a lot of money.

All testing of products is not effective, however, as Bob discovered. He found that the testing process can cause negative reactions to the product. One way he discovered this was through tryouts of the CD-ROMs by graduate business students. He learned that they disliked being tested or graded:

> BH: They [graduate business students] took the final examination on the disc, but their grades weren't reported. We also sent it to 700 students at Harvard who had to take the final examination and record it on the disc. We got some feedback from about 400 students there about little nit-picky things, that were so little, that we made very few alterations and changes to our disc. Since they were being graded they were more selective in their feedback, saying that some questions weren't worded properly. Same responses from IBM employees who used it on a stand-alone basis--and then we heard all kinds of problems--text, software, etc., none which amounted to a hill of beans. But it was the testing that caused this reaction of negativism.

2. Partnerships With Educational Institutions

As Bob and his company saw the need to test his CD-ROM with students to gather feedback, partnerships with educational institutions became very important in the evaluation process. The partnerships that he created with both students and professors were the most efficient means for him to obtain evaluative feedback on his products:

> BH: We have a professor in Canada testing out our CD-ROM to make it more instructionally effective--like a classroom lecture.

It was the partnerships that Bob created with professors that were used as a marketing ploy to sell his products to students as well. Once the professors were sold on the product, Bob knew that the students would follow:

MB: So your main contacts for evaluations are professors?
BH: Yes. But we don't get user reactions necessarily. Our instructional designers are good enough to make the product user-friendly. If it does the job, then it's got a functional instructional design.

You have to sell the professors first, then the students will follow and adoptions from these schools as well. We maintain good relations with these professors and try to develop new ones as well.

3. Target Audience (User) Tryouts

Another important feature of evaluations of CD-ROMs was user tryouts with the target audience. Bob and his company believed that the most important individuals to try out the disc were the targeted audience--students and teachers:

BH: We're selling to 2 audiences: one, professors who have to evaluate the material before they recommend that the students buy it (will it fit into the course I'm teaching); two, the students who know more about computers than the actual professors--they're more "computer-geeky" than their peers.
We're lucky since our disc is used for courses before you enter graduate school, so when we get an order from a college or university, we get feedback from the students who use it through the professors or from the students themselves. At one graduate school of business here in Virginia, we tested it on about 300 students who thought the subject-matter was good and that was easy to use.

At other times, Bob may call one of his professional teaching peers to get feedback on the disc, but that is often uncommon:

MB: What other methods do you use in your evaluations to target your market audience?
BH: Well, I'll call an accounting professor who I know and who is well-respected in his field, and he may give me some feedback like, "You could use a summary section here," but that's very minimal. I only do that if necessary.

4. Content Expert Review

Since Bob is a content expert in the area of accounting and works for a prominent graduate business school, the content of the CD-ROM is verified by him and is also

supported by the university:

> BH: I'm a subject expert, so we didn't really need to hire a lot of outside people...We, on the other hand, do it [evaluations] very casually, but because I'm the subject matter expert when someone says something to me, it has a whole lot more meaning to me than them--it's direct information. Our content is never challenged--it's backed by me and the university I work for as a visiting professor --no one will challenge that.

He does, however, pre-test most of the product's content to ensure its accuracy:

> BH: We do pre-test our content and how it is presented in pre-test exercises and problems, though.

5. Conferences and Seminars

Although Bob does not formally attend conferences and seminars to solicit feedback from students and professors, he often receives casual, positive comments on the CD-ROM:

> BH: I talk to everyone at the conferences and get their reactions--mostly accounting professors. Most of it is very positive.

Perhaps it is also these positive comments that Bob receives which leads him to believe that changes to his discs are unnecessary.

Overall Views and Uses of Evaluations

There were four sub-categories of overall uses and views of evaluations which Bob mentioned in his interview: financial constraints for evaluations; evaluation results used to make changes on CD-ROMs; how the development and evaluation processes differ between CD-ROMs and textbooks; and overall, that evaluations were helpful for the

product.

1. Financial Constraints for Evaluations

After conducting the interview with Bob and analyzing it, it is apparent that the small size of his company constricted him from performing more detailed evaluations of his products. He explains:

> BH: We can't really sell the products for a whole lot of money, we don't have the sources or facilities to create super teaching materials, so whatever we do has to be self-contained--we don't have a lot of money.

The constrains of time and money, so prominent in multimedia and software development, were also problems for Bob during the evaluations he conducted, particularly during the summative evaluation stage:

> MB: Are you mass-producing your disk and getting more feedback (summative evaluation)?
> BH: We'll test it internally, but if we send it out by the hundreds, it'll cost too much and take too much time...
> Time and money were our main constraints--we couldn't go the full evaluation route in that sense, although we would have liked to.

In small software companies, the presidents must often make the most out of all their employees by giving them diverse responsibilities. Bob used marketing representatives, an office manager, and himself as the critical workers for the evaluation, development, and production stages:

> MB: What employees helped with the evaluation process ?
> BH: It was the office manager and 2 marketing rep's.
> MB: What did they do in terms of gathering feedback?
> BH: Test the disc and answer some questions wrong and right and see if that process worked. To see if all the links were there.

To avoid financial constraints of being a small company, Bob finds ways to save his company more money. One way, as mentioned previously, is that he is the content expert on staff, so there is no need to hire anyone else to verify the content of the CD-ROM. Another way is that he avoids focus groups and surveys, using casual conversations with professors instead as a form of feedback. He explains how his company is different from the larger publishers who spend more money:

> BH: The difference between us and a larger publisher is that we have subject experts on staff here (me)--they don't. They totally rely on focus groups in which they ask professors questions. They do surveys--that's their approach to this market. We, on the other hand, do it very casually, but because I'm the subject matter expert when someone says something to me, it has a whole lot more meaning to me than them--it's direct information. Our content is never challenged --it's backed by me and the university I work for--no one will challenge that. While it costs hundreds of thousands of dollars to the big companies, it costs us very little the information we get from casual conversation.

The decision whether developers will make changes to their products is often decided by weighing the importance of the changes gathered versus their costs. So far, Bob has not seen an overwhelming need or suggestion that has caused him to dramatically change any of his software (which, again, hints at the skepticism he has for the effectiveness of evaluations):

> BH: None of those suggestions so far have been strong enough for me to make changes. We already have a clear edge in the market--so these suggested changes have to be cost-justified, and they're usually not. $1700-1800 changes like these cost us too much and aren't going to improve the disc any. We weigh the changes against cost and decide if they're really necessary to spend the money. Most of them aren't!

2. Evaluation Results Used to Make Changes on CD-ROM

Bob did not offer many examples of changes he made to his CD-ROM because

there were so few changes actually made:

> BH: We used the same approach as our Windows product because we knew that it worked. But we wanted to slick it up a little bit to get out ahead of the competition, so we decided to video lecturettes for about 15 minutes each summarizing each of the chapters. Each lecturette is me talking, just like being in a classroom--very slick.

As a few students realized that they were being graded during their beta tryouts of the disc, some feedback was elicited which caused Bob to make a few minor changes to the disc:

> BH: We also sent it to 700 students at Harvard who had to take the final examination and record it on the disc. We got some feedback from about 400 students there about little nit-picky things, that were so little, that we made very few alterations and changes to our disc.

3. How the Development and Evaluation Processes Differ Between CD-ROMs and Textbooks

Bob added a few responses supporting the view that the development and evaluation of CD-ROMs often differ from others, particularly those of textbooks, for several reasons. One way in which they are different is that textbook publishers rarely ask users for reactions about the textbook and often do not have content experts on staff. Instead, most textbook publishers, including those who venture into electronic publishing, use marketing representatives, who are often unable to explain to the user how to work the CD-ROMs they market. Moreover, unlike CD-ROMs, most textbooks require tutorials and have large printing costs which are higher than CD-ROMs. Bob best illustrates this by explaining the differences between a small company like his versus a larger publisher:

> BH: This is what separates us from the large text book publishers--if a professor is writing a text book for publisher, he receives no money up front usually--he'll only receive a royalty based on sales. The same thing if he takes that text book and makes it into a CD-ROM. It costs the big publishing company nothing up front, but it does cost a lot in marketing, printing, and royalty costs after the book is done. Then to develop a tutorial to accompany the text costs $150,000! Some of these texts don't translate well onto CD-ROM and often are embarrassing--why, because that was developed by a professor and it's free. These companies never ask questions about the software in their focus groups or surveys--they know it's horrible! They all admit to that, but they don't tell their customers that. They ask in focus groups or surveys what do you want in an accounting text, but they never ask reactions about the actual software. Their representatives aren't content experts either--they're just sales people--they don't even know how some of these discs work. We don't have to worry about that.

4. Overall, Evaluations Helpful for Product

This was one of the most difficult categories to place Bob in based on his responses in the interview. In fact, the researcher placed him in this category with some reservations. Although he does not wholeheartedly believe in the evaluation process, there were comments that Bob made which suggest that evaluations were helpful in some manner in improving his products:

> BH: We're lucky since our disc is used for courses before you enter graduate school, so when we get an order from a college or university, we get feedback from the students who use it through the professors or from the students themselves.

It may be inferred from his responses that the ultimate success of his products can be attributed in some fashion to the evaluation processes he used, such as testing and feedback from students who tried out the disc. Although Bob credits his company's success to its unique market niche, much of it could not have been achieved without evaluation of its products:

> BH: We have [accomplished] six figures in sales and 7 titles. We still get contacted from other publishers who want to put our name on their products and we may do that in the future. We have a very unique market niche. We target the largest and best business schools.

While it is apparent that Bob does not regard evaluation with high importance, it still appears that, overall, it was beneficial for him.

Analysis of Case Study

Bob's interview was very revealing. Although he conducts evaluations of his CD-ROMs, it appears that this process is treated more as an after-thought or one that, although essential for meeting market needs, is often costly, time-consuming and burdensome. Bob's overall view of evaluation, then, was not a completely positive one, unlike the other developers interviewed for this study. His adhering belief that his company was so far ahead of the market with its products suggests that evaluations were not so critical to the development and production of his CD-ROM. This belief also limited his reliance on a needs assessment for his product, since he contended that there was a proven need for it already:

> BH: But we're so out ahead of the market that we don't do so much in term of evaluations--it's mostly just in-house.
> So when we got ready to do our CD-ROM, we didn't have to test our market--we know there's a market for our product already and that we're the leader in that market (so a needs assessment wasn't needed). We wanted to keep the same design as our Windows product (i.e., there will be text and pre-and post-testing to assess what they read and learned -- a series of exercises and problems interactively, and then there will be a final examination). We used the same approach as our Windows product because we knew that it worked.
> ...98% of our audience doesn't use multimedia, and the 2% that does, we're forced to satisfy their needs.

It was evident that the changes made to the CD-ROM brought out in the evaluations were ones that Bob did grudgingly. This, again, may stem from his belief that evaluations, although helpful, were not an absolute necessity for his software (which may also account for why he does not perform summative evaluations of his products). Another acceptable theory may be that this was a result of the company's need to react to the market which dictated that he make the necessary changes:

> BH: We do these changes grudgingly, but it needs to be done--we have to react to the market.

One other reason that Bob avoided summative evaluation testing was because he found that glitches in his products were often caused by user faults or misunderstandings, rather than by a programming error in his company:

> MB: Are you mass-producing your disk and getting more feedback (summative evaluation)?
> BH: We often find that if it [CD-ROM] doesn't work, then it's because of the user's stupidity--they don't follow instructions--or the machine screwed up, something like that. At first we were scared that it wasn't working, but we found it was not fault of our's. We haven't had to make a lot of changes. We're dealing with a topic that has an extremely long shelf-life (accounting)--not much has changed in that field for a long time. So we don't necessarily upgrade the subject matter.

Moreover, the type of negative reactions he received when testing the disc obviously distorted his view of evaluation and testing:

> BH: ..Then we heard all kinds of problems--text, software, etc., none which amounted to a hill of beans. But it was the testing that caused this reaction of negativism.

Another important view that Bob held was equating marketing feedback and research with evaluation. He solicits this marketing feedback through several means as a

critical feature of his evaluation process: phone calls, informal meetings, and conferences, as he reveals:

> MB: Tell me more about how you gather these reactions?
> BH: 3 ways: One, we have two marketing people who solicit information through phone calls or who go out on trips to potential customers; and three we'll attend conferences and have a booth, people come by and look at our stuff--that's were you get that market research or evaluation feedback.

In analyzing the entire interview, Bob does not place a great deal of importance on evaluation, although it can be inferred that his use of certain evaluation processes, such as testing and user tryouts with students, can be regarded as vital evaluation processes that he used in producing the accounting CD-ROM his company markets.

Overall, then, some part of the company's success can be attributed to its evaluations, but perhaps the constraints of being a small company outweighed his desire to perform a more comprehensive evaluation. One quote, in particular, led to this inference:

> BH: Time and money were our main constraints--we couldn't go the full evaluation route in that sense, although we would have liked to.

Case Study 3: John

The Setting

John is a software developer and instructional designer at a state university medical school in Virginia. He has worked with various software publishers in developing and evaluating instructional CD-ROMs used for medical students. Most recently, he worked with a large publishing company in producing and evaluating an anatomy CD-ROM created for first year medical students and other professional students interested in studying anatomy. With 15 years of experience as an instructional designer and software developer and advanced educational degrees in these areas, his skills in developing interactive CD-ROMs uniquely qualified him to produce such a disc.

Before our interview began, we discussed the field of instructional technology, since we both had training in this area. After we exchanged a few more stories, our interview began. During the beginning of interview in his office, he explained how the concept of the disc began out of dissatisfaction for the ways in which anatomy was taught in a lecture and lab environment:

> MB: Let's talk about the CD-ROM that you created first.
> JJ: The idea was conceived of by two faculty members here to teach dissection on an interactive medium. While the students are doing dissection of cadavers, at the same time they're also studying radiology and learning to recognize anatomical structures and radiographs with a variety of imaging techniques. The purpose is to train them to give them a basic understanding of anatomical structures in X-rays, MRIs and CTs. This was formerly taught in the lecture hall with the use of slides and also in the laboratory during the course of dissections with X-rays boxes. This is ideal for a small group, but not for a large one, so there was dissatisfaction with the teaching methods and a very uneven teaching experience (not everyone got the same thing)--because there's a lot of different tutors in the lab and everybody has varying levels of expertise. Nearly all of them are not trained as radiologists. The

> idea was conceived to have all of these radiology films available through the computer with a set of interactive exercises, so that the students could study this on their own and it would be a very consistent exposure to the materials.

What John envisioned for his CD-ROM was to work with the publisher of the textbook in creating the CD-ROM:

> JJ: The other goal was to work with the publishers in matching the CD-ROM up closely with the textbook and the matching lab manual. Our goal from the beginning was to make this a very tightly integrated product that would eventually be marketed along with the textbook, so we planned from the beginning to work with the publisher to sell the product, through the electronic publishing division.

The two major categories that emerged from this interview are: 1) types of evaluation performed (formative and summative) as well as the methods that were used for each type of evaluation; and 2) overall views and uses of evaluations. Each major category has sub-categories that will be discussed as well.

Types of Evaluation Performed

Although John does not like to use the terms formative and summative evaluation, the methods he used in evaluating the CD-ROM will be separated into formative and summative stages to better understand their importance in study. John does not use the terms "formative" and "summative" because he believes that development and evaluation are contiguous and iterative processes which are constantly on-going throughout the life of the product:

> MB: Tell me about some of the evaluation processes that you went through with this project.
> JJ: The evaluation is part of the development process. Whenever you look at a textbook in instructional design, evaluation is always separated out as a separate loop--first you develop, then you evaluate, then you go back and re-develop and

evaluate again. That's fine for a textbook presentation but in reality, development and evaluation are a contiguous and iterative process. So the 2 phases were contiguous--completely overlapping.

MB: Any other evaluation stages that you go through? Tell me about the formative evaluation processes.
JJ: Here again, I don't use the terms formative and summative because evaluation is on-going and will continue to be on-going as long as the product is alive. So I don't make a distinction between formative and summative because a project like this is never complete. Software is never done unless you pull it off the market.

Formative Evaluation Methods Used

1. Product Testing

John shared his views on the importance of product testing for the CD-ROM he developed, specifically through the debugging phase. It was during this formative stage where he designed a comment section into the disc where users could type in their comments and feedback on the product as they were using it:

MB: Tell me about some of the debugging you went through.
JJ: For instance, a button wouldn't line up correctly over the structure or an arrow didn't point right to the area. It tends to be very X-Y coordinate oriented--where something is exactly on the screen. And I use the same comment approach to that --I give the CD to the two doctors and they run through it use the same comment buttons to leave their own comments. The same mechanism that is used for comments from the students is used by the content experts themselves.
MB: And you print out these files to use for yourself.
JJ: Yes, it's just a text file.

2. Partnerships with Educational Institutions

As an instructional designer and developer who works for a medical school, establishing partnerships within the school for evaluation purposes was very easy and useful for John. He used both doctors and students at the school to help him:

> JJ: We [the doctors and him] worked to develop each section incrementally, according to the regions of the body. When we developed a section we would then let the students use that for a semester.

During the summer breaks, John used faculty members and medical students between their first and second years of medical school, who had also just completed the anatomy course, to program the disc. Like the doctors themselves, the students became content experts in the area because they had finished the course:

> JJ: Another thing in terms of evaluation of this we used, one of the goals of the project was that we would have students and faculty members do all the programming. So one of the evaluation elements was trying to make the authoring process as simple as possible. So I developed a series of templates for different types of interactions and developed a method where those templates were very easily dropped into the existing programs. This took quite a bit of honing and work with the students who worked over the summer. I would then clean up their content files after the students completed them. It was definitely a lengthy evaluation process to learn how to best facilitate that so that we didn't have to do all the programming in my office.
> MB: How did you facilitate that?
> JJ: I would sit down with the students working on the project and give them something to work with and they would tell me how it worked out. Then they would be back on the phone to me for anything that didn't work. So the evaluation process really involved the number of phone calls that I got back on any given week. I knew when the phone call volume dropped off that the process was working well. I'd also drop in once a week to make sure that they weren't making any major errors and check their work.
>
> MB: Tell me about some of your other workers on this project. You mentioned you had student workers.
> JJ: The people that had the most input into it in terms of evaluation have been the students who work on it each summer. These are medical students between their first and second years and they're paid for by funds from the Radiology Department. These are the people who have become most intimate with the program and they will typically have a lot of comments on how things could work better, and the content. During that summer they spend a lot more time on the project than I do.
> MB: Sounds like they become the content experts.
> JJ: Typically, they almost become the content experts themselves because they've taken the course already and have a lot of work with the project.

3. Target Audience (User) Tryouts

As mentioned in the last sub-category, the use of students as both content experts to program the disc as well as users during tryouts was an efficient means by which John used the target market of the disc in two ways--to get feedback on the CD-ROM and also to ensure its accuracy:

> JJ: We showed it [CD-ROM] to a small group of about 3-4 students and we would get feedback, and then I would work on the interface some more, develop some additional features. That was how the initial concept was honed and continued throughout the entire project...When we developed a section we would then let the students use that for a semester. As part of the evaluation process we used focus groups of about 3-4 students who would actually use the product and then we would sit down and have discussions with them about how it worked, what were the problems and so forth.

Near the close of the interview, the researcher reviewed with John how he utilized his target audience during user tryouts of the disc:

> MB: Before we end, let's just summarize what you do with your evaluations of your CD-ROM, just to make sure everything is correct. So you basically had med students come in work on some of the content--that was one evaluation process. Another evaluation process was testing it out with students--testing for content, user-friendliness, what they liked, test it for bugs. Any others--you said you worked on it as well.
> JJ: Yes, and we did some focus groups.

4. Content Expert Review

The review of the accuracy of the disc by medical doctors was critical during its earliest stages and was conducted both by John and the publishing company he worked with:

> JJ: In the very first phase, in terms of development of the concept of how the product would work, I sat down with the 2 content experts (a Ph.D. and a medical doctor) and developed working prototypes, would show them that and then they would make comments on it and I would make revisions.

During the early stages of development, however, John found the publisher was not very helpful in evaluating the disc, which left him alone with the doctors at his school to evaluate the content of the disc:

> MB: What was your work like with the publisher?
> JJ: They haven't been very helpful, to tell you the truth. They've given us practically no input on this project. We've sent them preliminary versions of the projects and we have not heard anything back from them. There are some administrative changes/problems in the company which have stalled us. For reviewing it, they sent it out to a peer review committee of about 5 people (content experts), but we've gotten no formative feedback so far.

> MB: Tell me more about the evaluation processes of the publisher.
> JJ: They've reviewed a preliminary version of the CD-ROM. They sent it out to 5 peers, who would install it on their machines and make suggestions and revisions, like an editorial board. They're content experts as well (medical doctors). Peer review is the main evaluation process that they use.

As medical students programmed the disc during the summer session, John made sure that the professors (a medical doctor and a Ph.D. in the subject area) reviewed all their work, especially the content of the disc. This also helped to strengthen the partnerships he created with his school:

> MB: Tell me about the product cycle--a time frame.
> JJ: During the summer of each year we hire 1 or 2 new medical students to add additional content. That also gives us insights into the process, since the students who are hired have just finished the course, have used the CD-ROM, know what's on the disk, how their peers tend to study. Basically the students write up a lot of the practice questions and the faculty members choose which images are to be used and will write up the content (didactic material) for each of the images. These questions are reviewed by the faculty members.

Content expert review of a medical CD-ROM can be more comprehensive than other types of discs due to the nature of the topic and the impending threat of lawsuits if

the content is inaccurate. For this reason, the accuracy of the medical information on the CD-ROM was continually reviewed by several individuals on the project team: the designers, John, and the doctors. The disc, then, goes through three extensive reviews in the formative stage: content perspective, usability, and technical debugging, as John explains:

> MB: And you have some students who are designers...
> JJ: Yes, I typically have 2-3 graduate students who work in my office at any given time and I can think of at least 3 graduate-level instructional designers who've helped me on the project. In one of the evaluation phases, once the med student has finished putting in all the content for a given section, the material is given back to me and I will go through every single interaction, read everything that was written--every possible question, and so will my grad students. They'll look at it from a readability, understandability viewpoint and go back and do a lot of editing on the wording. We don't intend to change the content, but it's also reviewed by the two doctors who are experts in the field and they have their own ideas on how this could be better presented. This is the stage where we pick up the comments, "Hey, this question is too hard for a student at this level of education." So it goes through four reviews at that point in time: content perspective, and a grammatical, readability perspective, as well a technical debugging phase.

