A MODEL FOR EVALUATING THE EFFECTIVENESS OF COMPUTER-MEDIATED CONTINUING MEDICAL EDUCATION

A Thesis

Presented to

The Faculty of Graduate Studies

of

The University of Guelph

by

VERNON R. CURRAN

In partial fulfilment of requirements

for the degree of

Doctor of Philosophy

May, 1999

0-612-43254-8

ABSTRACT

A MODEL FOR EVALUATING THE EFFECTIVENESS OF COMPUTER-MEDIATED CONTINUING MEDICAL EDUCATION

Vernon R. Curran
University of Guelph, 1999

Advisor:
Professor Mark Waldron

Over the years, a variety of distance learning technologies have been used to address the continuing medical education (CME) needs of rural and remote physicians. Some of these have included audio teleconferencing, video conferencing, and slow scan imaging. Although the CME literature has documented these initiatives in detail, it is deficient of evaluative information which could advise key decisions regarding the effectiveness of these instructional technologies in providing CME.

In recent times, the emergence and growth of the World Wide Web and CD-ROM technologies have introduced new prospects for delivering continuing education to the rural medical practitioner. Consequently, the rationale for this study was to develop and fieldtest a model for evaluating the effectiveness of computer-mediated instructional courseware. This is a response to the need for comprehensive and systematic evaluation approaches to inform decision-makers, program developers, and future efforts in computer-mediated CME delivery.

An eclectic evaluation planning matrix was designed by selecting various elements and concepts from evaluation approaches identified in the literature. This planning matrix was used in developing the evaluation model. The model was then field tested by evaluating a computer-mediated courseware program "Dermatological Office Procedures" using a modified pretest-posttest control group study design. Evaluation was performed at formative and summative levels, and numerous methods were used to collect antecedent, reaction, cognitive, behavioral, and instructional transaction data.

A metaevaluation was formulated and conducted to assess the effectiveness of the fieldtest of the evaluation model. An evaluator self-report, a focus group with program developers, and interviews of stakeholders and decision-makers were used to gather data about the effectiveness of the courseware evaluation model. The main conclusion of the study was that the CME Courseware Evaluation Model was useful in collecting evaluative data to inform decision making and improve the instructional product. Recommendations for further research are presented, implications for educational practice are discussed, and a modified version of the evaluation model is provided reflecting the results of the study. Incidentally, the results of the fieldtest revealed that computer-mediated instruction was an effective means for delivering CME at a distance.

ACKNOWLEDGMENTS

Special appreciation is extended to the members of my doctoral advisory committee, Dr. Glen Filson, Dr. Wayne Pfeiffer, Dr. Don Richardson and Dr. Carl Robbins who shared their time and support during the proposal and analysis stages of this study. Special thanks to Dr. Mark Waldron, my doctoral Advisor, for the flexibility, support and advice extended to me during my studies and the completion of my research work. His continued guidance, friendship and support throughout this venture are greatly appreciated.

Special thanks are also extended to the CME Planning Committee Members (Dr. Tony Gabriel and Dr. Lydia Hatcher) the subject matter consultants and instructors (Dr. Wayne Gulliver and Dr. Ian Landells) the courseware development team (Dr. Ted Hoekman, Dr. Chet Michalski, Kelly Best, and Nicki Murray) for their time, assistance and commitment in planning and developing the courseware. Also, thanks to J. Mary Ryan, Patti Thistle and Sandy Collins at the Health Sciences Library, Memorial University of Newfoundland for their assistance in obtaining materials.

A sincere thank you is expressed to Gerry White who assisted me with the statistical analysis of the results of this study. His support and advice in analyzing the data not only attests to his professional qualifications, but further defines him as a committed and valued colleague.

Special appreciation is extended to the following agencies and organizations for their financial support of this project and study: CANARIE Inc., Stieffel Canada Inc., ICN Canada, the Human Resources Development Agreement (Government of

Newfoundland and Labrador) and the Telecom Applications and Research Alliance (TARA).

No words of appreciation can truly express the contributions of my wife, Lisa, whose loving support enabled me to complete this work. Her encouragement, patience and understanding were invaluable.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER I

INTRODUCTION TO THE STUDY

Introduction

The purpose of this study was to develop and validate a computer-mediated instructional courseware evaluation model. As a means of validation, a summative metaevaluation system was applied to the evaluation model after it was fieldtested in evaluating a hybrid continuing medical education courseware program on Dermatological Office Procedures. This dissertation describes the hybrid courseware system which was evaluated and reports on the methods, results, and conclusions of the fieldtest of the evaluation model and the metaevaluation. The procedures which were followed for designing the evaluation model and the metaevaluation system are also discussed and presented in the dissertation. The main elements of the study were the fieldtest and metaevaluation of the evaluation model as a possible solution to the need for a comprehensive and systematic approach to assessing the effectiveness of computer-mediated CME courseware.

A Hybrid Courseware Delivery System delivers self-paced and collaborative instruction at a distance through the integration of World Wide Web (WWW) and CD-ROM technologies. The system encompasses several instructional components which include:

- Interactive Multimedia Tutorials: the courseware Web site presents multimedia-enhanced instructional tutorials which integrate hypertext, images, video and sound files;

- One-to-One communications: electronic mail between learners and instructors is possible;
- Access to Remote Systems: links from the course Web site to library resources and other related clinical and academic sites;
- Asynchronous Group Communications: the primary tool for interactive dialogue is provided by asynchronous computer conferencing;
- Online Testing: the forms capability features of the Web enable the creation and grading of online self-assessment tests.

The World Wide Web, the fundamental content delivery mechanism for the hybrid system, is the front-end application that has stimulated the enormous growth of the Internet. The Internet is a transport mechanism consisting of thousands of interconnected computers and networks through which data and information can be transmitted. A Web "browser" uses a graphical user interface to view information transmitted over the Internet. Web browsers read Hyper Text Markup Language (HTML) and the "http:\\" addressing scheme to allow us to travel over the Internet to a Web site, where we can navigate our way through the various pages of data.

A World Wide Web site contains pages or documents. Web documents or pages are what is displayed by a Web browser. What is displayed is determined by an ASCII text file containing HTML codes called tags. These tags specify how text in the file is to be displayed, and can call for the contents of certain other types of files to be automatically displayed at specific spots in the document. These other files may include graphics or other specialized media files containing sound, animation, and film clips. Web pages contain many things not offered by the printed page. They offer entry fields (into which visitors can enter information); animation, video, and audio; buttons (hypermedia graphics linked to certain functions); and hypertext (colored or highlighted text that links the visitor to another section of the Website).

The core of hyperlinking and navigating World Wide Web documents is hypermedia. According to Yang and Moore (1995) hypermedia has two fundamental characteristics. One of these characteristics is hypermedia's non-linear association of information. Informational segments in the hypermedia environment are represented in chunks or nodes which a learner has the ability to navigate through by following paths or links that appear relevant (Tolhurst, 1995). This non-linear association encourages active individualized learning. The second characteristic of hypermedia is that it is represented and managed around a network of multiple information or multimedia formats. Text, charts, graphics, animation, video and sound are all possible in hypermedia systems. The use of these multi-format knowledge bases has several potential benefits:

1) Providing rich and realistic contexts for multichannel learning. A hypermedia system can present vivid information in various formats.

2) Accessing information non-linearly. Allows learners to access and explore information based on their needs, interests, or whims.

3) Focuses learners on the relationship of facts. The structure of hypermedia helps learners focus their attention on learning relationships among facts.

4) Encouraging active, student-centered learning. A hypermedia system is an interactive environment. Learners control learning speed, amount, and path based on their abilities and needs. (Yang and Moore, 1995, p. 5)

The World Wide Web also enables the use of multimedia for the presentation and representation of information. Multimedia refers to an information environment that uses computers to integrate text, graphics, images, video, and audio (Galbreath, 1992; Tolhurst, 1995; Shih and Alessi, 1996; Najjar, 1996). Interactive multimedia is an extension of hypermedia which provides learner-directed non-linear methods for accessing and presenting information.

Another important advantage of using the World Wide Web as an instructional delivery platform is that it enables the integration of asynchronous computer conferencing systems for facilitating online collaboration between learners. Computer conferencing is a computer messaging system which merges telecommunications and computers to allow geographically dispersed groups of people to communicate with one another by way of text-based messages (Tagg and Dickinson, 1995; Berge, 1996). It does not include various types of real-time, or synchronous communications such as chat rooms, voice-based teleconferencing or video conferencing. Participants use personal computers, a modem and Internet connection, and a WWW browser to communicate with a central host computer server running a computer conferencing software. They can call the host computer 24 hours a day, read comments left by their colleagues, and leave messages for other learners. More than one person can be connected with the host computer at a given time, but the ususal experience is asynchronous, that is the interactions occur independent of time and place. One participant may respond to a question posed by another several minutes or days later. The choice of when to contribute to the discussion is at the sole discretion of the participants.

Computer conferencing, and computer-mediated communications in general, transform the traditional classroom communication model of one-to-one or one-to-many interactions. They enable a discourse which is based on many-to-many interactions, including learner-to-learner and learner-to-instructor. Ruberg et al. (1996) have reported significant benefits of the increased peer interaction and collaboration resulting from computer conferencing. According to these authors, these benefits are well documented in both cooperative learning research and in constructivist theories of learning:

1) students are forced to confront each other's ideas;

2) students can enact complementary roles, provide mutual guidance and support, and can serve as scaffolding to help each other accomplish learning tasks that might otherwise be too difficult;

3) students can find a direct relationship with a real audience from which they can get meaningful feedback;

4) students can experience and construct new understandings and ideas in a peer discourse setting (Ruberg et al., 1996, p. 245).

In summary, there are several benefits to the creation of a distributed CME learning environment through the WWW. First, it facilitates a student centered, collaborative learning environment which encourages greater interaction among medical practitioners. Emphasis is placed on collaborative information sharing and discussion. A second benefit to this distributed learning system is convenience; instruction and learning are not confined by space or time. Learners and instructors can access online continuing professional education from their home or office. The hypermedia courseware system is also easy to use. WWW browser software integrates access to all types of Internet resources in one easy to use interface. Both instructors and learners can quickly and easily learn to use Web browsers to navigate the Web, access course materials, and communicate with colleagues.

As well, courseware development is relatively quick and easy, and resources are also readily available on the WWW. Subject matter media and material which have been programmed in HTML are easily updated and disseminated. The Internet provides standardized access and is readily accessible to almost anyone with a personal computer, a modem, and an Internet connection. Commercial Internet service providers are now available in most areas and software for accessing the Internet is readily available, either by downloading directly from the Internet itself, or by purchasing browser software from a book or computer store.

The underlying rationale for utilizing a hybrid delivery system, alternatively known as a "Web site on a disk," is that it overcomes bandwidth obstacles by transmitting multimedia information and HTML documents to a user's computer from a CD-ROM. One of the main challenges confronting developers of multimedia Web-based courseware is the efficient delivery of electronic material and media to learners living in low bandwidth geographic areas. Multimedia information is streaming (continuous), meaning it generates continuous bitstreams that must be sent over a network in the correct order with virtually no delays. If out-of-sequence data are encountered, application performance is severely degraded or completely lost resulting in poor image quality or loss of audio (Galbreath, 1995). According to Galbreath (1995) computer data performance on the other hand, is not susceptible to such conditions. If packets of data are delayed, or are not received in the correct sequence, they can be retransmitted with minimal interruption to the end user. Networked computer data by nature is not heavily dependant on timing. However, networked multimedia data is continuous and relies on timing and synchronization.

A major issue in transporting multimedia across the Internet is bandwidth. Bandwidth refers to the transmission capacity of a given medium, or how much data the medium can transmit in a given amount of time (Galbreath, 1995). In many rural areas, the existing telecommunications infrastructure is POTS or plain old telephone service. These lines are made of twisted pair copper wires and were originally designed to carry analog information (voice), however they are capable of transporting digital information using the Internet when a modem is used. The modem receives and sends digital information from one computer to another, from one computer to a server, and from one server to another computer. The problem with POTS lines is they have a very small bandwidth: information is sent and received very slowly. Therefore, the smaller the bandwidth, the less information can be sent at a given rate of time. POTS lines present significant obstacles for the transmission of networked multimedia data which requires large bandwidth capacities.

Another problem, and one that users in urban areas face even where upgraded digital telecommunications are in place, is that the more users that enter or use a network, the less the available bandwidth for other users. In essence, bandwidth decreases proportionateley with the number of users on a network (Galbreath, 1995). Given that multimedia information cannot tolerate delays and needs guaranteed bandwidth to achieve optimal performance, the capability to provide and deliver efficient online CME multimedia courseware to rural physicians in low bandwidth regions is difficult to achieve. An alternative delivery model is a hybrid CD-ROM/World Wide Web instructional technology delivery system. The hybrid system overcomes bandwidth obstacles by transmitting multimedia information and HTML documents to a user's computer from a CD-ROM. Figure 1.1 provides a schematic overview of the Hybrid CME Courseware Delivery System.

The Hybrid CME Courseware Delivery System uses CD-ROMs (Compact Disc Read Only Memory) as storage devices for HTML documents. CD-ROMs are optical discs which are read by a laser-based disk drive. CD-ROMS were originally designed as mass-storage mediums for computer-readable text. They have high capacity, are inexpensive when duplicated in quantities, easy to distribute, and are quite resilient. However, they are slower than other devices, one needs a CD-ROM drive for delivery, and they may require the transfer of some data to a hard drive. Each 4.72-inch disc stores approximately 650 megabytes (MB) of digital data. Today, CD-ROM discs store digital text, graphics, audio, and video images.

The Problem

Sustainable rural communities are defined by their ability to remain integrated and cohesive over time and to sustain social, economic, cultural and political organization and activities in light of external and internal stressors. A conceptual model, based on the signing of the 1992 Rio Declaration on Environment and Development, best describes the interrelationship of the social, economic, and environmental factors involved in sustainable development. This model, presented as Figure 1.2 demonstrates that sustainable development is the balancing of environmental, economic and social concerns

Figure 1.1 The Hybrid CME Courseware Delivery System

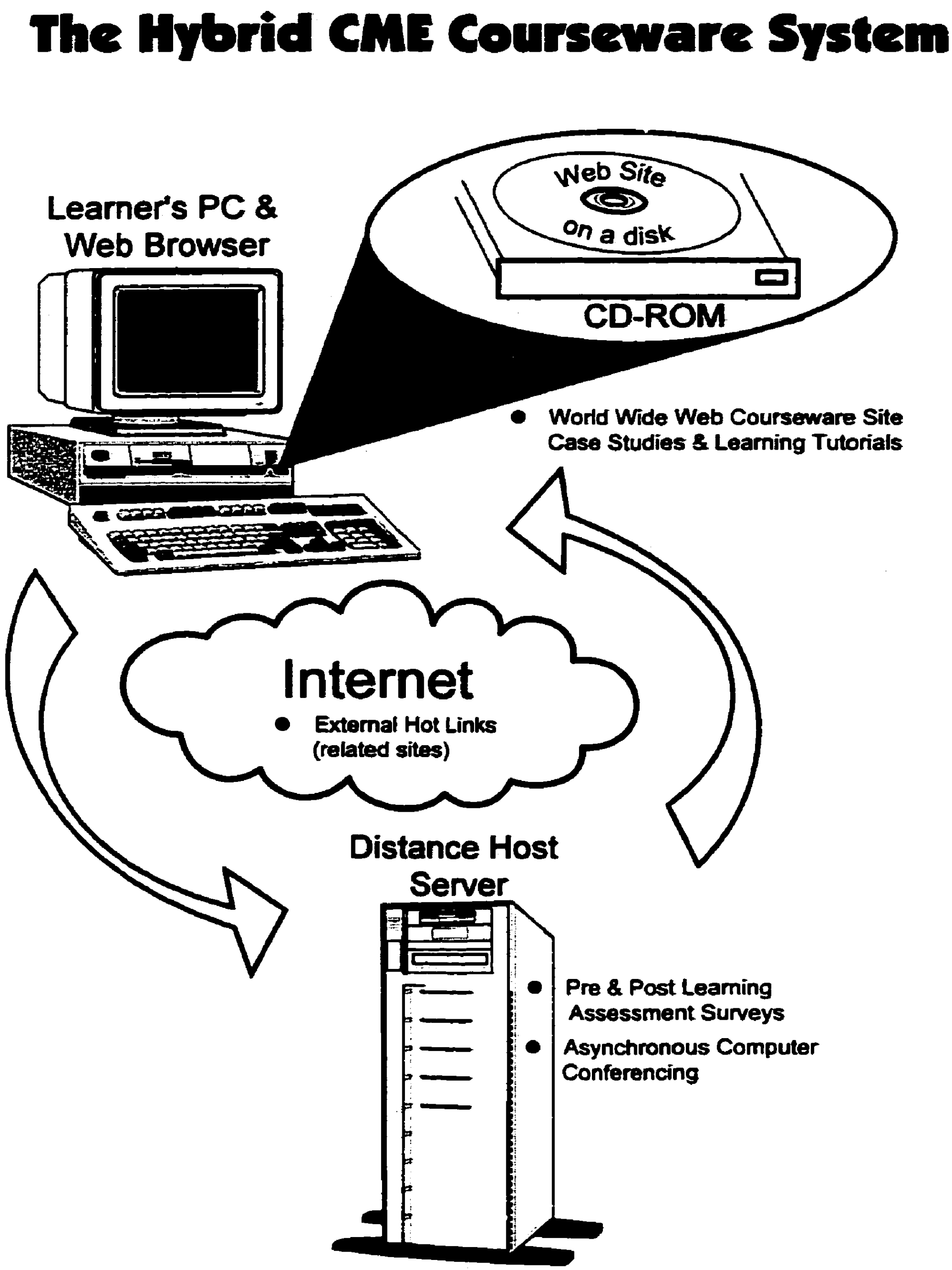
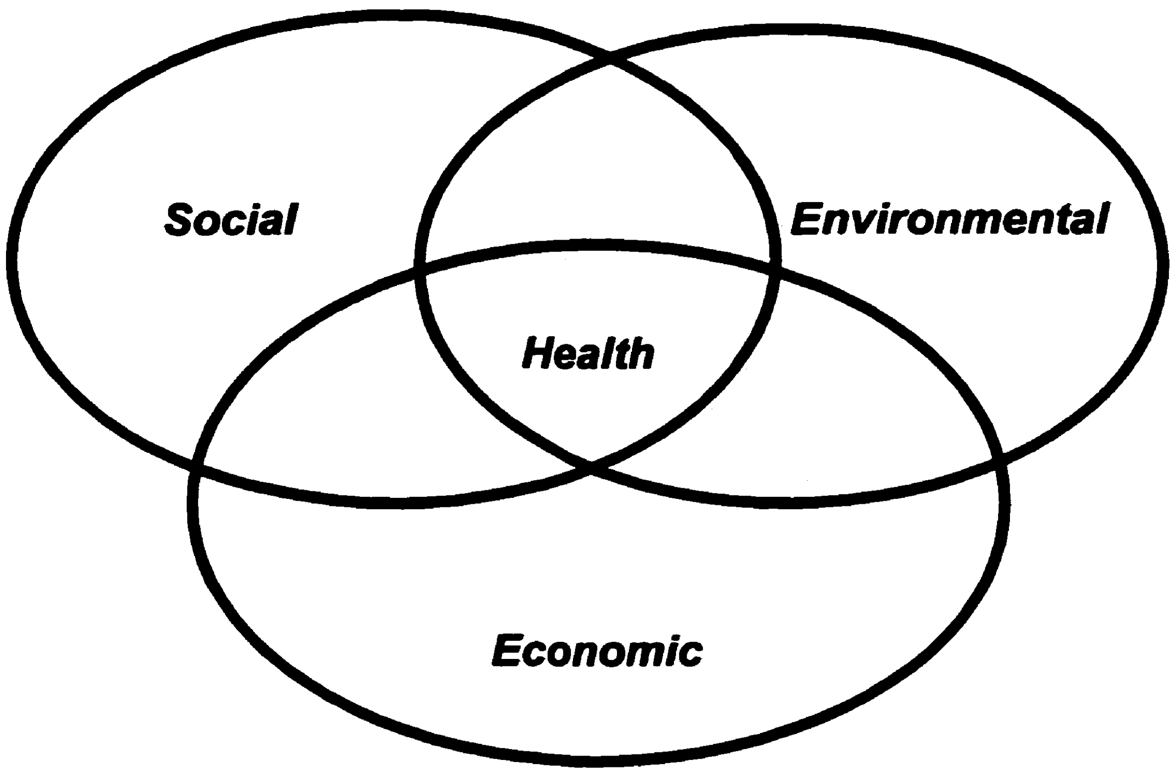


Figure 1.2 Sustainable Development Model (Rio Declaration on Environment and Development, 1992)



Source: Health Canada (1997)

to achieve health for the present generation, without sacrificing the well-being of future generations. Human beings are the center of concern for sustainable development and are entitled to a productive and healthy life in harmony with nature (Health Canada, 1997). This idea was also a major theme of the report of the World Health Organization Commission on Health and the Environment (1992) which stated that "the maintenance and improvement of health should be the center of concern about environment and development".

At the same time that the concept of sustainable development has been evolving, ideas about what makes human populations healthy have also been changing. Traditionally, the availability of conventional health services, including hospitals and physicians, was seen as the most important factor influencing health. However, over the last two decades, there has been growing awareness of the role other factors play in making people healthy. The concept "population health" encapsulates this concern with the living and working environments that affect people's health, the conditions that enable and support people in making healthy choices, and the services that promote and maintain health (Health Canada, 1997).

According to Health Canada's "Sustainable Development Strategy" the population health approach does not diminish the importance of the health care system and its contribution to a population's health, but stresses the importance of additional factors and the interactions among them. Determinants of health is the collective label given to the multiple factors that are now thought to contribute to the health of populations (Health Canada, 1997). They include such things as people's biological endowment and individual responses, the social and physical environment in which they live, the economic conditions of their society, and the accessibility and quality of the health care system.

The "population health" line of thinking, like our approximation of sustainable development and sustainable rural communities, suggests that many interrelated factors contribute to the health of a population. Among these factors, the quality of a community's health care services and programs, and the extent to which they are designed to maintain and promote health, to prevent disease, and to restore health and function are significant to the health of our rural communities. These elements have always been considered the cornerstone of health care systems. However, a shortage of rural and remote physicians currently challenges the provision of sustainable rural health care in many developed and developing countries, and subsequently the health and well-being of our rural and remote communities.

In Canada, an unbalanced distribution of physicians between urban and rural communities contributes to this problem. Although 23.5 percent of Canadians live in rural areas, including communities with populations of up to 10,000, only 17 percent of family physicians and 4 percent of specialists practice in these areas (Rourke, 1993). This apparent "under-servicing" has prompted many initiatives to improve the education, recruitment, and retention of rural physicians and has become a priority of rural communities, medical schools, medical associations and governments.

There are many obstacles with regard to recruiting and sustaining an adequate supply of rural health care practitioners. Common deterrents include lack of time for family and leisure, lack of work and educational opportunities for family members, professional isolation, lack of professional development opportunities, low salaries, poor locum support, underfunded hospital services and over-scheduling (Rourke, 1993; Rourke, 1994). Each rural setting has its own special challenges. In the smallest, most remote communities, help is a long time and distance away. This places immense strain on limited local resources and on the physician, particularly when emergencies occur. In larger rural communities with a small hospital there are different stresses. The rural doctor usually has a practice that includes house calls, nursing home visits and even extensive hospital-based medicine in addition to regular office practices. This means extra work including emergency medicine shifts, direct care of in-hospital patients, obstetric deliveries and sometimes general practice anesthesia.

Rural family medicine is a demanding and challenging form of medical practice. The rural physician frequently practices in an isolated environment with inadequate resources and limited or distant specialist back-up resources. This isolation necessitates a level of clinical competence beyond that of urban family physicians. As well, the rural physician is often expected to perform a generalist role in every aspect of clinical practice. Because of this, she must develop and maintain a special base of knowledge and technical skill in a variety of clinical areas -- particularly those related to rural medicine, including: emergency medicine, obstetrics and anesthesia (Rourke, 1988; Woolf, 1991; Kamien and Buttfield, 1990; Gill and Game, 1994).

Several studies have confirmed the existence of these unique and varied continuing medical education needs among rural physicians (Rourke, 1988; Woolf, 1991;

Kamien and Buttfield, 1990; Gill and Game, 1994). Some studies have also investigated the differences between the rural and urban physician's continuing education needs (Lott, 1995; Rosenthal and Miller, 1982; Woolf, 1991). These studies indicate significant differences in the continuing medical education needs of rural and urban medical practitioners. A majority of these studies also suggest that these differences are influenced by the nature of medical practice and, in some instances, by the distance of a rural medical practice from major urban areas. The further a rural physician is from an urban area and large urban health care resources, the more knowledgeable and competent he must be in a greater number of clinical areas.

Hays et al. (1994) working in Queensland, Australia, developed a "sampling framework" for rural and remote doctors and surveyed 311 of these doctors to compare their training and practice profiles with those of 142 urban doctors. They found that doctors who were more than 80 km (or one hour's travel time) from the nearest hospital and support services were more likely to practice a wide range of clinical and procedural skills. In a similar study, Bitt et al. (1993) surveyed 231 full-time Australian general practitioners. They found that rural physicians were more likely to be sole practitioners whose access to medical specialists and other support services was found to decrease relative to population.

It is no coincidence that rural physicians experience great difficulty participating in, and accessing, continuing medical education. The very factors which characterize rural medicine also present significant barriers for participating in CME activities. Geographic distance contributes to the cost of attending selected CME activities and increases the time required to be away from family and practice. Arranging the necessary locum coverage for their practice and hospital responsibilities also makes "getting away" difficult for rural physicians. These obstacles are of great concern for the rural physician who must maintain his skills in an ever-changing and developing field of medical practice.

Several authors have suggested that rural physicians perceive their opportunities for participation in traditional CME activity as inadequate (Lott, 1995; Gill and Game, 1994; Rosenthal and Miller, 1982; Woolf, 1991; Rubenstein et al., 1975). As well, Bhatara et al. (1996) have suggested that rural physicians' sense of professional isolation, because of a lack of continuing education opportunities, influences feelings of job dissatisfaction with rural practice. The result of this gap in access to, and participation in, CME is a lack of peer interaction and educational resources afforded by a large hospital staff and medical school, and an over-dependency on journal review and reading as the main method for addressing many continuing medical education needs (Lott, 1995; Rourke, 1988; Woolf, 1991; Gill and Game, 1994).

Several studies have examined alternative continuing medical education delivery methods and the rural physician's satisfaction with these forms of CME. Useful approaches to addressing continuing medical education needs, including regional CME workshops and the use of distance education technologies, have proven applicable for delivering CME in rural and remote practice areas (Rourke, 1988; Gill and Game, 1994; Rosenthal and Miller, 1982; Strasser, 1992; Wilson et al., 1982; Wise, 1994). Several studies have also focused attention on the use of modern telecommunications and computer technology for delivering CME to the rural and remote physician population. Of these technologies, several have been well documented in the CME literature. Audio teleconferencing, video teleconferencing, slow scan imaging, and videotape programs have been used for many years to deliver CME at a distance. (Black and Dunikowski, 1985; Dunn et al., 1980; Lindsay et al., 1987; McDowell et al., 1987; Oeffinger et al., 1992; Moore and Hartman, 1992).

One example of a distance CME program which has withstood the test of technological change and time is the audio teleconferencing program, Wednesday at Noon, delivered through the audio teleconferencing network of Memorial University of Newfoundland's Telemedicine Center. The Wednesday at Noon program has served an important role in the delivery of continuing medical education to rural and remote physicians practicing in isolated coastal communities of Newfoundland and Labrador. This program, initiated in the 1970s, provides weekly, one-hour CME audio teleconference sessions with clinical subject matter consultants using an interactive case-based instructional model. For over twenty years, rural physicians from around the province have logged-on to the weekly program to interact with specialists and their fellow colleagues.

In recent years, the landscape of the distance education field has been transformed. Significant advances in information and communication technology have enabled the rapid movement of information to almost anywhere in the world. Computer capacities and speeds have advanced to levels previously unimaginable, and present day innovations in multimedia and data compression capabilities are enabling the integration of voice, data, and images over computer networks. Delivery systems that use fiber optics or regular telephone lines are also making use of advanced telecommunication technologies (e.g., Integrated Services Digital Network) and these are revolutionizing the technologies used for delivering distance education programming.

The increased capacities of these information and communication technologies have also contributed to a movement away from traditional CME (Moore et al., 1994). Physicians who have installed computers in their offices can access a variety of distant databases. In some areas, electronic consultation networks have developed using e-mail technology and medical information systems which provide the rural physician with rapid access to assisted literature searches and other information sources. As well, direct telephone or even video consultations are rapidly evolving and shrinking the rural professional isolation problem (Kantrowitz et al., 1979; Verby & Feldman, 1983; Money, 1986; Moore and Hartman, 1992; Puskin, 1992). Telemedicine, using a combination of computer, multimedia and telecommunications technology is expanding throughout the world. According to Moore and colleagues (1994):

> "We believe that the forces that are currently changing health care........will shortly provide opportunities to create a new CME, one that will be more accessible, more convenient, and more relevant. This will be the new paradigm for CME." (Moore et al., 1994, p.11)

The modern computer technologies of the Internet and CD-ROM offer significant opportunities for addressing the CME needs of rural and remote physicians. However, being a new medium for CME delivery and a new area for medical education research there is little understanding of how effective the new computer-mediated learning technologies (Internet, World Wide Web (WWW), Computer Mediated Communications, Interactive Multimedia) are for the delivery of CME at a distance to rural and remote physicians. Yet, the effective transfer and implementation of computer-mediated learning technologies for the delivery of continuing medical education at a distance offers exciting prospects for:

- bridging and reducing the isolation of rural practice;
- enhancing and increasing the knowledge, skills, competencies and comfort of those practicing in rural areas;
- augmenting both recruitment and retention efforts of rural physicians;
- and, influencing the quality of health care provided in rural and remote areas.

Nevertheless, instructional technologies are only tools by which instruction and instructional material is facilitated and delivered. Telecommunication and computer technologies can solve many of the efficiency and access problems encountered by rural and remote health care practitioners, but issues surrounding the instructional effectiveness of continuing medical education programs at a distance are often neglected in reports of technological use. Researchers in educational technology have often warned that disillusionment with the use of technological innovation in instruction is largely due to too little consideration being given to questions of practical relevance of technology in educational settings. The educational effectiveness of certain technologies are usually not the driving force for using them to deliver distance education programming. Dawson (1991) suggests that many studies spend too much time asking questions involving "technology and media comparisons" rather than more "fruitful research areas on effectiveness, learner characteristics, learning task characteristics, instructional methods, and technology attributes" (p.26).

A majority of the studies in the continuing medical education literature concerning the efficacy and utility of distance education technologies are little more than a compilation of case studies, opinions, and advice. Many of the authors in this area have concerned themselves with describing the technologies used to facilitate CME at a distance, rather than assessing the quality and effectiveness of the technologies for facilitating learning, satisfying participants' learning needs, and enhancing or producing change in the performance of targeted clinical skills and competencies.

Evaluation research studies and evaluation models are needed if these new instructional technologies are to be successfully developed and effectively used in the continuing education of rural and remote health care providers. Rural communities need guidance before they invest scarce resources in expensive technologies that may not be appropriate or adequate for their needs. Governments, health care boards, and medical education providers need additional information to assure that new rural telemedicine and distance education projects are appropriate and effective. As well, reliable information about how well different technologies work for different purposes, the effectiveness of these technologies in achieving identified outcome measures, and the lessons that were learned by the pioneers in this field, could help others avoid the same mistakes and improve on the efforts of others.

Collis (1993) believes that considerable experience has already been accumulated with respect to the use of telecommunications in distance education, however the rapidly advancing possibilities of the distance education technologies are challenging those involved in the design, implementation, and evaluation of electronically distributed learning. One of the main challenges, now and in the future, includes the issue of how to predict and evaluate the educational impact and added value of new possibilities in telecommunications while they are still in evolution. In this regard, evaluation studies are particularly important.

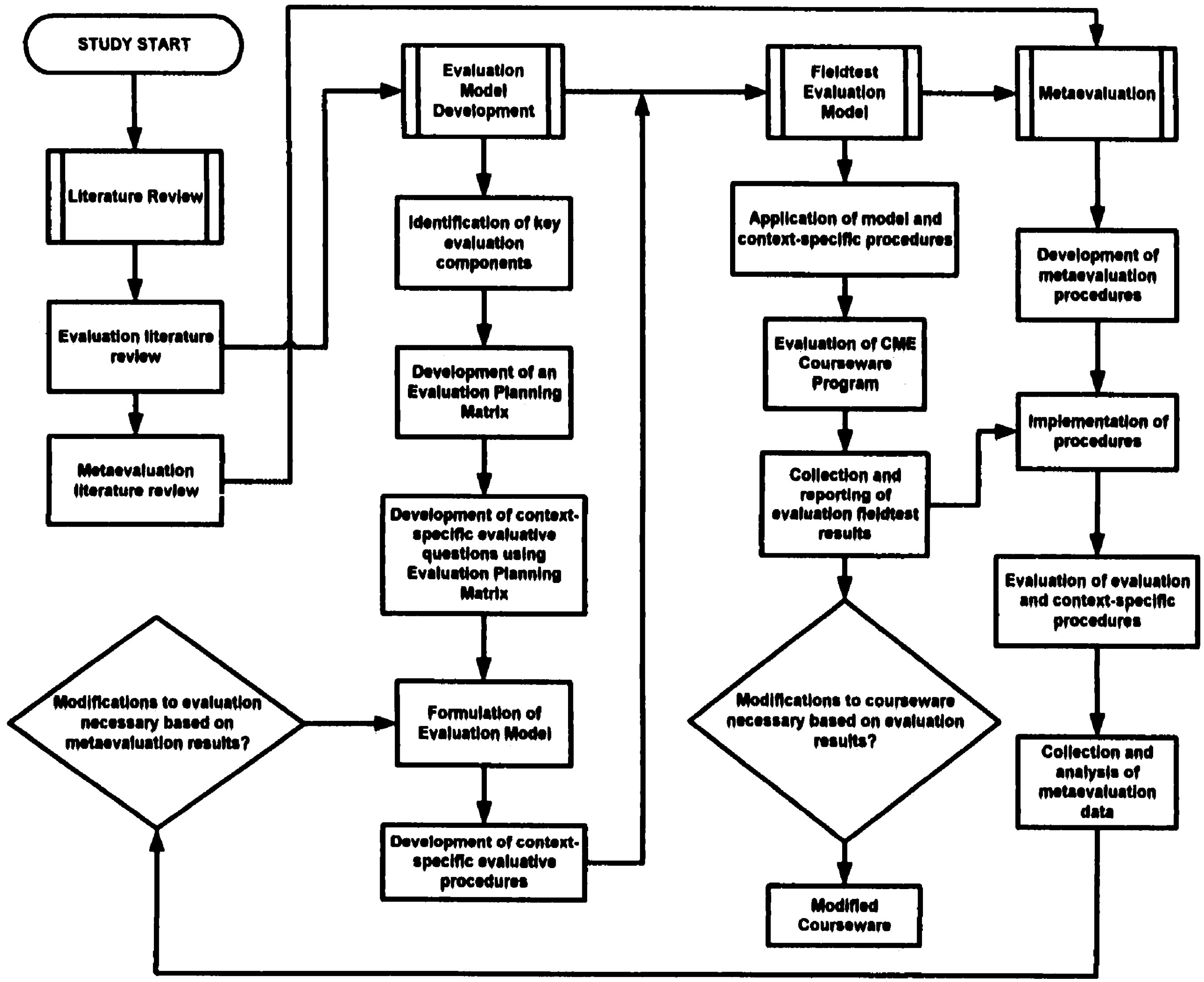
In light of the need to be more proactive in evaluating and reporting the effectiveness of continuing medical education at a distance, and with the absence of any evaluation models designed specifically for computer-mediated CME courseware, it is suggested that an evaluation model for computer-mediated courseware be designed. This model should provide a mechanism that allows its users to conduct comprehensive evaluations of the effectiveness of computer-mediated instruction. At the same time, the evaluation model should be flexible enough to be adapted to specific settings and to allow its users to be eclectic in operationalizing its application.

Purpose of the Study

The study reported in this dissertation was conducted to develop and validate an evaluation model for computer-mediated CME courseware. As indicated earlier, there is growing interest in the use of telemedicine and Internet technologies as a means for enhancing the health care services which can be provided to rural and remote communities. A great deal of time and resources can be expended on the design, development, and delivery of computer-mediated courseware programs. It is therefore imperative that the individuals involved in developing these programs, as well as key stakeholders within the health care field, be provided with meaningful information and evidence that these CME programs are effective in the continuing education of rural and remote health care providers. It is assumed that an evaluation model for computer-mediated instructional courseware would be of great interest to a number of these individuals.

Figure 1.3 provides a schematic overview of the conceptual elements which comprised this dissertation's study design and were used to address the purpose and research questions of the study. There were four main activities which directed the study. First, a literature review of evaluation models, frameworks and methodologies, and metaevaluation procedures. This literature is described and discussed in Chapter Two of this dissertation. Second, the design and development of the courseware evaluation model and the metaevaluation procedures which were used in the study. The process which was followed for designing and developing the evaluation model and the procedures of the metaevaluation are presented in Chapter Three. Third, the fieldtest of the evaluation model and the reporting of fieldtest results as they applied to the evaluation of computer-mediated instructional courseware. The results of the fieldtest are presented in Chapter Four. Fourth, the metaevaluation of the evaluation of the CME instructional courseware using the metaevaluation procedures developed for this study. The metaevaluation results and findings are also presented in Chapter Four.

Figure 1.3 Study Design Conceptual Flow Chart



Limitations of Findings

The evaluation model resulting from this study was designed and developed based on a review of various evaluation approaches, frameworks, models and methodology. The model was then applied to a computer-mediated instructional courseware product which was being developed and implemented. The evaluation conducted on the courseware served as the fieldtest of the evaluation model. The evaluator shared the results of the fieldtest with the various stakeholders involved in the design, development and delivery of the program. These stakeholders included: CME Planning Committee members; multimedia program developers; subject matter experts; Coordinator, Office of Professional Development; and, the Vice-Dean, Professional Development. A metaevaluation, an evaluation of evaluation, was then conducted on the fieldtest of the model as a means of validation.

The primary limitations of the study were concerned with threats to external and internal validity. External validity was a concern because the evaluation model was only fieldtested with one particular computer-mediated courseware program and one group of participants. The obvious limitation of this design relates to the extent to which the fieldtest findings could be generalized to other programs, populations or groups. The hybrid courseware delivery system evaluated in this study was an innovative means for delivering computer-mediated instruction. It utilized CD-ROM and WWW technologies, and integrated multimedia-enhanced self-paced instruction with asynchronous computer conferencing. The content of the instruction was dermatological office procedures. Therefore, the external validity of the fieldtest results may be limited, or the inferences which can be made beyond the fieldtest, to other courseware or distance education programs. However, while the fieldtest results may have limited generalizability to other populations and programs, there may be sufficient external validity to make inferences related to other courseware programs for physicians.

The internal validity of the fieldtest was addressed in the study by the use of a modified quasi-experimental evaluation study design. However, the sample size of the study groups suggest that the findings should be interpreted with some caution. As well, all possible precautions were taken throughout the process of developing and administering the evaluation instruments, and while scoring and analyzing the data to minimize the effects of internal validity concerns.

A great deal of time was devoted to the selection and construction of instruments for measuring the evaluation criteria of this study. The pre and post-learning achievement test items were developed by an instructional design specialist and the subject matter consultants, and were based on the ascribed learning objectives. The Computer Attitude Scale was found to have high validity and reliability, both for this sample of physicians and previous sample learner groups. The items of the Courseware Evaluation Survey were based on several studies and the survey was pilot tested among a group of instructional design specialists. As well, the Retrospective Pretest-Posttest Performance Survey was developed by the subject matter consultants and the instructional design specialists, and was based on performance objectives which had been ascribed by the CME Planning Committee. Commonsense, expertise, previous research findings, and the wisdom of the planning committee members guided the development of instruments and test items.

The inferences concerning the effectiveness of the evaluation model as a mechanism for measuring the effectiveness of computer-mediated CME courseware programs may be more limited. The evaluation model was not tested on other types of courseware products or programs with different subject matter. It was not compared critically to other evaluation models, frameworks, or guidelines which appear in the literature. Thus, inferences could be made only to the evaluation of similar programs with similar populations. As well, there were limitations to the metaevaluation findings based on the evaluator self-report procedure and the notion of net-benefit.

One of the procedures which was used in the metaevaluation process was that of evaluator self-reporting. As a component of the metaevaluation process the evaluator prepared a self-report that addressed the strengths and weaknesses of administering and analyzing the evaluation instruments which were developed for the fieldtest. This self-report constituted a critical and reflective analysis on practice and technique by the evaluator who conducted the fieldtest of the evaluation model. A procedure of this type is open to criticism because it may lack objectivity or introduce evaluator bias. However, Michael Collins (1991) is a strong advocate of such procedures and has made a call for the need for a more critical, reflective practice of adult education.

> "Careful reflection by adult educators on the nature of their own practice is integral to the quality of that practice. The term careful is intended to signify a kind of reflection in practice that is caring - care-ful - as well as attentive to detail and critical. Thus, while careful reflection does not offer an easy way out, it understands that practice entails failures. It identifies shortcomings but leaves us with the prospect of nurturing a principled vocation." (Collins, 1991, p.80)

A second limitation to the metaevaluation process relates to the notion of net-benefit. In evaluation, net-benefit refers to the value of evaluative information relative to the costs to collect and analyze it. The fundamental question underlying net-benefit centers upon the effort and resources expended upon evaluative activity and whether the benefits of collecting this information outweigh the expended time and resources. Will the information enhance or improve decision-making concerning the worth or value of an educational program or product? The limitations in the study concerning net-benefit relate to the fact that the courseware and the evaluation which was conducted were funded by external grants. This external funding may have created a context which overshadowed the actual resources and costs associated with conducting the various procedures of the evaluation model. The issue raised as a result of this context is whether the values and perceptions of the various stakeholders and decision-makers consulted during the metaevaluation would have been the same were the project not funded by external grants.

In summary, based on these limitations it is difficult to claim that the specific courseware program evaluated in the fieldtest would be similarly effective with a different sample. Likewise, it is difficult to propose that the evaluation model would be similarly effective with a program with different content or instructional strategies, if applied in a different context under contrasting circumstances, or that it is more effective than other evaluation models and frameworks. However, stronger conclusions and recommendations can be made concerning whether or not the courseware program was effective in this particular situation and whether or not the evaluation model was effective when applied to this particular instructional courseware product.

Research Questions

The following research questions were formulated to direct the study:

1. What were the strengths and/or weaknesses of the evaluation model?
2. Was the evaluation model useful in serving the information needs of the intended users?
3. Did the evaluation model follow practical and feasible means for collecting evaluative information?
4. Was the fieldtest of the evaluation model conducted in an ethical manner, with due regard for the welfare of those involved in the evaluation, as well as those affected by its results?
5. Did the evaluation model convey technically adequate information about the features that determined the worth or merit of the program being evaluated?

A metaevaluation process, an evaluation of an evaluation, modified from Sheets (1983) was developed to answer the research questions and assess the effectiveness of the evaluation conducted during the fieldtest. Using the Program Evaluation Standards (2nd ed.) established by the Joint Committee on Standards for Educational Evaluation (1994) as a guideline, the evaluator was able to evaluate the process, procedures, and results of the fieldtest of the evaluation model. Research questions two to five were based on the four categories of standards proposed by the Joint Committee: utility, feasibility, propriety, and accuracy. Metaevaluation is discussed in greater detail in Chapter Two of the dissertation while the actual metaevaluation procedures and results are described and presented in Chapter Three and Four.

Significance of the Study

The study described in this dissertation was exploratory in nature. Exploratory research is often used to learn and understand more about an area of investigation in order to design and execute more systematic and extensive studies of the same phenomena. This is a common purpose in research when a new interest is being explored or when the subject of the study itself is relatively novel and unstudied. Opponents of exploratory studies argue that they rarely yield definitive answers about a topic and are often difficult to conduct because there are few guidelines or rules which may direct the researcher.

Proponents of exploratory research studies, on the other hand, argue that such approaches to inquiry foster creativity and enable the investigator to adapt modes of inquiry to the item or object of study. Exploratory studies also allow the researcher to adopt an investigative stance, maintain an open mind, be very flexible, and explore all sources of information. Creative questions must be formulated and asked, and the researcher is encouraged to explore avenues of investigation which may uncover unexpected factors that have larger implications for the field in general.

An exploratory research approach was followed as the structure for inquiry of this study because the topic, the evaluation of hybrid computer-mediated learning systems, was new and relatively unstudied, and the broader principles surrounding it were largely unordered. The notion of evaluation is not a new concept to the field of education. A variety of frameworks, models, and approaches exist in the literature, however none of those which were reviewed offered an ordered system of comprehensive procedures or methods for evaluating instructional courseware, particularly courseware which merged a variety of new media learning technologies. The literature was also absent of any empirical studies which sought to compare the efficacy or efficiency of one framework versus another. A majority of the literature which critiqued the various evaluation approaches was based on non-empirical evidence. In an attempt to provide order to the abstract principles and concepts revealed in the evaluation literature an instructional courseware evaluation model was designed eclectically, fieldtested, and validated using a metaevaluation system. It is this process, described in detail in this dissertation, which makes a significant contribution to the development of theory in the field of continuing education and new media learning.

Theory goes beyond what one can observe and measure, and refers to a set of interrelated definitions and relationships that organize our understanding of empirical concepts in a systematic manner (Marshall, 1994). It is usually constructed for several reasons; because it expresses a new unifying idea about a phenomena, addresses previously unanswered questions, and/or provides new insights on the nature of the phenomena in question. Theories and models of evaluation are intertwined and are often used synonymously. A model is a reference to a system of abstract concepts and is often used to graphically depict theoretical concepts (Marshall, 1994). The essence of a model is that it requires researchers to engage theory. Models seek to simplify the phenomena in question, assisting us in conceptualization and explanation, and acting as an aid to complex theoretical activity by drawing our attention to concepts, variables and their interrelationships.

Education, in all its forms, is a practice discipline. Its practitioners are directly engaged in designing, developing, implementing, and evaluating instructional programs and systems in a variety of contexts, at various levels of complexity, and under diverse circumstances. This practice is also supported by another level of practitioners, those who educate students, administer educational programs and institutions, and develop and analyze knowledge for informing the practice of educating. This breadth of educative practice and the various undertakings which form its essence indicates that the field is a complex discipline. It is within this complexity that researchers develop theory in order to describe, explain, predict, and control educative practice.

The development of theory provides a way of identifying and expressing key ideas about the basis of educational practice. It is through theory development that this basis may be explored and explained in terms of general or more delimited descriptions of the learner, learning, instructional systems, and the educational system. As an example, the basis of practice may be studied in a delimited way by focusing on specific events that occur in specific contexts, such as the characteristics of adult learners who take part in distance education programs. In contrast, a more abstract notion of theory development could focus on an overall guiding delineation of the field of distance education. Despite the grandeur nature or delimited scope of theory development it is, in its developed and developing form, aimed at helping the educator or trainer to understand practice in a more complete, systematic, comprehensive, and insightful way.

The main contribution of this study to the development of theory would be at a level of practice theory. Practice theory focuses on the achievement of a goal and the prescription of actions or tasks for reaching that goal. In educational practice, the goal of evaluation is to collect and disseminate information for improving decision-making concerning the merit, worth or value of some educational program or system. The significance of this study to a theory of evaluative practice is it delineates a process for synthesizing evaluative theory, constructing a model for evaluation based on eclecticism, provides a model for guiding the assessment of computer-mediated instructional courseware, and offers a system for evaluating the effectiveness of the model which has been designed. Therefore, apart from the evaluation model itself, the process by which this model was eclectically constructed as well as the process for evaluating its validity are all contributions to the field.

According to Walker & Avant (1995) theory development may be needed when one of several situations exist. First, a variety of concepts about an area of knowledge might exist but there is no way to link them all together. A second situation may encompass circumstances where theory already exists and the theorist's purpose is to determine the strengths and weakness of that theory. A third situation in which theory development might be needed is one in which there is a body of literature, but it has not been successful in explaining a phenomena or in guiding practice. In this instance an analysis of current theories in the area of investigation might indicate inconsistencies in the traditional theory base. A synthesis of the current theory base and a derivation of new theory could provide a new structure for the concepts, unifying ideas about the phenomena of interest, and offering insights for informing practice.

At a general level, a review of the literature indicated that there were no models in existence which linked the various concepts and methods of evaluation into a unifying body of knowledge for guiding the effective evaluation of computer-mediated instructional courseware. The result of this analysis and synthesis of evaluation was the need for the development of an eclectic model of evaluation. The broadest definition of eclecticism is that it is a theory that selects what is best from among many theoretical stances. What is best means what works in a given context at a given time. Eclectic evaluation, in this manner, is guided by a certain level of pragmatism because the goal of the eclectic evaluator is to ascertain what evaluation questions, methods, and instruments are most effective for a specific program and under particular circumstances. Sceptics, on the other hand, would offer a more negative opinion of eclectic evaluation by saying that eclecticism does not adhere to a single viewpoint. They would argue that an eclectic evaluator can succumb to disillusionment because he takes a position that no one theory, model, framework, or approach is true or efficacious for evaluating educational and training programs. Contrary to this opinion, an eclectic approach to evaluation can be viewed to be just as effective as any other noneclectic framework until proven otherwise. Eclectic evaluation allows the extraction of key principles from many conceptual models without having to agree with the underlying components of the theory from which those principles were derived. The evaluator may, in these instances, use these principles to provide order and to capitalize on particular procedures for guiding evaluative activities in often unique and atypical circumstances from those for which the original theoretical concepts were intended.

A significant contribution of this study to current knowledge is found in the eclectic evaluation model which was designed and developed following such an approach. It is this notion of eclecticism which may best guide future evaluation efforts in the field of technology-based learning. Mirroring the rapid advances in the development and growth of information and communication technologies, the field of distance education is being transformed by the introduction of innovative technology-based delivery mediums. An eclectic approach to evaluation enables evaluators to draw on the strengths of a variety of theoretical frameworks in order to construct evaluation processes which take into account the intricate nature of the learning technologies, the instructional strategies they carry, and the interactive communications they facilitate.

Eclectic evaluation models can be used to evaluate the effectiveness of computer-mediated instructional courseware programming at formative and summative levels. At a formative evaluation level, program developers and instructional designers can use a variety of formative assessment procedures to make changes to technology-based learning systems while they are still in development stages. Summative evaluation procedures can be used to produce information which reveals the impact of technology-based learning systems on learning achievement, learner satisfaction, and performance change. Eclectic models also allow the evaluator to combine a variety of quantitative and qualitative methods and instruments in collecting information to guide decision-making. It is the flexibility inherent in eclectic evaluation that will enable future practitioners to adapt and utilize the merits of a variety of evaluative methods and approaches in conducting exploratory research on the effectiveness of the new media learning technologies.

Definition of Terminology

Asynchronous Computer Conferencing: Computer-mediated communications software that enables learners and instructors to communicate via textual messages at varying times and places. Using a WWW browser participants can access the software to read, create, edit, and publish messages to a common online discussion area residing on an Internet server.

Browser: A software program which uses a graphical user interface and allows users to read hypermedia or hypertext information on the Internet via the WWW.

CD-ROM: Compact Disc Read Only Memory. Optical storage disks with large density to store a variety of computer data and multimedia information.

CME: Continuing Medical Education. All formal and informal educational, training, and self-directed learning activities which are designed to enhance, improve and/or change the clinical and professional knowledge, skills, and/or attitudes of practicing and licensed physicians.