5. Focus Groups

The use of focus groups for evaluation was one of the most important approaches that John used in the formative stages of the disc. It was the feedback elicited during these focus groups that caused the most changes to the disc.

John combined two evaluation methods during focus groups which worked very well: he had user tryouts and then interviewed the users in focus groups once the tryouts had ended to gather their feedback. It was also during the focus groups that John decided to add an important evaluation feature of the disc--a built-in comment section where users could provide their own feedback. Once a focus group and user tryout ended he collected

all the comments in the form of a comment log. The comment logs were especially useful for reviewing the content of the CD-ROM:

> JJ: We worked to develop each section incrementally, according to the regions of the body. When we developed a section we would then let the students use that for a semester. As part of the evaluation process we used focus groups of about 3-4 students who would actually use the product and then we would sit down and have discussions with them about how it worked, what were the problems and so forth. Also, one of the features built into the program is a comments button, which, anywhere in the program, a student can click on that button and a window pops up and allows them to write whatever comment they want to leave--it could be about a feature, the content, a mislabelling or misleading questions, anything. We had comments on a wide variety of things. But the program also has a time/date stamp on it and also tells us exactly where the student was within the content, so we know exactly where they were in the program, even what question they were on. So every semester of use, we collect this log and I give the log back to the content experts, because 95% of all the comments are content-oriented, and they review it and make sure that all the comments made were authentic, because a lot of it boils down to misunderstanding the content. But if they're consistently misunderstanding a concept, then maybe it should be explained more thoroughly elsewhere. So that again is the concept of the iterative evaluation of the program.

Not surprisingly, then, John found the comment logs more useful than the focus groups themselves because they provided more direct and timely feedback during user tryouts:

> JJ: Actually, we found these comment logs being a more fruitful form of evaluation than the focus groups because you're getting comments right when people are using it not after the fact sort of reflective thing, so you're catching a lot more details.

In the earlier stages of the disc, user tryouts with the medical students in a "think-aloud" session were often very casual and provided valuable feedback once the focus groups began. This evaluation process also helped improve the interface of the CD-ROM, leaving John to change only the content:

> MB: Tell me about the testing procedures you did for the CD-ROM.
> JJ: In the very earliest stages whenever I created a working prototype I would give

it to a student, on an individual basis and then follow up with him in the focus group, and then just observe them when they used the software and asked them to think aloud. I promised myself not to say anything during the actual tryout. Afterwards I would have a focus group discussion with 3-4 students who had gone through that process. The interface has been fairly set for nearly a year-and-a-half, we're just changing the content now.

MB: How did you record the comments from these students in the evaluations?
JJ: I just made notes. Just jotted them down...I probably used 3 focus groups during the interface design stage. A lot of times I would just informally grab a student and say, "Hey, do you have 10 minutes? Sit down here and tell me what you think of this and lead them through it." Just students who came in and asked me questions.
MB: Tell me about some of the comments you got at that stage.
JJ: The focus at that time was on interface, not content. What I was soliciting were comments on interface issues. "Do you know what to do now? Is it obvious what the next step is? If you had to get back to another image you saw 10 minutes ago, would you know how to do that?" Ease of navigation, that sort of thing.

Overall Views and Uses of Evaluations

John mentioned four sub-categories of overall views and uses of evaluations in his interview: evaluation results used to make changes on CD-ROMs; how the development and evaluation processes differ between CD-ROMs and textbooks; overall, that evaluations were helpful for product improvement; and made suggestions for improving evaluations.

1. Evaluation Results Used to Make Changes on the CD-ROM

John offered many examples of changes he made on the CD-ROM as a result of the evaluation approaches he used. Many of the suggested changes from the user tryouts of the disc were interface changes:

> JJ: There's also been some interface changes as a result of the evaluations as well. Like for instance how the menus work--there was some confusion about how the menu structure worked, so that was modified according to input from the comment logs...One of the goals of this was interface simplicity--we have no instructional manual for it, but we do have on-line documentation with a help button they can click on anywhere in the program to help. And we found out that works well.

Another change made to the CD-ROM suggested in the evaluation was that students wanted a place to check off where they left the program so that they did not have to repeat any part which they had already finished:

> MB: So tell me what changes you made on the CD-ROM from those comments.
> JJ: One of the interface changes was to put a check mark by everything that they completed so they would remember what images they had worked on within a given session. The CD does remember what they've done in any given session.

One other comment from the medical students was to eliminate any computer sounds that told them if they had answered a question wrong. John eventually eliminated those sounds which the users found embarrassing:

> JJ: Another thing we discovered through the formative stages is that med students don't like sounds in the form of feedback. Typically, this software is used in group lab settings and they didn't want embarrassing sounds telling them they were wrong.

2. How the Development and Evaluation Processes Differ Between CD-ROMs and Textbooks

During our interview, John frequently expressed dissatisfaction with his publisher. One explanation of the discontent he holds for textbook publishers arises from their belief that sales are their most vital evaluation tool. John certainly disagrees with this view, as he describes how some publishers are in a state of turmoil:

> JJ: That's what drives the publishing business--sales. It's extremely important to them--it's probably their most important evaluation tool as well.
> MB: But for you would you say it's more important for interactions...
> JJ: Well, if you want to use the terms formative and summative evaluations, the summative evaluations are the sales--that tells you whether the product actually met the needs of the users. Of course, there are a lot of other things factored in--the marketing processes, were the clients informed--you get a lot of mixed information with sales figures because there's a lot more than just the quality of the product...Most of them [publishers] are in a state of disarray.

Moreover, John's publisher was not very helpful with the evaluation and development of his disc. In fact, they often causing deadlocks due to frequent disruptions in the company. He continued his response stating how textbook evaluations are limited and therefore would not work for CD-ROMs:

> JJ: They [the publishers] haven't been very helpful, to tell you the truth. They've given us practically no input on this project. We've sent them preliminary versions of the projects and we have not heard anything back from them. There are some administrative changes/problems in the company which have stalled us...we've gotten no formative feedback so far [from them]...Peer review is also the other evaluation process they use...but that's the only evaluations that they use--sales and peer review. That might work fine for a textbook, but not for a CD-ROM.

3. Overall, Evaluations Helpful for Product

John concluded that evaluations were helpful for improving his CD-ROM. This response was based on his belief that evaluation and development are a contiguous process:

> MB: Overall, if you had to look at your evaluation processes, have they helped you with the disc?
> JJ: Oh, yes. A I mentioned before, I really don't see evaluation a separate component--it is simply part of the development cycle. I cannot conceive of developing something without a constant flow of information from users and from other content experts. It's just a natural part of the process. I couldn't work without it.

4. Made Suggestions for Improving Evaluations

At the close of the interview, it became clear that John had definitive views of how evaluations processes could be improved. He described these improvements by explaining the mistakes made by some of his peers in the industry, especially when they ignored their target audience:

> MB: So you're using a lot of educational evaluation techniques that you've learned from your background in instructional technology and education.
> JJ: My background is in education. Software development evaluation is...I've never gained a lot from that methodology. They don't tend to work a lot with their own users...
> MB: They really don't work with their target audiences...
> JJ: Yes. They tend to think more in technical terms than in humanistic terms. Basically, taking educational design theory and matching it with real-life situations --Dick and Carey's book is pretty simple to use, but there's not a model out there that actually shows the processes that anybody uses. The processes in the textbook are not what's used--
> MB: The evaluation processes and the design processes--
> JJ: Yes, both.

Moreover, many of the evaluation models that currently exist, as will be examined in the final chapter of this study, are often insufficient to evaluate CD-ROMs and need to be modified to be helpful:

> MB: That's why I'm conducting this study-- to put this information out there. It's amazing that there aren't many standard models out there to use for evaluations...
> JJ: Well, there are models out there, but once you get out there and start working you find you have to modify them radically according to your situation.

John views the primary cause of this dilemma as the field of instructional design being in such a sad state, producing improperly trained designers, and companies that too often rely on sales as their primary means of evaluation:

JJ: The field of instructional design is in a pretty sad state, to tell you the truth. The research that is done in the field has practically no relationship to what's done in the field. It's interesting, I've read an article that asked if we're dictating ourselves out of existence. If this isn't a field of applied research, everyone will ignore it--and I think that's what happened. Most of the people doing multimedia development outside of universities working for small companies--I don't think they're coming out of instructional design programs, they're coming out of communications and computer science programs. Designers trained in art schools are also becoming useful.
MB: And I think this defeats the purpose of multimedia--the interactivity.
JJ: Yes, at least educationally. And those people designing interactive games, they're absolutely in the dark. There's no precedence of any field other than kinetic game theory. Those people are just groping around in the dark. Their evaluation methods are how many sales they make.
MB: I guess that's how publishing companies get feedback--in term of marketing and sales.
JJ: That's what drives the publishing business--sales. It's extremely important to them--it's probably their most important evaluation tool as well.
MB: But for you would you say it's more important for interactions...
JJ: Well, if you want to use the terms formative and summative evaluations, the summative evaluations are the sales--that tells you whether the product actually met the needs of the users. Of course, there are a lot of other things factored in--the marketing processes, were the clients informed--you get a lot of mixed information with sales figures because there's a lot more than just the quality of the product...Most of them [publishers] are in a state of disarray.

Analysis of Case Study

John's background in education evidently influences his opinion of the importance of evaluation in the production and development process. He considers these two processes to be contiguous and iterative, rather than separate. His 15 years of experience as a developer, along with his critical views of the field of instructional design, shows that he believes the field can be improved through increased humanistic training and by using their target audience during the evaluation process:

JJ: My background is in education. Software development evaluation is...I've never gained a lot from that methodology. They don't tend to work a lot with their own users...

MB: They really don't work with their target audiences...
JJ: Yes. They tend to think more in technical terms than in humanistic terms. Basically, taking educational design theory and matching it with real-life situations.

In the interview John clearly sees the role of focus groups, targeted user tryouts, the comment logs, and content expert review as vital components of evaluation, all which he mentioned several times. Since he works in a university, John wisely took advantage of his surrounding environment by using both students and doctors in his evaluation processes:

MB: Before we end, let's just summarize what you do with your evaluations of your CD-ROM, just to make sure everything is correct. So you basically had med students come in work on some of the content--that was one evaluation process. Another evaluation process was testing it out with students--testing for content, user-friendliness, what they liked, test it for bugs. Any others--you said you worked on it as well.
JJ: Yes, and we did some focus groups.
MB: And in those focus groups you had students who had already tested the CD-ROM so they could give you feedback?
JJ: There were different focus groups typically because we would get an evaluation during each year, during each stage of development. Once we basically settled on the interface, the focus was more on cleaning up the content.
MB: And that all came out in the evaluation process?
JJ: Right.

At the end of the interview John clearly stated his discontent with the publisher he was working with. This arises from his view that textbook publishers are usually in a state of disarray, unwilling to modify their royalty methods, and only use sales figures and peer review as their main evaluation approaches:

MB: Will you work with the publishing company in marketing the disc?
JJ: Oh, yes. We've also approached some other publishing companies as well. We'll see. But some want to use the old royalty method of paying off the authors (10% of sales goes to the authors), and that doesn't make sense. Most are in a state of disarray.

MB: I guess that's how publishing companies get feedback--in term of marketing and sales.
JJ: That's what drives the publishing business--sales. It's extremely important to them--it's probably their most important evaluation tool as well.

Since John does not believe that sales are the primary means of obtaining evaluation feedback once the disk is released, he has not conducted a "summative" evaluation of his CD-ROM. A more plausible reason for this fact is he does not use the terms summative and formative due to his belief that development and evaluation are contiguous, on-going processes. Today, John continues to evaluate his CD-ROM in various stages for its next revision.

Case Study 4: Gwen

The Setting

Gwen works for a very large, leading software publishing company in Utah which produces such software programs as word processing, spreadsheets, databases, and presentations. Her company also owns what is recognized as the most popular network software used by companies throughout the world. The project which we discussed was her development and evaluation of an instructional CD-ROM created to better train network users by increasing their literacy of the networking software. The targeted audience for this CD-ROM was those programmers, network administrators and information systems professionals who wished to earn certification from her company as a specialist in their field. In her role as Project Manager in the Education Department (under the Skills Assessment Division and Testing Group), she has worked on numerous educational projects, particularly throughout their evaluation and testing stages. Gwen's advanced educational degrees were helpful in her work at the company, particularly in the development and evaluation stages of their products.

Before our interview began, the researcher thanked Gwen for responding to the posting for help that he had placed on an electronic newsgroup for evaluators. Gwen stated that she was happy to help in any way with the study. She also mentioned that her boss, David, who has interests in technology and education, would be willing to help as well. We also briefly discussed our similar backgrounds in education. As our telephone interview from her office in Utah began, Gwen detailed her role at the company and the

CD-ROM which she created and evaluated:

> MB: Just tell me a little about your role at your company, how long you have been there, your title...
> GM: I've been with my company for a year-and-a-half. I'm a Project Manager in the Education Department under the Skills Assessment Division in the Testing Group-- Certification Testing. I handle all outside projects and contracts.
>
> MB: Using some examples, tell me about your evaluations processes and procedures for instructional CD-ROMs.
> GM: Right now, our group has one instructional CD-ROM out called the *Guide*. It is an instructional, educational testing exposure tool that we put out to increase literacy of potential certification candidates [for the networking software which her company produces]. It first came out at our international conference last September. We started development last March. It was an idea that David [her manager] and I worked up to help have a better educational training experience. The CD includes information on all of our certification programs, and all our certification levels. You need to take and pass 8 tests to become certified by us and the knowledge ranges from basic networking fundamentals through to various tasks--the CD takes you across the whole range of tasks. It also includes information on all our certification levels.

Gwen explains her idea of evaluation in relation to her disc, including a very thorough review process, which is not uncommon in a large company such as her's:

> MB: What does evaluation mean to you in terms of your CD-ROM?
> GM: We evaluated it from the conceptual point, we had a project approval committee that evaluates everything from the concept through to the business plan through the distribution plan to the end result. Conceptually, it's reviewed all the way along by upper management to make sure the messaging is right.

Two major categories emerged from this interview: 1) types of evaluation performed (formative and summative) as well as the methods that were used for each type of evaluation; and 2) overall views and uses of evaluations. Each major category has sub-categories that will be discussed as well.

Types of Evaluation Performed

Formative Evaluation Methods Used

1. Needs Assessment

The need for the *Guide* originated in the Education Department at Gwen's company as they saw the need to develop a product to assist students from around the world who were interested in earning networking certification by her company. The tests are usually very demanding, so test anxiety is often very high, especially among their foreign users:

> GM: David and I felt that we needed to put a tool out there would help students who are training by whatever method to take these tests (they're high stress tests) --we wanted to put something out there that would help ease that test anxiety, and to help them know what avenues they could take to better train and prepare for the tests and to give them exposure. So what we built are assessment tests. They're built along the same specifications as the certification exams for all the different tracks and tests that we have out there. We have a smaller test on the CD that candidates can take and it will tell you how prepared you are to take the actual examination. They're content-balanced to the certification test--so they're accurate assessments. They break it down by section and tell you where you missed the most, and where you got the most right, so that you can have this as a guide to help you study and train. It also helps you to know where to go from where you are right now. They're built to be valid assessments of the person's skill and knowledge at that time.
>
> MB: It sounds like your CD-ROM, in a sense, is like SAT preparation or review courses and discs that are out there now.
> GM: Kind of. It won't help you pass the test if you take it a hundred times and you'll get to know what the answers are--it's not like that. But it helps guide you to know where you need to study more. The tests look and act just like the actual certification exams.
> MB: And all the tests are written tests?
> GM: No, all computer-based. My project was to manage the building of a simulation which simulates the networking environment and we use that in the certification exam. It's a complete simulation of the exam--the assessments tests on the *Guide* are accurate tests. They're good practice sessions...
> MB: So, they're hands-on experience...
> GM: Exactly, the simulation engine is part of the CD. You can go in outside the

testing environment and you can explore all the functionalities of the network administrator. It helps a lot of people.

We do tests in 13 different languages, all around the world. So all are our tests are translated, and the simulation engine is translated as well. The Spanish and Portuguese cultures have a very high level of test anxiety and fear of failure, so they were a perfect market for this type of literacy tool.

2. Product Testing

Gwen and her company carefully reviewed and tested the CD-ROM in its earliest stages though beta and alpha tests and cross-platform and platform testing, both internally and externally:

> GM: As far as technical evaluation we have the programmers who help build the interface...there's a review process. We had a five day beta review where we sent CDs out to some outside people to take all the tests, make sure they scored right, make sure all the functionality worked, simulations didn't crash their machines, make sure that what they were hearing was really what we wanted to say. So we did internal review and evaluation as well external review and evaluation. We got information from all levels--upper management down to...we tried it out on probably 50 different machines just to make sure it was solid, then we burned it and put it out there. We still had a few bugs.

Since the objective of this study was to translate the terminology used by software developers and to apply it to the terms that the field of evaluation uses, the researcher asked Gwen if she ever uses the terms formative and summative for her evaluations, rather than beta and alpha testing:

> MB: Do you term alpha testing as internal testing then?
> GM: Yes, our alpha testing was internal on this project. Beta testing was internal and external. We actually paid people to go through item-by-item on the disc, to look at it very carefully, to make sure that it all functioned right, and that the content was good. From a test standpoint, that was important to us because we needed to make sure that the items were good measures and that they were valid and reliable. From a CD standpoint, it wasn't all pulled together at that point, but we still wanted to make sure that it functioned well and interacted well. We were interfacing with a lot of different programs on the CD, so we didn't want to put it

> out there without it going through a rigorous evaluation--we wanted to make sure it worked well.
>
> MB: Do you use the terms formative and summative evaluations for your procedures, or is it more alpha and beta testing?
> GM: We use alpha and beta testing as terms. Alpha testing was internal, like I said, where we just tested the functionality of the menus; beta testing we worked more with making sure that the test items worked well and that the driver interacted well with the menus and that all the programs worked well together. Then we did a *premium beta review* which was internal which checked all the program content. That's when we put it on a bunch of different platforms to make sure that we couldn't crash it anywhere.

The concept of *premium beta review* was a concept that the researcher had not heard of in instructional technology, nor was it in any of the multimedia books researched, so he asked Gwen to define it:

> MB: And premium beta would be almost the final version, or the final version?
> GM: No, it would be almost the final version except for bug fixes, all the content should be there, all the typos corrected...
> MB: So it's near final completion...
> GM: Right--just some final tasks to make sure that we haven't missed anything. Premium is what we call code-complete.

Gwen explains the importance of the beta review process as a primary means of evaluating the *Guide*:

> GM: The beta review process, in which the input came back on an individual basis, was the method we used to gather evaluative feedback.

3. Partnerships With Educational Institutions

One of the many ways that Gwen's company markets, distributes, and tests its products is through partnerships with authorized education centers around the world. These centers are sites which have established close ties with Gwen's company and are

authorized for training network professionals to earn certification through the company. Much like SAT test review centers, they help students prepare for exams:

> MB: Is there a charge for the CD-ROM?
> GM: No, it's free to customers. Our authorized education centers distribute it and there is a very nominal charge ($1, $1.50) to the education centers. It's free to our end-users.
>
> MB: What about if someone called up and asked for it?
> GM: If people call in and ask for it, then we'll send it to them and we also give it out at conferences. But the reason we wanted to have our education centers in the loop is that it will help people who really don't know the best way to be educated into the certified network program--the CD gives them that open door.

The authorized education centers and the advisory council also helped Gwen with the user testing of the disc, especially during its beta testing stage:

> GM: We also talked to the education centers and the advisory council. They reviewed it from the business plan point-of-view. They all looked at the beta copy to make sure that could use and that it had the functionality that they were looking for as an educational and marketing tool in their environment.
> We tried to customize this tool to fit our product into the needs of our authorized education centers (AEC)--one of the needs from their point of view for these assessment tests was that their students with lower levels of experience didn't get into advanced classes because that slows everyone's learning down.
> The advisory council is made up of various world-wide representatives of the authorized education centers, at length.
> MB: That sounds like a large group. How many were there?
> GM: They break down into regions, so the group that we talked to was probably about 15 people. We talked to them every month.

It was through the education centers that they also established ties with users of the products (the students):

> GM: From the testing group, our customer service people have contact with actual test takers. So our audience was two-fold: we wanted the education centers to have a tool that they could give students while they were training and we wanted a tool the test takers could use to give themselves to have a better training and testing experience. We had a broad range of users in the developmental stages.

4. Target Audience (User) Tryouts

User tryouts with the target audience of students were instrumental in gathering feedback on the *Guide*. Most of these contacts were established through the technical support and customer service representatives that work for Gwen's company. In establishing those contacts, Gwen obtained a wide range of testers, both experts and non-experts in the area of computer networking:

> MB: Did you actually target your intended market audience in your beta testings? Those people whose certification was aimed at, or was it kind of random selection?
> GM: Actually, it was a list of people. Some were certified and were experts so that we could get expert opinion on the content of the questions and the whole feel of the tool. Others were not certified or in the process of being certified. So we tried to get a wide range--we selected who we wanted to send it to, to make sure that we had a range of people that might possibly use the tool. All of them were people who we had contact with before, through either our certification, beta testing, or customer service issues. They were all people who had some kind of level of contact with our company. A lot of them were our technical support people, who work on our technical support line for our company. They have a lot of interaction with the customers and had a lot of good input about what questions were more relevant and how it might work better. All of our workers are certified as well, so they know the product well.

5. Content Expert Review

Content review was an important evaluation process for Gwen and her company. In fact, the in-depth review process of the *Guide* by different levels of the company is evidence of the importance of content review of the disc. Moreover, ensuring the functionality of the disc was critical because the disc's content changes so frequently as upgrades and advances in technology are made:

> GM: There was a lot of content review. Messaging is extremely important to our

company and to our Education Division, making sure that the information we put out is as accurate as it is humanly possible at the time of release. Our programs evolve and change quite a bit. We also built some functionality in so that any information that changes can be downloaded from the Web to your computer and then the CD would use the latest drivers, tests, interfaces, so that the CD is never obsolete and is self-updatable from the Web. We worked on that functionality quite a bit.

6. Conferences and Seminars

As with several of the developers interviewed in this study, Gwen uses conferences to distribute her disc, but does not attend conferences specifically for feedback on her products. Sometimes, however, she does receive good feedback from users at conferences which she does use to make changes to the disc:

> MB: Do you present your products (i.e., CD-ROMs) at conferences and seminars to gather feedback on them? If you do get feedback (evaluation) from these conferences or seminars, have you used any of them to make changes to your discs? Any specific examples?
> GM: We demo and give the CDs out at conferences, and people usually have input, which we use to improve the product on the next rev. But we do not specifically conduct reviews and gather input during the development cycle from conferences.