Collaborative Learning: Learning which occurs when individuals have the opportunity to share information and experiences with their peers, confront and discuss conflicting viewpoints, and collaborate in the completion of assignments or self-directed learning activities.

Common Gateway Interface (CGI): The protocol standard that enables electronic form submission and interactivity on the Web.

Computer-Mediated Instruction (Learning): The use of computers, computer networks, and electronic materials distribution mechanisms (such as the WWW and CD-ROM) to publish and distribute learning materials, and facilitate communications between instructors and learners studying at a distance.

Distance Education: The delivery of instruction and instructional materials to learners who are separated by place and/or time from the instructor or organizations providing the instruction. Learning materials are distributed and/or communication and instruction is facilitated through the use and application of various technological mediums such as print, computers, videos or telephone. Also known as distance learning.

E-Mail: A form of computer-mediated communications that uses special software programs for receiving, saving, creating and forwarding textual messages to other computer users via a communications network such as the Internet.

Evaluation: The determination of the worth, value or merit of something with the purpose of collecting and providing information to decision-makers and stakeholders for planning, rejecting, revising, and/or improving a particular program, product, and/or materials. In education and training, evaluation can focus on learning achievement (cognitive gain), participants' reactions and satisfaction with instruction and instructional materials, learners' behavioral changes (performance gain), and impact of instruction on broader outcome measures, such as health outcomes (impact gain).

Evaluation Model: A conceptual guide for designing, developing, and implementing an evaluation of education, training, and instructional materials.

Fieldtest: A step in the systematic development of a process or product in which the process or product is applied or implemented in a setting that approximates the context for which the process or product is intended to be used.

Formative Evaluation: Evaluative activities or procedures which are conducted during the instructional design and development stages of educational materials or products. The purpose is to identify and incorporate changes and revisions to instructional materials or products while they are still in a development state and before they are delivered.

Graphical User Interface (GUI): An interface system which uses icons and other visual and audio aids to guide a user in using a device, instrument or program.

Hybrid Courseware: A computer-mediated instructional system which integrates CD-ROM and WWW applications to carry instructional media, provide hyperlinks to resources on the Internet, allow for interactivity through CGI form submissions, and enable collaborative learning via computer mediated communications.

Hypermedia: Information presented via various media types (text, video, graphics) which are connected by special HTML tags called hyperlinks and allow access from one chunk of information directly to another.

Hypertext Markup Language (HTML): The program codes used to develop documents for publication on the World Wide Web.

Internet: A global network of computer networks that follows a common communications protocol (TCP/IP) and addressing scheme (HTTP) and allows computers connected to the network to link and share resources with each other.

Metaevaluation: "the process of delineating, obtaining, and using descriptive and judgmental information about the practicality, ethics, and technical adequacy of an evaluation in order to guide the evaluation and publicly report its strengths and weaknesses" (Stufflebeam, 1981, p. 151).

Summative Evaluation: Evaluative activities or procedures which are conducted after an educational program or product has been implemented. The purpose is to measure the impact or intended and unintended outcomes of instruction and educational products or materials.

Server: Computer hardware and software which are connected to the Internet and manage the various linkages among remote computers and enable communications, resource searching, and information sharing.

World Wide Web (WWW): The front-end application of the Internet which presents information distributed through the Internet via text, audio, video and other multimedia formats.

Summary

In summary, the purpose of this study was to develop and validate an evaluation model for evaluating the effectiveness of computer-mediated continuing medical education instructional courseware. A hybrid computer-mediated courseware program on Dermatological Office Procedures served as the product on which a primary evaluation was conducted. This primary evaluation functioned as the fieldtest of the evaluation model while a secondary evaluation, a metaevaluation, enabled the validation of the effectiveness of the evaluation model. The subsequent chapters of this dissertation will describe the processes followed for conducting the study.

Chapter Two will examine the literature related to the use of distance learning technologies in the delivery of CME, metaevaluation, evaluation models and methodologies, and antecedent evaluation variables. The material presented in this chapter served as the source of information for designing the evaluation model and metaevaluation procedures used in this study.

Chapter Three presents an overview of the processes followed for designing and developing the evaluation model which was fieldtested in this study. This chapter also describes the specific evaluation instruments which were developed using the evaluation model as a conceptual guide. The procedures which were developed for evaluating the evaluation model fieldtest are described as well.

The results of the fieldtest of the evaluation model and the metaevaluation are presented in Chapter Four. In Chapter Five the results of the fieldtest, as they apply to computer-mediated courseware instruction, and the metaevaluation are discussed and explained.

Chapter Six of the dissertation draws conclusions to the study and a discussion of the implications of the findings are presented. A modified evaluation model is also proposed because of the findings resulting from the metaevaluation. It is this modified model which has significance for future developers and providers of computer-mediated continuing education.

CHAPTER II

REVIEW OF THE LITERATURE

Introduction

This chapter serves as a review of the research literature which was pertinent to the design and validation of a model for evaluating the effectiveness of computer-mediated CME instructional courseware. The review examines various topics and concepts related to continuing medical education, the evaluation of continuing medical education activities which have been delivered at a distance, metaevaluation, evaluation models, evaluation methodology, and antecedent variables. The discussion includes an analysis of the advantages and disadvantages of the various models and methodologies, and summarizes the components which are most useful in an eclectic evaluation of computer-mediated CME courseware.

Evaluation and Continuing Medical Education (CME) at a Distance

Continuing medical education (CME) involves the ongoing process of maintaining professional competence through reading, consultation, group experience or other learning activities. Continuing professional competence is one of the more difficult problems which confronts members of the medical profession. Physicians are expected by society to provide a high standard of medical care which reflects the latest scientifically accepted diagnostic and therapeutic procedures.

In Canada, certification to practice as a family physician is governed by the

College of Family Physicians of Canada (CFPC). The principle mandate of the CFPC is to promote high-quality medical practice by Canadian family physicians through the coordination and promotion of education at undergraduate, postgraduate and continuing medical education levels. In 1969, the CFPC introduced certification in family medicine to recognize those members who could demonstrate acquisition of the knowledge, skills and attitudes integral to the practice of family medicine. From the beginning, it was intended that those physicians who were successful in attaining this standard would make a commitment to maintain it. The first Maintenance of Certification Program was introduced in 1977. Since then, certification has been considered a continuing process throughout one's medical career.

All departments of family medicine at the 16 Canadian medical schools are accredited by the CFPC to train residents in family medicine. Following completion of the family medicine residency programs, family physicians have the choice to remain active members of the CFPC. Though membership in the organization is voluntary, it is contingent upon a mandatory program of self-directed continuing medical education throughout one's career. This is done through a CME credit program whereby the CFPC accredits CME activities according to standards and principles that promote high-quality learning. Members are required to participate in a minimum amount of acceptable continuing medical education (50 hours of CME each year). This has been achieved through a CME credit system, in which one credit is assigned to the equivalent of one hour of CME.

In 1995 the CFPC introduced some important changes to its CME requirements.

These were brought into a single, integrated program called MAINPRO (Maintenance of Proficiency). MAINPRO has evolved to be flexible and fair, while also providing guidelines to ensure that selected CME activities are useful and meaningful. A central program planning component of the MAINPRO guidelines is evaluation. MAINPRO program providers are encouraged and required to utilize some method of evaluation as a means for assessing and improving the educational quality of CME programs. Evaluation is seen to play several important roles as part of the MAINPRO guidelines: as a measure of learning effectiveness; as a decision-making tool for program providers and future CME participants; and, as a quality assurance measure for CME Planning Committee members and other stakeholders.

A review of the literature on the evaluation of traditional face-to-face continuing medical education reveals that thousands of studies have been conducted since the 1950s. Numerous investigators have undertaken the evaluation task as a means for better understanding and decision-making regarding the affect of educational interventions on physicians' performance, and the health outcomes of their patients (Davis et al., 1984, 1992, 1995; Lloyd and Abrahamson, 1979; Stein, 1981). One of the main conclusions made by Davis et al. as a result of several meta-reviews of the evaluation literature was that CME is an effective intervention for enhancing and changing physicians' performance, and to a limited extent the health outcomes of their patients.

Lloyd and Abramson (1979) appear as the first investigators to perform an extensive meta-review of the CME evaluation literature. They reviewed the CME evaluation literature for the period 1966-1977 and found only 47 studies published in

English that utilized objective and systematic methods of evaluation. Twenty three of these studies demonstrated changes in physician knowledge, competence, or performance, or an effect on patient care. Stein (1981) also conducted a meta-review of eight CME evaluation investigations published during the 1970s which reported changes in physician behavior (and, in one, improved patient outcome). In each of these studies investigators reported systematic and planned efforts to evaluate effectiveness by means of: cognitive tests of knowledge gain; chart review; follow-up questionnaires; analysis of changes in referral patterns; attitudinal questionnaires; audience reaction questionnaires; and course or program completion records. The meta-review revealed that CME activities organized on sound educational principles were effective in producing knowledge gain and changing or enhancing physician performance.

The most extensive meta-review of the literature pertaining to the evaluation of CME has been performed by Davis et al. (1984, 1992, 1995). These authors reviewed the CME evaluation literature over a 30 year period, from the 1960s to the 1990s. In their first meta-review Davis et al. (1984) reviewed the literature evaluating continuing medical education and its impact on the satisfaction, competency and behavior of physicians, or on the health of their patients. The articles included in their review were from 1964 to 1982. Their review revealed various forms of study design in the evaluation of CME, including the most rigorous, the randomized controlled trial, as well as the controlled trial and the case control approach.

Of the 238 studies investigated, a majority (71 percent) were executed in the least rigorous fashion, that was either descriptive or before/after study design. Only 24 articles

(10 percent) reported randomized control trials. All of the randomized control trial studies, except one, showed behavior and/or knowledge changes. Of all the 238 articles, 86 studies assessed competency measures of a knowledge, skill or attitudinal type. Over 90 percent of these studies showed a positive change in competency as the result of a CME intervention. The authors' main conclusions from this meta-review focused on the need for more rigorous measures of effectiveness in future studies and the finding that CME interventions can and do have an effect on the participating physicians' competence and performance and (less consistently) on patient outcomes.

In their second meta-review Davis et al. (1992) reviewed the literature from 1975 to 1991. In this review they only selected CME evaluation studies which included a rigorous randomized controlled trial study design. They identified 777 CME studies, of which 50 met their criteria. Thirty-two of these studies analyzed physician performance; seven evaluated patient outcomes; and eleven examined both measures. The majority of the studies of physician performance showed positive results in some important measures of clinical resource utilization, counseling strategies, and preventive medicine (Davis et al., 1992). Of the 18 studies of health care outcomes, eight demonstrated positive changes in patients' health outcomes. The authors concluded that CME interventions were shown to consistently improve physician performance and in some instances, where measured, to improve health care outcomes.

In a third, and the most recent, meta-review Davis et al. (1995) reviewed the literature from 1992 - 1995. They utilized the same criteria as their 1992 study including a review of studies that followed a randomized controlled trial study design and objectively assessed physician performance and/or health care outcomes. Ninety-nine studies were reviewed by the authors. Seventy percent of these studies demonstrated a significant change in physician performance and 48 percent of the interventions aimed at health care outcomes produced a positive change. They concluded that there was a:

> "...robust body of research assessing the outcomes of physicians' clinical education, as evidenced by several thousand extant articles, including increasing numbers of the most scientifically rigorous variety, the randomized-controlled trial. In addition, this research has grown substantially over the years...during the three years between the current and the last review the number of randomized controlled studies has doubled" (p. 703).

The authors also suggested that the educational interventions that constituted the field of CME was growing and effective CME programs could comprise a variety of methods beyond the traditional short-course model. The main conclusion from their most recent meta-review was physician performance could be changed or improved by many forms of CME interventions and to a lesser extent, so could health care outcomes. Davis et al. offered several explanations as to why changes in health care outcomes lagged behind those of physician performance. These included: patients' not accepting physician recommendations; the socioeconomic and educational status of patients; and, frequently the limited effectiveness of the clinical interventions themselves (Davis et al., 1995).

Continuing medical education programs which are delivered using distance education technologies and methods, and planned in accordance with the MAINPRO accreditation criteria are eligible for continuing medical education credit. Distance education techniques have been used for many years to provide continuing medical education to rural and remote physicians. The earliest reports of CME being delivered at a distance date back to the 1960s. However, contrary to the rigorous evaluation studies reported in the literature on traditional face-to-face CME, the literature on continuing medical education at a distance suffers from a shortage of reports of systematic evaluation. A majority of the evidence supporting the effectiveness of CME at a distance is, for the most part, based on satisfaction measures or descriptive case reports.

Distance education occurs when an instructor and learner(s) are separated by geography and time, and instruction is mediated through either print, communication technology, computer-based technologies, or a combination of these technologies. Distance education delivery modes are distinguished according to the technologies and medium used to carry the learning materials and facilitate the two-way communication between participants and instructors. The four main categories of distance learning technologies are audio, video, computer (data), and print.

Instructional audio tools that use the spoken word or voice include the interactive technologies of audio teleconferencing and short-wave radio. These technologies facilitate synchronous communication and instruction. The technical components of a typical audio-only conference might include telephone hand sets, speaker phones or microphones, an audio bridge that interconnects phone lines and controls noise, and a speaker device to facilitate multiple interactions (Willis, 1995). Audio teleconferencing has been a continuing education delivery mode for health professionals, and in particular rural physicians, since the 1960s (Meyer, 1983). The use of audio teleconferencing as a distance learning technology has been reported in several continuing medical education studies from the University of Wisconsin-Madison, University of Alberta School of

Medicine, Memorial University of Newfoundland's Faculty of Medicine, the Ohio State University Medical School, Albany Medical College, and selected CME projects in Maine and Texas (McDowell et al., 1987; House et al., 1981; Parker and Baird, 1977; Gellman and Franke, 1996). In one study, audio teleconferencing was even used internationally, bridging the Emergency Hospital of Yerevan, Armenia and the Boston University School of Medicine to provide formal continuing medical education opportunities to practicing physicians in that European country (Screnci et al., 1996).

Several of the studies reported in the literature evaluated the instructional effectiveness of audio teleconferencing through the collection and summarization of participant reaction and registration data (Treloar, 1985; Lindsay et al., 1987; Lockyer et al., 1987). However, other than these few studies, the literature is very limited in terms of any systematic approaches to evaluating the instructional effectiveness of audio teleconferencing in delivering CME to physicians. Most of the studies were concerned with asking questions and providing answers on the technology used, and its implications for cost efficiencies. They offered little information that could influence meaningful decision-making regarding the effectiveness of the technology for producing learning gain or changing physicians' performance. In fact, in several of the studies noted, the researchers concluded that there was a need for more intensive study of the effectiveness of audio teleconferencing which sought to assess specific behavioral change in clinician performance or patient outcomes as a result of participation in distance CME (Lindsay et al., 1987; Parker and Baird, 1977).

Two-way radio technology has also been used as a means for delivering CME at a distance to geographically dispersed physicians. Two-way radio technology was first introduced by the Albany Medical College in 1955 as a means to broadcast lectures to six hospitals within a 50-mile radius of Albany (Ebbert, 1963). Using a FM (frequency modulation) radio station, staff at the Medical College delivered one-hour presentations to participants in several New England states. The two-way radio technology also permitted questions from one site at a time during the presentation which allowed for minimal site-to-site interaction. However, Ebbert (1963) did not report on any systematic evaluation attempts to measure whether the instruction occurring over the radio network actually enhanced the knowledge, skill, or performance levels of participants.

According to Willis (1995) print-based distance education has been a mainstay of distance education and distance learning materials, and is the basis from which all other delivery systems have evolved. Many of the first distance education courses were offered by correspondence study, with print materials and correspondence sent and returned to students by mail. Even with the arrival of newer advanced distance learning technologies, print-based materials remain an integral component of most distance education programs. They do not rely on advanced technologies for delivery, and are therefore easy to use. Print material production is cost-effective and easily edited. As well, print-based correspondence courses are not dependant on the technological infrastructure of an institution.

Several studies have reported attempts to provide CME at a distance through correspondence, print-based methods, and home-study (Wilson et, 1982; Marquis et al., 1984; Beaton, 1977; Evans et al., 1986; Engel et al., 1992; Boswell et al., 1994 ;

Nyarango, 1991). In several of these studies, the authors used participant feedback forms as a method for evaluating learners' perceptions of correspondence CME study. The findings from these studies suggested that most participants found print-based distance CME boring, offered little opportunity for interaction with other learners or the instructor, and required a great deal of commitment, time and learner motivation (O'Dochartaigh, 1971; Engel et al., 1992; Boswell et al., 1994). As well, the studies indicated that there were high attrition rates associated with CME courses delivered through home-study and correspondence formats.

Marquis et al. (1984) did find an increase in knowledge gain as a result of participation in a correspondence CME course which required the learner to study and review several patient management problem (PMP) activities. In this study, an experimental group which received repeated PMP correction materials showed improved performance over a control group that did not receive any instructional materials. The authors reported a knowledge transfer of 75 percent to the clinical practices of physicians.

Young et al. (1988) reported a study which provided a home study distance learning program for the continuing education of physicians on breast disease and early detection of breast cancer. In this investigation, an educational package was delivered in two formats: home study and a face-to-face workshop. Two questionnaires were used, one to assess knowledge, and one to measure the attitudes of physicians. For both formats, participants' knowledge and comfort in dealing with breast problems were measured before and after participation in the instructional programs. The results of the study indicated that there was a significant increase in both the knowledge and attitude levels of both the home study and workshop learning groups. As well, the authors found no significant differences between the groups on knowledge gain or attitude toward breast disease.

Evans et al. (1986) used print-based distance educational materials to instruct rural primary care physicians on hypertension management. These authors attempted to examine the educational impact of correspondence study by assessing the degree to which patient blood pressure levels were affected by physicians' participation in the CME activity. The results of the study were not positive, suggesting that the home study instruction did not influence clinical practices. However, as Cervero (1988) proposes, using patient outcomes as indicators of the impact of CME are, at best, challenging and questionable. Even though physicians may have actually learned and applied new knowledge and skill in their practice, they have little control over the compliance of particular patients to treatment. Therefore, while the physicians may have prescribed the correct treatment or management protocols, it is the patient who ultimately controls individual behavior and whether they will comply with the treatment protocols.

In a similar study, Engel et al. (1992) of the Wellcome Tropical Institute looked at a correspondence CME problem-based program for rural family physicians. The findings of this study suggested that problem-based learning activities produced significant increases in rural physician knowledge and understanding. The instructional models followed in this study provide a useful outline for the development and delivery of problem-based CME correspondence courses. However, Engel's report of significant gains in knowledge and understanding were based on subjective observations he and the other investigators made. The authors did not use any achievement tests to assess knowledge gain in order to evaluate the instructional effectiveness of their program.

Many of these studies are limited in scope, the evaluation methods they used, and the findings they reported. A main focus in some of the studies was the impact of the educational program in terms of knowledge, attitudinal or performance change. However, while impact studies are important for indicating the amount of learning or behavior change which has occurred, they overlook the importance of the process of learning and the characteristics of the learners which may have contributed to the learning or performance change which occurred. Several of the evaluations also collected participant satisfaction data, but they offered little information or recommendations on ways to improve print-based instructional materials or correspondence study formats in future CME programs. As well, the evaluation studies did not report any methods for formatively evaluating instructional materials during the design and development stage of instructional design. This is a major limitation of the evaluations which were reported, particulary since print-based instructional materials were being used.

Video-based distance learning technologies include interactive video conferencing and instructional television. Interactive video conferencing uses compressed digital video for the transmission of motion images over normal telephone lines or data networks such as high capacity Integrated Services Digital Networks (ISDN) (Willis, 1995). Interactive video conferencing may be facilitated as a point-to-point or a multipoint connection in which several locations are simultaneously connected for a real-time video conference.

Asynchronous instructional video tools include still images such as slides, slow scan video, and pre-produced moving images such as film and videotape. Slow-scan video transmission has been used in several projects as a means for delivering CME to rural health care providers. Slow-scan video enables the transmission of two-way audio and freeze frame images over two telephone lines. Images, slides, X-rays, EKGs, skin lesions, and video images of patients are transmitted over one-line, while simultaneous conversation between two or more persons at multiple sites is facilitated over another (Dunn et al., 1980; Dunn and Fisher, 1985; Sanders et al., 1995).

The Sioux Lookout region project in North Western Ontario was a leader in the use of slow-scan video as a continuing professional health education method (Dunn et al., 1980; Dunn and Fisher, 1985). The Sioux Lookout project used slow-scan video for facilitating formal and informal CME by using various presentation methods -- consultation, discussion/case presentation, and lecture. However, no systematic evaluation of learning effectiveness, either through achievement tests or follow-up analysis of behavior change, is reported in the literature.

Hampton et al. (1994) reported the findings of a study on the effectiveness of video conferencing for delivering CME to rural physicians. The study's results indicated a 21 percent increase in knowledge gain through a pre and posttest measure of participants' knowledge in a clinical CME program. In another study, Burleson and Sugimoto (1984) described the findings of a CME program delivered through video conferencing which indicated a majority of participants would recommend video conferencing as a method for participating in continuing education.

Whitten et al. (1988) also reported on a study they conducted to compare physician perceptions of the effectiveness of CME programs delivered by interactive television (ITV) with CME programs delivered in a traditional face-to-face setting. They used a post-CME satisfaction survey to measure participants' satisfaction with CME delivered by traditional means and by the interactive television delivery mode. The results of the study indicated that there was little difference between the perceptions of learners participating in either face-to-face or ITV. As well, ratings of the interactive television program were very high, suggesting participants were very satisfied with their learning experience through that technology. However, Whitten et al.'s study, like many of the others lacked a systematic approach to evaluating the effectiveness of the instructional technologies for enhancing learning or affecting the performance of physicians in their clinical setting. Whitten et al. (1988) did acknowledge this weakness and suggested that future evaluation efforts should focus on the examination of learning and retention among learners in ITV programming.

Satellite technology has also played a role in the delivery of CME to physicians in rural and remote regions. Chouinard's (1983) study of satellite technology, and its application in education and telemedicine, outlines the Anik A-1 and Hermes satellite projects of the 1970s in Canada. Launched in 1976, the Hermes satellite was, at the time, the world's most powerful communications satellite, providing telemedicine, tele education and direct broadcasting to Canada's northern, rural and remote areas. Several projects were piloted using the Hermes satellite, including numerous telehealth projects. One experiment provided an audio link between a remote nursing station in Kashechewan, Ontario and a base hospital in Moose Factory. A two-way audio and video link between the Moose Factory hospital and the University Hospital in London, Ontario also enabled the facilitation of continuing education and telehealth related communications and consultations (Chouinard, 1983). However, the experiences of the Ontario project and similar others in Newfoundland and British Columbia suggested that satellite technology was not always the most effective nor efficient means for providing distance learning and telehealth links. According to Chouinard (1983) satellite technology is often not an efficient means where other simpler and cost-effective modes of technology are available.

Computer applications in distance education are varied and may include computer-assisted instruction (CAI), computer-managed instruction (CMI), and computer-mediated learning (CML) (which refers to computer applications and interactive multimedia that facilitate the delivery of instruction). Several examples of the computer-mediated instructional technologies available for distance education include: electronic mail; synchronous and asynchronous computer mediated communication applications; distributed materials on the World Wide Web; and interactive multimedia applications on CD-ROMs (Compact Disk Read-Only Memory).

Personal computers offer rural and remote physicians many opportunities to utilize electronic databases and networks from their homes or offices without expensive travel to distant CME courses, and without loss of practice time. According to Manning and Petit (1987) computer and telecommunication technologies will open the door for new approaches to medical practice recertification methods, better access to clinical information, enhanced communication among physicians, and greater access to electronic medical databases. If Manning and Petits' predictions were correct, then this study is one of many which will see how electronic information handling and communication, and the exploitation of computer and telecommunication technology can be used effectively for addressing the continuing education needs of the rural physician.

The first experiments with CAI in medical education began in the 1960s at Ohio State University where computers were used as instructional tools for simulating patient encounters (Piemme, 1988). Since these early studies, CAI has grown and proliferated among medical schools and colleges. In the United States, the Association of American Medical Colleges has highly recommended the production and use of educational software in medical schools (Stocking and Benjamin, 1995). As well, many industry experts and developers have designed integrated CAI programming systems. These systems may be used as office and patient management automation tools, extracting necessary information to assist the physician in tracking performance patterns and formulating areas for future computer-based learning (Shortcliffe, 1983; Storey, 1983; McDonald, 1983).

Several studies have documented the development and use of CAI and clinical decision support systems as medical education tools (Scott, 1994; Locke and Rezza, 1996; Rosenblatt, 1984; Ganiats and Groveman, 1986). The results of these investigations suggest that CAI is used most effectively for simulating virtual clinical encounters, in which the computer actually replicates medical cases and encourages the user to choose among a series of diagnostic and therapeutic choices. Participants in these forms of CAI find the computer simulations challenging, instructive, and fun (Locke and

Rezza, 1996; Rosenblatt, 1984). In one study, Locke and Rezza (1996) reported the use of computers for enhancing clinical decision making and improving diagnosis by acting as a clinical decision support system (Locke and Rezza, 1996). The physician was required to manage a virtual patient utilizing current knowledge and using the guidelines as an educational adjunct to assist them throughout the process of diagnosis and treatment. The authors reported that physicians' reactions to the computer based training application were positive and that physicians were satisfied with the method, and the interactive multimedia (Locke and Rezza, 1996). However, the investigators' evaluation efforts were limited and apart from participant satisfaction data and the researcher's observations, there was little other information provided for making an informed decision on the effectiveness of the computer-based instruction.

In recent times the growth of the Internet and the World Wide Web have created new opportunities for providing distance education. Proponents of the Internet suggest that it will have far greater impact on global communications than any other previous communication technology development. Doyle (1996) predicts greater Internet developments in the future which will include continued improvements in speed of Internet access as ISDN (Integrated Services Digital Network) line developments increase the potential for downloading large files, such as real-time video. As well, new Web browsers have enabled the transmission of encrypted information to enhance security. This enhanced security has important implications for the storage and transmission of medical related patient information. Further, advocates of online CME are suggesting that in the future, CME workshops and courses will be delivered through the World Wide

Web (WWW) by accredited CME Web Service providers. Physicians will be able to pay for CME services using digital cash or credit cards, and submit on-line evaluations using CGI Web page forms.

The World Wide Web, still in its development stages, is undergoing continuous scrutiny and investigation. For educators, the full potential and use of the Internet and the World Wide Web is still uncertain. What is known is that the WWW can assist the physician in staying abreast of new clinical developments, communicating with colleagues, and accessing remote databases (Meissner and Vujicic, 1995). In the United States, CME is being delivered through the WWW by several university-based clinical divisions and CME offices.

At the University of Tennessee Medical Center, physicians can complete certifying examinations in fluoroscopy procedures placed on the WWW (Thompson et al., 1996). Physicians have expressed widespread satisfaction with the certification process, reporting ease of access and confidentiality. At Marshall University's School of Medicine, a CME Web-based course has been developed to improve physicians' clinical and history taking skills, and is accredited for one hour of CME credit (Hayes and Lehman, 1996). The system simulates an actual patient encounter, with the learner playing the part of examining physician and the program acting as patient. Pictures of the patient serve as image maps to which the user can point and click, and inspect more closely. Lab and radiologic studies can be requested and the learner can submit a diagnostic and treatment for evaluation and CME credit. The authors indicate that responses to this form of "virtual" CME have been very positive (Hayes and Lehman, 1996), however they do not provide any evaluative information on how effective the program has been in producing knowledge gain, enhancing behavior, or how future programs could be improved.

In another example, the University of Iowa College of Medicine has developed an online "Virtual Hospital" (Galvin et al., 1994). This WWW program includes multimedia teaching files, current diagnostic and therapeutic algorithms, patient simulations, historical information, patient instructional data, and on-line CME materials (Galvin et al., 1994). Response to the Virtual Hospital has been positive and new information is being added to the site every day. Similarly, an interactive educational multimedia program developed at the John Hopkins Medical Institute has been used to teach physicians about the role of computed tomography in detection and evaluation of splenic disease (Calhoun and Fishman, 1994). The program includes four sections: lectures with images, text and audio; a text section patterned on a journal article with an index categorized by pathologic process; a quiz with links to relevant text sections; and teaching files with selected patient histories and diagnoses (Calhoun and Fishman, 1994).

In several other studies, health sciences libraries and librarians have performed outreach functions using computer networks in attempts to address the information needs of rural physicians (Moore and Hartman, 1992; Pivalo, 1994; Jennet et al, 1990; Rankin, 1992; Dorsh and Lindwirth, 1993; Leist and Kristofco, 1990; Manning, 1990; Craig et al, 1992). In these studies, health sciences libraries have provided journal request and delivery services, electronic access to information, electronic communication devices for peer consultation, and continuing education opportunities. Several pilot "management information systems" programs have been reported in the literature, and interactive telecommunications technologies appear to be playing a larger role in bridging the information gaps that exist for rural physicians.

The Texas Tech MEDNET project (Moore and Hartman, 1992); the University of Illinois Library of Health Sciences medial information system project (Pivalo, 1994); the Faculty of Medicine, University of Calgary MIS pilot project (Jennett et al., 1990); and, the Georgia Interactive Network for Medical Information (GaIN) (Rankin, 1992) have reported initiatives to address the information needs of the rural and remote physician. The projects vary in their use and application of technology, yet encompass similar processes for improving access to quality medical information for rural physicians.

In a majority of the projects participant satisfaction surveys were used to evaluate the effectiveness of the information systems and the extent to which the projects addressed physicians' information needs. The findings of these surveys suggested a high level of satisfaction with the quality of the information and the services provided. In certain instances, although not intended as a measurement of effectiveness by the evaluators,

actual changes in clinical performance were reported on the satisfaction survey or through interviews as a result of information provided through the medical information systems (Moore and Hartman, 1992; Jennett et al., 1990; Money, 1986).

Several of the projects provided rural physicians with access to electronic networks and provided training on the use of online medical databases. The networks included 24 hour access to e-mail, MEDLINE databases, and online forum discussions.

However, usage of the systems was very disappointing (Pivalo, 1994; Rankin, 1992; Dorsch and Landwirth, 1993), and the authors found that literature searches were not conducted very often. Even with the provision of training, a majority of respondents to the evaluation surveys indicated that computer literacy was their main deterrent to use of the information systems (Pivalo, 1994; Rankin, 1992; Dorsch and Landwirth, 1993; Craig et al., 1992).

A major limitation of the studies reported in this literature review was the lack of systematic and comprehensive evaluation in assessing why programs succeeded or failed, how programs could be improved in the future, or how effective the technologies were in producing learning gain and improving behavior. Large amounts of time, effort and resources have been, and continue to be expended on the design and implementation of CME programs delivered at a distance. Most of the literature which was reviewed were descriptive case presentations of a specific application in a particular setting, or limited studies of the impact of CME delivered at a distance. Insufficient attempts were made by the evaluators to follow or report on systematic evaluation strategies for assessing the effectiveness of CME programming along a variety of evaluative criteria. Further, there was little evidence of the use of formative evaluation methods for improving instruction or instructional materials during the development stages.

With the exception of several audio teleconferencing and video conferencing studies the reporting of evaluation studies of CME programming delivered at a distance is not well documented in the literature. This gap in the literature is even more apparent for the use of the newer applications of the World Wide Web, asynchronous computer conferencing, and interactive multimedia. There is even less evidence in the literature that many of the reported distance education programs have been sufficiently evaluated using formative and summative methods.

Metaevaluation

Metaevaluation, the evaluation of evaluation, was first introduced in the evaluation literature by Scriven (1969) and Stufflebeam (1974). Scriven (1969) described the theoretical foundation of metaevaluation as the methodological assessment of the role of evaluation. Several years later, Stufflebeam (1974, p.68) elaborated on metaevaluation by summarizing it as "a procedure for describing an evaluation activity and judging it against a set of ideas concerning what constitutes good evaluation." Both Scriven and Stufflebeam were the earliest proponents of metaevaluation and since their early conceptions the idea of what metaevaluation is and how it is conducted has conjured different interpretations by various authors.

Baker et al. (1980, p. 6) described metaevaluation "as an attempt to discover how well an initial 'primary' evaluation was carried out by holding it up to the light of critical investigation." However, Baker et al.'s notion of metaevaluation, like many others, was of limited value because it did not denote a process nor methodology for actually conducting metaevaluation. According to Sheets (1983) the conceptual foundation and methodologies for conducting metaevaluation are very limited. Stevenson et al. (1979) also noted "there are as many potential conceptions of metaevaluation as there are of evaluation itself" (p.38).

As an example, Cook and Gruder (1978) limited their conception of metaevaluation to the evaluation of empirical summative evaluations or studies in which the data was collected directly from program participants within a systematic design framework. Gowin and Millman (1978) illustrated the process a little differently by describing metaevaluation as a process which encompassed three tiers. The first tier involved the educational product or program being evaluated. The second tier was the actual primary evaluation of the product or program and the reports and data generated as a result of the implementation of the evaluation. The third tier comprised the analysis of the items produced by the second tier activities and was considered metaevaluation.

> "Metaevaluation, the third tier, is supposed to produce added insights with respect to the evaluation documentation, the second tier." (Gowan and Millman, 1978, p.2)

Stufflebeam (1974) proposed two purposes for metaevaluation, formative and summative. At each level the metaevaluator was to assess the goals, designs, processes, and results of evaluation efforts. Formative metaevaluation was designed to guide the evaluator in designing and conducting the evaluation, thereby providing assistance with respect to evaluation decision-making. This level of metaevaluation usually resulted in recommendations, and the procedures recommended by Stufflebeam for conducting formative metaevaluation included ratings and rankings using Delphi techniques, standardized ratings of experimental designs as they already existed in the literature, logical analysis by the primary evaluator or metaevaluator, and the application of administrative checklists.

Summative metaevaluation on the other hand served accountability purposes by identifying the strengths and weaknesses of an evaluation. This level of metaevaluation was conducted retroactively to produce public judgments of the merits of the completed evaluation work. The outcome of summative metaevaluation was usually judgements on the worth or value of the evaluation and results. One of the procedures which Stufflebeam identified as being useful for this level of metaevaluation was summative case studies. This procedure usually resulted in an overall descriptive judgment on how well the evaluation design was implemented, its strengths and weaknesses, and what specific problems were encountered during its implementation.

In general, the literature on evaluation is limited despite previous claims of the same. The literature which does exist under the umbrella of metaevaluation is vague and the concept itself is misinterpreted by some writers. In many reports, most authors do not describe appropriate nor systematic procedures for the actual appearance of metaevaluation. And, in some instances other investigators appear to have confused the concept of metaevaluation with meta-analysis.

> "the state of the art of metaevaluation is limited in scope. Discussions of the logical structure of metaevaluation have been cryptic and have appeared in only a few fugitive papers... The writings on metaevaluation have lacked detail concerning the mechanics of metaevaluation... While some devices, such as technical standards for tests, exist the available tools for conducting metaevaluation work are neither extensive nor well-organized. Finally, there are virtually no published designs for conducting metaevaluation work. Overall, the state of the art of metaevaluation is primitive, and there is a need for both conceptual and technical development of the area." (Stufflebeam, 1974, p.4)

Further, reports on how to conduct a metaevaluation study and outcomes or results of such studies are inadequate. This would suggest a need for further development of the field.

> "The task of conducting a metaevaluation requires procedures or tools that can be used to perform the analyses. A fertile area for research in evaluation is to develop guides to doing competent metaevaluations." (Gowan and Millman, 1979, p.11)

Despite the limited work in this area, several authors have suggested guidelines and frameworks for metaevaluation studies. Gowan and Millman (1979) proposed that metaevaluators should use a "set of concerns" about evaluation as a guide for metaevaluation work. These concerns could be used to direct metaevaluation assessments of the critical features of an evaluation and evaluation documents. These concerns may encompass "standards for educational evaluation as the criteria against which to judge the original evaluation effort" (p.11).

Gowan and Millman also suggest that the metaevaluator can use a checklist and/or a set of questions that are designed to expose the logical structure of the original evaluation. This checklist can be used to rate the evaluation and to consider how soundly the evaluator judged the adequacy of the original "tier one" program or product according to the technical adequacy of the performance data and other indicators of program effect.

Cook & Gruder (1978, p.6) described the use of seven models for the summative metaevaluation of empirical data. The authors suggest that any one or combination of these variations can be conducted depending on the time of metaevaluation, status of the data, and the number of data sets involved. These seven models included:

- essay review of an evaluation report -- after-the-fact commentary on a single set of evaluation data that are not reanalyzed;
- empirical reevaluation -- manipulation of raw data about a program

to assess the validity of the conclusions from a primary evaluation;

- empirical reevaluation of multiple data sets;
- consultant metaevaluation -- attempts to judge and improve an evaluation while it is under way by means of feedback from one or more consultants who continuously monitor a primary evaluation;
- simultaneous secondary evaluation analysis of raw data -- analysis of the evaluation data by persons other than the primary evaluator;
- multiple independent replications -- commissioning of more than one evaluator to independently design and implement a study of a particular program's impact.

However, the authors' writings do suggest that their models were best suited for use with large-scale evaluations of program curriculum or other large-scale social programs.

Baker et al. (1980) used interviews with primary evaluation staff and consumers of evaluation reports as a metaevaluation method to evaluate evaluations of secondary school curriculum programs. The authors used the data collected from their interviews to assess the merit of the evaluations according to the standards for educational evaluation proposed by Stufflebeam: technical adequacy, utility, ethical, and practicality. Similarly, Sheets (1983) conducted interviews with program directors of a university department to assess an evaluation of a faculty development program. Sheets formulated and based his interview questions on the 1981 Standards for Evaluation of Educational Programs, Projects and Materials developed by the Joint Committee on Standards for Educational Evaluation.

Stufflebeam (1974), one of the earliest proponents of metaevaluation, believed that because metaevaluation was a form of evaluation its conceptualization had to be consistent with some of the more common premises of evaluation:

- Evaluation was the assessment of merit; thus, metaevaluation meant assessing the merit of evaluation efforts.

- Evaluation serves decision making and accountability; thus metaevaluation should provide information pro-actively to support the decisions that must be made in conducting evaluation work, and metaevaluation should provide retroactive information to help evaluators be accountable for their past evaluation work. Another way of saying this is that metaevaluation should be both formative and summative.

- Evaluations should assess goals, designs, implementation, and results. Thus, metaevaluation should assess the importance of evaluation objectives, the appropriateness of evaluation designs, the adequacy of implementation of the designs, and the quality and importance of evaluation results.

- Evaluation should serve all persons who are involved in and affected by the program being evaluated; hence, metaevaluation should serve evaluators and all persons who are interested in their work.

Stufflebeam (1981, p.151) proposed that metaevaluation was a process which involved "delineating, obtaining, and using descriptive and judgmental information about the practicality, ethics, and technical adequacy of an evaluation in order to guide the evaluation and publicly to report its strengths and weaknesses." This notion of metaevaluation was based on much of Stufflebeam's extensive experience and research on evaluation processes. This conceptualization was useful to the current study and was used as a guide for the development of the metaevaluation procedures of this dissertation. As well, Stufflebeam's (1978) definition of summative metaevaluation was also used to direct the actual metaevaluation methods and type implemented in this study:

> "...sums up the overall merit of an evaluation, and is usually done following the conclusion of a primary evaluation. It holds evaluators accountable by publicly reporting on the extent that their evaluation reports meet standards of good evaluation practice. Finally, summative

metaevaluations help the audiences of primary evaluations determine how seriously they should take the primary evaluation's reported conclusions and recommendations."

Stufflebeam's (1974) earliest work on metaevaluation saw the development of a list of standards for evaluating evaluations. If evaluation can be defined as the "assessment of merit," as Stufflebeam writes, then metaevaluation can be described as the "assessment of the merit of an evaluation." A metaevaluation evaluates the extent to which an evaluation is:

- Technically adequate in revealing the merit of a given object;
- Useful in guiding decisions;
- Ethical in dealing with people and organizations; and
- Practical in using resources.

In order for evaluation to be credible, a metaevaluation must judge the worth of an evaluation against these set of principles. Stufflebeam developed these four categories as broad areas of investigation containing a total of 34 standards that could be used for judging the worth of an evaluation. In 1981, the Joint Committee on Standards for Educational Evaluation (founded in 1975) built on Stufflebeam's original standards and published the renowned "Standards for Evaluations of Educational Programs, Projects, and Materials" (1st ed.). According to The Joint Committee (reported in Gould et al., 1995, p.4):

> "the Joint Committee foresaw several benefits from the development of sound standards: a common language to facilitate communication and collaboration in evaluation; a set of general rules for dealing with a variety of specific evaluation problems; a conceptual framework by which to study the often-confusing world of evaluation; a set of working definitions

to guide research and development on the evaluation process; a public statement of the state of the art in educational evaluation; a basis for self-regulation and accountability by professional evaluators; and an aid to developing public credibility for the educational evaluation field." (The Joint Committee, 1981)

In 1989, a decision was made by The Joint Committee to revise the 1981 standards in light of new developments in the field of evaluation. Following a series of Joint Committee Working Meetings, a rewriting process comprising a Panel of Writers, National and International Review Panels, Field Tests, and Public Hearings was conducted and the original standards were rewritten. In 1994, the new Program Evaluation Standards (2nd ed.) were published. The 1994 standards included thirty specific standards categorized into groups relating to four attributes necessary and sufficient for sound and fair program evaluation: utility, feasibility, propriety, and accuracy. The 1994 standards also saw the inclusion of a new standard:

> A12 Metaevaluation: The evaluation itself should be formatively and summatively evaluated against these and other pertinent standards, so that its conduct is appropriately guided and, on completion, stakeholders can closely examine its strengths and weaknesses.

According to The Joint Committee, the Standards were intended to:

> "provide a guide for evaluating educational and training programs, projects, and materials in a variety of settings. They are intended both for users of evaluations and for evaluators. People who commission or conduct evaluations, or who use evaluation results to improve education and training in schools, universities, medical and health care fields, the military, business and industry, the government, and law, will find the Standards useful. They have been developed for use by teachers, administrators, school board members, trainers, evaluators, curriculum specialists, legislators, personnel administrators, counselors, community leaders, business and educational associations, parents, and others. The Standards guide the design, employment, and assessment of evaluations of educational programs, projects, and materials" (p.1).

As well,

> "The Standards are not detailed technical standards, and they do not replace textbooks in technical areas such as qualitative and quantitative research design and analysis, measurement and data collection, data processing, and report writing" (p.2).

Stufflebeam's conception of the collection of descriptive and judgmental information according to predetermined criterion, and the 1994 Program Evaluation Standards of The Joint Committee served as the impetus for the design of the metaevaluation procedures used to evaluate the evaluation model fieldtested in this study.

In summary, the literature on metaevaluation is sparse and limited in terms of how to conduct metaevaluation in a systematic and methodological manner. A variety of authors have proposed different methods and instrumentation which can be used to evaluate evaluations. This study incorporated a metaevaluation design based on Stufflebeam's conceptualization of the purpose of summative metaevaluation, and Sheets (1984) and Baker et al. (1980) use of interviews of key stakeholders and decision-makers as a procedure for metaevaluation. The 1994 Standards of The Joint Committee served as the criterion against which the evaluation model and procedures fieldtested in the study were evaluated. Other methods which were used in the metaevaluation were also based on some of the literature just described. The actual design and methods of the metaevaluation followed in this study are described in Chapter Three and the results presented in Chapter Four.

Evaluation Models

An examination of the literature on evaluation indicates that there are numerous evaluation approaches, models and frameworks in existence. The early approaches to evaluation were largely influenced by behaviorally-oriented measurement perspectives. The emphasis of evaluation at that time was placed more on the behavioral outcomes of the teaching/learning process than on the actual process of how one arrived at that point. Practitioners followed an approach which was entrenched in the scientific paradigm. However, this line of thinking was soon challenged by the understanding that a positivist-oriented evaluation in itself only offered a shallow interpretation of the effectiveness of instructional programs. Many came to realize that the teaching/learning process was influenced by an array of concerns, claims, and issues which were not amenable to scientific measurement alone.

Educational evaluation has since evolved and newer approaches have tended to advocate a focus on the actors and stakeholders in educational programming, the process of teaching, and the intricacies of the instructional context. Kerr (1997) notes that "as evaluation has changed from an algorithmic to a more heuristic methodology, it has become situation-specific" (p. 18). The influence of the emerging naturalistic philosophies meant that evaluation approaches had to evolve to address the needs of a wide array of contextual concerns and circumstances. Therefore, evaluation approaches have constantly adapted and/or new ones been invented in order to meet the varying needs of programs and program audiences (Worthen and Sanders, 1987). This acclimation is evident in Thompson's (1987) observations when he noted that an eclectic approach was often required when evaluating distance learning materials because there were an array of different factors which had to be considered (factors not commonly found within other educational contexts). Melton (1992) also speaks of the evolutionary character of evaluation through the lens of a pragmatic philosophy. "Use whatever works to improve the decisions people make." According to Melton the purpose of evaluation is to provide a more rational basis for decision-making than would otherwise exist.

However, far from working against the prospective evaluator, the evolution of evaluation approaches and the array of evaluation models and methods from which an evaluator may choose can be used to a practitioner's advantage. Steele (1973) reports that experienced evaluators rarely follow a specific evaluation model. Rather, they are more likely to modify a model or models to suit a particular situation.

> "In many situations, rather than extensively adapting a particular approach, you might be better off to construct your own, borrowing the parts of other approaches that are most useful and building patterns and processes that are appropriate to your needs" (p. 55).

Thorpe (1993) also suggests that there is no consensus around a best or right model of evaluation. She points out that there are a variety of approaches and methods to choose from and evaluators can choose one or a combination of approaches to meet the purposes and resources available to them at the time. The procedures of an evaluation should be designed to address the local purposes for which it is being set up, which will obviously change according to the program being evaluated. Thorpe's remarks reiterate those of Suchman (1967) who emphasized that evaluators need to assess programs in relation to their practical settings and to use whatever research techniques are available and appropriate to the circumstances and needs of a particular evaluative study.

These authors' comments reflect the thinking which influenced the evaluator to examine a variety of evaluation models and then select those aspects of a model or models that best suited the evaluation of computer-mediated CME courseware delivered at a distance. Following an eclectic approach, several models were examined to determine which had components suitable for the purpose of designing an evaluation model to be used with computer-mediated CME courseware. As well, a number of other evaluation models, approaches, and frameworks were also reviewed and are discussed, with emphasis on their strengths and weaknesses.

House (1978) and Worthen and Sanders (1987) have used taxonomies to classify and compare several of the more significant evaluation approaches. And, while the authors' two taxonomies differ in the depth with which they compare the various approaches, their underlying purpose is the same -- to compare the approaches on principal components, audiences served, purposes and major outcomes, characteristics and methodologies, and the typical questions they seek to answer. The two taxonomies are modified, combined, and presented in Table 2.1. This combined taxonomy was developed because it provides a clear overview of the more prominent models discussed in the literature. Each approach presented in the taxonomy will be briefly discussed with an overview of its main advantages and disadvantages.

Table 2.1 Taxonomy of the six major approaches to evaluation (adapted and modified from House (1978) and Worthen and Sanders, 1987))

CATEGORY	Objectives-Oriented	Management-Oriented	Consumer-Oriented	Expertise-Oriented	Adversary-Oriented	Naturalistic & Participant-Oriented
PROPONENT MODEL	Tyler's Model	CIPP Model	Scriven Model	Connoisseurship Model	Judicial (Quasi-Legal) Model	Stake's Responsive Model
MAJOR AUDIENCES	Managers	Decision-makers	Consumers	Connoisseurs Consumers	Jury	Stakeholders Practitioners
OUTCOMES	Determine the extent to which objectives are achieved.	Provide useful information to aid in making decisions.	Provide information about educational products to aid decisions about purchases or adoptions.	Provide professional judgements of quality.	Provide a balanced examination of all sides of controversial issues or highlighting both strengths and weaknesses of a program.	Understand and portray the complexities of an educational activity, responding to an audience's requirements for information.
MAJOR CHARACTER-ISTICS	Specify measurable objectives, use objective instruments to gather data, search for discrepancies between objectives and performance	Provide rational decision-making, evaluate all stages of program development.	Use criteria checklists to analyze products, product testing, informing consumers.	Base judgements of individual knowledge and experience, use of consensus standards, team site visitations.	Use of public hearings, use of opposing points of view, decision based on arguments heard during proceedings.	Reflect multiple realities, use of inductive reasoning and discovery, first hand experience on site.
TYPICAL QUESTIONS	Are the students achieving the objectives? Is the teacher producing?	Is the program effective? What parts are effective?	What are all the effects?	Would a critic approve of the program?	What are the arguments for and against the program?	What does the program look like to different people?

Objectives-Oriented Approach

The main proponent of the objectives-oriented approach was Ralph Tyler and its main purpose is to determine whether a program's objectives have been achieved, and to draw conclusions on how successful a program has been based on these findings. According to Stufflebeam and Shinkfield (1985) an objectives-oriented evaluation assesses this success by measuring the discrepancy between stated program objectives and program outcomes. The methods used in objectives-based studies involve the collection and analysis of performance data relative to specified objectives. Objectives are specified as the actual behavior(s) that a learner is expected to competently demonstrate as a result of instruction. If a desired behavior can be stated or described, it is then amenable to measurement with some sort of achievement test. The approach is largely influenced by B.F. Skinner's behaviorism and has had a significant impact on the development of past and present-day instructional systems design theory and practices.

The objectives-oriented approach is undoubtedly the most prevalent evaluation style in educational program assessment. It has good common sense appeal, as educators have a great deal of experience with the use of behavioral objectives and standardized testing (Stufflebeam and Shinkfield, 1985). Worthen and Sanders (1987) note that this is one of the greatest strengths of the objectives-based evaluation model; "it is easily understood, easy to follow and implement, and produces information that educators generally agree is relevant to their mission" (p. 72). This is also advantageous to the evaluator and the instructor, because it simplifies the development of achievement tests and instruments for assessing the extent to which program objectives have been achieved.

However, the Tylerian approach is widely criticized as an evaluation model because it portrays evaluation as a terminal event, allowing final judgements only (Madaus et al. 1983). It presents information only after an educational program has been delivered. Everything tends to be viewed from the point of view of the students and their change in behavior (which if positive is viewed as achievement and program success). Therefore, the objectives-oriented model for evaluation offers little opportunity for taking a more flexible open-ended look at other aspects of program effectiveness. According to Worthen and Sanders (1987) objectives-oriented evaluation:

> "(1) lacks a real evaluative component (facilitating measurement and assessment of objectives rather than resulting in explicit judgements of merit or worth), (2) lacks standards to judge the importance of observed discrepancies between objectives and performance levels, (3) neglects the value of objectives themselves, (4) ignores important alternatives that should be considered in planning an educational program, (5) neglects transactions that occur within the program or activity being evaluated, (6) neglects the context in which the evaluation takes place, (7) ignores important outcomes other than those covered by the objectives the unintended outcomes of the activity), (8) omits evidence of program value not reflected in its own objectives, and (9) promotes a linear, inflexible approach to evaluation (Worthen and Sanders, 1987, p. 73).

Tyler's approach to evaluation is clearly based on behavioral objectives and the assessment of whether they have been achieved by the learners. Although this is a desirable measure of program impact, additional information is required to indicate whether learners in the computer-mediated courseware are satisfied with the content of the program, experience difficulties in using the instructional software, and are likely to change their clinical practices as a result of the instruction they receive. Tyler's approach is central to the task of evaluating the impact of the courseware program, but his approach is not comprehensive enough because it fails to consider other important factors related to the effectiveness of computer-mediated instruction.

Management-Oriented Approach

The management-oriented approach is designed to enable managers and administrators to make critical decisions about how a program can be improved, or whether and to what extent a program should be continued, or both (Worthen and Sanders, 1987). The most widely known and used application of the management-oriented approach is Daniel Stufflebeam's CIPP evaluation (Context, Input, Process, Product).