7. Surveys and Questionnaires

One way in which Gwen utilized the partnerships with the education centers was by establishing exit surveys with test-takers as they completed the certification exam. All these surveys were not paper based, but instead computer-based. As test-takers finished the exam they had the opportunity to provide feedback on the *Guide*, if they used it:

> MB: It sounds like the partnerships that you created outside the company were very helpful. What other plans do you have with upgrades of the next version. Or do you have any new plans?
> GM: We do actually. I'm in charge of a couple of projects that are going to help us with that. One of them, we have exit surveys on our certification exams, that after

you completed your exams, you take the survey and we ask if you used the *Guide*, how it could be more helpful to use. This is done on the computer--the surveys are connected to the test itself. It was directed at the person who had actually taken the test.

Another way that Gwen is soliciting feedback for the disc is by establishing an on-line research group where users can fill out surveys on the Internet:

> GM: I am implementing right now an on-line research group representing the whole spectrum of our customer base--from the cream-of-the crop (best-scoring CNE's and CNI's) all the way down to the people in the program having trouble and we will be doing some on-line surveys to get their input on how well these tools work for them.
>
> I also talk to our Education Centers frequently.

She also plans to create an interactive survey that would give the company instant feedback on the disc:

> GM: At the end of this month, we're hoping to have an interactive survey package that I've designed that will allow people to give comment on individual test questions and also on the programs and tools that we have available, so we'll get comments. So it's not just a "rank your satisfaction" kind of survey, though we have those too--they're quicker to answer. But now we'll have a comment field to use that can be downloaded--actual opinions. It will give us intimate contact with our certification candidates--and they'll benefit from it and we will also.

What Gwen envisions next for the *Guide* is to establish a link to the company's home page from the CD-ROM to get feedback on the disc:

> MB: You said that there are feedback for users of the *Guide*--are these open-ended questions mainly?
> GM: It's pretty much open-ended, although when we're thinking about new functionalities we'll put questions out to an audience, "What do you think of this?" Getting some kind of link on the CD-ROM that would actually lead to our home page would be the next step.

Summative Evaluation Methods Used

1. Market Feedback

Gwen has summative evaluation techniques such as post-certification exams, surveys, and customer satisfaction levels for the *Guide*. She explains how customers went through the disc once it was released and how the post-release review was done internally:

> GM: Our marketing group went over it, upper management went over it. Customers looked at it at the conceptual points--from usability and functionality points of view (would it work for them).
>
> After it was out, we had people tell us how it didn't work, how it would work better, what they really liked, what they hated. We reviewed it in every way possible.
>
> We had post-release review, too. We actually had people internal to our company use it who are on the development group to tell us how they would do it better (programmers, etc.). They're especially helpful when you don't have a deadline looming over you.

In the next version of the CD-ROM, Gwen intends to use suggestions from customers who call into the company for technical support and offer their feedback on the disc:

> GM: And we get comments from customers who have the CD and call in for support. We get good evaluative information from these calls that will be applied to the next rev. [revision].

Overall Views and Uses of Evaluations

There were three sub-categories of overall uses and views of evaluations which Gwen referred to in her interview: evaluation results used to make changes on CD-ROMs; new synergies created between divisions of the company for evaluation; and overall, that evaluations were helpful for the product.

1. Evaluation Result Used to Make Changes on the CD-ROM

Gwen offered several instances of changes made to the *Guide* that were elicited in the evaluations. During several cross-platform and platform testing sessions for the disc, she discovered that some of the users' computers were locking up. Gwen learned that this was due to user misuse and misunderstanding. From these testing sessions, she made graphical interface and performance changes to the CD-ROM :

> GM: The types of feedback we got-- "I have a 486 and it's got a sort of funky graphics setting" and one of the programs we put on there that talked about the adaptive testing (a demo) worked fine, but if you put it on anything higher than a 586, it would lock up your system, because it's a DOS based program. We got input that lots of people who didn't have a lot of experience taking the test, would launch a test and then go, "Oh, I didn't want to do this" and they wouldn't close out of it, and it would lock the system up. So we put up data lock boxes that would pop up when you tried to do that saying that you already have a session running and to close it and start again.
>
> Graphic issues--we built everything to a basic 800x600 standard [screen resolution], but when you put it on a 640x480, it looked awful, so we had to scale everything down. We did a lot of that Q/A stuff for look and performance.

During the tryouts of the disc, Gwen realized from the target (user) audience that it needed to be more specific in its direction through the program because some users got lost during their tests. She resolved this by adding six more tests that were broader, but easier to navigate through:

> GM: We've also had a lot of input from people who have used it that has helped us make some changes.
> MB: Tell me about some of those changes.
> GM: We simplified some of the program information, got more specific on some other information on how to contact people to help you decide on where you want to go. We added a bunch of new tests that represented a broader...we had 12 tests on the first version, and we added 6 more tests.
> MB: Are all of these tests simulations?
> GM: Some of them are multiple-choice tests, but the network tests are mostly simulations. Some of the other tests are scenario-based, some are multiple choice, but they don't all have simulation applications.
> MB: So it's a variety of testing measures.

GM: Right.

2. New Synergies Created Between Divisions of Company for Evaluation

This category became most evident in Gwen's interview. It was clear that working in a large company was beneficial for Gwen as she utilized all of its resources in evaluating the CD-ROM. In the production and development of the *Guide*, there were many employees from different divisions of the company who assisted her in the evaluation, clearly creating a synergy within the company. These included programers, a partner of the company, marketing, communications, customer support, technical support, and others:

> MB: Tell me about some other employees who help out with the evaluation processes... You mentioned Marketing...
> GM: Yes. We had a test programmer who built the test and programmers who built the interface, the 3 of us made sure the program worked well. We had Corporate Communications and Education Communications who reviewed anything that had any messaging in it--as far as content--to make sure it was accurate and make sure it was professional and fit with our corporate standard. Then we had the marketing groups and the Marketing Program Managers for each of the programs (CNE, CNI, etc.) to make sure that their courses were accurately represented and that all the information was correct. Tech support people did a lot of the question review and the beta testing, some of the customer service reps did that as well.
>
> MB: So tech people and customer support personnel were very helpful in relaying what the customers really wanted.
> GM: Right.
> MB: So it's definitely a team project among many divisions of your company.
> GM: Right.
> MB: Your Education Group then would be the driving force behind the evaluation
> GM: Right, the Testing Group within Education.
>
> MB: You didn't really have to hire any new or additional employees then for your evaluation procedures.
> GM: We worked with one of our partners to help us develop the items. Actually this project began with no budget and 3 programmers and myself. We worked some long hours.

By working for a large company that is a leader in its field it became apparent in the interview that Gwen's evaluation of the CD-ROM adhered to a very comprehensive, organized review process that was representative of how well the company is run. A standard in the evaluation process was sign-off sheets in each division and at every level of production and development:

> GM: As a general policy, our education department has a very standard review process starting with the project approval committee and ending with people specified that they have to sign off and approve anything that goes out to the field. So a lot of it from a content, look and feel, messaging standpoint is dictated to us. Its very important both to the company and in my department that products go out looking the best they can and working well. It's standard for us to adhere to those guidelines.
>
> MB: So at the formative (beginning) part of the evaluation and the project, you have kind of set standards by upper management as to how the project should progress...
> GM: Right
> MB: And what they have in mind for it...
> GM: Maybe not so much that, but as it evolves, they need to review and make sure through all the phases (pre-alpha, alpha, pre-beta, beta, premium beta) that they approve the changes. So there are sign off sheets.

3. Overall, Evaluation Helpful for Product

Toward the end of interview, Gwen was asked to consider the importance of evaluation for the *Guide* and for her company. As with the other five developers interviewed for this study, she clearly believes that the evaluation processes have helped her to improve her products. Since the evaluation of the first revision was so successful, Gwen already has extended her evaluation of the next revision of the disc, which attests to how high she regards evaluation as a vital feature of the production and development process:

MB: So would you say that overall, then, evaluations have been helpful for your product?
GM: Yes. In fact, we probably went further and broader with the evaluation of this next revision because the information was so valuable from the first revision.

Analysis of Case Study

As the researcher conducted and analyzed this interview, it became more evident that Gwen and her company performed one of the most comprehensive and efficient evaluations of the six developers interviewed for this study. That may be a result of the large size, talented employees, and enormous resources of the company she works for, but it is also a reflection of the company clearly positioning itself well as a leader in its market as well as the strength of its organizational structure. Clearly these strengths are evidence of how efficiently run and well organized its evaluation process are, and also how important and beneficial they are to the entire company.

At the close of the interview, Gwen was asked to review the evaluation processes that she used for the *Guide*. After hearing her responses, the researcher realized how intensive and complete her evaluation was, both internally and externally:

MB: If you could just summarize, before I leave, what your evaluation procedures are.
GM: The evaluations we went through are functionality, content (both from the test point of view and the programs and marketing information point of view--to make sure that what we were saying was what we wanted to say and that the content of the test was accurate and that the measurements were accurate and valid).

We did cross-platform testing to make sure it functioned across the variety of machines and platforms and environments.

We did evaluations where people had no exposure to any of the tools that are on this and we just handed it to them, saying, "Can you install it, can you get through it, find networking information for me" to see how easy it was to use--to see how user-friendly, how functional it was.

We tried to make it crash in every scenario.

> Our marketing group went over it, upper management went over it. Customers looked at it at the conceptual points--from usability and functionality points of view (would it work for them).
>
> After it was out, we had people tell us how it didn't work, how it would work better, what they really liked, what they hated. We reviewed it in every way possible.
>
> We had post-release review, too. We actually had people internal to our company use it who are on the development group to tell us how they would do it better (programmers, etc.). They're especially helpful when you don't have a deadline looming over you.

There were only three categories in which Gwen's responses did not place: focus groups, financial constraints for evaluations, and suggestions for improving evaluations.

Gwen did not use focus groups in her evaluation because the feedback which she received from customers who called the company through its customer support and technical support lines and from those who used the disc during its beta stages, were extremely useful for evaluative feedback:

> GM: We have not had focus groups, or any kind of meetings concerning this product. The beta review process, in which the input came back on an individual basis, was the method we used to gather evaluative feedback. And we get comments from customers who have the CD and call in for support. We get good evaluative information from these calls that will be applied to the next rev.

The partnerships that Gwen and her company had established with educational institutions and with customers were enormously beneficial in the evaluation process:

> MB: Tell me about the overall effectiveness of your evaluation. If you really believe it has helped you to improve your product and made it better for your market.
>
> GM: It did. Obviously, it's important to make sure that any information gets out as accurately as possible. So for that point of view it was critical. But the evaluation we did with the outside groups--the education centers and the other marketing groups was critical in the developmental stages because we are a little bit removed form the customer arena and they helped us make sure that we built it to be something that they could use.

Although the large size of the company and its enormous resources were helpful for Gwen in her evaluations, she still made sound use of the employees of the company as vital resources. With their assistance, Gwen did not have to spend a large amount of money in the evaluation, even though her company could afford to do so. In fact, she was so successful in the production, development and evaluation of the *Guide*, that her company awarded her a budget for its next revision:

> GM: Actually this project began with no budget and 3 programmers and myself. ... Since there was no precedent for this project, there was no funding--we had to do it the best way that we could. We had no budget at all. It's pretty impressive considering that.
> MB: So you really had no financial constraints at all?
> GM: We didn't actually spend any money building it, only people money. We only spent money in the actual production of the CDs. Since it was so well-used by our customers after the first revision, we received a budget for the second revision, but we haven't hired any new people--we're just using internal resources.

Case Study 5: Jack

The Setting

Jack is a tenured business professor at a leading state university in Virginia as well as a multimedia developer. He also holds an endowed chair in banking within his department. Most recently Jack worked with a large, prominent educational textbook publisher in producing a multimedia based textbook on the field of marketing. This was the publisher's first venture into electronic publishing, so there were some obstacles to overcome throughout the development process. The project was developed onto an instructional CD-ROM and accompanied by a textbook for use by students and professors both in the classroom as well as at home. As co-author of the marketing textbook and CD-ROM, Jack played a prominent role in the development, programming, and evaluation of the CD-ROM which he shares in this interview. During our interview, he also explained the evaluation processes used by the publisher as well.

The researcher had known Jack for nearly two years, as we both worked for the same school. Before our interview began, we sat in his office and discussed what we had accomplished during the last year or so, since we had not seen each other for nearly that long. When the interview began, I first asked Jack to describe his target audience for the CD-ROM and its accompanying textbook and also asked him how the disc was authored:

> MB: The CD-ROM is intended for undergraduates, graduates?
> JL: Both. It's used in undergraduate programs and in graduate MBA programs. The discs were authored in Toolbook...Plus, a textbook was created to accompany the disc.

The two major categories that Jack referred to in this interview are: 1) types of evaluation performed (formative and summative) as well as the methods that were used for each type of evaluation; and 2) overall views and uses of evaluations. Each major category has sub-categories that will be discussed as well.

Types of Evaluation Performed

Formative Evaluation Methods Used

1. Needs Assessment

By explaining how the Marketing CD-ROM developed, Jack offers insights of how the disc was needed in colleges as an exciting, new way of computer-based teaching:

> MB: Tell me a little about the background of creating a marketing textbook onto CD-ROM--how you approached it.
> JL: I was teaching an advertising course and I would use examples of current ads in class. I would actually sit and edit a tape together of all the commercials that I wanted to show. So I would write a lecture and figure out commercials that demonstrated the points of the lecture--and I would make a tape of those. It worked fairly effectively, but it was impossible in class to get back from ad to ad on the videotape to make comparisons. I had hundreds of commercials and there was no way to index them, other than by topic, which still took a while to access.
>
> IBM had given a million dollars each to four schools to integrate computers into the classroom. Looking at them, they were pretty unadvanced. So I took the challenge of developing a laserdisc of all these commercials. It holds about 2 hours of video on it--that's a lot of 30 second commercials. It worked well, but I knew it had a lot more potential. I was approached by several publishers who asked me if I was interested in doing a marketing textbook, but I was interested in doing a cutting-edge textbook onto a CD-ROM. Then the publishers came back to me with proposals. I chose the publisher I'm working for now because they assured me of its priority and gave me an advance for it plus a budget that I could draw from for future royalties. They told me they wanted to be first in the market.

Another way in which Jack sought to give his product more recognition and visibility, as

well as helping himself with its content, was by co-authoring it with another professor who is well-respected in the marketing field:

> JL: ...But I needed help so I brought on another professor, the co-author, from South Carolina. He is probably considered the number one name in consumer behavior in the marketing field nationwide--so that's a big draw.

2. Product Testing

As the Marketing CD-ROM was being developed, Jack sought to beta test the disc with a number of colleges throughout the country:

> JL: In the early stages, we took beta copies (prototype or test copies) to 22 schools and we gave them the CD-ROMs and asked for their feedback.

Another means of product testing used by the publisher which Jack worked with was hiring a software company to evaluate the disc. This company tested the disc on various platforms and tested it for bugs as well:

> JL: We had a software company evaluate the disc. My publisher sent it out to this company and said, "crash the disk!"
> MB: Tested it on cross-platforms...
> JL: Various computers--IBM platforms--486, 386, pentiums, different memories, all that kind of stuff. So that's evaluation also...It was just bug testing.
> The publisher also evaluated it in-house. They sat down and had their computer people go over it, and their arts and graphics people.

3. Partnerships With Educational Institutions

Since Jack works for a university, he realized the potential in having professors and students from 22 colleges and universities throughout the country evaluate the disc. This became one of the most cost-saving and effective means for him to get feedback on the disc. These reviews by students and professors were paid for by the publisher Jack

worked with:

> JL: In the early stages, we took beta copies (prototype or test copies) to 22 schools and we gave them the CD-ROMs. We also gave the printed materials, the CD-ROM material, and the faculty the CD-ROM presentation card to these 22 schools--they were all paid by the publishers. It was a paid review. The schools were colleges and universities such as Notre Dame and other community colleges.
> MB: So it was very diverse.
> JL: Very diverse and across a lot of levels--state universities, land grant universities, public and private schools. They gave the material to the faculty members then the students got to use the CD-ROMs in the classroom and then they all filled out questionnaires (which I'll give to you).

The reactions from the students and professors was very good, according to Jack. They realized the potential of a new learning vehicle:

> MB: That's great. What were the reactions?
> JL: We got a really good reaction because it was different. Students thought it was very fun and the fact that they could see a commercial right there was enamoring to them. The product changed significantly from then--there were a lot of things that changed.
> MB: And those were brought out in the evaluation?
> JL: Yes.

4. Target Audience (User) Tryouts

As previously mentioned, both students and professors at the 22 schools tested the disc during user tryouts:

> JL: They [the publisher] gave the material to the faculty members then the students got to use the CD-ROMs in the classroom...

As a means of corroborating Jack's use of his target audience during the evaluation, the researcher asked Jack if he could provided him with an example. Jack offered a two-page untitled questionnaire that was used during the class testing of the CD-ROM. This form asked students two questions related to their course preferences and

whether they thought they would actually use the product:

1. Would you prefer a CD-ROM text to a traditional hard bound, paper text?
2. Would you like to use this product for your principles of marketing course?

5. Content Expert Review

Jack and his co-author, both marketing professors, are content experts, so it was their responsibility primarily to review the content of the disc and the textbook:

> JL: Terry [the co-author] and I are the subject experts, so there was no problem in verifying the content of the disc. We both worked together in reviewing the content.

Through the partnerships that Jack had created with colleges throughout the United States, he asked the professors to also verify the content of the disc and the textbook, as well as to evaluate the overall technology. The feedback obtained at this stage would be used to make changes to the next version of the CD-ROM. Since this content and technological review was done after the disc was released, it can be considered as a summative evaluation process:

> JL: As soon as the product came out (January, 1996), we went out for evaluation--to validate everything (summatively).
> MB: Tell me about that.
> JL: The publisher hired 4 professors whose job was to evaluate the product so that we could decide what to do with the second edition.
> MB: Tell me what they evaluated.
> JL: They evaluated the content of the book (which I may be able to share with you) and they evaluated technology.

In reviewing the content of the disc, professors throughout the country who agreed to participate in the evaluation were paid an honorarium for their work. The publisher also sent them a letter thanking them for their participation in the review of the

CD-ROM and textbook. This two-page letter, given to the researcher by Jack, also asked them to structure their review around several questions based on both the student and instructor versions of the disc and book. Most of these questions asked professors ways in which they thought the disc and textbook could be improved, as well as basic strengths and weaknesses questions. These review questions included:

> 1. Is the organization of the this text appropriate for your course in its present form?
> 2. Do you find the level of the material appropriate for your students?
> 3. Please discuss three strengths and three weaknesses of the Student CD-ROM. How does this software compare to others CDs in this area? Any suggestions for improvement?
> 4. Please discuss three strengths and weaknesses of the Instructor's CD-ROM. How does this software compare to others CDs in this area? Any suggestions for improvement?
> 5. How does the Instructor's Kit (User's Guide, Instructor's Manual, Test Bank, Instructor CD-ROM) compare to competing ancillary packages? Any suggestions for improvement?
> 6. How effectively did the Student Kit (textbook, Study Guide, & Student CD-ROM) work together in and out of the classroom? Any suggestions for improving the effectiveness of this project?.

It appears that these questions were simple enough to allow the professors to structure their review around the critical contents of the disc.

6. Focus Groups

Focus groups were also used by Jack and his publisher during the formative stages of the CD-ROM. Most of the feedback obtained at this time were suggested changes for the user interface:

> MB: Tell me about the focus groups you used for evaluation.
> JL: I used these at the very beginning to clarify the kinds of buttons and features that students thought might be useful. These groups consisted of two focus groups

of undergrads that had just finished the principles (Marketing) course. Each group consisted of approximately 15 students.

They were used again when we began to develop screens, buttons and features. This time, only one group was used, but with fourth year marketing concentrators. Only 8 students were in this group.

Jack also used informal and casual conversations with students in marketing classes taught by his peers, which were similar to focus groups. The students made suggestions that were very useful for Jack:

> MB: Were there informal meetings and informal conversations with users during the evaluation process.
> JL: I did a number of these with John's [another professor in his department] classes. They were done very similar to focus groups. They did make suggestions which we used. All of John's classes were significant. They helped me more than anything because they were testing the product as it went along. They also understood that their suggestions were being used to shape the product, so they were very willing to provide me with feedback. Specific changes that came about because of these sessions included: 1)glossary - terms with definitions on the screen; 2)interactive multiple choice testing; 3)background (college degree, etc.) of the Captains of Industry [a series of studies on people in the industry].

7. Conferences and Seminars

Jack has used conferences and seminars, both educational and professional, to gather feedback on the disc both in its early development stages as well as when the CD-ROM was about to be officially released. The suggestions from the conferences were extremely beneficial for him because they came from his peers. Some will be used for the second edition of the CD-ROM:

> MB: Did you or the company present the CD-ROM at a conferences or seminars to solicit feedback as part of your evaluation? If you did get feedback, was any of it used to make changes on the disc?
> JL: This will take me some time to reconstruct but off the top of my head: DePaul Conference on advanced technology, National AMA conference, Virginia

> Educators Conference, Advisory Board, Capital Campaign Kickoff, Parents Day. Yes, some changes were made from these suggestions at conferences. The DePaul conference and the AMA conference (attended by marketing professors throughout the U.S.) suggested that behavioral objectives be developed at the beginning of each chapter. They also suggested that we figure out how to clearly read any advertisements on the screen. Both of these changes were significant. This first conference took place in the early fall of 1994 and while other professors were presenting ideas of what they would like to do in multimedia, I presented a prototype of the final product. My session was quite lively because they could see what I was talking about rather than trying to imagine. That summer I presented at the AMA Summer Educators Conference. I was on a panel with two other multimedia developers. Again, I had a prototype so my portion of the program provided a lot of feedback. The same suggestions were presented at this conference plus the idea of putting in case law for the legal section of the environmental analysis. While I did not make this change for the first edition, I will make it in the second.

8. Surveys and Questionnaires

As part of the tryouts and focus groups with students and professors, one way in which Jack and the publisher sought more information was through the use of questionnaires and surveys, both in the formative and summative evaluation phases of the disc:

> JL: The publishers gave the material to the faculty members then the students got to use the CD-ROMs in the classroom and then they all filled out questionnaires.