Stufflebeam's (1971) CIPP approach to educational evaluation is based on the view that the most important purpose of evaluation is not to prove but to improve (Stufflebeam and Shinkfield, 1985). In Stufflebeam's opinion evaluations should foster improvement, provide accountability, and promote increased understanding of the events under analysis. Evaluators are encouraged to use the process of evaluation and evaluation methodologies as tools for helping make programs work better for the people they are intended to serve.

The CIPP evaluation model is oriented toward a systems view of education which focuses on providing ongoing evaluation services to educational decision-makers, rather than formulating or establishing the path of an individual study (Stufflebeam and Shinkfield, 1985). Essentially, the use of the CIPP model is intended to promote growth and to help educators systematically gather and use feedback to identify and address needs, or at least, do the best they can with available resources. As a result, a thorough educational evaluation will consider all of the following:

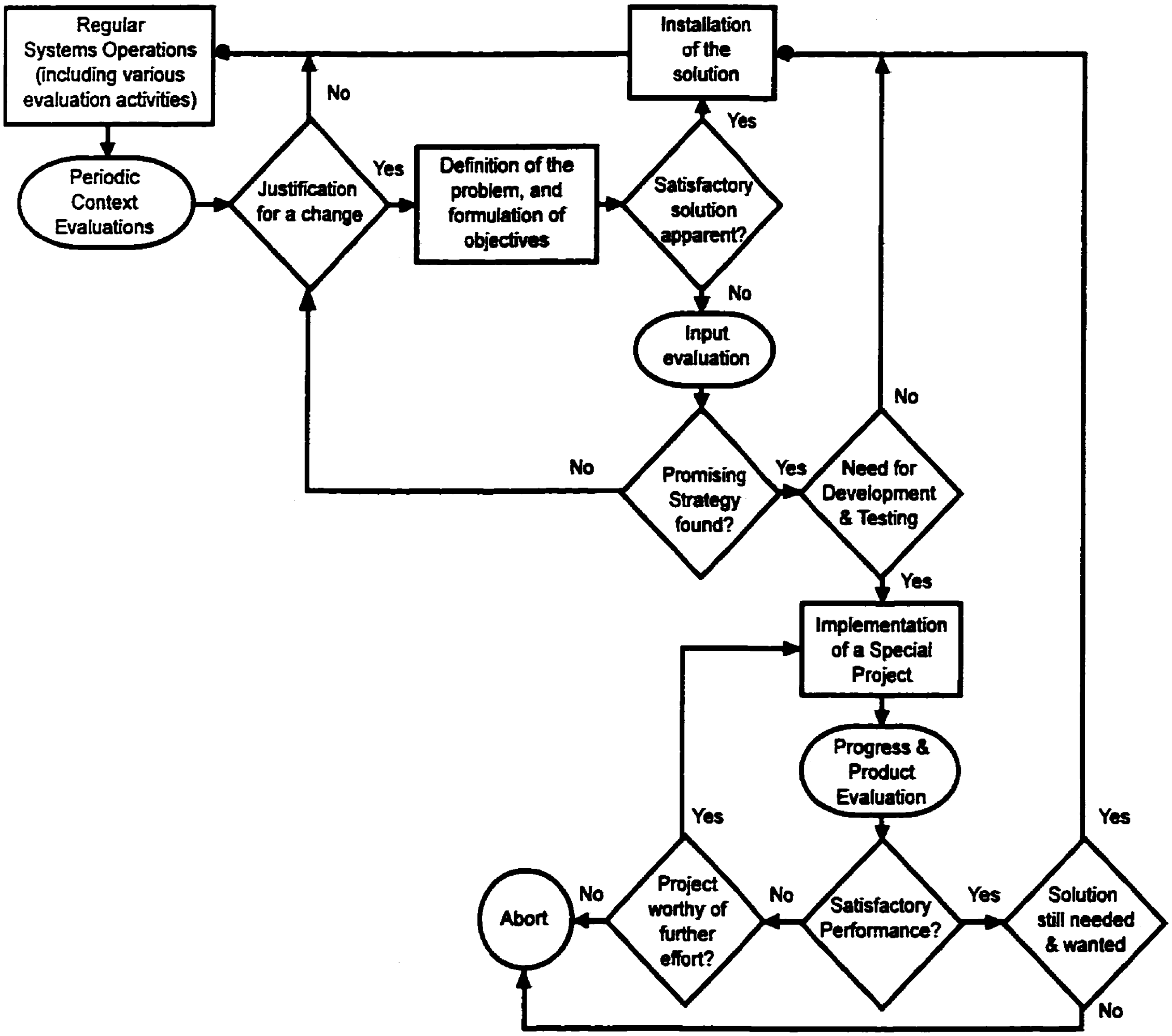
- the *context* in which a product or an object will be used;
- information or *input* about the changes necessary to improve products;
- the *processes* involved in creating and using products;
- and the value of the final *product* in comparison with the needs detected in the context assessment stage of an evaluation.

The main value of this theory is that it provides a general framework for isolating relevant components in an evaluation (context, input, process, and products) (Stufflebeam and Shinkfield, 1985). Figure 2.1 presents Stufflebeam's flowchart model for the facilitation of a management-oriented evaluation strategy.

The CIPP evaluation has several advantages. The CIPP approach includes an assessment of the context in which a program need arises, the input of stakeholders within an organization, and it also evaluates the way in which current programs operate. Therefore, it has an improvement-oriented focus and is very useful for informing the concerns of practitioners. The CIPP approach is also very comprehensive and has been used extensively in the evaluation of education programs (Madaus et al., 1983). As well, the model may lead to findings that include information related to unintended program outcomes.

However, the CIPP approach is usually directed by the information needs of decision-makers, therefore the questions that are considered relevant reflect the interests and needs of a narrow group. This can be seen as a disadvantage because the evaluation

Figure 2.1 Stufflebeam's Flowchart for the Implementation of CIPP Evaluation Strategies



Source: Stufflebeam & Shinkfield (1985) may be distorted from the start, favoring the interests of management and ignoring the value perspectives of other stakeholder groups. The approach can also be costly and time consuming to implement and Guba and Lincoln (1985) imply that the model makes assumptions about the rationality of decision-makers and the sincerity of the decision-making process.

The components of the model related to process and product evaluation were particularly applicable to an examination of computer-mediated instructional effectiveness and impact since these components focused on program activities and outcomes. The instructional process of computer-mediated instruction is highly structured and dependant on the design and development work which occurs before a product is delivered. Therefore, an understanding of how well the process of instruction functions in the actual context of learning is important. As well, Stufflebeam's systems view of evaluation included the concept that information should be provided to decision-makers. This was an important element of this study as the CME Planning Committee had an important stake and interest in the success of the program. However, the context and input components were not of similar value since they were more concerned with the planning of evaluation, thus negating the use of the model in its entirety.

Consumer-Oriented Approach

Scriven's (1967) consumer-oriented evaluation approach, also known as the "goal-free evaluation model," was developed as an alternative to the goals-based models that dominated during the 1960s and 70s. The foundation of the goal free model is that evaluators purposely ignore the goals or objectives intended by program developers, and attempt to uncover all program effects irrespective of its intentions. However, according to Madaus et al. (1983) the evaluation approach is more of a philosophy for educational evaluation than a model which prescribes a framework for the use of particular methods or instruments.

One of the advantages of the consumer-oriented approach is that it can produce findings about unintended effects. It is also advantageous in that it attempts to control the level of evaluator bias by ignoring the intentions of program developers. According to Scriven (1987) the model may be particularly useful for external evaluators with a need to control biases which will enable them to judge a program's value without being constrained by the established objectives of a program and the expectations of managers and administrators. However, the usefulness for external evaluators is a disadvantage for internal evaluators who are usually unable to avoid the influence of intended outcomes.

Kettle (1994) notes that some evaluation idealists believe that it is important for evaluators to discover the underlying value perspectives among program audiences and any hidden agendas which influence program processes. However, these are sources of information which consumer-oriented evaluators must avoid because they influence the evaluation of the intended program goals. Another criticism of the model is that it has serious methodological problems and offers no mechanism for measuring the validity of an evaluator's judgement (Worthen and Sanders, 1987). According to Worthen and Sanders it is not well suited for less experienced and novice evaluators. This approach was not appropriate for this study because the concerns of all stakeholders were important, including the program developers. As well, one of the purposes for developing the evaluation in this study was to ensure that it was based on a practitioner-as-evaluator rationale, something that the consumer-oriented approach is not best suited.

Expertise-Oriented Approach

The expertise-oriented approach, alternatively known as the connoisseurship model (Eisner, 1975) is based on a tradition of criticism which is inherent to the field of art and literature (Burnham, 1995). Art and literature criticism are based on the insights of individuals who have expertise in a given area and because of this there are no hard and fast rules to adhere to, and no specific step-by-step process to follow. The evaluation process is centered in the evaluator. The evaluator is an expert who is in effect the instrument or measuring device.

The rationale behind the approach is that educational program appreciation, like art and literature appreciation, is considered to be a complex thing. Therefore, an expert with broad and extensive expertise in the area being evaluated is needed to make judgements about the merit and worth of a program (Burnham, 1995). The education critic does not provide the methods used to arrive at the conclusions nor does he have to provide reasons for the judgements which are offered. The evaluator as "critic" is respected because of his or her reputation.

According to Kettle (1994) one of the major advantages of the art criticism model is that "it exploits an evaluator's superior, esoteric level of understanding about a specific area for the benefit of a lay audience" (p. 78). This may be useful for programs where the expert is highly respected and the stakeholders hold confidence in that person's expertise. However, this superiority may be disadvantageous because it relies on the sole expertise of one person, which places the value perspectives of the evaluator in a questionable position. Another disadvantage is related to the availability of an expert in the given program area to act as the sole critic of the object being evaluated. The focus on expertise also negates the value perspectives and opinions of a range of stakeholders who may also be able to describe the perceived strengths and weaknesses of a program, and because of these factors was not an appropriate approach for this study. As well, the esoteric nature of this approach was not well suited to the study.

Adversary-Oriented Approach

The adversary model, also known as the quasi-legal model, is based on summative decisions, primarily whether the program being examined has negative or positive outcomes. It was developed as an evaluation approach by T.R. Owens and R.L. Wolf and is modeled after the judiciary system in that two opposing sides are organized; an advocacy team to substantiate and argue for the positive results and an adversary team to do the same for the negative results (Kettle, 1994). Each team collects evaluative information to support their position and then presents these findings to a jury for judgement. The objective for an adversarial evaluation is to make an informed judgement about a program by analyzing the rigorously researched pros and cons.

In comparison to other models discussed thus far, the adversarial model is the most democratic (Worthen and Sanders, 1987). The model is described as possessing both participatory and transaction qualities (Kettle, 1994). Transaction evaluation models involve people through negotiation, while participatory models attempt to engage the participation of people who have direct involvement in the program. In the adversary model, this is achieved by allowing program participants representation at a mock-trial (Stufflebeam and Shinkfield, 1985). In many ways it takes a qualitative approach to investigation and is subjective in terms of the research methodology the advocacy and adversary teams may use.

According to Worthen and Sanders (1987) one of the model's greatest strengths is that it provides decision makers with information which is of a comprehensive nature. This occurs as a result of its process, which involves at least two separate individuals or groups of people researching and building a substantive argument for their opposing positions. Another advantage is that the adversary-oriented approach is effective at uncovering any assumptions that may exist about the various functions of a program (Kettle, 1994). The opportunity to hear a position based on substantiated evidence and to question either side enables the jury to make an informed judgment of the merit and worth of a program.

The model can be disadvantageous because it is contingent on the competencies with which both sides collect their evidence, articulate their position, and question the opposition. The model is also an expensive and time-consuming approach and often places programs on a continuum of good or bad, positive or negative, reject or accept. Such a dichotomous choice may not always apply and in this study certainly did not. This study required an approach to evaluation which was improvement-oriented and enabled decision-makers to see the strengths and weaknesses of the program, while also allowing program developers to identify areas where improvements could be made and the program enhanced. The fundamental nature of the adversary-oriented approach was also beyond the resources of the study.

Naturalistic & Participant-Oriented Approach

Participant-oriented approaches, also known as the transaction models, are described as the most democratic of all the major evaluation models (Worthen and Sanders, 1987). They reflect a naturalistic epistemology and advocate qualitative-inquiry approaches. The key proponent of a participant-oriented approach to evaluation has been Robert Stake. Robert Stake's (1967) responsive evaluation emphasizes the needs of the client and proposes that evaluations respond to audience requirements. According to Stake's model, evaluation issues and concerns are advanced organizers, and are compiled and aggregated from discussions with learners, stakeholders, program sponsors, and program staff (Madaus et al., 1983). Stake (1977) says:

> "An evaluation is responsive evaluation (1) if it orients more directly to program activities than to program intents, (2) if it responds to audience requirements for information and (3) if the different value-perspectives present are referred to in reporting the success and failure of the program" (p.163).

One of the key features of responsive evaluation is its sensitivity to the pluralistic views of stakeholders (Kettle, 1994). This pluralistic view implies that certain value systems may conflict with one another during the issues identification stage. This can be a constructive element for sponsors interested in understanding a program's merit, worth or value from the perspectives of multiple realities. This focus on pluralistic values also distinguishes responsive evaluation from previous evaluation models which tended to take a singular value perspective.

The responsive approach may also be valued for its flexibility. Evaluators are not committed to pre-defined evaluation methodologies or plans. Therefore, it does not confine evaluators to "preordinate" evaluation models based on the positivist inclination to use behaviorally-oriented objectives and achievement tests. Responsive evaluation is also useful for evaluating programs in which needs have not been clearly identified. According to Kettle (1994):

> "...because the responsive evaluation approach considers the client's needs as no more or less important than the needs of other program audiences, and because the approach is designed to allow the needs of all program audiences to emerge during the evaluation process and still provide useful results, it is beneficial for use in evaluating programs for which an evaluation is deemed necessary, but which lacks defined standards and criteria for making an evaluative judgment" (p. 95).

According to Neuman (1989) the naturalistic paradigm offers promise for producing the detailed, context-bound information necessary to understand the effectiveness of computer-based instruction (CBI). One of the major potential strengths of CBI is that it provides a high degree of individualization which fits perfectly with the naturalistic assumption of individual constructions of reality. Naturalistic research which focuses on the individualized nature of learner's interactions with courseware is more likely to yield information that can be translated into principles for improving courseware design. Table 2.2 summarizes the various naturalistic techniques which Neuman says can be used in an evaluation of CBI.

Table 2.2 Techniques of Naturalistic Inquiry for the Study of Computer-Based Instruction (Neuman, 1989)

Techniques	Source(s)	Results
Prolonged and persistent observation	Teachers' and learners' interactions with courseware, including introductory and follow-up activities	Insights into: presentation of learning task, response strategies, effective and ineffective feedback, relative power of reinforcement techniques
Informal interviewing	All participants (individual and group)	Verification of insights, elements of effectiveness, strengths and weaknesses, suggestions for improvements
Document analysis	Courseware displays, documentation, teachers' records, students' products and records	Verification of insights, contextual information

Nevertheless, several weaknesses are inherent in the responsive approach. Stufflebeam and Shinkfield (1985) argue that it is less precise in measurement than preordinate and less qualitative approaches. Another weakness of the model is related to its responsiveness to the needs of various stakeholder audiences. It is possible that the needs of some stakeholder audiences may outweigh those of others because they are more able to explain their value perspectives. The methodology of the responsive approach is also considered heuristic in nature, which means that evaluators do not have a procedural step-by-step guideline to follow (Kettle, 1994). This may be problematic for novice evaluators who can be inundated with conflicting value perspectives.

Several other models were also reviewed because they entailed components or strategies which were applicable to the development of an eclectic model for evaluating the effectiveness of computer-mediated CME courseware. The first models which are presented can be grouped under a common category of "experimental" approaches to evaluation. These evaluation approaches depend upon the assignment of subjects to control or experimental treatment groups, achievement and performance tests, comparisons, and concern with the control of internal and external validity.

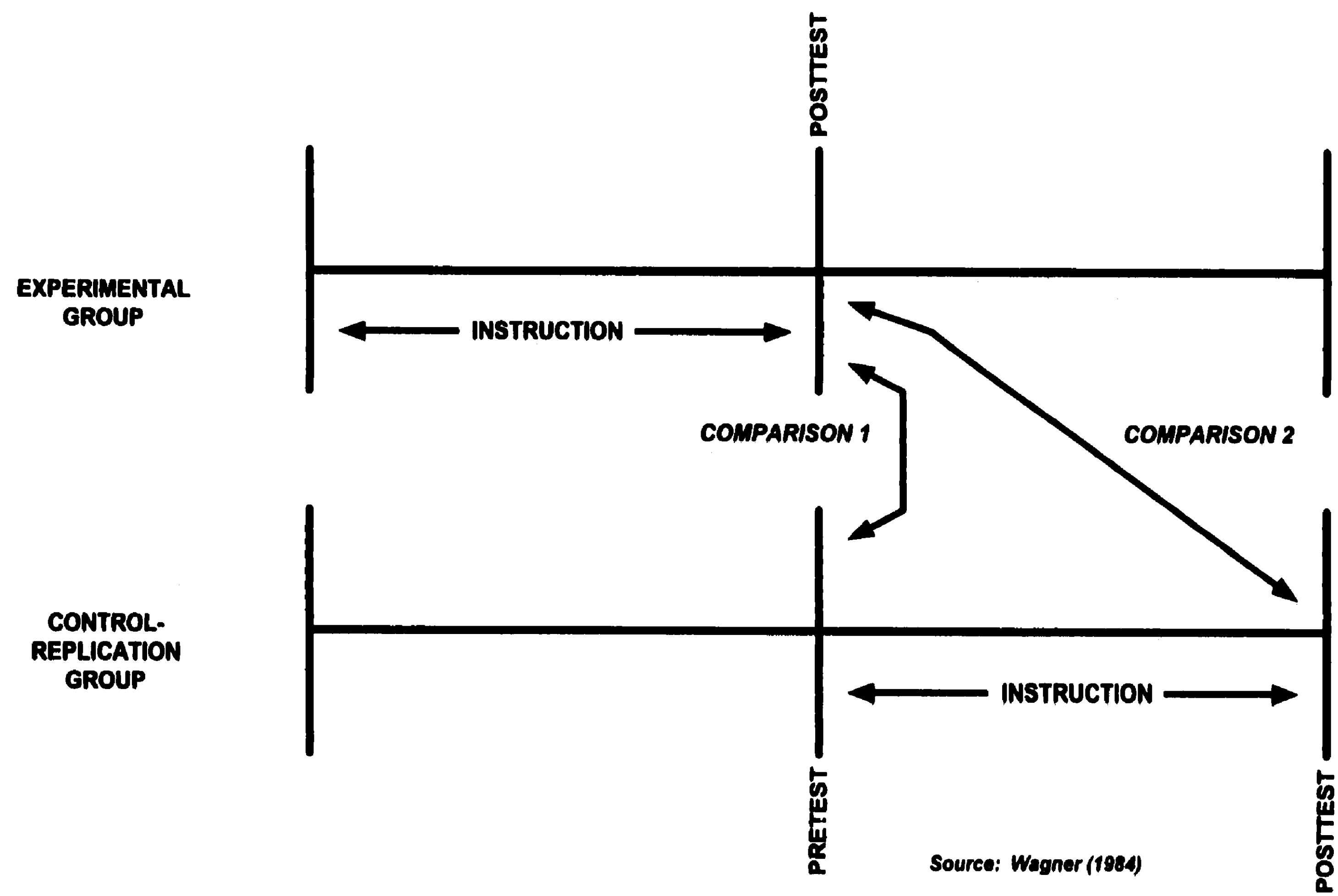
According to Babbie (1992) in order to conduct evaluation research, investigators must be able to operationalize, observe, and recognize the presence or absence of the outcomes that are to be produced. A key variable in evaluation research is the outcome or response to some form of treatment. If a training or educational program is intended to result in some specific outcome, the researcher must be able to measure that outcome.

Babbie (1992) has suggested several recommendations for the implementation of an effective evaluation research design. In evaluation research it is often appropriate and important to measure aspects of the context within which the a program is conducted. These variables are external to the program itself, yet they may affect it. It is also necessary to measure the affect of the educational intervention itself, the treatment. In part, this assessment may be conducted through the assignment of a sample of subjects to experimental and control groups.

Wagner (1984) proposes the use of the Staged Innovation Design as a means for ascertaining the instructional effectiveness of an educational program. This approach makes use of an experimental and a control-replication group and allows the posttest for an experimental group to be compared with both the pretest and posttest scores of a control-replication. This design controls for several threats to internal validity and is useful for situations in which all subjects must be exposed to a treatment, but it is not necessary that all subjects receive the treatment simultaneously. Figure 2.2 provides a schematic overview of the methodology. In the first stage, the experimental group is provided with instruction and posttested while the control-replication group is pretested. The results of the experimental group posttest are compared to results of the pretest administered to the control group in order to ascertain the degree of effectiveness of the instruction. In the second stage, the control-replication group is exposed to the treatment. Their posttest results are compared to the posttest results of the experimental group, thus replicating the study.

The main advantage of this model is that it enables the evaluator to use a second group of learners who are completing the same program of study as a control-replication group. The achievement scores of this group are then used as a comparison with the scores for an experimental group. This is advantageous for the evaluator because of the convenience of using a comparable group of learners as a control group, which in many instances are readily available. However, the main disadvantage of the study design is that it does not control for the influence of history, because the two groups are expected to complete the course at two different times. As well, there are no methods for assessing the learner's satisfaction with the instructional program or for measuring the impact of the training on the learner's performance. In this respect, the model is very dependent on quantitative measurement (achievement outcomes) and does not offer the evaluator any flexibility in terms of utilizing qualitative methods for examining the instructional process.

Figure 2.2 The Staged Innovation Design



Muller (1995) suggests that the use of experimental and quasi-experimental research designs are the best methods for collecting empirical evidence of student performance in educational software evaluation. The chief advantages of these designs is that they control for sources of invalidity (Muller, 1995).

> "If we can study the effects of a program under controlled conditions, we theoretically can conclude with some certainty that any effect is a real one." (Muller, 1995, p.27)

The research designs which Muller proposes include the pretest-posttest control method in which learners are randomly assigned to one of two groups; an experimental group receiving treatment, or a control group not receiving treatment. The single-group pretest posttest design is often employed when randomization of subjects is difficult to obtain. The nonequivalent control group design which is identical to the pretest posttest control group design except the subjects are not randomly assigned to groups. As well, the time series design which involves the measurement of a group of learners at periodic intervals. Each of the research designs has advantages and disadvantages, and some are more rigorous designs than others. However, the decision to select and implement one design over another is entirely dependant on the instructional software under investigation, and the conditions under which the evaluation is occurring.

Muller (1995) was also wise in recommending that the information obtained by the implementation of these designs not be used alone in making evaluative conclusions on the effectiveness of instructional courseware. Rather, the information should be used as a complement to a range of evaluative data collected on a particular software product. According to Muller (1995) there are four broad areas of concern when evaluating an educational software program. These include:

(1) program content: the suitability of materials for the learners and the objectives, and the accuracy of the content;

(2) pedagogy: the nature of the program's feedback, the types of learning modes used;

(3) program operation: the control that users have when using the program, the program's quality and the quality of the documentation;

(4) student outcome: the degree to which learners learn what the program intends to teach (Muller, 1995).

Muller's research designs are used for collecting information in the fourth category of this evaluative data list, student outcome data. Although experimental rigor, internal validity, and control are important factors in designing a study to measure and compare student outcomes, the purpose of evaluation in this study was to provide information and feedback to an array of stakeholders in the program.

According to Dillon and Gunawardena (1992) the outcome determination phase in classic experimental models often fails to answer the questions important in the study of distance education. They suggest that post hoc assessments, which measure differences between control and experimental groups after the administration of a treatment, may determine whether the outcomes have been achieved, but provide little information regarding the role of the program in achieving the objectives. As well, they believe that the random selection required by classical experimental design is often impossible in distance education settings due to the necessity of using intact groups.

Student outcome data alone would not have fulfilled the information needs of the developers of the courseware, the subject matter experts, or the decision-makers.

Therefore, the use of experimental or quasi-experimental designs alone was not appropriate. Nevertheless, Muller's proposal that other areas of instructional courseware effectiveness be evaluated supports the notion of an eclectic approach to evaluation.

Another evaluation approach which has gained attention in recent years is Guba and Lincoln's (1989) Fourth Generation evaluation which is similar in many ways to Stake's Responsive Evaluation approach. The Fourth Generation model is a form of evaluation in which the claims, concerns and issues (CCI) of stakeholders serve as "organizational foci" for determining what evaluative information is needed to understand the strengths and weaknesses of a program. The evaluation model is implemented within the methodological doctrine of the constructivist inquiry paradigm and is an emergent approach to evaluation because it seeks to include the countless contextual elements that underlie all human endeavor.

The authors have identified the approach as fourth generation because they see it as "moving beyond its predecessors, preexisting generations of evaluation models and theories, characterized as measurement-oriented, description-oriented, and judgement-oriented, to a new level whose key dynamic is negotiation" (p. 38). Fourth Generation evaluation asserts that measured outcomes are not true descriptions of the way things are. Rather, outcomes are the "meaningful social constructions" that individual or groups of actors construct to interpret the contexts in which they are located (Guba and Lincoln, 1989). These outcomes are uncovered, deciphered and negotiated through an interactive process that involves the evaluator and all program stakeholders.

Guba and Lincoln suggest that the emergent form of evaluation inherent in the

Fourth Generation model serves to empower marginalized groups in ways that other models cannot. They argue that previous evaluation models operate to enfranchise or disenfranchise stakeholding groups by way of: "the selection process of who will be involved and consulted; the questions to be asked; the methods to be employed; and the interpretation of the findings" (p. 9).

Advocates of naturalism, interpretism and constructivism argue that the main advantage of Fourth Generation evaluation is its departure from the methodology proposed by previous evaluation models (which Guba and Lincoln have referred to as Generation One, Two and Three). These earlier "generations" are built upon scientifically grounded ontological beliefs in positivism that an objective reality exists which is driven by irrefutable natural laws. They are also driven by an epistemological assumption of the existence of "a duality between observer and observed" which enables the evaluator to objectively measure and interpret the arena of the observed while neither influencing nor being influenced by it (Guba and Lincoln, 1989).

Guba and Lincoln (1989) describe the methodological guidelines of Fourth Generation as follows:

"1. Identifying the full array of stakeholders who are at risk in the projected evaluation.

2. Eliciting from each stakeholder group their constructions about the evaluand (the program being evaluated) and the range of claims, concerns and issues they wish to raise in relation to it.

3. Providing a context and a methodology through which different constructions, and different claims, concerns and issues can be understood, critiqued and taken into account.

4. Generating consensus with respect to as many constructions and their related claims, concerns and issues as possible.

5. Preparing an agenda for negotiation on items about which there is no, or incomplete, consensus.

6. Collecting and providing the information called for in the agenda for negotiation.

7. Establishing and mediating a forum of stakeholder representatives in which negotiation can take place.

8. Developing a report that communicates to each stakeholder group any consensus on constructions and any resolutions regarding the claims, concerns and issues that they have raised (as well as those raised by other groups that appear relevant to that group).

9. Recycling the evaluation once agin to take up still unresolved constructions and their attendant claims, concerns and issues" (p.73 - 74).

There are several disadvantages inherent to Fourth Generation evaluation. First, an adherence to a responsive constructivism evaluation implies that the evaluator must give up control over the process of evaluation (Guba and Lincoln, 1989) because stakeholders are seen as playing equal roles with the evaluator and the client throughout the process. This "loss of control" may have serious methodological consequences if the individuals who do not posses expertise in methodological issues become decision makers (Guba and Lincoln, 1989).

The model is also time consuming, extravagant, abstract, and ambiguous. A considerable amount of the evaluator's time can be spent identifying the stakeholder audience and then interacting with them to adequately interpret their claims, concerns, and issues. The process can become further complicated and onerous when the amount and types of claims, concerns, and issues brought forth by this interaction and negotiation extends beyond those contemplated by the evaluator. In this manner, the model is unfocused because there are no grounds for planning, preparing or forecasting time and schedules for its completion. According to Guba and Lincoln (1989) this aspect of the evaluation model "....may seem to render fourth generation evaluation impractical" (p.55). And in many instances and contexts it would be.

The evaluation process becomes more complex if consensus cannot be reached by the stakeholders on the claims, concerns and issues they have brought forth for negotiation. This stage of consensus-building may serve to jeopardize the program evaluation process by accentuating stakeholder animosity, uncovering issues unrelated to the evaluation task at hand, and entangling the evaluator as a facilitator of good-will. One question raised by this is: When does the evaluator actually engage in the collection of meaningful information to improve the program or service?

Although they did not describe models of their own, several authors' views of evaluation and evaluation models were of interest and value. Thorpe (1993), in particular, does not propose so much an approach to evaluation, but rather a vision for the role of the practitioner in evaluation, a vision of practitioner-as-evaluator. In this approach those with primary responsibility for the program should also take responsibility for its evaluation, development and improvement. Evaluation should become something which is applied to practice and used to make a difference to that practice by leading to better ways of doing things. If evaluation is to be useful in practice then it needs the involvement of the practitioner, and must be applied to practice if it is to serve the interests of learners and practitioners.

The main aspects of Thorpe's Practitioner-as-Evaluator Vision:

- Evaluation should operate within the constraints of the open learning unit concerned;
- Be robust enough so that it does not break down under pressure from peaks in work load;
- Flexible and capable of being adapted to different courses;
- User friendly so that it makes realistic demands of staff and resources;
- Be a form of in-service staff development which makes practitioners more aware and better informed;
- Lead to procedures which improve practice, even before the findings are applied.

Another useful evaluation model was proposed by Bland et al. (1984). These authors advocated a user-centered approach for increasing the results from the internal evaluation of continuing medical education activities. They describe internal evaluation "as an information-collection system, internally managed and designed to collect information about program activities and outcomes so that interested persons can use the information" (p. 54). The overriding emphasis of the user-centered approach to evaluation is the utility of the resultant data. Three features characterize the evaluation approach: an ordered set of steps with usefulness as the primary concern at each step; delineation of evaluator and decision-maker roles; and attention to the general communication aspects of evaluation (Bland et al., 1984).

There are four steps in a systematic, user-centered evaluation design:

(1) find out who wants the information, that is who must make the decision -- the information users or decision makers;

(2) identify the decisions for which data can and should be collected;

(3) determine how the data are best collected for the identified decision makers (the evaluator provides information on appropriate study designs and data collection strategies);

(4) choose the deadline and the format for presenting the data to the given information users.

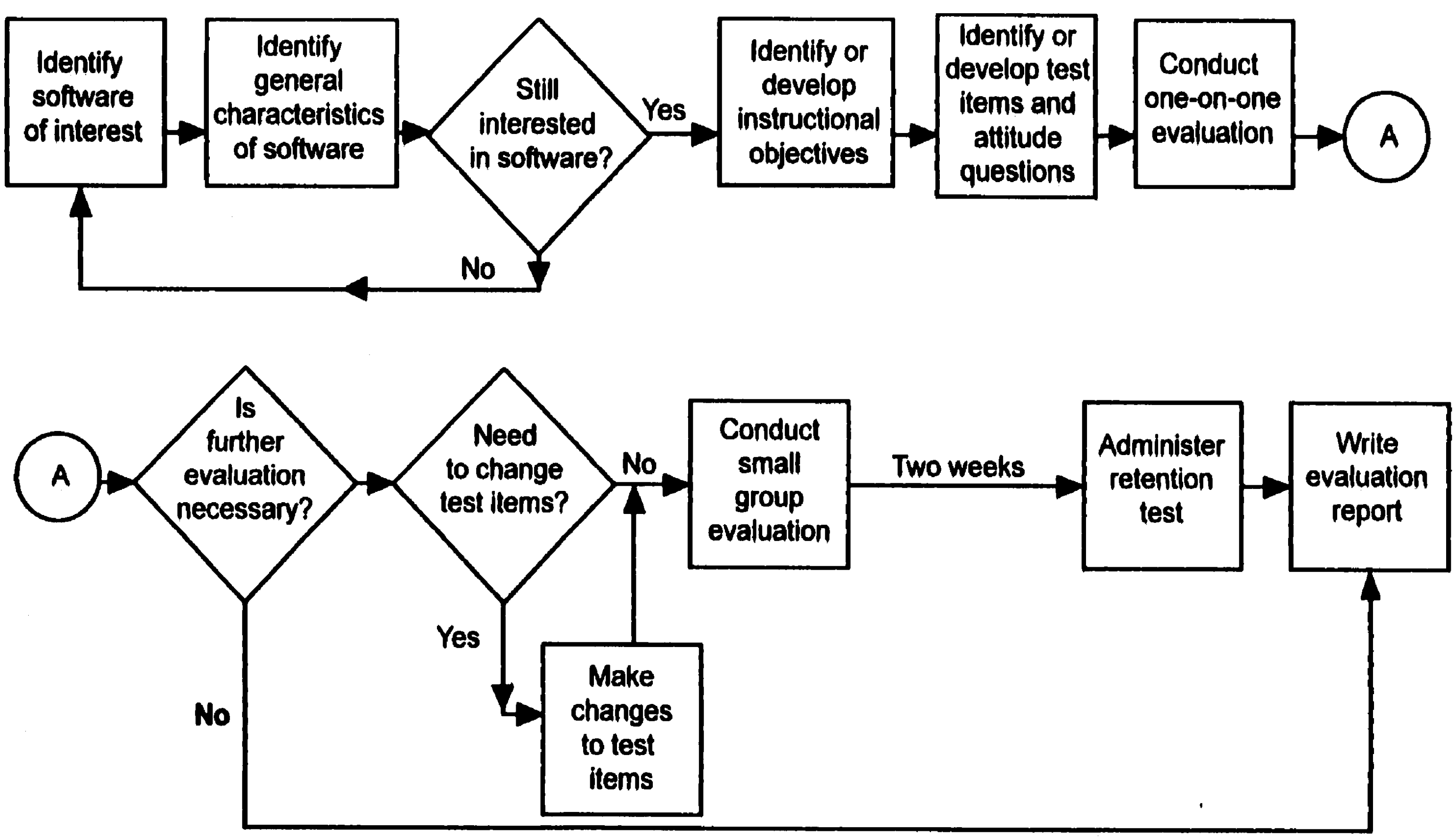
Several authors have also discussed and debated the role of evaluation in measuring the effectiveness of instructional software programs. A main recommendation made by these authors is that learners should serve as participants in the evaluation process and that evaluators should measure what students achieve or learn as a result of participation in courseware instruction (Reiser and Kegelmann, 1994; Gill et al. 1992; Dillon and Gunawardena, 1992). These outcome measurements should include the testing of students achievement before and after they complete a software program. They also recommend that the learners' perceptions and experiences in using instructional courseware should be examined, allowing in-depth understanding of individual learner interactions with instructional courseware. Reiser and Dick's (1990) instructional courseware evaluation model exemplifies the use of systematic and comprehensive summative strategies and methodologies in measuring the instructional effectiveness of computer courseware.

Reiser and Dick propose that instructional courseware effectiveness should be measured according to several evaluative criteria: learner satisfaction; attitude change; and learning achievement. The evaluation procedures (methods and instruments) which are used to measure these criteria include: pretesting learners; observing learners as they work through a software program; posttesting learners to identify what they have learned; and the collection of information on learner attitudes towards software instruction.

Figure 2.3 presents Reiser and Dick's (1990) instructional software evaluation model. The model recommends that evaluators should initially review the instructional software, identifying or developing the performance and learning objectives the courseware is designed to teach. Following this phase, objectives and performance-based tests and learner attitude surveys are developed. During the one-on-one phase, individual learners are observed by an evaluator as they work through the program. The learners are then tested on their knowledge and skills, and questioned about their attitudes toward the instruction. These observation and testing phases provide data about the effectiveness and use of the program, and about the quality of the test items. The next phase, the small-group phase, involves 8 to 20 learners simultaneously going through the program in a realistic learning context. These learners are given an immediate posttest that assesses learning and attitudes. The same learners are then tested for retention two weeks later.

In this model, final evaluative decisions about the software are made at three points. The first decision point comes after the initial review of the software package. If the software is clearly not adequate or not appropriate for current needs, the software is rejected and the evaluation ends (Gill et al., 1992). The second decision point comes after the one-on-one evaluation. If observational, attitudinal, and performance data collected during the evaluations provide enough clear evidence for either adoption or rejection, evaluation ends (Gill et al., 1992). The third decision point comes after the small-group evaluations. The final decision is based on the cumulative data which is

Figure 2.3 Reiser and Dick's Instructional Software Evaluation Model



Source: Reiser & Dick (1990)

documented in an evaluation report (Gill et al., 1992).

Reiser and Dick's model is useful for measuring the effectiveness of instructional courseware which has already been developed. It combines the use of summative achievement data with attitudinal information which are collected from the users of the program. However, the model does not focus on any formative evaluation methodologies which could be used during the design and development of software instruction. Therefore, the model does not provide an approach which is comprehensive enough to serve the interests of program developers and stakeholders alike.

For a variety of reasons, no single established evaluation model was well suited to the task of formative and summative evaluation of the effectiveness of computer-mediated CME courseware. Some models were too complex or costly to use. Others were too narrow in focus. However, several of the models contained components or presented concepts useful for the evaluation of CME courseware. The concepts taken from the models that were most useful for this study included practitioner-as-evaluator, utilization and decision-orientation, naturalistic-orientation, and outcome-orientation.

Practitioner-as-evaluator was a useful concept because it suggests that evaluation should be flexible enough to be integrated within the normal instructional design, development, and implementation tasks of the practitioner. The concepts of decision and utilization-oriented evaluation were appropriate because of the role of the CME Planning Committee in identifying learning needs and designing the courseware. A naturalistic-orientation was essential because it enables the evaluator to explore effectiveness and the interaction of learners with instructional software from the perspective of the user of the instructional system. This can provide the evaluator with a wealth of information on the strengths and weaknesses of a program which may otherwise be missed when utilizing quantitative instruments such as achievement tests and performance surveys. Finally, the focus on outcome-determination was relevant because it allows the collection of data on how well individuals learn and apply the concepts and subject matter from a particular instructional program. This information is essential in judging the effectiveness of instructional materials.

A number of evaluation models and approaches were presented in this section of the literature review. The relevance of these models and approaches to the task of evaluating the effectiveness of computer-mediated instructional courseware was discussed, and strengths and weaknesses of each were identified. Finally, those components and concepts most useful for this study were recognized and discussed. These components and concepts are reflected in the planning of the CME Courseware Evaluation Model presented in Chapter Three.

Evaluation Methodology

A prevalent theme found in the literature on evaluation methodology was that quantitative methods have traditionally dominated research and evaluation studies in the past. However, the current consensus among many authors is that quantitative and qualitative methodologies should be combined to provide a more comprehensive view of the effectiveness and impact of educational programs. Several authors (Cronbach, 1980; Patton, 1980; Burnham, 1995; House, 1991; Guba and Lincoln, 1985; Willis, 1993;

Willis, 1994; Thorpe, 1993) suggest that evaluation studies should be designed combining the two approaches rather than relying solely on one approach or the other. Patton (1980) notes that "on many occasions -- indeed for most evaluation problems -- a variety of data collection techniques and design approaches will be used" (p. 18).

According to Willis (1993) "quantitative evaluation relies on a breadth of responses and is patterned after experimental research which focuses on the collection and statistical manipulation of relatively large quantities of data" (p. 63). Quantitative methods require the use of predetermined categories so that the experiences of all subjects are limited to certain responses. Evaluative conclusions emerge from the resulting statistical analysis. The quantitative approach's reliance on statistical analysis improves its objectivity and permits greater comparison and statistical aggregation of the data, and is particularly useful when there will be large numbers of respondents for whom more in-depth, personalized approaches are not feasible.

However, the preordinate nature of quantitative research methods can be an obstacle to evaluation studies when more in-depth information from the learner's perspective is required. Quantitative approaches, by definition and design, offer respondents a limited number of possible response options, particularly when forced choice surveys are used. As well, statistical analysis often results in an illusion of precision that may be far from reality (Thorpe, 1993). In contrast, qualitative evaluation focuses on depth of response, highlighted by the gathering of detailed descriptions of situations, events, interactions and anecdotal data from typically a smaller group of respondents. The focus in qualitative methods is on the quality of the response and the study of particular issues in greater depth and detail.

Cronbach (1980) suggested that designing an evaluation study is as much art as it is science. The art of evaluation involves formulating and constructing a design and gathering information that is appropriate for a specific situation. There are no ideal models or methods and no preferred nor ideal measures. Deciding what and how much information to compile in an evaluation study involves making difficult decisions about trade-offs. Compiling large amounts of data usually takes longer and costs more. Getting less information takes less time but reduces confidence in findings. It is important, therefore, to establish boundaries on data collection in the early stages of program evaluation.

House (1991) and Guba and Lincoln (1985) suggest that "effective evaluation" should encompass methods from both the scientific (quantitative) and the naturalistic (qualitative) paradigms, with emphasis placed upon the interactions between the program and its context. Similarly, Thorpe (1993) recommends a strategy of "applied common sense" in which you pragmatically take what you need from both (quantitative "have the goals been achieved" and qualitative "what has been going on") and from any other approaches that emerge or exist.

Willis (1994) states that educational program evaluation has undergone "a number of refinements since it was proposed as a discipline in the 1950s, from an emphasis upon tests and measurements, to behavioral objectives, to supplying information to decision-makers, to an emphasis upon judgment and values" (p. 100). Willis notes that a current consensus favors a more balanced view of the use of qualitative and quantitative forms of inquiry.

> "Recognizing that reality is not exclusively numerical or categorical, current consensus thinking stresses a more balanced approach, emphasizing that both are important and that both should be incorporated in some way to provide an adequate description of reality...both qualitative and quantitative techniques can be incorporated into a single model for use in evaluating distance education" (p. 100).

According to Burnham (1995):

> "Most competent researchers and evaluators use both approaches to complement one another. A thorough evaluation will help stakeholders understand both a feeling or a problem and how widespread it is" (p. 83).

Cronbach (1980) listed the following among his "Ninety-Five Theses."

> 59. The evaluator will be wise not to declare allegiance to either a quantitative-scientific-summative methodology or a qualitative-naturalistic-descriptive methodology.
>
> 60. External validity - that is, the validity of inferences that go beyond the data - is the crux; increasing internal validity by elegant design often reduces relevance.
>
> 95. Scientific quality is not the principal standard; an evaluation should aim to be comprehensible, correct, and complete, and credible to partisans on all sides." (pp. 7-11).

Thorpe (1993) suggests that the integration of methodology serves an important purpose in the evaluation of distance education programs because different stakeholders often have a need for evaluative information for different purposes. Subject matter experts will want to know if the content they presented was appropriate and addressed the learning needs of the students; program developers will want information that allows them to judge the quality of the materials and how it can be improved; and administrators will want to know how effective the program is and the return they are getting for the money being invested. Therefore, these different needs have implications for the types of information collected and the types of data collection methods which the evaluator uses.

Several other authors supported the use of multiple methods and criteria to evaluate the effectiveness of continuing education programs. Ronald Cervero's (1988) approach to the evaluation of continuing professional education programs exemplifies the use of multiple methods to evaluate multiple criteria of program effectiveness. According to Cervero (1988) the issue with evaluation is "not whether programs should be evaluated or not, but rather to what extent evaluative information should be collected systematically" (p. 131). Cervero proposed a framework of seven types of evaluative categories (Figure 2.4) which can be used for the systematic evaluation of continuing professional education. He suggested that no one category is inherently better or more useful than another, and that information collected at one level does not infer program success at any other level. The seven categories include: program design and implementation; learner participation; learner satisfaction; learner knowledge, skills, and attitudes; application of learning after the program; the impact of application of learning; and program characteristics associated with outcomes.

The likelihood that a program evaluator or provider will ask evaluative questions from all seven categories is largely dependent on five further questions which program providers must ask themselves:

> "First, what is the purpose of the educational program? Second, who needs what information for what purposes? Third, what are the practical and ethical constraints related to the evaluation effort? Fourth, what

Figure 2.4 Cervero's CPE Evaluation Categories

	Program Design & Implementation →	Learner Participation →	Learner Satisfaction →	Learner Knowledge, Skills & Attitudes →	Application of Learning After the Program →	Impact of Application of Learning →	Program Characteristics
Evaluative Questions	• Learning activities • Instruction • Learning Environment	• Number and type of participant • Quality of participation	• Satisfaction with subject matter • Satisfaction with instructor • Cost to attend • Satisfaction with educational experience	• Change in cognitive, affective or psychomotor competence	• Degree of knowledge, skill, attitudinal transfer to the workplace	• Extent to which health of public has improved	• Characteristics of the CPE program • The individual professional • Nature of proposed change • Social system
Evaluative Methods	• Questionnaires • Interviews • Observational rating scales	• Observation • Registration data • Demographic surveys	• Feedback or reaction survey	• Paper-and -pencil tests • Pretest — post-test	• Self-reports • Observation • Archival Analysis	• Questionnaires • Interviews • Chart Audits	• Surveys • Interviews

resources are available to conduct the evaluation? Fifth, what values and preferences of the educator impinge on the evaluation?" (Cervero, 1988, p.146)

Cervero notes the framework is not all inclusive, and educators and program evaluators should pay close attention to the unique aspects of their program and ask questions which are most appropriate to the requirements of that activity, and the nature of their educational setting. Similarly, Kirkpatrick (1976) has suggested that evaluation criteria should be of four types. Knowles (1980) describes the four criteria this way:

1) *reaction evaluation*: which ideally takes place periodically during a program and provides data to the program managers about how the participants are feeling about the program -- data can be used to make changes in design, methods, personnel, facilities, and the like, as the program moves along;

2) *learning evaluation*: which provide data, ideally through pre and posttests about what knowledge, skills, attitudes, and values have been acquired by the participants;

3) *behavior evaluation*: which provide data to show what changes in actual performance have been produced;

4) *results evaluation*: which provides data about the tangible results of the program in terms of reduced cost, improved quality, increased productivity, lowered accident rates, and the like" (p. 202).

Abrahamson (1984) has suggested that the criteria by which CME has been evaluated has undergone an evolutionary process (Table 2.3). According to Abrahamson the earliest reported measures of the success of CME programs were attendance records and the level of satisfaction expressed by participants about a given course. However, questions were raised about the significance of such findings and what they actually meant, because almost invariably the findings of such measures were positive and presented little information about how much or to what extent learning occurred as a

Table 2.3 Abrahamson's Evolutionary Steps of CME Evaluation Criteria

Abrahamson's Evolutionary Steps of CME Evaluation Criteria
1. Attendance
2. Happiness
3. Knowledge
4. Competence
5. Performance
6. Patient outcomes

result of the CME program.

The testing of learning achievement then became the norm and in some instances took on the form of pre and post-program testing. As well, an important extension of this technique was the administration of a posttest at some time following the completion of a CME program, in some instances up to six months following an event (Abrahamson, 1968). Most programs evaluated using testing methods showed significant cognitive gain and some even showed higher scores on the delayed achievement tests than at earlier periods. These results usually led to conclusions that participants were acquiring new information. But, this raised a further question, did they become more competent?

"Competence" in this case had an implicit meaning: capability to perform (Abrahamson, 1984). Another step in the evaluation of CME then evolved, to ask whether the participants really implemented changes in their practice. According to Abramson (1984) a person could become more competent as demonstrated in some form of assessment immediately at the conclusion of a continuing education program, but did that carry over in actual performance? Thus, another evaluation technique introduced into the CME evaluation repertoire was the direct observation of physicians' performance and the chart review.

These methods usually meant having access to the physician and the opportunity to assess what a physician did in everyday activities. Some evaluators often used interviews or questionnaires to examine physicians' self-reports of changes in performance, despite the notion that these approaches were often questioned for their objectivity and validity. Other techniques which were used to evaluate physicians' behavior change included chart review in hospitals and exploration of the office records of the practitioner. Nevertheless, Abrahamson (1968) suggested that many challenged the extent to which observed or reported changes could actually be linked to a program of continuing medical education:

> "Even after the fact, is it possible for medical educators to attribute documented changes to the specific stimulus of a course, series of conferences, symposium, closed-circuit television program, etc? There are unfortunately so many intervening and contaminating variables that even the tightest experimental research design often leaves us frustrated in our attempt to establish cause-effect relationships." (Abrahamson, 1968, p.628)

The next logical step in the evolution of evaluation in CME then emerged: If the physician's performance changes, are his patients better off? However, Abrahamson (1984) suggests that with the evolutionary progression of CME evaluation methods investigators were beginning to find the dependant variables more and more difficult to study:

> "...the connection between continuing education of the physician and

> patient outcomes is hardly a simple cause-and-effect phenomenon. The continuing education effort provides the physician with knowledge or with skill; the knowledge or skill is applied in the physicians practice; the practice behavior of the physician is perceived by the patient; the patient's perceptions brings about a reaction; that reaction has an impact on the patient problem. For any one of the linkage components there must be a number of potentially intervening variables. One example lies in the whole area of patient compliance and the "extraneous" variables influencing that factor: empathy of the physician, health habits of the patient's family, and attitudes towards health care" (p. 12).

Caplan (1973) echoed this ambivalence towards the utility or reliability of using measurements of patient health outcomes as a reflection of the effectiveness or impact of a continuing medical education program. Caplan described an evaluation study on a CME program in pediatrics and obstetrics which reported that infant mortality rates had failed to improve in a particular region in spite of the CME programs being offered. His main criticism of this suggested relationship was that the measure "decreased mortality rate" was an inappropriate indicator of the educational effectiveness of the program. He used an analogy to describe this issue more clearly:

> "Even the massive amount of water added to the ocean during a hurricane would likely be undetected by a careful scientist measuring the water level at a safe distance. But is it justifiable then to conclude that hurricanes have no effect? The evaluation measure must be precisely appropriate to the intention of the learning experience." (Caplan, 1973, p.1151)

Caplan suggests that anecdotal information, by way of physician self-reports following participation in a CME program provide results which offer important information on the success of a given CME activity. This information must not be overlooked simply because it does lend itself to objectivity or strong validity.

As a further example of these complexities, Ferguson et al (1984) surveyed participants in a CME workshop on their self-reported reactions to the recommendations of a CME intervention, as well as their stated reactions for adopting or not adopting the recommended behavior. The results of the survey revealed that physicians who did not use the procedure prior to the workshop and had an unfavorable attitude toward it were more likely to not adopt the recommendations. Physicians who were not using the procedure but had a favorable attitude towards the recommendation indicated that they had adopted the procedure or were trying it. The authors concluded that attributing changes in behavior to CME efforts presents significant challenges. Such changes often result from accumulated experience and knowledge rather than a single exposure to a new idea. In addition, attitudinal and environmental variables may prevent physicians from adopting the desired behavior (Ferguson et al., 1984).

The literature also proposes a distinction between two types of evaluation -- summative and formative. Summative and formative evaluation are differentiated by the objects they assess, the information they provide, and the audiences they are designed to serve. Guba and Lincoln (1985) suggest that the aim of formative evaluation "is to modify and improve the design of an innovative program of curriculum while it is still under development" while summative evaluation "is to critique a completed entity in terms of professional or expert standards so as to be able to certify and warrant merit" (p.51).

Generally, summative evaluation occurs after an instructional program has been developed and delivered to the learner. It is meant to collect information that enables decision-makers to judge the impact or effectiveness of a program in terms of a variety of outcomes. These outcomes could include the amount of learning which has been affected, the influences of instruction on an individual's performance, the reaction or satisfaction of learners with instruction, the effect of a training program on the organization, or a cost-benefit analysis (Dick and Carey, 1985).