To verify that Jack uses evaluation instruments, the researcher asked him if he would provide several examples. As previously mentioned, Jack and his publisher used a two-page questionnaire designed for class testing of the CD-ROM. In that questionnaire created by the publishers, they used a simple ranking scale by asking the students to rate the various software features of the disc. Question 1 was: "Please mark the following in terms of their 'usefulness' on a scale of 1 through 4, with 4 being 'very useful' and 1 being

'not very useful.' For all items given a 4 or 1 mark, please provide your reasons why in the comment section below." The software features mentioned on the questionnaire included: buttons, index screen, full screen, note pad, bibliography, discussion questions, chapter review, chapter objectives, marketing highlight boxes, print ads, and video ads. Five other questions related to the disc's content, interface, user preferences and navigability were asked on the questionnaire:

1. What did you like the most about using this CD-ROM product?
2. What did you like the least about using this CD-ROM product?
3. Were you able to use the CD-ROM without referring to your text? Why or why not?
4. Would you prefer a CD-ROM text to a traditional hard bound, paper text
5. Would you like to use this product for your principles of marketing course? Why?

Summative Evaluation Methods Used

1. Market Feedback

Jack and his publisher used several summative evaluation methods for the CD-ROM, most of it to obtain market feedback from professors and students. The summative evaluation methods he used were: conferences, peer review, and surveys.

As mentioned in the previous section, Jack presented the final version of his CD-ROM at various conferences and seminars at which he obtained feedback on how to improve the next edition of the disc.

Once the prototype disc was improved through the formative stages and finally released, as part of the content and technology peer review by professors, Jack and the publisher solicited summative information on how the next version of the CD-ROM and textbook could be improved:

JL: After the prototype I met with the publisher to make the final changes--format changes, graphic changes. Terry and I worked together in reviewing the content. It took about 2 years to get it to the market. As soon as the product came out (January, 1996), we went out for evaluation--to validate everything (summative).
MB: Tell me about that.
JL: The publisher hired 4 professors whose job was to evaluate the product so that we could decide what to do with the second edition.

Most recently, Jack met with his publisher to address the second edition of the CD-ROM and the text. In that meeting they also evaluated the competitor's product:

JL : We met just recently to discuss the second edition. We reviewed comments from previous evaluations we had conducted in order to make some changes. We also evaluated the competitor's product--it just came out. It's quite a different product.

The questionnaires and surveys which were given to the professors as part of the peer review of the disc were also distributed once the disc was released to obtain both formative and summative information, as explained in the previous section of this paper:

JL: They gave the material to the faculty members then the students got to use the CD-ROMs in the classroom and then they all filled out questionnaires.

Another summative evaluation process used by the publisher and Jack was by using the marketing representatives of the company to obtain feedback from customers as they sold the disc and textbook. At one meeting, Jack met with the publisher and the marketing representatives to discuss feedback from professors and students:

JL: The book rep's were queried about what were the problems associated with selling the book (adoptions) and what kinds of changes would they make in the next edition to improve that. About 10 book representatives from the publisher met with us for 4 hours. The rep's discussed the feedback they had heard from professors throughout the country about the disc.
MB: What were some of those comments?
JL: Schools that didn't have technology (multimedia computers) couldn't purchase the book. Also, some professors felt that it might create two levels of students--

> those who had [multimedia computers] and those who don't. So even with the printed book and the CD-ROM having exactly the same concepts, the same words, there were still those who had computers (the rich kids) and those who don't (who had to read the bare-bones words)-- I can understand how that would happen.

Overall Views and Uses of Evaluations

This second primary category uncovered four sub-categories which were insightful for understanding Jack's overall opinions and uses of the evaluation process. The four sub-categories were: evaluation results used to make changes on the CD-ROM; how the development and evaluation processes differ between CD-ROMs and textbooks; overall, that evaluations were helpful for the product; and made suggestions for improving evaluations.

1. Evaluation Results Used to Make Changes on the CD-ROM

Jack offered many examples of changes he made to the CD-ROM from the evaluations. One instance of a change he made to the glossary of the disc was obtained during student tryouts:

> MB: Tell me more about your evaluations--more how information you received from the evaluation which you used to improve the CD-ROM.
> JL: The glossary in the textbook is a complete glossary. What we found students doing from the evaluation is that they would actually go to the glossary [on the disc] and as you click each word, a definition pops up. They would study that way. Now, what we did was we took that information, pulled back and set up our glossary chapter by chapter so that the students don't have to go through all the words in the glossary. They can only go through the words in the glossary from that chapter.

Another change to the disc obtained during the tryouts was that students often would try to "cheat" through the chapters by taking the multiple choice questions at the

end of each chapter. The students, then, would not read the entire chapter. To correct this in the next edition, Jack planned on making the questions more difficult and longer:

> JL: Here's another one: these are behaviors that we observed from the evaluations. The first thing students would do when they opened up a new chapter is go do the multiple choice tests at the end of the chapter. The multiple choice tests are interactive, so when you click on something and get it right, it tells you you've got it right; if you get it wrong a little button pops up at the bottom that says "click to review." And when you click it you go to the page in the textbook that has the answer. So what the students would do is take the multiple choice tests, figure out what they didn't know, and then they would go and study those sections of the chapter and breeze through the other sections. It was an incorrect form of learning. That's something we learned only because of these evaluations. We had no idea this was happening before.
> MB: So it was defeating the purpose of learning--it was an incorrect way of learning.
> JL: Yeah. However, we can't stop that, but what we can do is make longer, better, more comprehensive multiple choice questions at the end of every chapter. We're going to take the bad behavior and capitalize on it, and push them in the right direction. It also means that we have to raise the levels of the questions at the end of each chapter, because I want them to miss more now. When we first designed those questions they were designed at the lowest level of difficulty. They were completely definitional and not applicational. It will force the students to a higher level.
> MB: And probably read more of the text...
> JL: Yes, because they will miss more.

As students used the disc during the evaluation, Jack discovered that most of them hard a difficult time reading all the textbook on a computer screen. In fact, it became mundane for them to go through the entire text, constantly scrolling down to read all of it. The next edition of the disc promises exciting new changes:

> MB: Any other examples?
> JL: Students have a tough time sitting in front of the computer, reading all the textbook. What that means to me is probably all the words in the book will not be replicated onto the next CD-ROM. We'll probably have more exercises, more graphics, more like the Web. We'll take some of the concepts and bring them back to the students with the Web being the application concept of all of these things. We'll actually have links on the new disc to Web Sites of companies and

government agencies.

Since students had such a dificult time reading the text throughout the CD-ROM, Jack intends to make both the CD-ROM and the textbook similar in the next edition. He envisions more exercises for the students as well. All of this feedback was communicated to him through the marketing representatives of the company, who had obtained it from the professors and students:

> MB: So what changes will be made in the second edition?
> JL: By making the textbook four-color hardbound, we're moving the book closer to the CD-ROM so that what's on the CD-ROM graphically (pictorially) will be closer to the new book. So, that's a big change. I'm not positive what the CD-ROM will look like at this point--I do know that it won't have all the words, but it will have more bullet points, shortened phrases, exercises such as statistical analyses for the students...
> MB: More of an interactive presentation.
> JL: Yes. We're also adding two more chapters to the textbook. The instructor's edition will also be different--the examples to be used in classrooms by professors will be developed separately for the instructors and separately for the students. I think that's going to make a more powerful instructor product. For example, we may put in the idea of "guerilla warfare" (used in marketing in the instructor's edition), but not likely in the student's book--because it gives the instructor, then, some other things to talk about that aren't in the book, that are well laid out and structured.
>
> MB: So all the feedback that these rep's had received was from the professors, who also communicated feedback from the students as well?
> JL: Yes--both from the professors and their students.

2. How the Development and Evaluation Processes Differ Between CD-ROMs and Textbooks

Jack had chosen his publishing company because he believed that they would give him the necessary support, skills, and marketing exposure that he needed for his CD-ROM and its accompanying textbook. However, as the development of the disc progressed, Jack

found that he disliked the methods his publisher was using for the evaluation. He discovered that they were treating the evaluation of his CD-ROM like that of a textbook. One plausible reason for this was the fact that this was the publisher's first venture into electronic publishing--they had produced only textbooks for so many years and the development and evaluation of a CD-ROM certainly posed new challenges for them. Jack stated his dissatisfaction with his publisher's evaluation through his concern that they were treating his CD-ROM as if it were a textbook:

> JL: I think they [the publishers] don't understand the product well enough that they can get people to evaluate it at the level it that it has to be...But they're trying to be sort of arms length, including like we don't know who's on other end--I know who's on the other end!...They're approaching the evaluation like that of a textbook, as they always have, but we have something more than that here.

Jack's discontent with the publisher's evaluation of his CD-ROM, along with his statement that he would like to manage the evaluation, show his passion for producing a superior product and is testament to the importance which holds for evaluation. Although Jack did not fully detail how CD-ROM evaluations were different than textbook evaluations, during the development of his disc he realized that the evaluation of a CD-ROM could not be handled like that of a textbook.

3. Overall, Evaluations Helpful for Product & 4. Made Suggestions for Improving Evaluations

These two sub-categories are combined because Jack agrees that evaluations were helpful for his product, but he simultaneously offered suggestions on how he believed they could be improved. Most of his suggestions for improvements of evaluations arose from

his desire to perform a more in-depth evaluation of his CD-ROM. He explained that informal meetings and conversations with the users would be a more efficient method of evaluation:

> MB: So have evaluations helped you to improve your product?
> JL: I think so, but it would be better if they let me do it--because I think they don't understand the product well enough that they can get people to evaluate it at the level it that it has to be...If they really want a good evaluation then what they should be doing is send the disc out to schools, let them talk, and then we'll let you guys all go out to dinner after the national meeting or something and let you talk about it.

Analysis of Case Study

As both a developer and a professor, Jack realized the importance of the partnerships he had established with peers in his field. As business professors, he and his co-author used their close ties with other marketing instructors throughout the country to gather evaluative feedback on the CD-ROM. Through these close partnerships with educational institutions, he also developed ways to obtain feedback from students as well, which was sometimes communicated through the professors or the marketing representatives themselves. In addition, Jack also utilized the students and professors from his university to help him with the evaluation of his disc.

Jack envisions a more comprehensive evaluation for his CD-ROM than the publisher has conducted so far. His aspirations are to have control of the entire evaluation himself, as explored in the last section of this study. So far, he has not yet been able to secure full management of the evaluation from his publisher. One reason Jack offered for wanting control of the evaluation is his vested interest in the product and his knowledge of it--he was the primary author and had control of most of its development; therefore, he

knew the product better than anyone else. Moreover, his knowledge of marketing principles and of the educational market may be additional strengths that he believed would be helpful to market the disc.

It is also apparent that, through the many changes Jack made to the CD-ROM and to the textbook, he views evaluation as a vital means of improving his product. Moreover, his desire to manage the entire evaluation of the CD-ROM suggests the importance which he places on evaluation and how it can be used effectively for product improvement.

Case Study 6: Mary

The Setting

Mary is President of a small multimedia software firm in Virginia which she created in 1985. Originally, the company began from a department which she formed in the School of Education at a leading research university in Virginia. She took that department and developed it into a private company in 1991. She has developed, programmed and evaluated numerous instructional CD-ROMs for both public and private companies and universities. We discussed several of these CD-ROMs at length in our phone interview, but also focused on a current project which she was developing on breast cancer for medical students and other health care professionals at a university medical school. Mary's background as an educator, computer programmer, and as a Ph.D. recipient in Curriculum were vital skills that she brought to her firm to develop interactive multimedia. She holds a Bachelors degree in Computer Science and two advanced degrees in Education. Mary has also taught at both the high school and college levels.

Before our interview began, the researcher and Mary discussed the field of instructional design, as we both had training in this area. We also spoke about her dissertation which detailed the development of a CD-ROM which she had worked on for her company. After we exchanged more information about our backgrounds, the interview began.

As a professional trained in both education and technology, Mary had specific views of how project development and evaluation should work together. She also

explained that development and evaluation are contiguous processes:

> MM: All throughout product development we're involved in evaluation...We evaluate the pieces as we go along.

Mary is more interested in the affective domain, rather than just the cognitive domain during the development of a project, which is why she performs impact evaluations, as she will further explain later in this case study and briefly details here:

> MM: As a designer, I'm always interested in designing an evaluation to learn what I most need to know. I'm interested in the design of materials which have an impact on the affective domain, not just the cognitive domain.

There were two major categories that emerged from this interview: 1) types of evaluation performed (formative and summative) as well as the methods that were used for each type of evaluation; and 2) overall views and uses of evaluations. Each major category has sub-categories that will be discussed as well.

Types of Evaluation Performed

Formative Evaluation Methods Used

1. Needs Assessment

Mary's training in education and programming, along with her experience in running a small company, has influenced her to discern at the start of the project whether the client has realistic goals and expectations. She does this by asking her clients several questions before she agrees to accept the project:

> MM: I am, number one, an educator. We have an obligation--we are wrestling to affect our educational system. We have an obligation to evaluate our products and see whether we made a difference and then use what we learned from those projects to integrate for our future projects.

> We need to meet the needs of the funder--I make the funder write down, as I ask them, "What will have to happen with this project so that when it's over, you will feel that it's successful." That tells me right away what is foremost important in their minds.
>
> For me to take a project, I have to believe that a client has realistic expectations that can be accomplished in a realistic time frame. Those two things are vital to me. That's why most projects fail--unrealistic expectations and unrealistic time frames for impact from the client.

Mary also realizes that you can learn from past mistakes made in evaluations:

> MM: As evaluators, we have preconceived ideas--we want to design an evaluation to validate our beliefs, and sometimes that doesn't always happen. You have to be willing not to have your beliefs validated--to learn from your mistakes, to say, "I was wrong there."

Part of the needs assessment process for Mary when she is interviewing clients about their expectations for the evaluation is telling them that she believes in conducting an impact evaluation to assess the long-tem effects of the product:

> MM: I have a test that I ask people when I'm working with them on the evaluation in their setting. I ask them to write down what it is they most want to know. It's interesting--many people don't really know what they want from an evaluation. I learned from an evaluator that you should only evaluate those things that you care about otherwise, you begin to trivialize your own evaluation. No matter who I deal with, they all come up with lousy questions and I reprimand them to write better ones. It's not an issue to me whether this [multimedia environment] is a more effective learning environment or not--it matters what the impact of the disc (product) is.

2. Product Testing

Since Mary believes that development and evaluation are contiguous processes, formative evaluations of her products are constantly being conducted during development. One of those formative evaluation processes is product testing with users, specifically beta, alpha, cross-platform and platform testing:

MB: Tell me more about your testing procedures.
MM: We do a number of different kinds of testing. There is a point which we bring users into our office and they use the product and we take notes or videotape them and we answer their questions. We continue to refine the product at that level. There is another level where we reach a certain point in the beta where we let them take it home and then bring it back. We're looking for two problems: one, we eventually hire testers who try to "break" the product, so we do robust testing across platforms, probably on about 25 different kinds of computers and 25 different kinds of printers, not which we all have in our office--we go out and hire people to test it. The ones which we're evaluating learning, we have them take it home and then come back in and we're after impact--what did you learn, what did you see, what was of interest, what did you see, does this make sense? So we kind of quiz them verbally. And trying to get at what the problems were. We pay them for their attention to detail. They're very eager, and they give us a lot of information.

MB: And certainly, then, testing is part of your evaluation?
MM: Oh, absolutely.
MB: Because there is a difference is terminology between educational evaluation and evaluations of technology.
MM: Well, we really do formative, summative, and impact evaluations.

3. Partnerships With Educational Institutions

Mary realized the tremendous benefits of working with schools in performing her evaluations, particularly with students. Coming from a university environment, she found these partnerships easy to establish, especially in getting target audience members:

MM: We do all the formative evaluation in-house then we do what we call "localized formative evaluation" with people in the [university town] that seem to be an appropriate audience.

In forming her company, Mary even hired many of the same employees from the department she previously worked with at the university:

MM: I went into business with some of my employees from the university I used to work for.

4. Target Audience (User) Tryout

This was perhaps the most important category for Mary, as she mentioned it frequently throughout the interview. During our conversation, she confirmed to me the importance of using authentic users in the evaluations as a vital factor which, as explored in Chapter 2 of this study, are often ignored:

> MM: One of the biggest criticisms of evaluations is that they are done in too narrow a group for too short a period of time, with something less than authentic users--authentic users are the users who this product is actually made for.
> MB: It's interesting that you should mention that because I've found that most evaluations of CD-ROMs (whatever little there are in print) often ignore their target audience.
> MM: Many of them really don't care about the success of the product. It's not just about numbers. I'm not interested in whether a student thinks a dictionary is informative--it matters more whether it's of use to them and their chosen field. I want to know if our products add value to the user.

Just as the researcher had found that videotaping users during tryouts of the disc and asking them to think aloud worked well during the formative stages of the product, so did Mary. She also tracks on-line activity with the authentic users as a means of evaluation:

> MM: At some point we take the product in a formative stage to authentic members of the audience--if 5 is enough, we take students and run them through using the applications and we videotape across their shoulders. We run the video, we don't video the face, we videotape the actions and capture their words. We do the think-aloud kind of structure and will prompt them to think aloud if they're quiet. I want them to focus on the content, but also want to know what their thought process is. So we run about 5 students through that kind of evaluation and then we come back and transcribe the videotape sequence, with their words and comments.
> MB: Fully or partially with the pertinent ideas...
> MM: Since we only do about 5, we do all of them. We have other people to review them as well. We look for places where the learner looks lost or has questions. I put them in charge of their environment and that lowers the risk of failure and the minute I lower their risk of failure I increase their motivation to continue to work with the project. We evaluate that aspect of it.

> We also track on-line activity. By asking them to hypothesize answers to questions on the disc, that gets them to interact with the case more. On the exit part of the disc, before you end, you have to identify the answers to the questions--what did you learn and why. That is captured by the computer and we go back to the users and verify what they said.

As part of her tryouts with authentic users, Mary also seeks to capture how the environments in which the discs are used affect the attitudes of the learners. By doing this, she gathers information which is used to modify the disc to meet the needs of both the users and her clients:

> MM: We also gave the disc out to all medical students to use in their own environments--we're interested in the environmental factors of the project. I want to capture data about where they actually interacted with this project--at home or in the lab. They can also keep it or turn it in at the end of the class. And the questions is, "Have you chosen to keep the disc, pass it on to a friend, or did you return it?" That states something about the disc's value. We're very interested in the attitude of the learner--pre and post attitudes and beliefs.
>
> MB: So you base you designs on both your clients needs and beliefs as well as the intended audience's needs and beliefs--is that true?
> MM: Right, I want to know what my client believes to be the most critical factor--I have to be able to evaluate something to give my client data that they believe--they have to have beliefs that are realistic.

5. Content Expert Review

Mary strongly believes in evaluating the accuracy of the content of the products on which she is working. Currently, she is completing a project for a research hospital by designing an instructional CD-ROM on breast cancer, so the importance of assuring the medical accuracy of the disc's information is vital to the project's success. Part of this process is getting all the medical staff involved in the disc's design to sign off that they have reviewed the accuracy of the disc's content:

> MM: We're currently doing a project on breast cancer. We meet with authentic users. We do all the formative evaluations as we go along. We evaluate what we call the high level design--we basically pass right by the higher-ups in the organization to get their approval. Your content must be accurate--it doesn't matter how slick your project is, how cool it is--if your content is wrong, that's the easiest way to fail. At the high level of design, we make all the higher-ups sign off on this content that it's accurate. I tell subject matter experts that your job is make sure that I always tell the truth and my job is to tell it effectively.

During the earliest design stages of the CD-ROMs that she develops, Mary also reviews the content of the disc with authentic users (e.g., medical students and health care professionals) to verify that the content makes sense by having them review the scripts:

> MM: All throughout product development we're involved in evaluation. We start at a very low level evaluation as far having someone else read the scripts...We evaluate the pieces as we go along--we evaluate scripts as they're on paper, we read them, and have others read them and have them read them out loud to others and ask them what makes sense, what was confusing.

6. Focus Groups

Mary uses focus groups before, during, and after user tryouts to gather evaluative feedback on her products. Focus groups are also a important part of her initial planning stages for the product, since she intends to eliminate the barriers to learning for her audience:

> MM: One of the things I do before I ever begin a project is I start with a focus group of the authentic users. I meet with them and I pose a number of questions to them. What I'm after is attitudinal problems. I basically have to identify all of the barriers to my audience accepting this information. What are the barriers to them accepting it. If you don't believe that smoking causes lung cancer, for instance, then that's a barrier--an attitudinal barrier. I want to identify every reason that someone in the field would look at my product and discount it--whether it be a piece of the lesson, the facts, or the attitude. So I meet with 10-12 people representative of this authentic audience.

7. Conferences and Seminars

Although Mary presents her products at conferences and seminars throughout the country, she only gains a limited amount of usable feedback from them. She was very adamant that the reactions of users at conferences are often quick and unqualified. Instead, Mary relies on more reliable, qualified data:

> MB: Do you attend seminars and conferences to present your products (i.e., CD-ROMs) and to solicit feedback from your target audience? If so, have you used feedback to make changes to your products (and can you give specific examples of changes if you did use them)?
> MM: Yes I do. I make changes based on what is found in evaluations. People at conferences make comments for a number of reasons (probably a good dissertation topic) often those who "appear" to be experts have done nothing similar. While I always listen to comments of my colleagues--some of them are valuable and some of them are bull. I am not about to design to please the "masses" which attend conferences. I guess I would be surprised that anyone would do that. Who at a conference has spent enough time with a product to pass a fair judgement? While I can walk a trade show and give suggestions for design changes, most often I notice the design did not meet the basic requirements of good design. Perhaps some people who don't really engage in evaluation rely on the audience in a conference presentation to lead their designs--I simply don't. I have more reliable data to use.

8. Surveys and Questionnaires

As part of the focus groups that Mary conducts, she often uses exit interviews, as well as surveys and questionnaires, to obtain evaluative feedback:

> MM: In our evaluations, we almost always do triangulation methods, exit interviews, written evaluations, we also evaluate electronically--what they did, where they went, what their attitudes were, what they reported. We evaluate records even if there are no records being reported to that instructor. As a designer, I'm interested in how long the user spends and what they reported (through exit interviews, questionnaires, or through written formal evaluations).