Formative evaluation, on the other hand, has a more immediate purpose. It is the systematic collection of information for the purpose of informing and determining the quality of instructional materials while they are in the design and development stages (Northrup, 1995; Golas, 1983; Chinien & Hlynka, 1993; Dick & Carey, 1985; Weston, 1987; Thompson, 1987). The original proponent of formative evaluation was Scriven (1967) who envisioned its role as a means for increasing the effectiveness of an instructional product while it was still in a development stage and capable of being revised.

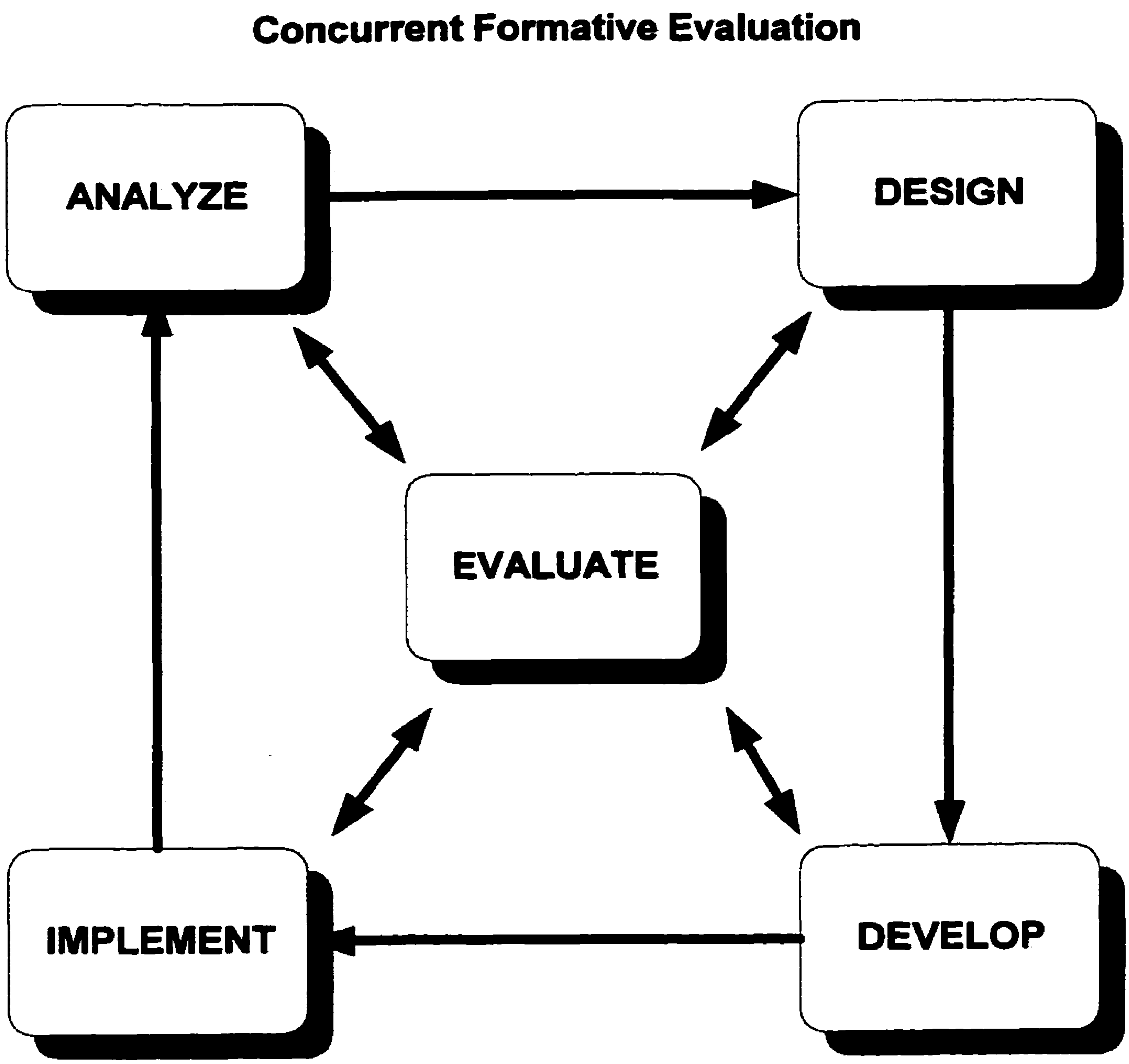
Russell and Blake (1988) suggest that formative and summative evaluation may be applied to both people (learners) and products (instructional materials). The primary purpose for the formative evaluation of instructional products is revision and improvement, while the purpose for formative evaluation of learners is diagnosis and remediation. Summative evaluation of an instructional product, on the other hand, is used to determine effectiveness or efficiency. And, summative evaluation of the learner is often used to assign grades and to certify competency.

There are three broad questions which are usually addressed by formative evaluation. According to Chinien & Hlynka (1993) the first question relates to the content; the second relates to the technical quality of the material; the third pertains to its learnability. Each question can be addressed by collecting information from a source, which for content and technical quality is usually an expert for verification and revision, and for learnability assessment typically involves students for providing feedback data. A majority of authors suggest that a combination of experts and learners should be involved in the formative evaluation of materials due to the different feedback which each can provide (Weston, 1987). Consulted experts might include content or subject matter experts, pedagogical experts, instructional design experts and media experts, among others. Each type of expert tends to focus upon that aspect of the material which falls within his/her own area of expertise.

Dick and Carey (1990) have suggested that there are three stages of formative evaluation. The first is the one-to-one or clinical evaluation stage. In this initial phase the designer works with individual students to obtain data to revise the materials. The second stage of formative evaluation is a small group evaluation. A group of ten to twenty students who are representative of the target population study the materials in an approximate "real-life" setting to collect the required data. The third stage of formative evaluation is usually referred to as a field evaluation and the emphasis is on the testing of the procedures required for the installation of the instruction in as real a situation as is possible.

Northrup (1995) has suggested that the formative evaluation stages identified by Dick and Carey, as well as others, may be conducted concurrently during the instructional systems design process (ISD) (see Figure 2.5). As well, Northrup recommends that evaluators should use a variety of approaches throughout the various stages of ISD in

Figure 2.5 Northrup's Concurrent Formative Evaluation Process



order to improve and ensure the quality and effectiveness of instructional materials or products. These different approaches are presented in Table 2.4.

During the analysis stage of instructional systems design the purpose of formative evaluation is to confirm findings of the needs analysis and the proposed subject matter to be used in the instruction. Information is gathered to confirm accuracy, currency, relevance and comprehensiveness. At the design stage, after several brainstorming sessions on the instructional approach, a storyboard is created of the instructional product and this is evaluated by the key stakeholders and/or learner audiences. Formative evaluation during the development stage involves the rapid development or authoring of a prototype lesson on whatever tool is being used. The intent during this stage is to test the approach with the target audience.

Traditionally, there have been four basic methods used in formative evaluation: expert review; one-to-one evaluation; small group evaluation; and, the field test evaluation. However, Tessmer (1994) suggests that these traditional methods and tools may be replaced or complemented by several alternative evaluation methods that are not well known. Time or resource pressures, geographically distant experts, or complex learning tasks could serve as the impetus for using these alternative methods. As well, computer and electronic communication technologies have enabled the use of new tools for gathering information (Tessmer, 1994). Table 2.5 provides an overview of the alternative methods which Tessmer recommends, including their advantages and disadvantages.

Several authors have taken an interest in comparing the effectiveness of

Table 2.4 Northrup's Concurrent Formative Evaluation and Data Collection Criteria

Guideline	Group Involvement	Data Collection Strategies	Information to Collect
ANALYSIS Review findings with clients to verify correctness	SMEs key stakeholders policy makers	INTERVIEWS: one-to-one, two-to-three, three-to-one panel discussions, focus groups, mass mailouts	accuracy, currency, relevance, lack of bias, comprehensiveness
DESIGN Create a prototype of design approach	clients key stakeholders end users	think-aloud-protocol	attitudes: approach used, screen layout and design, general interactivity, user interface
Conduct a paper-based segment with rapid prototype	key stakeholders SMEs	guidelines for review instrument or checklist written feedback directly on instructional strategies	content, accuracy, currency, relevance, comprehensiveness, lack of bias
Conduct a storyboard evaluation	target audience	one-to-one by interacting with user using paper-based or computer generated storyboard	clarity of information, instructional strategies and procedures, logic, currency, accuracy, relevance, ease of use, lack of bias
DEVELOPMENT Conduct a prototype evaluation	target audience	small-group try-out	user performance-prototype assessment, paper/pencil task, online questions, problem-solving
IMPLEMENTA-TION Conduct a product evaluation	target audience	instrument to measure organizational data	user perceptions of instruction, technical and cosmetic adequacy including: motivations, interactivity, learner control, user interface, screen layout
		outcome data, field trial	confirmatory user performance and attitudes

Table 2.5 Tessmer's Advantages and Disadvantages of Alternative Methods of Formative Evaluation

Advantages and Disadvantages of Tessmer's Alternative Methods of Formative Evaluation		
Method	**Advantages**	**Disadvantages**
Two-on-one	learner dialogue learner agreement possible time savings	no pace/time data no individual opinions dialogue distracting
Think-Aloud Protocol	data on mental errors learning process data	learning intrusiveness awkward to use
Computer Interviewing	access to remote subjects continuous evaluation	time-consuming analysis training & equipment
Self Evaluation	easy to conduct insider's viewpoint	not rigorously conducted "forest-tees" problem
Panel Review	expert dialogue negotiated agreements	may move off task less independence
Evaluation Meetings	amount of group information quick tryout	only easy changes made
Computer Journals	continuous evaluation cost/time effective	equipment & software user literacy levels no evaluator present
Rapid Prototyping	assess new strategies assess new technologies	time & cost to develop undisciplined design

the various formative evaluation methods. The results of their studies have been mixed. Wagner (1983) for instance, compared the relative effectiveness of the first two stages of formative evaluation: the one-to-one and small group. The results of this study concluded that it might be possible to eliminate either the one-to-one or small group stage of formative evaluation and maintain effective revisions of instructional materials.

Owston and Wideman (1987) compared the effectiveness of panel reviews and field test methods for the formative evaluation of software. Their results suggest that field testing could bring to light technical and design limitations that are not obvious to teacher reviewers; provide more accurate information on the ease of use of the software; and give a clearer indication of the suitability of software in meeting learning needs. However, the overall level of agreement found between the panel and field tests in the study suggested that systematically conducted panel evaluations will usually be able to assess a program's quality with reasonable accuracy.

Some authors have also discussed the concept of planning as an important component of evaluation. In particular, Thorpe (1993) suggests that a plan, in and of itself, is a model for evaluation and serves to provide guidelines for evaluative action. Thorpe recommends that the components of an evaluation planning model include:

- The purpose for the evaluation? (program planning, policy-making, program improvement, program justification, or accountability);
- Who is the audience of the evaluation?
- What are the issues to be addressed by the evaluation?
- What resources are available to do the evaluation?
- What kinds of evidence will be useful in pursuing the issues?
- What data gathering methods will be used?
- What type of analysis will be performed?
- How will the findings be reported?

As well, Coldeway and DeLisa (1986) have emphasized the importance of evaluation planning as soon as the training needs for the target learning population have been identified. They argue that this enables program planners to place a greater emphasis on linking the evaluation plan with the design and development stages of the educational program. Coldeway and DeLisa propose an evaluation decision-making flowchart, Figure 2.6, which allows evaluators and CME program providers to determine what kind of evaluation is appropriate for a given CME event.

The first decision-point focuses on recognizing the type of evaluation to be used -- formative or summative. Decision-points beyond this first level are used for determining the depth of summative evaluation information that a program provider wishes to collect. Summative information at these levels allow CME planners to make decisions about the success of the program in terms of "learner satisfaction, change in learner knowledge, change in performance, and change in impact on the patient or health care environment" (Coldeway and DeLisa, 1986, p. 181). Once decisions about the type of evaluation and the kind of information to be collected have been made, the evaluator can then focus attention on five other evaluation concerns:

(1) When to evaluate?

(2) What kinds of evaluation questions are important to ask?

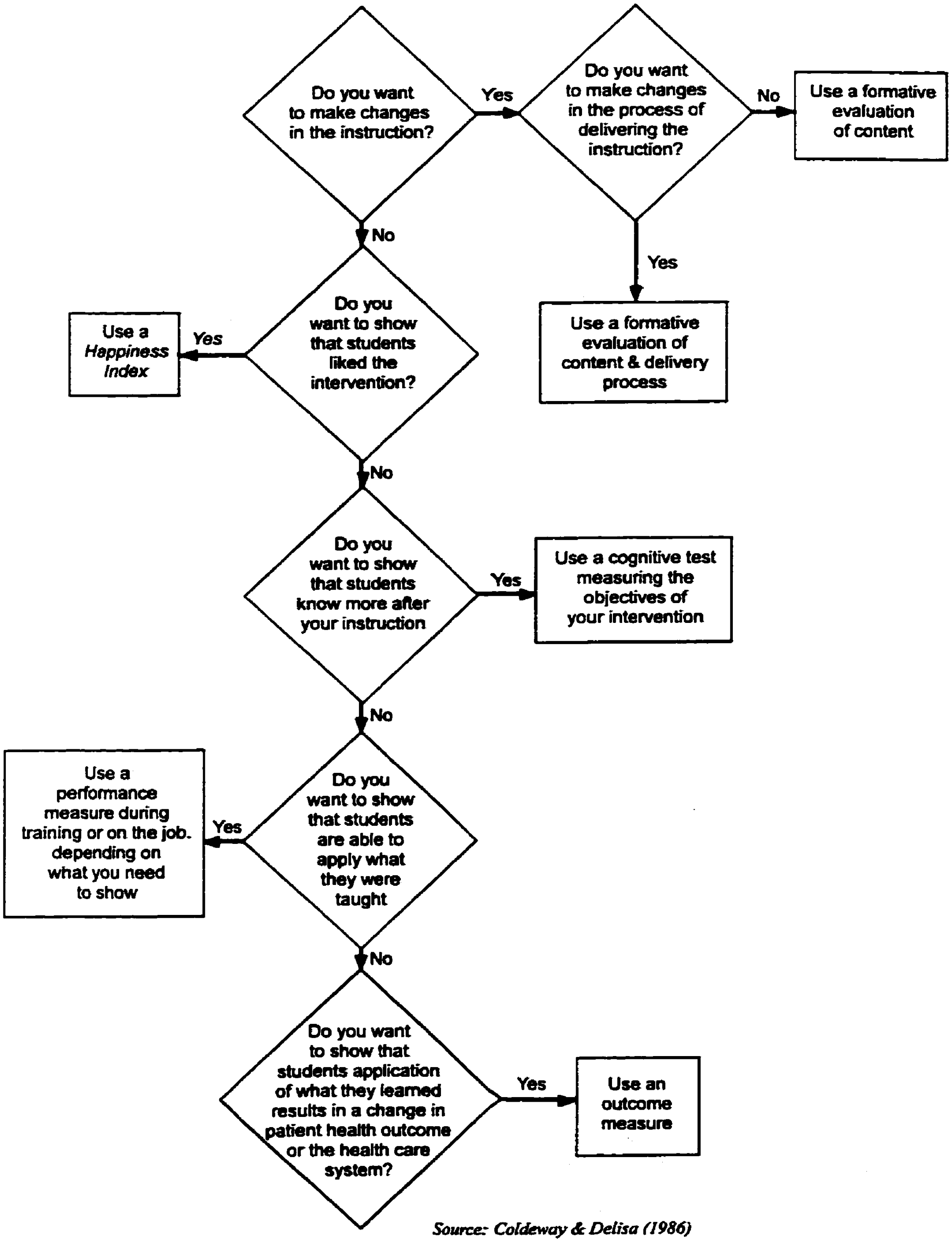
(3) What are the sources of evaluation data?

(4) What are the methods for evaluating?

(5) What are the ways to summarize and analyze the data? (Coldeway and DeLisa, 1986)

The sources of evaluation for an assessment of the effectiveness of a CME program are dependant on the type and level of evaluation that is being conducted. According to Coldeway and DeLisa (1986), typically for summative data, either learner

Figure 2.6 Coldeway and DeLisa's Evaluation Decision-Making Flowchart



Source: Coldeway & Delisa (1986)

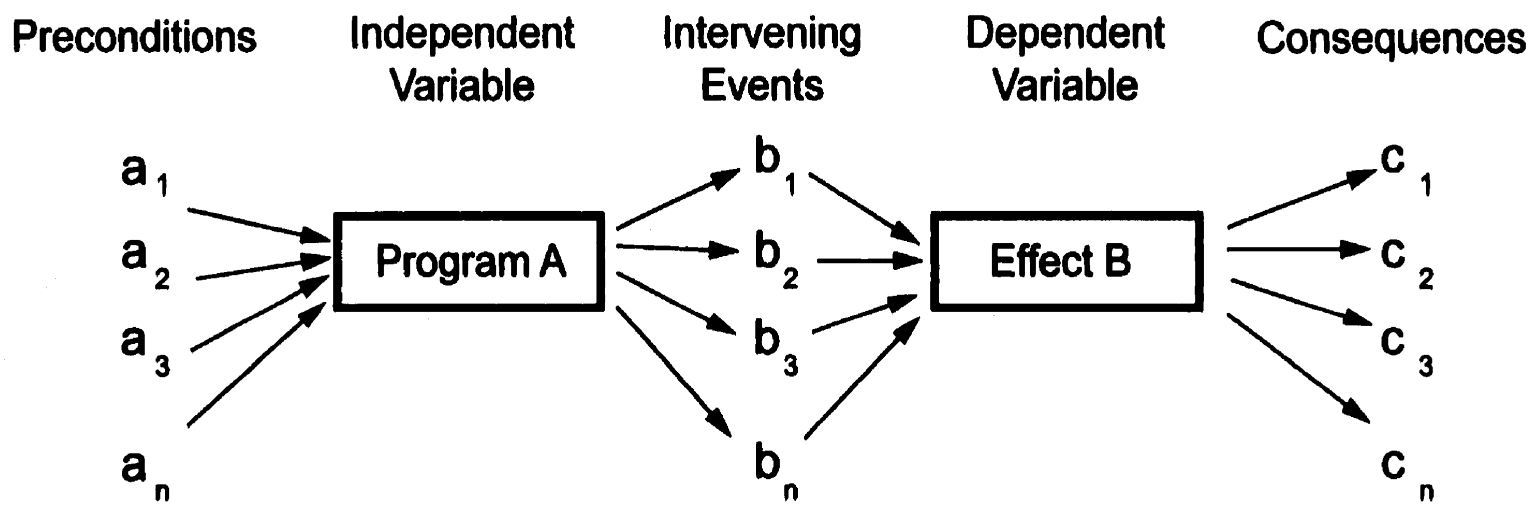
satisfaction surveys or knowledge change data are collected directly from the learners. Measures of change in performance may be collected by contacting the learner, supervisors, or their colleagues. And, to assess the impact of a CME program, it may also be possible to audit medical records and to collect information from the learners and their patients.

In summary, the literature on evaluation methodology suggested various options related to the selection of evaluation procedures. These options included recommendations regarding the types of methods to use, the value of various measures and sources of information, and the utilization of multiple criteria for determining the effectiveness of an educational program.

Antecedent Variables

Another theme which has relevance for the evaluation of computer-mediated and distance education programs in general is the collection of information on what Babbie (1992) calls antecedent conditions. Babbie suggests that one of the most important categories of evaluative information on which program success should be determined is antecedent conditions. Antecedent conditions may refer to the context within which a program is implemented or the characteristics of the learners for whom the program is directed. Suchman (1967) as well, emphasized the need for evaluation research to be conducted in terms of different categories of effect. Suchman's multi-causal model (see Figure 2.7) suggests that the intended causal consequences of an evaluation study are only one of many possible sets of events which influence and lead to outcomes. According to

Figure 2.7 Suchman's Multi-causal Concept of Evaluation



Source: Stufflebeam & Shinkfield (1985)

Suchman programs are inevitably evaluated within the context of other programs or events that affect the desired outcome. Therefore, an evaluation study's design must make an important provision to consider the preconditions that could influence the initiated program activity and outcomes.

Melton (1992) also recommends the measurement of the individual differences and learner characteristics which evaluators believe are relevant to the interpretation of outcomes. Melton suggests that various types of information might be collected to assist the evaluator in the meaningful interpretation of educational outcomes. This data could include: certain background variables such as schooling; motivation and attitudes; as well as measures of personality traits such as locus of control. Melton believes that such measures serve an important role, particularly in the formative experimentation of interactive multimedia as an instructional tool.

According to Busch (1995) attitudes towards computers and perceived computer self-efficacy expectations represent an important issue in the area of computer-mediated instruction. In this study, several factors surrounding attitudes toward computers and individual learner characteristics were explored because they have been shown to affect computer interest, enrollment for computer-mediated learning courses, and the motivation and persistence of learners to succeed in a computer-mediated learning environment. Research in this area, particularly its effect on secondary and post-secondary level students, is quite extensive. It focuses on two dimensions -- perceived self-efficacy and attitudes toward computers. Self-efficacy is defined as the belief in one's ability to carry out successfully a certain course of behavior (Busch, 1995). Research suggests that self-efficacy will influence the choice of whether to engage in a task, the effort expended in performing it, and the persistence shown in accomplishing it (Busch, 1995).

There has been significant research on the effect of computer attitudes on learning achievement and its relationship with gender, prior computer experience, and computer ownership (Nickell and Pinto, 1986; Busch, 1995; Shashaani, 1994; Woodrow, 1994). The literature is supportive of the existence of persistent gender-differences in attitudes towards computers. In the majority of these studies male secondary and post-secondary age students are found to have more positive attitudes toward computers than their female counterparts. Nickell and Pinto's (1986) study of the computer attitudes of college students found a significant difference between the computer attitudes of males to females. The authors found that males have significantly more positive attitudes towards computers than their female counterparts. Similarly, Busch (1995), Shashaani (1994), and Woodrow (1994) conducted surveys of secondary students' attitudes towards computers. The findings suggest that gender-differences in computer experience have a direct relationship to computer attitudes, and that males usually show more positive attitudes toward computers than girls.

Several studies have also suggested the existence of a relationship between computer experience and an individual's attitudes toward computers. Both Busch (1995) and Shashaani (1994) found computer experience to be strongly correlated with computer attitudes, suggesting that students with high levels of computer experience tend to have more positive computer attitudes. Shashaani (1994) assessed the extent of students' experiences with computers through the use of several survey items: whether or not students had taken any computer courses; the number of courses taken; home computer ownership; usage of computers per week; intention to take computer classes; where students first learned about computers, and the areas where students would like to use computers. Using a Computer Attitude Scale of 39 items to measure students' attitudes toward computers, a positive correlation between computer experiences and computer attitudes suggested that more exposure to computers was associated with more positive attitudes toward computers.

Fann et al. (1989) also found that students with more computer experience were more likely to have positive attitudes toward computers than those with less experience. As well, they found that students with a more positive attitude were also more likely to use computers more often for completing assignments than students with less positive attitudes. Similarly, Malaney and Thurman's (1989) study of the relationship between computer familiarization and computer attitudes also found that prior familiarization with computers was a determinant of positive attitudes toward computer use.

Computer experience is also believed to be a predictor of an individual's perceived self-efficacy or level of confidence in using a computer. Ertmer et al. (1994) suggest that individuals with less confidence or poor self-efficacy beliefs of computer use are more likely to underachieve in computer-based learning environments. The authors investigated the affect of computer experience on attitudes toward computers and judgements of confidence, or self-efficacy. The Computer Technologies Survey (Ertmer et al., 1994) was used as a pretest and posttest measure to assess students' attitudes toward computers and perceived self-efficacy. According to the authors, students often experience reactions toward computers that may either enhance or slow down, and in some cases even prevent the development of effective computer skills (Ertmer et al. 1994). If individuals are to embrace computer technologies as a learning tool and medium for instruction, they must feel confident and comfortable using them.

Ertmer et al. (1994) suggest that there are three types of computer attitudes that appear to have a significant effect on student achievement on computer tasks: anxiety, liking, and confidence (self-efficacy). The construct of self-efficacy refers to an individual's confidence in his/her ability to organize and implement actions necessary to attain designated levels of performance (Ertmer et al., 1994). Their research in learning achievement settings indicates that students' efficacy beliefs influence such achievement behaviors as choice of tasks, persistence, effort expenditure, and skill acquisition. They also suggest that the amount of experience a person has with computers appears to be a significant factor in an individual's judgements of self-efficacy for computer-related tasks (Ertmer et al., 1994). Anxiety, as well, is generally thought to be produced by a lack of familiarity. Therefore, as students become more familiar with computers, anxiety decreases and confidence increases.

Fann et al. (1989) have also studied the influence of computer efficacy on computer use. They describe computer efficacy as a belief in one's ability to use computers. The concept stems from research completed by Bandura and his associates about the role of personal efficacy and phobias (Bandura and Schunk, 1981). According to Bandura, negative attitudes and a loss of perceived control can influence people's ability to master a particular behavior (Bandura and Schunk, 1981). The authors surveyed a sample of university-level students completing a business communications course. They assessed attitudes towards computers, perceived computer self-efficacy beliefs, and the socio-demographic characteristics of their subjects. The results of the study suggested that personal efficacy is a strong predictor of subsequent adoption of computer technology. If individuals do not believe that they can interact successfully with computers, they most likely will avoid computers no matter how useful the computers may be (Fann et al., 1989).

Computer ownership and age appear to influence computer attitudes as well. In one study, Shashanni (1994) found that students with access to home computers demonstrated greater interest in participating in computer-related activities, higher attitudes towards computers, and greater knowledge of computers. As well, Nickell and Pinto (1989) suggested that age is a determinant of attitudes toward computers. In their study of computer attitudes, younger people tended to have more positive attitudes towards computer technology than older student cohorts. Jones and Wall (1989) also found that age was significantly related to computer anxiety in that older students experienced stronger anxieties toward computer use than younger students.