For the current CD-ROM that Mary is developing on breast cancer she intends to

correlate the information from the surveys, questionnaires, and exit interviews to learn more about how she can improve the product. In creating user questionnaires, she wisely reviews the questions with the client to improve them:

> MM: There are 3 forms of evaluation for this product: written evaluation for all the students, a tracking for all students on the computer, and we will have exit interviews and surveys of 10% of the students at every one of the 5 sites. So we will take that material and correlate it and see what we know.
> MB: That's very interesting. Tell me more detail about each of these 3 forms of evaluation.
> JJ: For the written parts, we work with the client in answering their questions. For instance, one questions was, "Did you learn anything in this program that you are apt to apply to your own life?" That wasn't a good question and we remodified it--"Did you learn anything here that you intend to integrate--to change your behavior. Did it talk to you?" The very last question my client had written was, "Write any additional comments here." We remodified that to a statement that works--"Is there anything more you would like us to know." That's always a question that I normally end a written evaluation with. Every word has to be carefully written.

Summative Evaluation Methods Used

1. Market Feedback

While Mary uses summative evaluation as a vital means of obtaining market feedback, she still maintains that impact evaluation is most vital to the success of her projects. She believes that summative evaluation focuses more on immediate impacts of the project such as knowledge gained and user reaction, while impact evaluation would assess longer-term effects of the product on the user and therefore be more valuable:

> MB: Tell me more about your summative evaluations--once your product is out there.
> MM: Well, the same things. Summative has a neater end to it than impact. Summative is an assessment of the immediate impact of the total project. So when the project is finished, the class is finished, you can finish your summative evaluation in 2 weeks. Impact evaluation may go on for 6 months to a year--you're looking at attitudinal and behavioral shifts, not just knowledge gained. Summative

> is often focused on knowledge gained and immediate reaction to the material. Summative has less interest to me, because I know that I can get you engaged long enough to teach you something in the short run. I really want to know whether or not I can change your behavior. That's kind of my personal preference.

Another summative evaluation method that Mary uses is product testing. She does this to ascertain that the final product can still work on various computers and platforms. Mary discovered that bugs in the CD-ROM are often a result of user misunderstanding or an error with the computer itself. She emphasized the importance of capturing all of these user reactions during pre- and post-testing of the product:

> MM: At the end you design an objective to decide whether a project met its goals and you look at methodologies. And also you design into your evaluation an opportunity to gather the kind of data that would have had an impact on that one way or another. If people don't appear to like it, why not? It may not have anything to do with the project-- it may have to do with the fact that the computers didn't have audio cards and they may not realize that. One woman said the project didn't do anything for her and as an evaluator I said, "Tell me more," not "Oh, really. What's the problem!" All you say is, "Tell me more." Eventually I found out that her computer didn't have an audio card and that was the problem. Every product we have done since then, when you start the disc, big letters come across the screen saying "this product requires audio--check this and go here if you don't."
>
> MB: I've learned that in some projects, it's not always the product that was bad, it may have been some kind of misunderstanding from the user.
> MM: If you see that a barrier to your successful evaluation, you better go get that data--"What machine did you run this on, was it a 386, 486, pentium, was it a single or three speed CD-ROM player?" If that could impact your evaluation, then you better capture that data. So you evaluate to find out what were the objectives, but you also are clever enough to capture the variables that will impact how someone viewed that project.

Overall Views and Uses of Evaluation

There were four sub-categories of overall uses and views of evaluations which

Mary mentioned in her interview: financial constraints for evaluations; evaluation results used to make changes on CD-ROMs; overall, that evaluations were helpful for the product; and made suggestions for improving evaluations.

1. Financial Constraints for Evaluations

The small size of Mary's company can be constraining in performing more comprehensive evaluations; however, she spends money very astutely during product development and evaluation, so she has fewer financial problems. She accomplishes this because she asks the client up-front what their expectations and time frames are, and makes certain that they are realistic:

> MM: We need to meet the needs of the funder--I make the funder write down, as I ask them, "What will have to happen with this project so that when it's over, you will feel that it's successful." That tells me right away what is foremost important in their minds.
>
> For me to take a project, I have to believe that a client has realistic expectations that can be accomplished in a realistic time frame. Those two things are vital to me. That's why most projects fail--unrealistic expectations and unrealistic time frames for impact from the client.

2. Evaluation Results Used to Make Changes on the CD-ROM

During the interview, Mary offered several ways in which the evaluation information she received was used to make changes on the CD-ROM on breast cancer. One of these changes was completed by modifying the user interface:

> MM: Overall, we took 60 students using the project...We evaluated their initial hypothesis and their ending hypothesis, as well as all their comments. What came out of the initial evaluation was that the students didn't like having to do go through a lot of sections (background information) about cancer--so we redesigned the interface so that there wasn't a required section up front about the natural history of cancer. You could go right into the case and then if you needed

> reference materials you could go back to the reference materials and learn it. Now, we'll take all those changes, modify it and we're putting it to test in 5 medical schools--giving it to medical students.

In user tryouts with the CD-ROM, Mary was told by medical students that she simply had to present the material in an organized way on the disc--that they would learn better that way:

> MM: Administrators still want to know if students feel that it's a better than a lecture. Medical students, as you know, are different learners. We found that some of the students fretted over was, "Look, you just have to tell me that which I have to know and I'll learn it"--they fretted over what it was that they had to learn. They used the term "tell me the bottom line." From that, we organized the disc in an appropriate fashion.

3. Overall, Evaluations Helpful for Product

Toward the end of the interview, the researcher asked Mary whether she believed overall that evaluations were helped in improving her product. She strongly emphasized that they were critical in the development process:

> MM: Of course our evaluation efforts are helpful! Why would anyone do something that is not helpful in the development of a product? Evaluation is an exercise that is completed to make a product better--regardless of the product. Anyone who cares about the quality of their work should engage in evaluation. If you do not openly evaluate your work, others will. It is far better to know how your work holds up and compares to others in the field before someone on the outside tells you. Evaluation is helpful.

4. Made Suggestions for Improving Evaluations

It was clear to the researcher during and after the interview that Mary's emphatic statement about the usefulness of evaluations for her products supported her overall view that evaluations were critically important. Moreover, her background as an educator

influenced her opinion of how evaluations could be improved. Critical in that view was satisfying the needs of both the client and the user:

> MM: A project has multiple objectives and different layers. Too often, I think young developers misrepresent the objectives of a project in a very narrow focus. We need to evaluate the product in such a manner that our client can learn what they need to know. As a designer, I'm always interested in designing an evaluation to learn what I most need to know. I'm interested in the design of materials which have an impact on the affective domain, not just the cognitive domain. It's much harder to design a product in the affective domain. We have neglected the affective domain because it is very difficult to evaluate whether or not we can change attitudes.
>
> When we evaluate a product we're most interested in the impact of the product on the learner's attitudes as well as the impact on their knowledge.
>
> We've done a poor job of evaluating--we have evaluated for our own needs and not the needs of the learner.

In her years of experience as an instructional designer and educator, Mary realized that she had to meet the needs of the users, rather than selfishly meeting her own needs during the project:

> MM: As evaluators, we have preconceived ideas--we want to design an evaluation to validate our beliefs, and sometimes that doesn't always happen. You have to be willing not to have your beliefs validated--to learn from your mistakes, to say, "I was wrong there."
>
> MB: Certainly your background in education and in computer science helps with your evaluations.
> MM: Plus strong motivation.

Mary consistently emphasized the need to test the product with the intended (authentic) user audience. In that role, talking to users and getting their feedback and opinions were critical to the success of the project. Foremost in this approach was the impact of the product over time:

> MB: Are there specific evaluation methods that work best for you, that you almost

always use or that work best for you?
MM: Talking to the end user. I think that in one way or another you have to communicate with them. They have to find a way to express their concern either in writing or verbally. You have to find a way to talk with the person who is using the product. And then you have to begin to look at the impact over time--I'm interested in impact evaluation--looking at attitudinal change.

Analysis of Case Study

It is evident that Mary's training in education and computer science were enormously influential in how she conducts and views evaluation. She was perhaps more critical of the evaluation process than any of the other five developers interviewed for this study.

Near the end of the interview, the researcher took into context all that Mary had said during the interview and concluded that modification was one of the keys to her evaluation approach. Mary was asked her if she agreed with this view:

MB: The heart of your evaluation process is, then, modification...
MM: We modify our approaches based on our clients' and users' needs. I don't think that there's any one question that you have to ask or any one treatment that you have to use. I don't think you have to record people on video or interview them, there may be appropriate uses for that, there may be appropriate uses to track the data electronically, and there may be times when it doesn't make any sense.

In summarizing her evaluation approaches toward the close of the interview, it became clearer that the goals of both the project and the client were most influential in her projects:

MB: Before we go, let's just summarize your evaluations--your key points of what you do in your evaluations, step-by-step--your main goals.
MM: I always try to address my personal goals, but my personal goals must be secondary to the project's goals. The project's goals must be clearly articulated early on--I want to know the stated goals and the unstated goals. I want to know

those things on the table and that aren't on the table. You have to be very astute--the most difficult project is one that is driven by a committee because they all look for something different, so I say I want to work with just one spokesman from the committee.

The objective of the evaluation should be to assess how well the project met its goals. Therefore, depending what the goals were, the project should be adjusted accordingly. On one project the objective was to provide motivation for students to learn about chemistry, not to teach them about the periodic chart. The objective was can we excite them about chemistry. It is much easier to evaluate knowledge gained--can you add to knowledge. Since we evaluate attitudes, we had to write an evaluation which uncovered attitudinal changes--one of them being how long it took them to interact with the product, given the freedom to come and go--how long will the student use it. It's harder to evaluate attitudinal changes.

At the end you design an objective to decide whether a project met its goals and you look at methodologies. And also you design into your evaluation an opportunity to gather the kind of data that would have had an impact on that one way or another. If people don't appear to like it, why not? It may not have anything to do with the project-- it may have to do with the fact that the computers didn't have audio cards and they may not realize that.

Ultimately, Mary believes that most projects fail due to unrealistic time frames and unrealistic expectations. For this reason, she always asks the client at the start of the project what they have in mind regarding these two views. This approach continues to work well for her:

> MM: For me to take a project, I have to believe that a client has realistic expectations that can be accomplished in a realistic time frame. Those two things are vital to me. That's why most projects fail--unrealistic expectations and unrealistic time frames for impact from the client.

Reflections on the Six Cases

In reading through the cases, it is apparent that the six developers interviewed for this study often equate many of the following terms and concepts with evaluation or as vital components of their evaluations, frequently using them interchangeably: marketing

research (marketing feedback); product testing; and product development.

By exploring the major categories that emerged in the interviews, this chapter detailed, in-depth, the opinions and views of the six multimedia developers interviewed for this study. Through a cross-case analysis, Chapter 5 further analyzes the cases presented in this chapter by examining categories that were mentioned by all six developers as well as the categories not mentioned by all six developers. It will also discuss some limitations of the study, how evaluators and developers can learn from each other, and make some recommendations for improving evaluations of CD-ROMs.

CHAPTER 5

CROSS-CASE ANALYSIS, DISCUSSIONS AND IMPLICATIONS

Introduction

This chapter presents a cross-case analysis of the cases in Chapter 4 and examines, in greater depth, the categories that were mentioned by all six developers as well as the categories not mentioned by all six developers. The researcher went back and forth between the categories which emerged from the cases studies and also reviewed the research conducted for this study to further analyze all the categories. During the content analysis, a cross-case analysis matrix was created to detail and summarize all the categories and to show how the responses from the individuals fit into those categories.

The limitations of this study are also enumerated. Two ways in which program evaluators and multimedia developers can establish mutual learning partnerships are examined: one, through recommendations and suggestions for increasing the utility of formative evaluations for instructional media; and two, through providing examples of translating existing educational evaluation approaches to evaluating CD-ROMs. Finally, conclusions of this study and suggestions for further research are provided.

Cross-Case Analysis

Through a cross-case analysis, this section will focus on the categories mentioned by all six developers as well as the categories not mentioned by all six developers. Table 7

presents a summary of the two major categories, their sub-categories, and their distribution among the participants.

Table 7:
Cross-Case Analysis Matrix

EMERGENT CATEGORIES	Jackie	Bob	John	Gwen	Jack	Mary
TYPES OF EVALUATION PERFORMED						
1. Formative Evaluation (Methods Used):	✔	✔	✔	✔	✔	✔
Needs Assessment				✔	✔	✔
Product Testing (Beta, Alpha, Platform and Cross-Platform)	✔	✔	✔	✔	✔	✔
Partnerships with Educational Institutions	✔	✔	✔	✔	✔	✔
Target Audience (User) Tryouts	✔	✔	✔	✔	✔	✔
Content Expert Review	✔	✔	✔	✔	✔	✔
Focus Groups	✔		✔		✔	✔
Conferences and Seminars	✔	✔	✔	✔	✔	✔
Surveys, Questionnaires	✔			✔	✔	✔
2. Summative Evaluation (Methods Used):	✔			✔	✔	✔
Market Feedback (Interviews, Surveys, Post-Release Testing, Conferences)	✔			✔	✔	✔
OVERALL VIEWS AND USES OF EVALUATIONS						
Financial Constraints for Evaluations	✔	✔				✔
Evaluation Results Used to Make Changes on CD-ROM	✔	✔	✔	✔	✔	✔

Table 7: **Cross-Case Analysis Matrix**						
OVERALL VIEWS AND USES OF EVALUATIONS (Continued)	**Jackie**	**Bob**	**John**	**Gwen**	**Jack**	**Mary**
New Synergies Created Between Divisions of Company for Evaluation	✔			✔		
How the Development and Evaluation Processes Differ Between CD-ROMs and Textbooks	✔	✔	✔		✔	
Overall, Evaluations Helpful for Product	✔	✔	✔	✔	✔	✔
Suggestions for Improving Evaluations	✔		✔		✔	✔

This table provides an overview of how the responses from the six developers applied to each of the categories and their sub-categories. An overall look at the table shows that the participants' interpretations and construction within these categories differ somewhat. Therefore, it is important to explore the categories that all of the participants mentioned in their interviews, and the categories that were not mentioned by all six participants.

The Two Primary Categories

Types of Evaluation Performed

Formative Evaluation Methods Used

During the actual production and development of a CD-ROM, many of the developers interviewed in this study used various formative evaluation methods to gather

feedback about their discs which were used to improve them. Some of these approaches were similar, while others were not, although all six developers mentioned that they conducted formative evaluations of their products.

Summative Evaluation Methods Used

Another important evaluation method used by the six developers was summative evaluation--getting feedback on their products once they were already released. This type of feedback is essential for making revisions on the next product upgrade and for overall product improvement. Interestingly enough, most of the developers view market feedback as the most important source of summative evaluation feedback, thereby equating marketing with evaluation. The methods used by the developers and their companies for summative evaluations were interviews with users, surveys, post-release testing, and feedback at conferences. Several developers in this study stated that publishing companies also use sales figures as a summative evaluation method, as this chapter will explain. Overall, four of the six developers mentioned that they used summative evaluative approaches for their CD-ROMs.

Overall Views and Uses of Evaluation

This primary category was influential in exploring the opinions and intended uses of the evaluations conducted by the six developers. Of the six sub-categories, each of the six developers mentioned two sub-categories, while four sub-categories were not mentioned by all six developers.

Categories Mentioned By All Six Developers

Eight categories, within the two major (primary) categories, which all six developers mentioned in their interviews emerged from a cross-case analysis. Some have a host of sub-categories that provide an in-depth look at the particular topic. A total of eight sub-categories emerged in which all six participants made a reference. In the first primary category, there were five responses which all the developers mentioned: 1) formative evaluation (specifically, the use of: 2) product testing, 3) partnerships with educational institutions, 4) target audience used for evaluations, 5) content expert review, and 6) conferences and seminars). In the second primary category, the six developers all referred to: 7) the use of evaluation results to make changes on the CD-ROM, and 8) an overall view that evaluations were helpful in improving the company's products.

1. Formative Evaluations

Each of the six developers interviewed believed in the importance of formative evaluations for their products. They all recognized it as an integral method of improving the CD-ROMs that they had developed. This helps to verify the importance of formative evaluations for the development of CD-ROMs as was hypothesized in the literature review section of this study. The most decisive commentary on the importance of formative evaluations came from Mary who stated:

> MM: Of course our evaluation efforts are helpful! Why would anyone do something that is not helpful in the development of a product? Evaluation is an exercise that is completed to make a product better--regardless of the product. Anyone who cares about the quality of their work should engage in evaluation. If you do not openly evaluate your work, others will.

While all the developers interviewed believed in the importance of formative evaluation, there were only several methods which they commonly used: product testing, partnerships with educational institutions, target audience (user) tryouts, content expert review and conferences and seminars. Each method that was similar is described below:

2. Product Testing

As the researcher had hypothesized, all six developers used product testing in their formative evaluations. These included beta, alpha, platform, and cross-platform testing. The testing process was essential for the developers to verify that the CD-ROMs were free of bugs and would not crash the user's system. Testing the product in the formative stages holds obvious importance in ensuring the disc's functionality. Moreover, testing can save the developer a great deal of money and embarrassment in the long run.

Gwen used various forms of product testing to certify that her CD-ROM worked properly and that the content was accurate. She even defined a term (premium beta testing) which was used to test the disc at another level beyond the beta stage. This comprehensive product review, both internal and external, attests to the importance she has for product testing:

> MB: Do you term alpha testing as internal testing then?
> GM: Yes, our alpha testing was internal on this project. Beta testing was internal and external. We actually paid people to go through item-by-item on the disc, to look at it very carefully, to make sure that it all functioned right, and that the content was good. From a test standpoint, that was important to us because we needed to make sure that the items were good measurers and that they were valid and reliable...we still wanted to make sure that it functioned well and interacted well. We were interfacing with a lot of different programs on the CD, so we didn't want to put it out there without it going through a rigorous evaluation--we wanted to make sure it worked well...we did a *premium beta review* which was internal

> which checked all the program content. That's when we put it on a bunch of different platforms to make sure that we couldn't crash it anywhere.

3. Partnerships With Educational Institutions

Since the six developers interviewed had produced instructional CD-ROMs, which are intended for an educational audience, it was critical that they work with educational institutions in testing, developing, and evaluating their products. It was also beneficial to them because these partnerships helped the developers and their companies save an enormous amount of money by testing their products with students and instructors in high schools and colleges. Most developers, in fact, worked with these institutions to gather feedback in their evaluation processes as a marketing ploy, which also enticed user interest in the product. Jackie used this approach very effectively, resulting in word-of-mouth advertising by students and teachers:

> JK: What we really do is use this [student and teacher tryouts] as a marketing ploy because if people are involved in giving you feedback at an early stage, they're the people that are going to be talking about it, they're the people that are most likely to buy it.

Gwen worked with Education Centers throughout the world in establishing partnerships. These centers distributed her disc to students for free and also helped the students to learn more about the company's certification program. Through her contacts with students, Gwen would eventually get their feedback on the disc:

> GM: Our authorized education centers distribute it [the CD-ROM] and there is a very nominal charge ($1, $1.50) to the education centers. It's free to our end-users...the reason we wanted to have our education centers in the loop is that it will help people who really don't know the best way to be educated into the our certification program--the CD gives them that open door.

4. Target Audience (User) Tryouts

Another critical formative evaluation method used by all six developers was getting their target audiences (users) to tryout the product. While some developers do use members who are not involved in the target audience, they realized that the most critical feedback would come from the actual targeted users of the product. Additionally, the developers knew that by getting their target audience to tryout the CD-ROMs, it would simultaneously arouse interest in their company's products. Through user tryouts, the developers seemed more intent on getting their users to like the CD-ROM, rather than to learn from it.

In all of the six cases, the target audience of instructional CD-ROMs, as expected, were students and teachers, and sometimes even parents of students. Mary was most fervent in her belief that intended users of her products were essential in the evaluation process, noting that most evaluations had ignored their target audience and were too narrowly focused. She emphasized that using students during the evaluation helped to add value to her products:

> MM: One of the biggest criticisms of evaluations is that they are done in too narrow a group for too short a period of time, with something less than authentic users--authentic users are the users who this product is actually made for...Many of them really don't care about the success of the product. It's not just about numbers. I'm not interested in whether a student thinks a dictionary is informative --it matters more whether it's of use to them and their chosen field. I want to know if our products add value to the user.

While Jackie's primary target audience was teachers and students, she saw the opportunity to market her products to parents of students as well. She hoped that some of these contacts with parents would be established through the teachers themselves, through

public relations campaigns, and through the use of the Internet:

> MB: Let's just briefly summarize what your company does in terms of its evaluations. You go out to teachers, professors, high schools and community colleges, give them your product for a few weeks and get their feedback. You have people on payroll--content coordinators, retired teachers, designers. You have focus groups...anything else?
> JK: Right, we do focus groups, we bring in other people throughout different stages of development. We might be sending out paper, we might be sending out a prototype, we might be sending out an Alpha version with free lessons on it. There's definitely different stages, so it's really those market people. Once we have a demo of something that has been knocked around that we want to show people then we send it out with our p.r. campaign and try to get people to write about it.
>
> MB: Mostly you send it out to teachers.
> JK: Yes. As far as after that once the product is completed and done for a version upgrade or for just establishing a presence, we set up field testing and getting a lot of students feedback as well as teacher feedback. And that ultimately is what is important.
>
> What we're also trying to do is with our Java-based stuff is getting into parents places and sites, education sites where...a lot of that is growing on AOL, we're trying to work at it from both ends. Also, ultimately getting a high school teacher using it or saying, "It's a good program, and I evaluated it"--that's for parents of kids who don't do well in math--to go to the math instructor and say "What else can I buy for my child to help them?" That word-of-mouth from teachers to parents is what we're going for as well.

5. Content Expert Review

All of the six developers would not risk the threat of having their content go unverified before they released their discs. To verify the accuracy of the content of their products, the six developers all worked with content (subject) experts in the area of the discs that they were producing: Jackie worked with math education experts; Bob limited his content review to a few accounting professors externally (since he was a content expert himself, he verified the accuracy of the CD-ROM); John worked with medical doctors at the university hospital; Gwen had employees within her company that were

experts to help her assess the disc's accuracy; Jack is a content expert himself, so he, like Bob, sought only a few professors to verify the content of the CD-ROM; Mary worked with medical students and the medical doctors at the university for which she was developing the disc.

It was noted that those professionals who were content experts in their areas or who had other experts on staff (i.e., Bob, Gwen, and Jack) helped save the companies they worked with a considerable amount of money since they did not need to hire external content experts to verify the content of their discs. Gwen explained the importance and advantages of having content experts in-house:

> GM: David, who is my boss, is a Ph.D. in Psychology and is a psychometrician who makes sure that the tests are accurate, valid measurements in that they relate well to what people do out in the real-world. Out test developers are industry experts--they're content experts and then we have trained them to write good test items. They are full-time employees of our company.
>
> There was a lot of content review. Messaging is extremely important to our company and our Education Division, making sure that the information we put out is as accurate as it is humanly possible at the time of release.
>
> As a general policy, our Education Division has a very standard review process starting with the project approval committee and ending with people specified that they have to sign off and approve anything that goes out to the field. So a lot of it from a content, look and feel, messaging standpoint is dictated to us. It's very important to our company and to Education that products go out looking the best they can and working well. It's standard for us to adhere to those guidelines.

It was also revealing how others (i.e., Jackie, John, and Mary) utilized their partnerships with educational institutions to verify the content of their products, while also saving money for both their company and the client. Jackie, for example, used a limited number of honorariums with teachers and focused more on gathering feedback from instructors who were interested in her product, offering them free discs for their insights:

JK: What we did find even in our process with getting table of contents feedback and prototype feedback is that a lot of people were requesting honorariums of $50-100. But we, being a small company...couldn't afford to do that at the stage where we were. So basically, we sent it out not saying we'll give you a $100 if you do this. We basically sent it out to 50 people that were people that we had good contact with that had liked our *Trig* program, or had liked what they had seen, or had mentioned to us somehow that they were interested in testing our *Algebra* program...Or we'd work something out with them to get a free copy of it when it's released.

...We might not actually have the resources to implement all their suggestions, but that's what they're getting--they're getting involved in creating a new multimedia program that teaches algebra in a way that would suit their needs. That, as well as it's free for however long it takes for them to get it on their network to get their student looking at it.

6. Conferences and Seminars

Each of the six developers gathered evaluative feedback on their products when they presented them at seminars or conferences, even if that was not their original intention. Since each of the six developers are experienced professionals who are involved in a number of professional organizations and institutions, it was natural for them to present their CD-ROMs at conferences sponsored by these organizations. Moreover, presenting their discs at conferences was a good promotional and marketing vehicle for them, especially for Jack and Jackie, whose companies sought various strategic means of marketing their CD-ROMs. Jackie, in fact, made sound use of one conference she attended by conducting a videotaped focus group of users as they tried out the disc:

JK: We had a focus group at NCTM (National Council of Teachers of Mathematics) in April and it was 10 people and we paid them. This is what we paid them $100 each for participating--both in terms of doing a pre- and post-survey as well as talking to us for about 3 hours. We have it on videotape. We showed them a prototype, got their feedback.

However, Bob, John, and Mary rarely sought to gain evaluative feedback on their discs at

conferences and seminars. In fact, all three of these developers made minimal changes to their products from user suggestions at conferences. Although John, for example, used conferences to present his products, he still received evaluative feedback on the disc at conferences, even though he believed it was not the ideal place:

> JJ: I've shown the program to my own peers and at conferences with poster sessions...So, I try to solicit input and take notes from those peer demonstrations. At these types of conferences you receive input from instructional designers and doctors.
> MB: Have you done this before and got feedback on this disc?
> JJ: ...It's not the ideal situation to get a lot of feedback at a poster session because people are grazing looking at many posters. I've gotten some feedback generally positive like, "It was very easy to use" and "When will I be able to get a copy?" It has not been of tremendous use from the evaluation standpoint.

Mary was even more resolute that conferences were not an ideal place to collect evaluative feedback on your products, even though she, too, used some of these comments to make changes to her CD-ROMs:

> MM: People at conferences make comments for a number of reasons...often those who "appear" to be experts have done nothing similar. While I always listen to comments of my colleagues--some of them are valuable and some of them are bull. I am not about to design to please the "masses" which attend conferences. I guess I would be surprised that anyone would do that. Who at a conference has spent enough time with a product to pass a fair judgement?...Perhaps some people who don't really engage in evaluation rely on the audience in a conference presentation to lead their designs--I simply don't. I have more reliable data to use.

The casual conversations that Bob had with users at conferences often helped him with some feedback on his CD-ROM, although he primarily used this information to determine if consumers liked his disc or not:

> BH: I talk to everyone at the conferences and get their reactions--mostly accounting professors. Most of it is very positive.

In the primary category of overall uses and views of evaluations, there were two sub-categories which all six developers mentioned: evaluation results used to make changes on the CD-ROM and overall, that evaluations were helpful for the product:

7. Evaluation Results Used to Make Changes on the CD-ROM

A primary goal of this study was to explore the utility of evaluations for CD-ROMs. To accomplish this, the researcher asked the six developers to give examples of some changes that they had made to their CD-ROMs which were elicited during the evaluations. Each of the six were willing to provide examples to verify that the evaluation results were used.

John was most specific in offering examples of changes he had made to his CD-ROMs from evaluations. One of these changes allowed students to bookmark where they left off in the program as they progressed through it:

> JJ: One of the interface changes was to put a check mark by everything that they completed so they would remember what images they had worked on within a given session. The CD does remember what they've done in any given session.

In his evaluations, Jack had discovered two critical ways in which students might have problems in learning the content on the disc. One was that students were only using the glossary to learn, and another was that they continued to answer only the multiple choice questions to test their knowledge of the subject:

JL: What we found students doing from the evaluation is that they would actually

> go to the glossary on the CD-ROM and as you click each word, a definition pops up. They would study that way. Now, what we did was we took that information, pulled back and set up our glossary chapter by chapter so that the students don't have to go through all the words in the glossary. They can only go through the words in the glossary from that chapter.
>
> Here's another one: these are behaviors that we observed from the evaluations. The first thing students would do when they opened up a new chapter is go do the multiple choice tests at the end of the chapter. The multiple choice tests on the CD are interactive, so when you click on something and get it right, it tells you you've got it right; if you get it wrong a little button pops up at the bottom that says "click to review." And when you click it you go to the page in the textbook that has the answer. So what the students would do is take the multiple choice tests, figure out what they didn't know, and then they would go and study those sections of the chapter and breeze through the other sections. It was an incorrect form of learning. That's something we learned only because of these evaluations. We had no idea this was happening before...However, we can't stop that, but what we can do is make longer, better, more comprehensive multiple choice questions at the end of every chapter. We're going to take the bad behavior and capitalize on it, and push them in the right direction. It also means that we have to raise the levels of the questions at the end of each chapter, because I want them to miss more now.

Mary had discovered that some novice computer users were not familiar with sound cards and other types of hardware, nor did they realize that they needed them to make the CD work properly. Since receiving this feedback, she has made one specific modification to every disc she has developed:

> MM: One woman said the project didn't do anything for her and as an evaluator I said, "Tell me more..." Eventually I found out that her computer didn't have an audio card and that was the problem. Every product we have done since then, when you start the disc, big letters come across the screen saying "this product requires audio--check this and go here if you don't."

8. Overall, Evaluations Helpful for Product

This sub-category was similar to the previous one--to explore whether the six developers truly believed that evaluations were helpful and useful for their products. All of

the developers in this study believed that, overall, evaluations were helpful for the product. This opinion attests to their view of the importance of evaluations for product development and improvement.

Mary was most adamant about evaluations being helpful for their products. She also offered a possible answer to a problem raised in Chapter 2: why evaluations results from software companies are infrequently released to the public--because they fear being evaluated and criticized themselves by the public (by both those in and outside of the software industry):

> MM: Of course our evaluation efforts are helpful!..Evaluation is an exercise that is completed to make a product better--regardless of the product. Anyone who cares about the quality of their work should engage in evaluation. If you do not openly evaluate your work, others will. It is far better to know how your work holds up and compares to others in the field before someone on the outside tells you. Evaluation is helpful.

Gwen and her company were so pleased with evaluation of the first edition of the *Guide* that she has already broadened her next evaluation:

> MB: So would you say that overall, then, evaluations have been helpful for your product?
> GM: Yes. In fact, we probably went further and broader with the evaluation of this next revision because the information was so valuable from the first revision.

Since John views evaluation and development as contiguous and iterative processes, he strongly believed that user and expert feedback were critical and natural parts of the process:

> MB: Overall, if you had to look at your evaluation processes, have they helped you with the disc?
> JJ: Oh, yes. As I mentioned before, I really don't see evaluation a separate component--it is simply part of the development cycle. I cannot conceive of

> developing something without a constant flow of information from users and from other content experts. It's just a natural part of the process. I couldn't work without it.

Categories Not Mentioned By All Six Developers

Of the two primary categories that emerged from the cases, eight sub-categories emerged which were different among each individual interviewed. The four sub-categories within the first primary category that were not mentioned by all six developers were: 1) needs assessment, 2) focus groups, 3) surveys and questionnaires (all three being formative evaluation methods used) and 4) summative evaluation. The four sub-categories within the second primary category that were not referred to by all six developers were: 5) financial constraints for evaluations, 6) new synergies created between divisions of the company for evaluation, 7) how the development and evaluation processes differ between CD-ROMs and textbooks, and 8) suggestions for improving evaluations.

1. Needs Assessment

A needs assessment was only conducted by three of the six developers. In large part, this was due to the fact that three developers (Jackie, Bob, and John) already knew that there was a market for their products, which made a needs assessment unnecessary. As Jackie explained, there was such a strong need for any math product that would help students with the basic concepts, that the need for the product had long existed and continues to exist, especially due to the high failure rates of math students:

> JK: ..Actually what we're also providing is people that are fed up with 50% of their students failing Introductory Algebra at the college level. It's ridiculous that a successful department has a 50% failure rate--that students are just fed up and

> looking for other things to do. So, if you can put them in front of a math product and provide that, then that's our sale.
>
> ...We haven't thrown out big concepts mainly because there is such a tremendous need for anything that helps in this marketplace.

Bob declared that the market for his accounting product was already established as well:

> BH: So when we got ready to do our CD-ROM, we didn't have to test our market--we know there's a market for our product already and that we're the leader in that market (so a needs assessment wasn't needed).

Gwen, Jack, and Mary all did conduct a needs assessment because they believed it was necessary to know what the market was like for their products, as well as to ascertain whom their target audience would be. Once they had discovered who their target (user) market was, they vigorously sought ways to get those users involved in their evaluations, particularly with user tryouts and focus groups.

2. Focus Groups

Although some of the six developers believed that focus groups would be useful for their evaluations of the CD-ROMs they developed, others did not. Bob and Gwen said that they did not conduct focus groups for their products, but only Bob did not offer an explanation of why. Gwen, on the other hand, mentioned that the beta review process and phone calls from customers were so successful in her evaluation process, that focus groups were not needed:

> GM: We have not had focus groups, or any kind of meetings concerning this product. The beta review process, in which the input came back on an individual basis, was the method we used to gather evaluative feedback. And we get

comments from customers who have the CD and call in for support. We get good evaluative information from these calls that will be applied to the next rev.

Jackie, John, Jack, and Mary all stated that they had conducted focus groups, often in a very casual atmosphere which allowed users to speak freely about their opinions of the CD-ROMs. Some developers (such as Jackie, John, and Mary) used focus groups after user tryouts, so that they would have immediate information on the users' assessment of the disc. John explains how he performed these together:

JJ: As part of the evaluation process we used focus groups of about 3-4 students who would actually use the product and then we would sit down and have discussions with them about how it worked, what were the problems and so forth.

3. Surveys and Questionnaires

Of the six developers, only Bob did not use surveys or questionnaires in his evaluations. One explanation for this is that Bob gathered enough critical feedback on his disc from user tryouts and from casual conversations that he had with both students and professors. He did not see the need to develop evaluation instruments such as these to get feedback on his products, as he says:

BH: They [large publishers] totally rely on focus groups in which they ask professors questions. They do surveys--that's their approach to this market. We, on the other hand, do it very casually, but because I'm the subject matter expert when someone says something to me, it has a whole lot more meaning to me than them--it's direct information.

These companies never ask questions about the software in their focus groups or surveys--they know it's horrible! They all admit to that, but they don't tell their customers that. They ask in focus groups or surveys what do you want in an accounting text, but they never ask reactions about the actual software. Their representatives aren't content experts either--they're just sales people--they don't even know how some of these discs work. We don't have to worry about that.

Jackie, John, Gwen, Jack, and Mary all used some form of questionnaires and surveys for their evaluations. They recognized the need to have users give their written feedback on the CD-ROMs so that this information could be used as historical data whenever they needed it. Jackie, for instance, collected responses from user surveys, had them typed up, and distributed them to all employees of the company. She would also distribute and discuss these user suggestions from surveys at company meetings:

> MB: Which employees or divisions of your company conduct evaluations of your products?
> JK: We're a small company, so it's really a combination of both marketing and development that are leading these [evaluation] efforts--our field test coordinators getting feedback to see what's out there. It's really to ensure that our developers are getting all the feedback.
>
> Once we get all these surveys, what we do is have our office manager type in everybody's responses into one word documents so that we could see...depending upon when they come in, you kind of toss them around...Once you see them all together you're able to say, "Oh, these are the competitors' products that they're using; this is what they really liked about it; this is what our strengths are." And this really helps the selling process.
>
> ...What we did was we put together all of the surveys and basically it's a meeting where everybody takes a look at all of this stuff and gets to see it on their own. But that's definitely handed out--every single survey is handed out to every employee involved in development.

4. Summative Evaluations

Not all of the six developers used summative evaluation methods in their evaluations. Bob and John did not see the need to conduct summative evaluations, basing their belief on the fact that only textbook publishers used summative evaluations as a means of gathering market feedback on their products once they were released. Moreover, they both believed that these publishers used sales figures as their only means of summative evaluation, often neglecting other forms of evaluation feedback. John did not

use the terms summative and formative, stating that evaluation was a contiguous process:

> JJ: I don't use the terms formative and summative because evaluation is on-going and will continue to be on-going as long as the product is alive. So I don't make a distinction between formative and summative because a project like this is never complete. Software is never done unless you pull it off the market.

Soon after this, however, he stated:

> JJ: Well, if you want to use the terms formative and summative evaluations, the summative evaluations are the sales--that tells you whether the product actually met the needs of the users. Of course, there are a lot of other things factored in--the marketing processes, were the clients informed--you get a lot of mixed information with sales figures because there's a lot more than just the quality of the product...Their [publishers] evaluation methods are how many sales they make.
>
> MB: I guess that's how publishing companies get feedback--in term of marketing and sales.
> JJ: That's what drives the publishing business--sales. It's extremely important to them--it's probably their most important evaluation tool as well.

One other possible reason why Bob did not perform a summative evaluation may have been due to the financial constraints of his company. Without more financial resources, it may be impossible for him to conduct a more extensive evaluation of his products, particularly a summative evaluation, which is not as critical as formative evaluation.

Jackie, Gwen, Jack, and Mary all performed some type of summative evaluation for their products, sensing the need to gather more critical feedback on the CD-ROMs once they were released. Although Jackie worked for a small company, she still found the means to conduct a summative evaluation of her CD-ROMs. She also realized that there was a point where evaluation could be overwhelming at the end, both financially and time-

wise:

> MB: At this point (summative evaluation) do you find, then, for your company because it's so small, that it's too costly and time-consuming to perform--that's what it sounds like you're saying.
> JK: Yes, although it's definitely important, in order to sell more. But as far as just manageability of how much evaluation you want to do--there's definitely a point where it doesn't make much sense to throw more money into it even if you have all the money in the world. But...it's just a kind of a given in the school market that you have to do that. But one problem we're dealing with is that we don't want to be giving away our products in order to get feedback from everybody in order to get the sale.

Gwen used summative evaluations for the *Guide*, such as post-release testing, as a means of making improvements to the next revision of the product:

> GM: ...We had post-release review, too. We actually had people internal to our company use it who are on the development group to tell us how they would do it better (programmers, etc.). They're especially helpful when you don't have a deadline looming over you.

Company employees were, again, extremely helpful for assisting her with the summative evaluation:

> GM: ...So we tried to get a wide range--we selected who we wanted to send it to, to make sure that we had a range of people that might possibly use the tool. All of them were people who we had contact with before, through either our certification, beta testing, or customer service issues. They were all people who had some kind of level of contact with our company. A lot of them were our technical support people, who work on our technical support line for our company. They have a lot of interaction with the customers and had a lot of good input about what questions were more relevant and how it might work better. All of our workers are certified as well, so they know the product well.

In working with his publisher, Jack and the marketing representatives of the company solicited feedback from professors who used the disc:

> JL: The book rep's were queried about what were the problems associated with selling the book (adoptions) and what kinds of changes would they make in the

> next edition to improve that. About 10 book representatives from the publisher met with us for 4 hours. The rep's discussed the feedback they had heard from professors throughout the country about the disc.

He also used comments from other evaluations and evaluated his competitor's product as two primary summative evaluation methods:

> JL: We met just recently to discuss the second edition. We reviewed comments from previous evaluations we had conducted in order to make some changes. We also evaluated the competitor's product--it just came out. It's quite a different product.

Even though Mary does not view summative evaluations as vitally important, she did perform them. Primarily, her assessment of whether the project met its goals, as well as the modification of these goals, were her primary summative evaluation approaches:

> MM: The objective of the evaluation should be to assess how well the project met its goals. Therefore, depending what the goals were, the project should be adjusted accordingly.

5. Financial Constraints for Evaluations

Not surprisingly, the three developers who worked at small publishing companies (Jackie, Bob, Mary) all experienced financial constraints during their evaluations. This is a result of smaller budgets and fewer employees in these companies. Bob expressed his need to save his company money in the evaluation process:

> BH: We can't really sell the products for a whole lot of money, we don't have the sources or facilities to create super teaching materials, so whatever we do has to be self-contained...We have another product for writing that we allowed graduate business students to go through it to test it--to find errors. It doesn't cost us anything--we just give it out and solicit feedback from the students--we don't have a lot of money.

John, Gwen, and Jack all worked with larger companies that had many more financial and physical resources than the others. Jack, in fact, secured both an advance and a favorable budget from his publisher which were helpful for the disc's evaluation and development:

> JL: They [publishers] assured me of its priority and gave me an advance for it plus a budget that I could draw from for future royalties. They told me they wanted to be first in the market.

John utilized the resources of the university which he works for, and allowed the doctors who were writing the content of the disc pay for the student workers, while others freely volunteered their time:

> MB: And did you work on a specific budget--in terms of your costs.
> JJ: No, we pieced it together as best we could. In terms of my costs for my own services, that's never charged nor do I have to account for it--that would be very difficult to do. Costs aren't accounted for--one of the doctors came up with funds for the student workers. Everyone else contributed their own free time.

Gwen employed the resources of the company in evaluating and developing her disc, engaging the employees of the company to help her:

> GM: ...We had no budget at all. It's pretty impressive considering that.
> MB: So you really had no financial constraints at all?
> GM: We didn't actually spend any money building it, only people money. We only spent money in the actual production of the CDs. Since it was so well-used by our customers after the first revision, we received a budget for the second revision, but we haven't hired any new people--we're just using internal resources.

6. New Synergies Created Between Divisions of Company for Evaluation

This category was one of the most intriguing ones. Through two interviews, a pattern emerged--the creation of new synergies in the participant's company which had

never been used before; that is, different employees from different divisions or departments all working together on the evaluation. While this may not be surprising to some companies, it was surprising in these interviews because it had long been assumed by professionals in the field that it was commonplace to have only a few employees from one division within the company working on evaluations of their products. This, however, was not the case with Jackie's or Gwen's companies, as field test coordinators, programmers, office managers, marketing and development, and other employees worked together to evaluate their CD-ROMs.

In further analyzing this category, it is apparent that it varied by each respondent. John, Jack, Mary, and Bob did not offer any responses that hinted at new synergies being created between the divisions of the companies they worked with. Both John and Jack worked with large textbook publishers as well as with their own organizations in evaluating their CD-ROMs, while Mary worked primarily with her own company as well as that of the clients'. The small size of Mary's company was probably one factor why new synergies were not created--it was simply too small to involve a large number of employees, and furthermore they did not have many divisions within the company. John and Jack worked with larger textbook publishers in addition to their own educational institutions, but it did not appear that new synergies were created in either organization for evaluations. It appeared at first that new synergies were created in Bob's company as well, but since there are no divisions or specific job titles in his company (other than himself as President) and because of its small size, it would be difficult to conclude that new synergies were created.

Jackie and Gwen, however, made comments which led the researcher to conclude that a number of employees from different divisions of their companies were involved in the evaluation, thus creating new synergies. Jackie's company, although relatively small, still had several divisions (such as programming, marketing, and testing) which were all involved in the evaluation processes of the CD-ROM. Clearly, evaluation was a combined effort of all the employees. It was most apparent in Gwen's company that new synergies had been created in evaluating her product. Gwen astutely utilized all of the company's resources, at times using employees from technical support, customer support, programmers, and even a partner of the company, to evaluate her product. The clear, organized, and considerable divisions of the company, moreover, were enormously resourceful for Gwen as she saved a great amount of money for the group she worked for (Education). In reviewing all the employees of her company who assisted her with the evaluation, Gwen came up with a long list:

> MB: Tell me about some other employees who help out with the evaluation processes... You mentioned Marketing...
> GM: Yes. We had a test programmer who built the test and programmers who built the interface, the 3 of us made sure the program worked well. We had Corporate Communications and Education Communications who reviewed anything that had any messaging in it--as far as content--to make sure it was accurate and make sure it was professional and fit with our corporate standard. Then we had the marketing groups and the Marketing Program Managers for each of the programs to make sure that their courses were accurately represented and that all the information was correct. Tech support people did a lot of the question review and the beta testing, some of the customer service reps did that as well.
>
> MB: So tech people and customer support personnel were very helpful in relaying what the customers really wanted.
> GM: Yes.
> MB: So it's definitely a team project among many divisions of your company.
> GM: Right.
> MB: Your Education Group then would be the driving force behind the

evaluation...
GM: Right, the Testing Group within Education...We also worked with one of our partners to help us develop the items...We have a corporate localization service here that handles the translation and the review of the translation...to make sure that the terminology is easily translatable, that it won't translate ambiguously or offensively...We built it into whatever it's supposed to be--a test or the *Guide*, give it back to them and then they review it and its format for each language again to make sure its right.

7. How the Development and Evaluation Processes Differ Between CD-ROMs and Textbooks

Four of the six developers in this study mentioned ways in which the development and evaluation processes of CD-ROMs differed from those of textbooks. They were Jackie, Bob, John, and Jack. Not surprisingly, two of these developers (John and Jack) had worked with textbook publishers in developing and evaluating their discs, and they were clearly unhappy with the way that these publishers conducted evaluations. Both John and Jack realized that their publishers were mistakenly treating the evaluation of their CD-ROMs like that of a textbook. Perhaps this stems from the fact that their publishers had traditionally produced textbooks for so many years, this being their first venture into electronic publishing. Apparently, these publishers were unaware of the new challenges that interactive media posed, particularly with evaluation and development. Moreover, the publishers continued to treat their authors' discs as if they were textbooks. Jack explained this dilemma stating that his publisher discounted the fact that he was more aware of what the product needed, especially since he was the one who was most involved with its content and development:

JL: I think they [the publishers] don't understand the product well enough that they can get people to evaluate it at the level it that it has to be...they're trying to

be sort of arms length, including like we don't know who's on other end--I know who's on the other end!

They're approaching the evaluation like that of a textbook, as they always have, but we have something more than that here.

Although Jackie did not have to work with an external publisher, she had a clear understanding of her market for instructional CD-ROMs and knew how to make an impact in the market by drawing upon the strengths of her disc's graphics and interactivity. She had thoroughly researched her competitor's products and knew them well enough to realize what type of mistakes they had made in development and production. Some of those mistakes were made during the evaluation stages of the disc. One competitor, she realized, had been in textbook publishing for so long that they handled the evaluation and development of a CD-ROM just as they had with their textbooks. Clearly, Jackie thought this was a big mistake:

> JK: I think that the best textbook authors obviously make the worst multimedia authors...Harper Collins has a program called *Summit* which is an Algebra tutorial program and they... that's basically their textbook on the subject--it's the same thing and it's pretty clear that they don't know how to write multimedia content. They think, "Oh, this is something big that we should enter this market program"--that's where we stand out with our videos, graphics and learning and knowing how to get around the program. That has definitely been our strength. Even if we can't match up with being able to have unlimited customer problems...there's definitely scoping of features that we have to do as a small company, even if we a list of--yes, we need audio versus not putting audio in there.

8. Made Suggestions for Improving Evaluations

In my conversations with the six developers, some of them were very open in expressing their opinions of how evaluations could be improved, while others were not. As expected, most of the developers who came from an educational background or

worked with an education institution (and therefore were well trained in meeting the educational goals of the project), were more critical of some evaluation approaches than others. For instance, John, Mary, and Jack were more uncomplimentary about some evaluation approaches than others and the ways in which they were conducted by developers, evaluators, and publishers. John mentioned that software developers, as an example, did not work with their target audiences and were more technical than humanistic in their evaluation and development approaches:

> JJ: My background is in education. Software development evaluation is...I've never gained a lot from that methodology. They don't tend to work a lot with their own users...
> MB: They really don't work with their target audiences...
> JJ: Yes. They tend to think more in technical terms than in humanistic terms. Basically, taking educational design theory and matching it with real-life situations --Dick and Carey's book is pretty simple to use, but there's not a model out there that actually shows the processes that anybody uses. The processes in the textbook are not what's used...
> MB: The evaluation processes and the design processes...
> JJ: Yes, both.

Mary was similarly critical of how evaluations were approached by some developers and evaluators, offering ways that they could be improved. She maintained that many projects often ignored the needs of the users and the clients, and instead selfishly focused on their own needs instead:

> MM: I learned from an evaluator that you should only evaluate those things that you care about otherwise, you begin to trivialize your own evaluation. No matter who I deal with, they all come up with lousy questions and I reprimand them to write better ones. It's not an issue to me whether this [multimedia environment] is a more effective learning environment or not--it matters what the impact of the disc (product) is. We've done a poor job of evaluating--we have evaluated for our own needs and not the needs of the learner.

Clear, articulated goals and modification are the most important means by which an evaluation should be conducted, according to Mary:

> MM: I always try to address my personal goals, but my personal goals must be secondary to the project's goals. The project's goals must be clearly articulated early on--I want to know the stated goals and the unstated goals. I want to know those things on the table and that aren't on the table.
>
> The objective of the evaluation should be to assess how well the project met its goals. Therefore, depending what the goals were, the project should be adjusted accordingly.

Jack had definitive ideas of proper ways that evaluations could be conducted. He proposed that the publisher let him manage the entire evaluation, rather than just part of it, because he has more knowledge of the product. In his opinion, more meetings and discussions with the schools would help a great deal in their evaluations, instead of treating the CD-ROM as a textbook, which it is not:

> MB: So have evaluations helped you to improve your product?
> JL: I think so, but it would be better if they let me do it--because I think they don't understand the product well enough that they can get people to evaluate it at the level it that it has to be...If they really want a good evaluation then what they should be doing is send the disc out to schools, let them talk, and then we'll let you guys all go out to dinner after the national meeting or something and let you talk about it. But they're trying to be sort of arms length, including like we don't know who's on other end--I know who's on the other end!
>
> They're approaching the evaluation like that of a textbook, as they always have, but we have something more than that here.

Performing better evaluations was more of a financial constraint than any other for Jackie, due to her company's limited finances. She regrets these limitations, but realizes the importance of evaluation. Also, she hopes to conduct more evaluations once her company earns greater profits.

> JK: Evaluation is extremely important both from, is our product meeting market

> needs as well as establishing a market presence in the first place. Everything that we're doing we couldn't have been making money without doing that first, especially that we're not a big company. Evaluation, in general, and validating what we do is very important for us--even paying others to look at it, but our funds are limited...And we have a Web Site, although it's pretty dated. When we have more money, we'll support it.

In examining the financial constraints of her company, Jackie realized ways in which her evaluation approaches could be improved:

> JK: When we do hire a big marketing staff we can get a customer on the phone and it's easiest to start with, "What do you think about the program?" That's definite information that can help development--ok, we need this and that--we need Windows 3.1, they're not interested in Internet products. Just to get that information about things we really don't know --we don't have a good sense of that sometimes--of what can actually help the development process. The manager of the Marketing Department is just me, right now and I can think about what people need and formalize it. There are a lot of tele-marketers or sales force people --if you're actually talking to people as much as I can, I want to know about that--about what type of interaction you had.

In analyzing the categories that were mentioned and not mentioned by all six participants, it is apparent that some of the developers had similar thoughts and comments about evaluations, while others did not. It was interesting, however, to compare their responses with each other to determine what they regarded as the most important and least important ways to evaluate and develop their CD-ROMs, while also exploring their overall views and uses of evaluations.

The next section of this chapter examines several limitations of this study.

Limitations of the Study

This study explored the opinions, knowledge, and expressions of the six individuals

interviewed. While this is not a comprehensive sample, it is still a purposeful one, as each participant was chosen based on his or her credentials and experience in the field of instructional CD-ROM production, development, and evaluation. While a comprehensive study of more CD-ROM producers and developers could have been chosen, the researcher believes that this would not have dramatically changed the results, validity, reliability, or credibility of what was discussed in this study. Furthermore, the use of case studies for this project provided a rich look at each individual interviewed. As Patton (1986) states:

> Case studies become particularly useful when one needs to understand some particular problem or situation in great depth, and when one can identify cases rich in information--rich in the sense that a great deal can be learned from a few exemplars of the phenomenon in question.

A major limitation in carrying out this study was the semantic confusion caused by terms with multiple meanings, multiple terms for the same meaning, and evolving connotations for the same terms. Definitional precision for technical terms relating to educational evaluation is still a weakness to this day and is even more difficult to define when these terms are compared to the interpretations of today's developers. When viewed as a whole and with relation to evaluation and software development, a clearly defined technical and evaluation vocabulary is difficult to achieve. This study will help to clarify many of the evaluation terms used by developers by not only defining the terms themselves, but also by viewing them through the voices and interpretations of six multimedia developers as they employed the principles of these terms during their development and evaluation of CD-ROMs.

How Software Developers and Program Evaluators Can Learn From Each Other

To explore how program evaluators and software developers can create a mutual learning environment and a mutually beneficial partnership, two topics are examined: recommendations and suggestions for increasing the utility of formative evaluations for instructional media, and translating existing educational evaluation approaches to evaluating CD-ROMs.

Recommendations and Suggestions for Increasing the Utility of Formative Evaluations for Instructional Media

One further question remains about how formative evaluations can achieve their desired goal--to inform the decision making process during the design, production and implementation stages of an educational program with the purpose of improving the program (Flagg, 1990). That question is a utilization one: how can one increase the likelihood that formative evaluation results will be used by decision makers? Flagg (1990) and Cambre (1978) contend that interpersonal relations and communication are key elements in answering this question.

To improve interpersonal relations, a working dialogue and trusting relationship between formative researchers and decision makers is critical. Establishing this type of relationship will lead to the eventual utilization of results. Besides producing results that are timely and relevant to the needs of developers, evaluators must be able to communicate the results in an understandable, nonthreatening, and brief manner. Long

printed evaluation reports (which may include the results of such evaluation efforts as surveys, questionnaires, user tryouts, focus groups, interviews, and product testings) written in the language of evaluation are not easily assimilated by those who work with audiovisuals in a pressured environment (Flagg, 1990). It is crucial that results be summarized and presented in an easily consumable form. During the evaluation of the *Cases in Cardiology* CD-ROM, the researcher constantly modified his evaluation summary reports to make them more legible for the entire project team and the stakeholders. Before meetings with employees of her company, Jackie made sure that the results from her evaluation surveys were typed and organized into concise, one-page reports which each employee could easily understand and use. It is even more critical that these evaluation results be shared and published in critical literature such as journals and magazines related to the field, so that they are readily accessible and easily read.

Effective formative evaluation requires a relationship of trust that is both receptive and responsive. To improve the development of instructional media during the formative stages, partnerships between evaluators and developers can be strengthened if both parties listen not only to their own needs and uncertainties, but also to those of the entire project team as well. Mary mentioned in her interview that some developers selfishly evaluate for their own needs rather than the needs of the client and the user. Both evaluators and developers should respond sensitively with timely, relevant, credible, and lucid information. Finally, one might even ask the decision makers to give evaluative feedback in the usefulness of the formative evaluations, so the next research effort can be better (Flagg, 1990).

Another possible means of increasing the utility of evaluations would be to have multimedia developers themselves evaluate their own products throughout the development process. They may be able to do this more effectively with training in educational programs such as instructional technology. This training could provide them with critical evaluation skills which could be used to improve their products. Both John and Mary offered this suggestion in their interviews--they regretted that many developers today lacked formal training in education and in the humanistic design of development.

Translating Existing Evaluation Approaches to Evaluating CD-ROMs: What Approaches Might Work

The scarce availability of literature on evaluations of CD-ROMs precludes an evaluator, software developer, or instructional designer from following a fixed, standard design for a CD-ROM evaluation, yet one may translate the existing methods from the fields of program evaluation and instructional technology and apply those same principles to evaluating CD-ROMs. In this sense, Scrvien's (1974) consumer-oriented product evaluation method is a common standard which can be emulated. Scriven published a product checklist which, in some items, is very similar to those checklists of early film, television, and radio evaluations. In this checklist, Scriven stresses thirteen items of importance for the consumer-oriented evaluation: need; market; performance--true field trials; performance--true consumer; performance--critical comparisons; performance--long term; performance--side effects; performance--process; performance--causation; performance--statistical significance; performance--educational significance; cost-

effectiveness; and extended support. However, even Scriven (1984) admits there are weaknesses in some of the applications of this approach, even stating that *Consumer Reports* may provide a decent model to follow:

> We should add a word about what may seem to be the most obvious of all models for a consumerist ideologue, namely *Consumer Reports* product evaluations. While these serve as a good enough model to demonstrate failures in most of the alternatives more widely accepted in program evaluation, especially educational program evaluation, it must not be thought that [I] regard them as flawless. Although *Consumer Reports* is not as good as it once was and has now accumulated even more years across which the separatist/managerial crime of refusal to discuss its methodologies and errors in an explicit and nondefensive way has been exacerbated many times...*Consumer Reports* is still a very good model for most types of product evaluations (p. 75).

Although an evaluator or software developer may be hesitant to use *Consumer Reports* to guide a CD-ROM evaluation, a quick review of these magazines periodically offers convincing guidelines to follow for product evaluations. In fact, consumers of educational products often use product evaluations done by others. Worthen and Sanders (1987) claim that the sale of educational products in the United States alone exceeds $500 million annually. This suggests that as competition in the industry grows,

> Marketing strategies become more creative, but often are not calculated to serve the best interests of the consumer or student. For this reason, some educational evaluators have actively urged consumer education, independent reviews of educational products patterned after the *Consumer Reports* approach, and requirements for objective evidence of product effectiveness" (Worthen and Sanders, 1987, p. 87).

Two typical instruments used to measure product effectiveness and strength of design in product evaluations are checklists and rating scales, which, as discussed previously in this study, were used in instructional films in the early to mid-1900s. These certainly would be helpful to evaluating CD-ROMs as well.

The consumer-oriented evaluation approach had several advantages indicated by Worthen and Sanders (1990) that are notable for evaluating CD-ROMs: one, it has made available evaluations of educational products (such as CD-ROMs) as a service to educators and who may not have the time or information to do the job thoroughly; and two, it has advanced the knowledge of educators about the criteria most appropriate to use in selecting educational products.

All in all, many different evaluation approaches may be translated and refined to evaluate CD-ROMs. The approaches discussed here are ones that the researcher believes would work best, but ultimately, this decision will always be a personal choice for professional evaluators and software developers to decide themselves. Knowing which approach is best for a given situation is often a subjective decision. The lack of an adequate empirical base is probably the single most important impediment to development of a more adequate evaluation theory and models. In the absence of relevant evidence about which model works best under which circumstances, adherence to any one model rather than another is largely a statement of philosophy or a profession of faith (Worthen and Sanders, 1987). As Scriven (1976) muses:

> There has been a good deal of work on "evaluation models" which are hybrids between ways of conceptualizing evaluation and reminders as to how to do it...Each can, I believe, contribute something of value to most clients, but beyond that I can hardly make a dispassionate judgment (pp. 28-29).

Antonoplos (1977) furthers this thought:

> There is little agreement or data to support the efficacy of one definition, model or approach over another and...empirical data are needed to determine the extent to which the various models are theoretically and operationally different and the particular goals or purposes for which they are best suited (p. 4).

In the actual CD-ROM evaluation that the researcher designed and managed during his doctoral studies, he found that one of the most important goals for the project team was to field-test (or tryout) the CD-ROM with the targeted audience. The attainment of that objective gave the team the most critical feedback during the evaluation, much of which was used to improve the didactic, user-friendliness and overall design features of the CD-ROM.

A practical evaluation approach is to videotape the targeted audience during these tryouts or test-runs and clearly document users' suggestions, comments, and attitudes into a legible summary report, which can be easily read and understood by the entire project team. Both Jackie and Mary made efficient use of videotaped tryouts, as discussed in the previous chapters. Keeping a project team constantly informed, as Flagg (1990) emphasized, is an important communication and evaluation vehicle that cannot be understated. If there is an open, friendly, and honest atmosphere to work in, the project obviously will be easier, more beneficial and more pleasant. Moreover, the videotapes will serve many useful purposes: most notably, they act as historical, documented, and convincing evidence of feedback from the tryouts, which can always be reviewed at will by any member of the project team; they also can document exactly at which part of the case glitches or other errors occur on the CD-ROM, which is particularly useful for programmers and developers to refer to when making corrections on the disc.

There are two realistic constraints to consider, those of time and money. Deadlines for productions usually put extreme pressures on a project team. The age-old saying that "patience is a virtue" cannot be over emphasized. Project time lines for evaluation and

production are often stretched repeatedly beyond deadlines due to the time-consuming tasks involved in the programming materials. These can cause much confusion and frustration in the production and development stages.

Budgets are often over-extended as well during production and development of instructional technologies. The project team must be more aware of and more realistic in facing these financial constraints during the entire lifetime of the project. Mary offered one possible solution to the enduring constraints of time, money, and unrealistic expectations. She meets with her clients before the project begins and verifies that they have realistic time frames, budgets, and expectations for the project before she accepts it. This process is often a negotiation between the developer and the client.

Conclusions and Suggestions for Further Research

This study was primarily conducted to serve as a valuable resource and reference tool to be used by those professionals interested in evaluating instructional CD-ROMs. To accomplish this goal, the study sought to detail and refine what is currently known or theorized about how software developers evaluate CD-ROMs by exploring the methods used by six software developers and by translating their evaluation concepts into terms that are easily discernable.

The interviews with the six developers revealed that developers are performing evaluations of CD-ROMs and that they have integrated various evaluation approaches into the development process of these discs. But the evaluation language that developers use is

frequently different than the language used by program evaluators. This disparity in language has inhibited communications and partnerships between evaluators and developers. It also poses a new challenge for evaluators to learn the language of developers if they are to be helpful to software development teams.

The responses from the developers conveyed their view that evaluation is an integral component of the development process and that it is also critical for product improvement. Numerous quotes from the developers suggest that they wanted data about many different types of outcomes in order to improve their products. However, it is evident that, even though the six developers were content with their evaluations, they were ultimately preoccupied with only two aspects of evaluation for their CD-ROMs: getting the discs to run (product testing) and user preference issues ("Will you buy it?," "Do you like it?"). Even though they produced educational products, the developers were not as concerned overall with academic practice and did not choose to evaluate what the users had learned from the discs. Instead, the partnerships with educational institutions were viewed by the six developers as vital sources for feedback and, ultimately, for selling their products, rather than being viewed as a means by which students could improve learning and increase their knowledge of the subject area through the use of these discs. This scenario was most evident with the textbook publishers that John and Jack worked with who seemed more intent on sales figures rather than the educational value that the discs could provide. Ultimately, this practice may defeat the very purpose of instructional CD-ROMs--to have students learn.

As financial resources for evaluations of CD-ROMs and other multimedia dwindle

and as competition in the computer software industry intensifies, both the quality and quantity of these evaluations may follow (Flagg, 1990). While no single evaluation method is considered to be the "best" for evaluating instructional CD-ROMs, perhaps professionals who conduct future evaluations will learn how to better utilize scarce resources. This may be accomplished through various methods of evaluations, such as those offered by the six developers in this study. Flagg (1990) suggests that the financial constraints which developers are facing may cause them to look for quicker rewards and to possibly ignore the importance of evaluation:

> The resources available for the development of instructional materials have diminished over recent years and, along with that change, has come a reduction in the amount and quality of formative evaluation. In a marketplace looking for a fast return on a minimum investment, reducing evaluation is a certain cost savings. Although this "penny-wise and pound-foolish" philosophy may have short-term benefits for producers, it is detrimental to the quality and utility of instructional products. Perhaps the training efforts of business and industry can show education the value and cost effectiveness of thoughtful evaluation. Until then, we need to look to some of our work in television and explore how to transfer economically the lessons we have learned to the development of instruction using the computer (p. 81).

If it is true that "history is the best teacher," then certainly the field of evaluation has a lot to learn from its predecessors and from its past mistakes. Mary was one developer in this study who mentioned that she had learned much from her past mistakes. From these mistakes, she learned how to increase the utility of her evaluations for future projects.

Competition between software publishing companies will likely continue to intensify. As this occurs, product improvement will become more valuable to maintain their market share which, in turn, may result in project managers, developers, market researchers, programmers, evaluators, and other employees in these companies working

together more closely and creating external partnerships (especially with educational institutions) to ascertain the needs of their target audience. As a result, new synergies may be created within software companies, as was witnessed with the companies where Jackie and Gwen worked. This also suggests a growing dependence upon the target user audience for vital evaluative feedback during the planning, design, production and implementation phases of CD-ROMs. Much research still remains to be done in the area of marketing and the relationship of the user of instructional media to the evaluation process. However, some of the approaches that the six developers in this study used in establishing partnerships with educational institutions and with their intended users offer practical examples which should be considered.

After analyzing the interviews with the six developers, it is evident that there remains a disparity in the evaluation language used by program evaluators and software developers. Some of the developers in this study utilized similar formative and summative evaluation concepts that program evaluators use such as needs assessment, testing, focus groups, surveys and questionnaires, and feedback. However, there are concepts that developers and evaluators do not mutually understand. For example, software product testing terms such as beta and alpha testing and platform and cross-platform testing are not terms used or readily understood by program evaluators. A routine examination of any evaluation/research dictionary or thesaurus, such as Scriven's (1996) or Vogt's (1993), will confirm this. This is a contributing factor in how ineffectively evaluators and developers communicate with each other. The disparity in language between developers and evaluators, moreover, can disrupt the development of any product.

Evaluators and developers must look to each other to establish mutual learning environments by which their strategies, concepts, and findings are shared. When both parties agree to help each other, this type of partnership can be established. It is, however, incumbent upon evaluators to take the first step in creating this partnership since many program evaluators are unfamiliar with the type of language used by developers. If evaluators have anything to offer to developers, primarily it will come from learning the language that developers use during product development. Once this language is mastered by evaluators, both parties will have stronger working partnerships and more effective means of communication. This also suggests that the partnerships will become mutually rewarding and that evaluators will be providing greater value to the development process of instructional media, such as CD-ROMs.

There are many lessons that can be learned by evaluators and developers. A number of evaluation approaches used by the six developers in this study can be emulated by program evaluators. Establishing partnerships with educational institutions was a sensible method that all six of the developers not only used to improve their evaluations and to entice the interest of their target user audience, but also was used to save a great deal of money. It would appear that since a large proportion of today's evaluators are from universities and colleges, partnerships such as these would be simple to develop. Another innovative way that two developers in this study engaged existing resources (using employees from different divisions), was by breaking down the barriers between themselves and their co-workers and creating new synergies between them. Clearly, evaluators who work in organizations or for institutions can also establish these kinds of

working synergies between themselves and their peers. These synergies would be mutually beneficial and rewarding.

It is interesting to note that even though the six developers were extremely satisfied with their evaluations, it became more apparent in the interviews that several of them wanted to perform more comprehensive evaluations of their products, but were constrained by their company's limited finances and project deadlines. These constraints were most evident in the smaller companies for which Jackie, Bob, and Mary worked. Even though each of these developers had found methods to save money (notably through their partnerships with educational institutions), they still found themselves limited by the modest finances and resources of their companies. Cambre (1978) and Flagg (1990) believed these constraints were similar reasons why evaluations of instructional media are not consistently performed, even though evaluations have long been known to be effective for improving products.

Although it is apparent that evaluators have much to learn from developers, in particular the language that developers use, there is one strategy that developers may learn from evaluators to increase the utility of their evaluations. Even though the interviews with the six developers confirm that they are performing evaluations of their products and indicate that the developers believe their evaluations were useful and beneficial, interestingly enough, multimedia developers do not organize their evaluations the same way that evaluators do. To a program evaluator, the organization of an evaluation plan is implicit at the onset of a program. This type of evaluation planning before development was not so evident in the strategies of the six developers. Perhaps, again, the constraints of

time and money offer a plausible explanation, since most evaluators only used the evaluation methods that they could afford. It is also possible that the six developers believed that their evaluation designs would emerge as the development of their CD-ROMs progressed. If developers can learn one important characteristic from evaluators it is how to improve the utility of their evaluations by structuring the evaluation design at the onset of development.

The worlds of evaluation of instructional CD-ROMs and of computer-based teaching and design are still in their infancy. What was revealed in this study can help to define many of the evaluation and educational issues involved in these emerging fields. Additionally, this study will provide future researchers, evaluators, software developers, instructional designers, project managers, and others with a valuable resource by which to conduct evaluations of their own products. It is hoped that new partnerships will be created between program evaluators and software developers to share their evaluation approaches with each other and to disseminate this information through published articles and studies that are readily accessible to all individuals.

Partnerships between program evaluators and software developers could best be facilitated when both parties mutually agree and understand how much they can help each other. Bridging the evaluation language barriers between developers and evaluators, which was one of the primary intentions of this study, may help facilitate this partnership. Moreover, a mutual agreement and understanding between program evaluators and software developers that development and evaluation are (as Jackie, John, and Mary suggested), contiguous processes, not independent ones, might also expedite this process.

This understanding may also help correct the mistaken view which some publishers hold that evaluation and development of CD-ROMs can be treated like those of textbooks while, in reality, textbooks and CD-ROMs are truly two distinct forms of instructional media as Jackie, Bob, John, and Jack attested and as was examined in the literature in Chapter 2.

The findings of this study advance the view that CD-ROMs and other forms of multimedia may make significant contributions to education, corporations, and to the profession of evaluation as a whole, to make learning and teaching more fun. Furthermore, the use of this study will assist those individuals who chose to undertake evaluations of instructional CD-ROMs and may be a catalyst by which program evaluators and software developers will establish working partnerships.

REFERENCES

Allen, R.J. (1996). The return on investment of computer-based training. *CD-ROM Professional*, 9(10), 35-45.

Allendoerfer, C.B. (1969). *Principles of mathematics*. New York: McGraw-Hill.

Antonoplos, D.P. (1977). *Evaluation reconsidered: Pied pipers of pedagogy*. Paper presented at the annual meeting of the American Educational Research Association, New York.

Becker, H.J. (1986). The computer and the elementary school. *Principal*, 64(5), 32-34.

Bewley, B.R. (1992). Reorganizing medical education. *British Journal of Hospital Medicine*, 47(1), 40-43.

Block, D.G. (1996). The returns of CD-ROM. *CD-ROM Professional*, 9(6), 62-69.

Calhoun, P.S. & Fishman, E.K. (1994). Interactive multimedia program for imaging the spleen: Concept, design, and development. *infoRAD*, 14(6), 1407-1414.

Cambre, M.A. (1978). *The development of formative evaluation procedures for instructional film and television: The first fifty years.* Unpublished doctoral dissertation, Indiana University.

Cambre, M.A. (1981). Historical overview of formative evaluation of instructional media products. *Educational Communication & Technology Journal*, 19(1), 3-25.

Carpenter, C.R. (1948). A challenge for research, *Educational Screen*, 27, 119-121.

Carpenter, C.R., Eggleton, R.C., John, F.T., & Cannon, J.B., Jr. (1950). The film analyzer. *Technical Report SCD, 269-7-15*, The Pennsylvania State College, Instructional Film Research Program.

Conn, R.B. (1992). Can continuing medical education prepare the current practitioner for the 21st century? *Archives of Pathology & Laboratory Medicine*, 116(6), 602-604.

Coutant, F.R. (1963). Determining the appeal of special features of a radio program. *Journal of Applied Psychology*, 23, 54-57.

Cronbach, L.J. (1975). Course improvement through evaluation. Reprinted in Payne, D.A., & McMorris, R.F. (Eds.), *Education and psychological measurement.*

Morristown, NJ: General Press, 243-256.

Devereux, F.L. (1933). *The educational talking picture*. Chicago: The University of Chicago Press.

Dick, W. & Carey, L. (1996). *The systematic design of instruction.* New York, NY: Harper Collins.

Doane, D.C. (1936). What makes a good educational film? *Educational Screen*, 15, 203-307.

DuBois, P.H. (1970). *A history of psychological testing*. Boston: Allyn & Bacon.

Ekulona, A. (1985). Videodisc simulations need Level III. *EITV*, 17(9), pp. 35, 50, 64, 65.

Ellson, D.G., Harris, P. & Barber, L. (1968). A field test of programed and directed tutoring. *Reading Research Quarterly*, 3(3), 307- 367.

Feibus, M., & Silverthorne, S. (1995, December). Measuring multimedia. *PC Week*, 33-35.

Finn, J.D. (1953). Professionalyzing the audio-visual field. *AV Communication Review*, 1, 6-17.

Flagg, B.N. (1990). *Formative evaluation for educational technologies*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Folberg, R., Dickinson, L.K., Christiansen, R.A., Huntley, J.S., & Lind, D.G. (1993). Interactive videodisc and compact disc-interactive for ophthalmic basic science and continuing medical education. *Ophthalmology*, 100(6), 842-850.

Gliessman, D., & Williams, D. (1966). Stimulus films: Films made from pretested scripts as a medium for teacher education. *Audiovisual Instruction*, September, 1966, 552-554.

Goodman, D.J. (1941). Criteria for the production and selection of visual aids. *Education*, 61, 359-363.

Greenhill, L.P. (1957). *Final report instructional film research program*. Port Washington, New York: U.S. Naval Training Devices Center. In Cambre, M.A. (1978), *The development of formative evaluation procedures for instructional film and television: The first fifty years.* Unpublished doctoral dissertation, Indiana University.

Gropper, G.L. (1967). Does programed television need active responding? *AV Communication Review*, 15, 5-22.

Guba, E.G. (1962). *Evaluation and the new media.* Address to the Ohio State University Annual Conference on Modern Media, July 6, 1962. In Cambre, M.A. (1978), *The development of formative evaluation procedures for instructional film and television: The first fifty years.* Unpublished doctoral dissertation, Indiana University.

Gussin, L. (1996). How multimedia supports primary education. *CD-ROM Professional*, 9(5), 56-68).

Hagenbruch, H. (1994). American memory--History meets the age of technology. *Library Software Review*, 13(1), 35-38.

Hedberg, J.G. (1994). *Information landscapes and exploratory user interfaces: Redesigning to improve learning outcomes.* Paper presented at the National Convention of the Association for Educational Communications and Technology, Nashville, TN.

Helgerson, L.W. (1992). *CD-ROM: Facilitating electronic publishing*. New York, NY: Van Nostrand Reinhold.

Herther, N.K. (1996). New, fast CD-ROM drives--A buyer's guide. *Online*, 20(4), 80-85.

Hipps, J. (1992). *New frameworks for judging alternative assessments*. Paper presented at the annual meeting of the California Educational Research Association.

Hoban, C.F. (1942). *Focus on learning: Motion pictures in the school*. Washington, D.C.: American Council on Education.

Hoban, C.F. (1956). Research and reality. *AV Communication Review*, 4, 3-20.

Hylton, J.B. (1988). *Case studies of ten teachers in the Virginia beginning teacher assistance program*. Unpublished doctoral dissertation, University of Virginia.

Jennings, J.J. (1994). Content of continuing medical education courses in obstetrics and gynecology. *Obstetrics & Gynecology*, 83(5), 789-791.

Komoski, P.K. (1974). An imbalance of product quantity and instructional quality: The imperative of empiricism. *AV Communication Review*, 22, 357-386.

Kroeker, K.L. (1996). Multimedia training multiplies in corporate classrooms. *CD-ROM Professional*, 9(3), 14-15.

LaGuardia, C. (1992). Virtuous disc selection or, how I learned to stop worrying and love to buy CD-ROMs. *CD-ROM Professional*, 16(1), 82-84.

Lashley, K.S., & Watson, J.B. (1921). A psychological study of motion pictures in relation to venereal disease campaigns. *Social Hygiene*, 7, 181-219. In Cambre, M.A. (1978), *The development of formative evaluation procedures for instructional film and television: The first fifty years.* Unpublished doctoral dissertation, Indiana University.

Lashley, K.S., & Watson, J.B. (1922). *A psychological study of motion pictures in relation to venereal disease campaigns.* Washington: U.S. Interdepartmental Social Hygiene Board. In Cambre, M.A. (1978), *The development of formative evaluation procedures for instructional film and television: The first fifty years.* Unpublished doctoral dissertation, Indiana University.

Levonian, E. (1960). Development of an audience-tailored film. *AV Communication Review*, 8, 62-68.

Levonian, E. (1962). The use of film in opinion measurement. *AV Communication Review*, 10, 250-254.

Levonian, E. (1963). Opinion change as mediated by an audience-tailored film. *AV Communication Review*, 11, 104-112.

Lincoln, Y.S., & Guba, E.G. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.

Lovelace, V., Schwager, I., & Saltzman, E. (1984). *Responses to death on "Sesame Street" from preschoolers and parents*. Paper presented at the meeting of the American Psychological Association, Toronto, Ontario, Canada.

Lowry, F. (1995). Computers a cornerstone of evidence-based care, conference told. *Canadian Medical Association Journal*, 13(11), 1636-1639.

Lumsdaine, A.A. (1947). Experimental research and the improvement of teaching film. *Educational Screen*, 26, 254-255.

Lumsdaine, A.A. (1964). Educational technology, programmed learning and instructional science. In E.R. Hilgard (Ed.), *Theories of learning and instruction*, Sixty-Third yearbook of the National Society for the Study of Education, Chicago: University of Chicago Press, 371-401.

Markle, D.G. (1967). *The development of the Bell system first aid and personal safety course*. Report No. AIR-E 81-4167, Palo Alto: American Institutes for Research.

Mayne, K. (1994). Practice-linked continuing medical education. *Medical Journal of Australia*, 161(10), 630-632.

Nicholls, P.T., Han, I., Stafford, K. Whitridge, K. (1990). A framework for evaluating CD-ROM retrieval software. *Laserdisk Professional*, 3(2), 41-46.

Nicholls, P.T. (1997). Hybrids enable new business models for multimedia. *Computers in Libraries*, 17(1), 64.

Orr, D.B. (1966). The evaluation of televised instruction. *AV Communication Review*, 14, 363-370.

Palmer, E.L. (1974). Formative research in the production of television for children. In D.E. Olson (Ed.), *Media and symbols: The forms of expression, communication, and education*. Chicago: University of Chicago Press.

Papert, S. (1992). *The children's machine; rethinking school in the age of the computer.* New York: BasicBooks.

Patton, M.Q. (1986). *Utilization-focused evaluation*. Newbury Park, CA: Sage.

Patton, M.Q. (1990). *Qualitative evaluation and research methods*. Newbury Park, CA: Sage.

Reid, S. & Woelfel, N. (1941). How to judge a school broadcast. *Evaluations of school broadcasts*, Bulletin No. 2, Ohio State University Bureau of Educational Research.

Reiser, R.A. & Kegelmann, H.W. (1994). Evaluating instructional software: A review and Critique of current methods. *Educational Technology Research and Development*, 42(3), 63-69.

Richards, T. & Robinson, C. (1993). Evaluating CD-ROM software: A model. *CD-ROM Professional*, 6(5), 92-101.

Report of radio activities 1938-1939 (1939). Station WBOE, Cleveland, Ohio Board of Education. Cleveland: Cleveland Public Schools.

Rockman, S., & Auh, T. (1976). *Formative evaluation of "Self Incorporated" programs.* Agency for Instructional Television, Research Report #30, Bloomington, Indiana: Agency for Instructional Television.

Rosen, M.J. (1968). *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. (ERIC Document

Reproduction Service No. ED 025 156).

Rothkopf, E.Z. (1963). Some observations on predicting instructional effectiveness by simple inspection. *Journal of Programmed Instruction*, 2(2), 19-20.

Rutting, C.S. (1967). *A study of the form and function of instructional television curriculum committees*. Unpublished doctoral dissertation, Michigan State University.

Saettler, P. (1968). *A history of instructional technology*. New York: McGraw-Hill Company.

Schroeder, E. (1995, August). Training CD-ROMs make the grade. *PC Week*, pp. 32-33.

Scriven, M. (1967). The methodology of evaluation. *AERA Monograph Series on Evaluation*, No. 1. Chicago: Rand McNally, 39-83.

Scriven, M. (1974). Standards for the evaluation of educational programs and products. In B. R. Worthen & J.R. Sanders (1987), *Educational evaluation* (pp. 88-89). White Plains, NY: Longman.

Scriven, M. (1976). *The intellectual dimensions of evaluation research*. Paper presented at the fourth annual Pacific Northwest Research and Evaluation Conference, Seattle, Washington.

Scriven, M. (1984). Evaluation ideologies. In B. R. Worthen & J.R. Sanders (1987), *Educational evaluation* (pp. 89-90). White Plains, NY: Longman.

Scriven, M. (1996). *Evaluation thesaurus*. Newbury Park, CA: Sage.

Sesame Street Research (1983). *Study on show #1839*. Unpublished manuscript, Children's Television Workshop, New York.

Stake, R.E. (1981). Case study methodology: An epistemological advocacy. In W.W. Welch (Ed.), *Case study methodology in educational evaluation* (pp. 31-40). Minneapolis: University of Minnesota Research and Evaluation Center.

Strommen, E.F. & Revelle, G.L. (1990). Research in interactive technologies at the Children's television workshop. *Educational Technology, Research and Development*, 38(4), 65-80.

Thurstone, L.L. (1920). What is an educational motion picture? *Visual Education*, 1(2), 24-28. In Cambre, M.A., *The development of formative evaluation procedures for instructional film and television: The first fifty years*. Unpublished doctoral dissertation, Indiana University.

Twyford, L.C. (1951). Film profiles. *Technical Report S.D. 269-7-23*, The Pennsylvania State College, Instructional Film Research Program.

Twyford, L.C. (1954) Profile techniques for program analysis. *AV Communication Review*, 2, 243-261.

UNESCO (1992). *UNESCO CD-ROM. Prototype Evaluation Summary*. ERIC Document Reproduction Services No. ED 373 783.

Vaughan, T. (1994). *Multimedia: Making it work.* Berkeley, CA: Osborne McGraw-Hill.

Vogt, W.P. (1993). *Dictionary of statistics and methodology: A nontechnical guide for the social sciences*. Newbury Park, CA: Sage.

Wilson, K.S. & Tally, W.J. (1990). The "Palenque" project: Formative evaluation in the design and development of an optical disc prototype. In B.N. Flagg (Ed.), *Formative evaluation for educational technologies* (pp. 83-98). Hillsdale, NJ: Lawrence Erlbaum Associates.

Worthen, B.R. & Sanders, J.R. (1973). *Educational evaluation: Theory and practice*. Belmont, California: Wadsworth Publishing Company, Inc.

Worthen, B.R. & Sanders, J.R. (1987). *Educational evaluation: Alternative approaches and practical guidelines*. White Plains, NY: Longman.

Yetton, L. (1996). CD-ROM networking yields a wealth of info. *Computer Technology Review,* 16(8), 41.

Zink, S.D. (1991). Toward more critical reviewing and analysis of CD-ROM user software interfaces. *CD-ROM Professional*, 4(1), 16-22.

Zuckerman, J.V. (1951). *Testing with a pre-release filmstrip as a means of predicting factual learning from a training film*. Washington, D.C.: U.S. Air Force Human Resources Research Laboratories, HRRL Memo, Report No. 14, November, 1951. In Cambre, M.A. (1978), *The development of formative evaluation procedures for instructional film and television: The first fifty years* (pp. 113-115). Unpublished doctoral dissertation, Indiana University.

Zuckerman, J.V. (1954). Predicting film learning by pre-release testing. *AV Communication Review*, 2, 49-55.

Appendix A: Sample Interview Questions

1) Tell me about the evaluation methods you use for your CD-ROMs (do you use the term "evaluation" for this process or perhaps another term?).

2) What aspects of CD-ROMs do you evaluate--e.g., learning, user-friendliness, customer satisfaction, efficacy of (a market for) your product?

3) What approaches/methods are you using to evaluate CD-ROMs? Do you conduct summative or formative evaluations (or both) of CD-ROMs?

4) Which employees/which divisions of your company conduct evaluations of these products--does it go through the R&D or Marketing Departments or another one?

5) If evaluations have been conducted by your company, have they been effective--i.e., have they helped improve the product or helped sell more of them?

6) How costly are these evaluations? Are evaluations conducted in-house (internal) or are they outsourced (external)?

7) What implications are there for your company in evaluating CD-ROMs--i.e., greater costs, improved products, trying to market the targeted audience, creation of new departments or divisions, new employees?

8) Are you utilizing similar or different evaluation approaches than educational program evaluators use?[offer an example of an educational evaluation approach] (do they use existing evaluation approaches or have they created their own?).

Appendix B: Definition of Terms

It is important to define several terms used by software developers in their evaluation, production and development processes for instructional multimedia, particularly for CD-ROMs. This appendix defines some of the terms used by the participants in the interviews to better understand their importance to this study. A definition of terms also helps to facilitate one of the main purposes of this study--the translation of the evaluation concepts used by multimedia developers so that both developers and evaluators can better understand the nature of each other's work, thereby promoting closer working relationships.

Multimedia Terms

Vaughan (1994) defines *multimedia* as "any combination of text, graphic art, sound, animation, and video delivered by computer or other electronic means" (p. 4). CD-ROMs are one of the most popular forms of multimedia. Other multimedia and software terms include:

End User, Interactive Multimedia, and Hypermedia

When an *end user* (referred to as *users* or as *target audience* in this study)--the viewer of a multimedia project--is allowed to control what and when the elements are delivered, it is called *interactive multimedia.* For instructional CD-ROMs, the end users would primarily be students, teachers, parents, and schools--they are the potential "purchasers" or "buyers" of these discs. When a structure of linked elements though which the user can navigate is provided, interactive multimedia becomes *hypermedia*. (Vaughan,

1994, pp. 5-6).

Multimedia Developers, Projects, Titles

The people who help make multimedia computer tools and technologies work together and who weave multimedia into meaningful tapestries are *multimedia developers*. For this study, the researcher interviewed six educational multimedia developers who have a wide range of responsibilities and experiences in evaluating, developing, marketing, and producing CD-ROMs. The software vehicle, the messages, and the content presented on a computer or television screen constitute a *multimedia project*. Once the project is shipped or sold to consumers or end users, typically in a box or sleeve, with or without instructions, it is a *multimedia title* (Vaughan, 1994, p. 5).

Scripting or Storyboarding

Determining how a user will interact with and navigate through the content of a project requires great attention to the message that is conveyed, *scripting* or *storyboarding* describing the parameters of the project, the artwork, and the programming. An entire project can be ruined with a badly designed interface. A project can also be broken with inadequate or inaccurate content (Vaughan, 1994, p. 5).

Authoring Tools, Human Interface and Platform or Environment

Multimedia elements are typically sewn together into a project using *authoring tools* (the two most popular authoring tools today are *Director* and *Toolbook*). These

software tools are designed to manage individual multimedia elements and to provide user interaction. In addition to providing a method for users to interact with the project, most authoring tools also offer facilities for creating and editing text and images, and they have extensions to drive videodisc players, videotape players, and other related hardware peripherals. Sounds and movies are usually created with editing tools dedicated to these media, and then the elements are imported into the authoring system for playback. The sum of what gets played back and how it is presented to the viewer is the *human interface*. This interface is just as much the rules for what happens to the user's input as it is the actual graphics on the screen. The hardware and software that govern the limits of what can happen are the multimedia *platform* or *environment*. (Vaughan, 1994, pp. 6-7).

Product Testing Terms

Product testing (also known as *field testing* or *bug testing*) is a vital component of CD-ROM evaluation for multimedia producers. The prototype disc must be tested for several important reasons: to ensure it is bug free and accurate, that it is operationally and visually on target, and that the client's requirements (even if the client is your own company) have been met (Vaughan, 1994). In addition, there are other sound reasons to test the product:

> It is important to test the work before it is finalized and released for public or client use. A bad reputation earned by premature product release can destroy an otherwise excellent piece of work representing thousands of hours of effort. If you need to, delay the release of the work to be sure that it is as good as possible. It is critical that you take the time to thoroughly exercise your project and fix both big and little problems; in the end, you will save yourself a great deal of agony! (Vaughan, 1994, p. 440).

Alpha and Beta Testing

There are two testing terms that need to be defined to better understand their role in this study. Vaughan (1994) states that the terms *alpha* and *beta* testing are used by software developers to describe levels of product development when testing is done and feedback is sought. *Alpha* releases are typically for internal circulation only, and are passed among a selected group of mock users. These versions of a product are often the very first working drafts (*prototypes*) of the project, and are expected to have problems or be incomplete. It is for this reason that they are only circulated internally.

Beta releases, on the other hand, are sent to a wider but still select audience with the same caveat: the product may contain errors and bugs, but is now shown to and tested outside of the software company by other mock users, who often have little or no involvement with the company. Because the product is being the shown outside of the company, its reputation will begin to take form during the beta phase. Vaughan (1994) maintains that if testing is to be successful in the evaluation, the beta testing group should be representative of real users, and should not include persons who have been involved in the project's production. For this reason it is important that beta testers have no preconceived ideas, as they are vital for providing objective commentary and reports on product improvement.

Platform and Cross-Platform Testing

Currently, there are two major computer platforms: Macintosh (which has its own operating system) and PC (which uses Windows or DOS as its operating system). There

are many different brands of computers on the market, all with various types of hard drives, random access memory (RAM), microprocessors, hardware and software. *Platform testing* is considered to be testing of a product (e.g., a CD-ROM) on the same platform (e.g., an IBM personal computer), although different types of memory, microprocessors (386, 486, pentium, pentium pro), and hard drives are used. *Cross-platform testing* would be testing a product across both of the platforms. Vaughan (1994) emphasizes the importance of platform testings for multimedia developers:

> One of the major difficulties you face in testing the operation of your multimedia project is that its performance depends on specific hardware and system configurations. If you cannot control the end user's platform, or if the project is designed to be shown in many different environments, you must fully test your project on as many platforms as possible.
>
> As any element of a computer's configuration may be the cause of a problem or a bug, you will spend a good portion of testing time configuring platforms, and additional time reproducing reported problems and curing them. It is very difficult for even a well-equipped developer to test every possible configuration of a computer, software, and third-party add-on boards (pp. 440-441).