In summary, a number of studies have explored the relationship between computer attitudes, individual computer experience and demographic characteristics, and achievement in computer learning environments. The findings of these studies suggest that computer attitudes may influence the opportunities for children and adults to deploy computer technology. Several individual learner characteristics in particular appear to influence attitudes toward computers -- computer ownership, computer experience, age, and gender. However, there is wide disagreement about how the various variables may interact and influence perceived computer self-efficacy and achievement in computer-mediated learning environments. What is apparent from these studies is that for maximum learner benefit to be derived from investment in computer-mediated learning programs and computer technology, it is important for program providers to be aware of the obstacles and challenges to participation in computer-mediated learning.

Summary

The literature on evaluation and CME at a distance, metaevaluation, evaluation models, evaluation methodologies, and antecedent variables (computer attitudes) has been reviewed and discussed in Chapter Two. This literature review revealed that no single evaluation model was properly suited for the formative and summative evaluation of the effectiveness of computer-mediated CME courseware. Based on the review of the literature on evaluation models and methodology, several evaluation concepts, procedures and methods suitable for the evaluation of computer-mediated CME courseware were identified. These methods and procedures were utilized in the design of an evaluation planning matrix that is presented in Chapter Three. In addition to the evaluation planning matrix, the evaluation model which was developed and fieldtested, and the metaevaluation procedures are also presented in Chapter Three.

CHAPTER III

METHODOLOGY

Introduction

This chapter focuses on the processes and procedures which were used in the planning, development, and fieldtesting of the CME Courseware Evaluation Model. Based on the extensive literature review presented in Chapter Two, an eclectic evaluation planning matrix was developed by the evaluator. The components of the eclectic planning matrix are presented in this chapter as they were used in planning and developing the CME Courseware Evaluation Model. The instructional courseware program which was evaluated during the fieldtest of the evaluation model is described, including an explanation of the evaluation model as it was developed from the evaluation planning matrix. As well, the study design, instruments, and analysis methods used during the fieldtest are also presented in this chapter.

The final section of the chapter is devoted to a description of the metaevaluation, an evaluation of an evaluation, of the fieldtest. The research questions which were originally stated in Chapter One are presented again and the procedures followed to address each of the study's research questions are outlined.

An Eclectic Evaluation Planning Matrix for Computer-Mediated CME Courseware

The eclectic evaluation planning matrix described in this chapter was designed and developed to provide a tool for formulating an evaluation of computer-mediated instructional courseware. The process for devising a matrix was modified and adapted from Sheets (1983). An eclectic evaluation planning matrix is a set of conceptual organizing components and guidelines which can be used to design, develop, and implement an evaluation. The main purpose of a courseware evaluation model developed using the eclectic evaluation planning matrix is to provide information to decision-makers and program developers for planning, implementing, disapproving, and/or improving computer-mediated instruction.

The eclectic evaluation planning matrix was designed and developed based on information gathered during the review of the literature. Several factors which influenced formulation of the design included:

- evaluation based on a practitioner-as-evaluator orientation;
- evaluation which integrates quantitative and qualitative methods and instruments;
- evaluation of multiple evaluative criteria;
- evaluation which is improvement and decision-making oriented;
- evaluation as a concurrent ISD process;
- evaluation which is formative and summative.

The six major components of the evaluation planning matrix are presented in Table 3.1. These components were acquired from the literature reviewed in Chapter Two and are discussed in greater detail in subsequent sections of this chapter.

1. Evaluation Study Design

The choice of pre-experimental, quasi-experimental or experimental study design

Table 3.1 Components of an Eclectic Evaluation Planning Matrix

Component		Source	Rationale
1.	Evaluation Study Design	Muller (1995) Babbie (1992)	Control and logical rigor establishes evidence for causality.
2.	Evaluation Type	Scriven (1967) Guba & Lincoln (1985) Dick & Carey (1985)	Evaluation serves to improve instructional products during development and assess the impact of implementation.
3.	Evaluation Criteria	Cervero (1988) Kirkpatrick (1976) Abrahamson (1984) Stake (1967)	Multiple criteria are required to conduct a comprehensive examination of instructional effectiveness.
4.	Evaluation Method	Reiser & Dick (1990) Willis (1993) Thorpe (1993) Guba & Lincoln (1985)	A combination of quantitative and qualitative data gathering methods should be used depending on the circumstances.
5.	Evaluative Element	Northrup (1995) Coldeway & DeLisa (1986)	The element of the evaluative effort must be identified to facilitate the process.
6.	Evaluation Questions/Rationale	Cervero (1988)	Evaluation questions/rationale direct the formative and summative evaluative purposes.

influences the control an evaluator may exert over various sources of external and internal invalidity. This component of the evaluation planning matrix and the evaluation model fieldtested in this study sets it apart from other evaluation approaches, models, and frameworks. The extent of experimental rigor that an evaluator wishes to apply in an evaluation is dependent on several conditions: the resources available; access to learners for sampling selection and assignment to experimental groups; the audience of the evaluation; the context, situation, and program being evaluated; and, the characteristics of the learners in the program being evaluated and their willingness to participate in a rigorous study. All of these factors must be considered in selecting an evaluation study design which will fit an evaluator's unique circumstances.

2. Evaluation Type

Formative evaluation should occur concurrently during the instructional design and development process to assess the quality of instructional materials and products while they are still in a format by which improvements and revisions can be made. Summative data is more often used by key decision-makers for rejecting or accepting a program. Summative evaluation may also include the collection of information to serve the information needs of program developers and thereby assist them in improving the instructional quality of learning materials and products. Both types of evaluation should be part of a comprehensive and eclectic evaluation of instructional courseware programs. This component of the evaluation planning matrix and the evaluation model also differentiate it from many of the evaluation approaches, models, and frameworks reviewed in the literature.

3. Evaluation Criteria

Based closely on Cervero's (1988) notion of multiple evaluative criteria for evaluating the effectiveness of continuing professional education programs, Babbie's (1992) antecedent variables, and Stake's (1967) observation of instructional transactions, the five types of summative criteria which can be assessed when using the planning matrix are:

- Antecedent (precondition) data
- Reaction (satisfaction) data
- Cognitive (learning) data
- Behavioral (performance) data
- Instructional Transaction data

The criteria upon which formative data should be collected are based on Northrup's (1995) formative data collection criteria:

- Analysis
- Design
- Development

Many of the evaluation models and approaches reviewed in the literature rarely addressed evaluative criteria along both formative and summative evaluation types. Therefore, this component of the planning matrix and the evaluation model distinguish it from many of the evaluation frameworks identified in the literature review. However, these evaluative criteria are not inclusive and evaluators can select any other appropriate variables which may fit their circumstances.

4. Evaluation Method

An evaluation of instructional courseware should integrate quantitative and qualitative research methodologies. Multiple data-gathering methods enable the evaluator to gather both quantitative and qualitative data which can serve a variety of information needs for stakeholders and program developers. Various methods may be used to evaluate computer-mediated CME courseware at formative and summative levels, collect data to address a single evaluation question, or one method may be used to answer more than one evaluation question. These methods have been discussed in Chapter Two.

5. Evaluative Element

The evaluative element is concerned with the object, item, or individual being evaluated. The information gathered from this study's evaluation model was ultimately concerned with the assessment of content, instructional strategies and design, media and instructional materials, instructional components, learning achievement, learner satisfaction, and performance gain. In the model, the participants are the object of cognitive, satisfaction, and behavioral data collection while the courseware product is the object of formative evaluation, reaction, and satisfaction data collection.

6. Evaluation Questions

Evaluation questions help guide the evaluator and are included as a component in the planning matrix and the evaluation model because of Cervero's (1988) use of similar questions in the evaluation of continuing professional education programs. The specific evaluation questions used for designing the evaluation model for this study are presented in Table 3.2.

The Computer-Mediated CME Courseware Evaluation Model

Based on the various components of the evaluation planning matrix the evaluator designed a model for the evaluation of computer-mediated CME courseware. The matrix of that model, the plan it is based upon, appears in Table 3.2 and represents the evaluation questions and activities that occurred during the fieldtest of the evaluation model. A schematic overview of the evaluation model is presented in Figure 3.1 and the methods and instruments which were used are described in greater detail in the following sections.

A unique feature of the CME Courseware Evaluation Model fieldtested in this study is its focus on the comprehensive and systematic collection of data along formative and summative evaluation types to provide information for improving the instructional courseware product and decision-making. At a formative level, evaluation was seen as a process to improve and revise the courseware product during the analysis, design, and development stages. By contrast, information which was collected at a summative level enabled decision-makers to assess whether the finished program was effective and represented a significant and innovative advancement in CME instruction. However, summative data can also be useful in providing information to program developers, yet formative data may not have the same value to decision-makers.

The model is distinguished by the two evaluation types: formative evaluation and summative evaluation. The formative evaluation stage encompasses three main

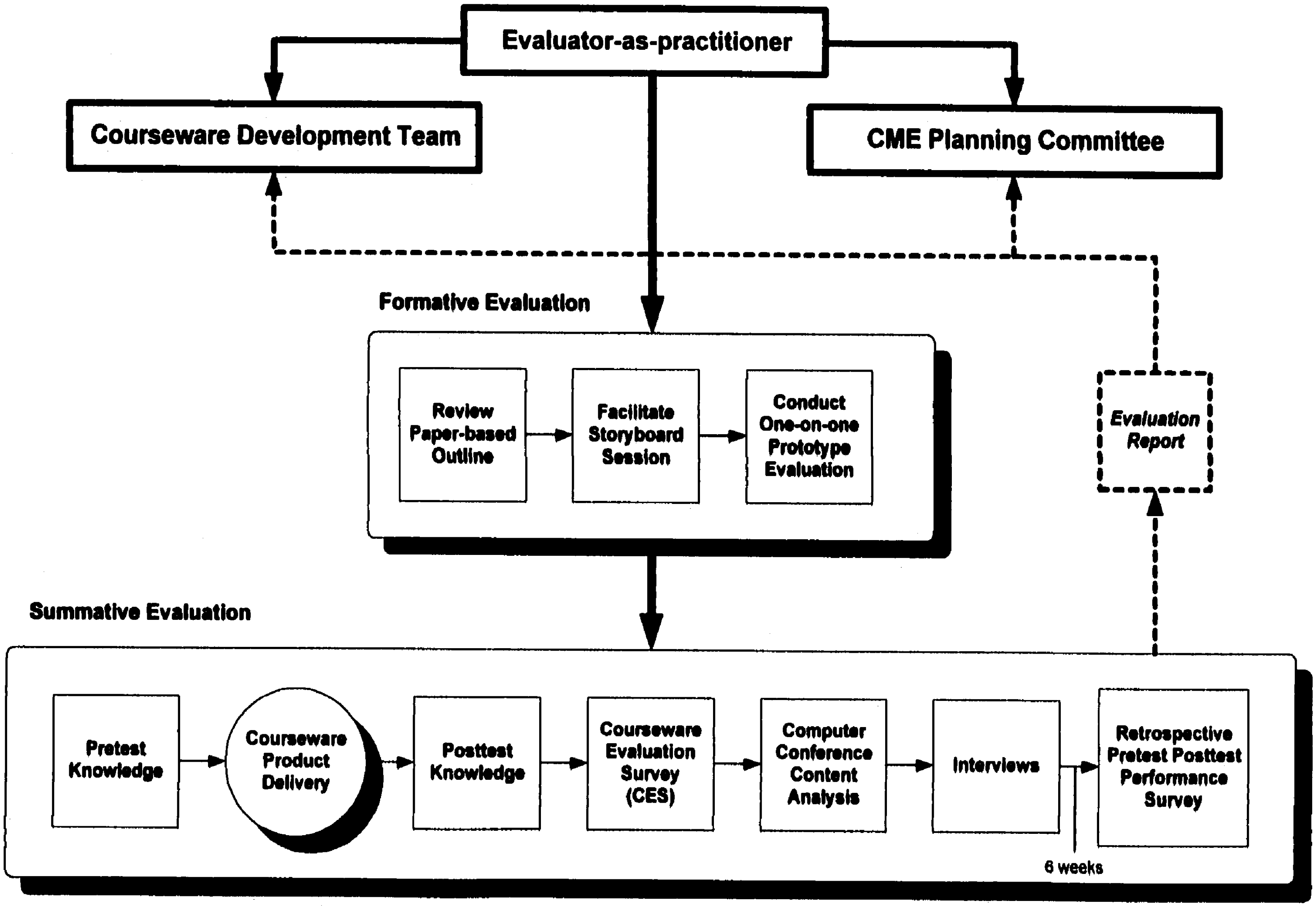
Table 3.2 Eclectic Evaluation Planning Matrix as Applied to the Development of the CME Courseware Evaluation Model

Evaluation Study Design	Evaluation Type	Evaluation Criteria	Evaluation Method	Evaluative Item	Evaluation Questions/Rationale
Modified Quasi-Experimental (Pretest-Posttest Control)	Formative	Analysis	Paper-based Instructional Outline	Courseware Product	CME Planning Committee members' perceptions of the accuracy, relevancy, comprehensiveness and bias of the learning and performance objectives, and the subject matter.
		Design	Paper-based Storyboards		CME Planning Committee members' perceptions of the clarity of information, the instructional strategies and procedures, logic, currency, accuracy, relevance, ease of use, and bias.
		Development	One-on-one		User's perceptions of instructional quality, technical and cosmetic adequacy including: interactivity, learner control, user interface, and screen layout.
	Summative	Antecedent (Preconditions)	Computer Attitude Scale	Participants	1. How did the participants': self-reported years of medical practice; gender; practice location; and practice type influence their attitudes toward computers.
					2. How did the participants': self-reported computer usage; computer access; and computer experience influence their attitudes toward computers?
			Demographic Profile Survey		3. How did the participants': self-reported computer software and Internet application usage influence attitudes towards computers?
			Pre and Post-learning Achievement Tests		4. How did the participants' attitudes toward computers influence learning achievement in a computer-mediated instructional environment?
					5. How did the participants': self-reported years of medical practice; gender; practice location; and practice type influence performance on a pretest of dermatological office procedures?
					6. How did the participants': self-reported years of medical practice; gender; practice location; practice type; computer usage; computer access; computer experience; computer software; and Internet application usage influence performance on a posttest of dermatological office procedures?

Table 3.2 (continued) **Eclectic Evaluation Planning Matrix as Applied to the Development of the CME Courseware Evaluation Model**

Evaluation Study Design	Evaluation Type	Evaluation Criteria	Evaluation Method	Evaluative Item		Evaluation Questions
Modified Quasi-Experimental (Pretest-Posttest Control)	Summative	Cognitive (Learning)	Pre and Post-learning Achievement Tests	Participants	7.	Was there a significant difference in knowledge gain of dermatological office procedures between participants in a computer-mediated instructional treatment group and a no-CME control group?
					8.	Was there a significant change in the level of cognitive knowledge of dermatological office procedures from pretest to posttest for participants in computer-mediated instruction?
					9.	What was the participants' level of cognitive knowledge of dermatological office procedures at the beginning of the program?
		Reaction (Satisfaction)	Courseware Evaluation Survey Interviews	Participants	10.	How satisfied were the participants with the subject matter, the instructional courseware, computer conferencing and the overall use of computer-mediated CME instruction at a distance?
		Behavioral (Performance)	Retrospective Pretest-Posttest Performance Survey	Participants	11.	How did the participants rate the extent of knowledge gain or performance improvement they experienced as a result of participating in a CME computer-mediated courseware program?
		Instructional Transactions	Content Analysis	Courseware Product	12.	How well did asynchronous computer conferencing function in facilitating collaborative learning and establishing a learning network?

Figure 3.1 The CME Courseware Evaluation Model



procedures for collecting data to inform and improve instructional design and development strategies. It is during this stage that instructional design and formative evaluation concur to improve design and development strategies. The formation of a CME Planning Committee to serve as product testers and content reviewers is an integral part of the evaluation model. This committee serves an important role during the initial two phases of formative evaluation. Committee members should include one to two subject matter consultants, and two to three family physician representatives. These members may or may not have any prior knowledge or experience with the development or use of courseware instruction. The formative evaluation process is led by the practitioner-as-evaluator.

During the analysis stage of formative evaluation the learning needs of the target audience are reviewed by the CME committee members. These needs will usually have been identified through a needs assessment strategy. These learning needs identify the discrepancies in physicians' knowledge, skills, or attitudes which may be addressed using courseware instruction. CME Planning Committee members serve an important function in interpreting these needs and assisting the evaluator-as-practitioner in writing performance and learning objectives which address the learning needs of the target population. The performance and learning objectives enable the evaluator-as-practitioner and the subject matter experts to identify a proposed content scheme which addresses the objectives and audiences' needs.

The content is then transformed into paper-based instructional strategies. Paper-based instructional strategies provide an overview of the proposed courseware performance and learning objectives, elaborations of content presentation flows, and the recommended media to present the courseware content to the learners. In the CME Courseware Evaluation Model these outlines are reviewed and critiqued by the subject matter experts and the CME Planning Committee members for content accuracy, relevance, comprehensiveness and bias. One to three CME Planning Committee meetings may be needed to review the paper-based outlines and to gather feedback from CME Planning Committee members. These meetings are facilitated by the evaluator-as-practitioner as small-group working sessions in which discussion of the outline and elaboration on the part of the subject matter experts will be warranted.

The paper-based outline meetings are followed by one to two storyboard evaluation meetings. These meetings are also held during the formative evaluation stage. Paper-based storyboards are developed by the evaluator-as-practitioner, in collaboration with the multimedia developers, and they provide an overview of the courseware's screen layouts, interface elements, and content presentation strategies. Presentation software may also be used at this stage to present the storyboards in an electronic format.

The storyboards serve as the preliminary draft of instruction and can be tried out with the CME Planning Committee members. The storyboard analysis includes a walk-through of the storyboards, the actions which are to occur, and an overview of navigation and courseware interface features. The CME Planning Committee should be utilized as much as possible at this phase, for trying-out the storyboards and for providing critical feedback on the features of the draft courseware. A small-group working session will usually work best at this phase. If the storyboards are created in a presentation software program this can serve to closely simulate the actual finished courseware product. Feedback from committee members should focus specifically on clarity of information, instructional procedures and strategies, logic, relevance, ease of use, and bias. The data collected by this procedure may result in significant revisions to content and instructional strategies.

The storyboard sessions are followed by a period of prototype development and then a prototype evaluation. A prototype is a completed segment of courseware instruction, possibly a single lesson, using the courseware delivery platform (World Wide Web -- HTML in this study). The prototype evaluation procedure is implemented by selecting a learner from the target audience and conducting a one-on-one learner try-out and interview. The learner is instructed to try-out the prototype lesson in its computer-based format. During the one-on-one interview, the learner works his or her way through the lesson using a talk-out-loud procedure. The talk-out-loud procedure allows the learner to identify and express confusion and difficulty where they are experienced. As well, the evaluator asks questions about the screens and the courseware:

- Did you understand how to navigate through the instruction?
- Was the instruction confusing? How?
- Were you able to make choices to view information that was relevant to you?
- Were you able to find your way back to previous screens?

Following the prototype evaluation phase and any changes which may have been made to the courseware, the entire courseware product can be developed with the assurance that several of the design flaws should have already been discovered. The Courseware Development Team and the CME Planning Committee are two separate working groups throughout the process. However, feedback and information sharing regarding what is working and what is not, what can be used or done to make things work is continuously distributed between and among members of both groups through the evaluator-as-practitioner. This division, delegation, and sharing of responsibility along planning, evaluation, and product development tasks ensures that the process does not become embroiled by individual differences and enables an efficient work flow.

The formative evaluation stage is followed by the second stage of the CME Courseware Evaluation model, the summative evaluation. The summative evaluation stage, like formative evaluation, may also reveal any further problems that still exist with the organization and presentation of instruction. This stage may also reveal several issues related to interactivity, learner control, screen layout and design. As well, this evaluative stage will enable the evaluator to assess the instructional outcomes of the courseware in terms of learning achievement and performance improvement.

In many ways this stage of evaluation resembles the field trial phase of Dick and Carey's (1990) instructional software evaluation model. Two main types of information should be collected at this phase, organizational and outcome data. The organizational data can be collected using a courseware evaluation instrument, such as a participant satisfaction survey, and interviews with a sample of participants. The survey method allows for the collection of quantitative data on learners' perceptions of the instructional, technical, and aesthetic adequacies of the courseware. Key questions should also be asked of the user regarding motivation, interactivity, learner control, user interface, and screen layout and design.

Interviews are an important part of this evaluation stage and are conducted shortly after the courseware has been completed by participants. The interview procedure is important because it allows evaluators to delve into the individual experiences of learners' interactions with the courseware in their particular settings and it serves to confirm and elaborate on the findings from the quantitative participant satisfaction surveys. The interviews provide qualitative data which enable the evaluator to achieve a deeper understanding of the factors which impede or facilitate effective learning during normal instruction. This detailed investigation of what works and what doesn't work with individual students is a logical step in courseware evaluation. The interviews are effective tools for ascertaining the truth about the courseware and incorporating that truth into future courseware design.

Interview questions should focus on a number of aspects of effective CME courseware instruction. Were the learning materials and instructional tasks presented clearly? How effective was the content of instruction? Was the media clear? Were the navigation and interface features easy to use? Interviews can also be used to find out about the unique circumstances of the learners, including their time limitations and constraints. Were they able to complete the courseware and the instructional requirements in the time allotted? Did the courseware meet their continuing education needs? How could the various courseware components be made more understandable?

Pre and post-learning achievement tests are important phases of the summative evaluation stage as well. These tests enable the evaluator to collect outcome data and determine the extent of learning which results from the instructional courseware. The pre and posttest phase may be implemented using any of the experimental or quasi-experimental research designs discussed in the research design literature. Several study designs which are appropriate for instructional courseware evaluation research have been presented in Chapter Two. As discussed, the execution of any particular design is dependant on the resources afforded the evaluator, the type of courseware or instructional software being evaluated, the population of learners, the purpose of the evaluation study, and the conditions under which the study is being conducted. The evaluation design which was implemented in the fieldtest is described later in this chapter.

A content analysis of the discussion which occurred during the computer conference is also performed at the summative evaluation stage. A unique advantage of computer conferencing systems is they automatically provide an electronic transcript of discourse interchanges which can be analyzed using qualitative and quantitative methods. A thorough analysis of these interaction patterns enables an assessment of the types and levels of the many-to-many interactions which occur. There are two objectives in performing an analysis of computer conferencing discussion. First, to determine how effectively the asynchronous computer conferencing functions as an instructional component of the courseware system. And second, the content analysis should include an assessment of the types of engagement, interactions, and participation which occurs in the online discussion. In summary, is computer conferencing being used as an effective collaborative learning tool?

A final phase of the summative evaluation stage includes the distribution of a retrospective pretest-posttest self-reporting performance survey. This survey is intended to collect information on learners' perceived changes in clinical performance as a result of participation in the courseware instruction. The self-reporting survey includes a list of the ascribed courseware performance objectives and learners are asked to rate, in a retrospective manner, their perceived change in knowledge, skills, and attitudes before and after their completion of the courseware. This survey should be conducted 6-8 weeks after the completion of the courseware.

The results of the summative evaluation stage are reported back to the CME Planning Committee and the Courseware Development Team. The report should document the areas where improvements to the instructional design of the courseware may be implemented. The report also enables an interpretation of the effectiveness of the courseware in producing cognitive (learning) gain and affecting the self-reported clinical performance behaviors of the learners. As well, the participant satisfaction survey provides valuable feedback to the Courseware Development Team regarding areas where improvements may be needed.

The CME Courseware Evaluation Model is an integrative framework because it occurs concurrently during the analysis, design, and development of courseware programs. The initial phases of formative evaluation ensure than major design flaws or inconsistencies are addressed, and the presentation of content is appropriate and addresses the learning needs of the target population. The model also ensures cost-efficiency and cost-savings by introducing modifications and changes during the development phases of the instructional courseware, before substantial development resources have been devoted to production.

A main feature of the model relates to it being utilization-oriented. It is concerned with the development and improvement of the courseware under evaluation. It does not serve to compare courseware with other media or other delivery mediums. This philosophy appears to be in sharp contrast with previous evaluation research efforts in the instructional technology field. According to Lee et al. (1997) research on computer-aided distance learning has been extensive, though not all of it particularly helpful. Some researchers have fallen into the trap that wasted much of the efforts of earlier studies which simply compared outcomes of learning from different media.

Spencer (1991) suggests that much of the available research on instructional technologies has dealt with comparisons between a particular medium (or method/approach) and the traditional way of teaching the subject. For many of these research comparisons this consists of the new medium being used to deliver most of the course, and then this is compared with the usual way of teaching the material, which may be chalk and talk or may include alternative media (e.g., teacher plus overhead projector).

Dillon and Gunawardena (1992) advocate evaluation research which focuses on the improvement and development of quality instructional products. They suggest that most of the evaluation research literature in distance education is descriptive or prescriptive, rather than research based upon systematic inquiry. In a majority of these studies evaluation research is used to compare a new distance delivery medium with the traditional face-to-face way of teaching, or it compares the effectiveness of one distance delivery medium over another. A majority of these studies have found no significant differences in the level of learning or achievement between the delivery mediums. According to Dillon and Gunawardena this is not surprising since what is generally occurring in these studies is a comparison of media pertaining to overall instructional impact.

Clark (1983) offers a typical commentary on this state of affairs:

> "The best current evidence is that media are mere vehicles that deliver instruction but do not influence student achievement any more than the truck that delivers our groceries causes changes in our nutrition. Basically, the choice of vehicle might influence the cost or extent of distributing instruction, but only the content of the vehicle can influence achievement." (Clark, 1983, p.449)

Clark (1983) is an advocate of evaluation research which relies upon the naturalistic methodologies and considers both the context in which the findings occur and the ability of the learners to reflect upon their experiences. According to Clark (1983) much of the existing literature, particularly comparison studies, is vulnerable to a variety of uncontrolled effects, including varying teaching methods, instructor characteristics, and instructional tasks.

Courseware Program and Subjects

The courseware delivery components have been discussed in Chapter One. In summary, they entailed the use of a hybrid delivery system which integrated CD-ROM, distributed multimedia materials in hypertext markup language (HTML), asynchronous computer conferencing, and electronic testing and survey applications (CGI forms). The courseware program evaluated in this study was accredited by the Office of Professional

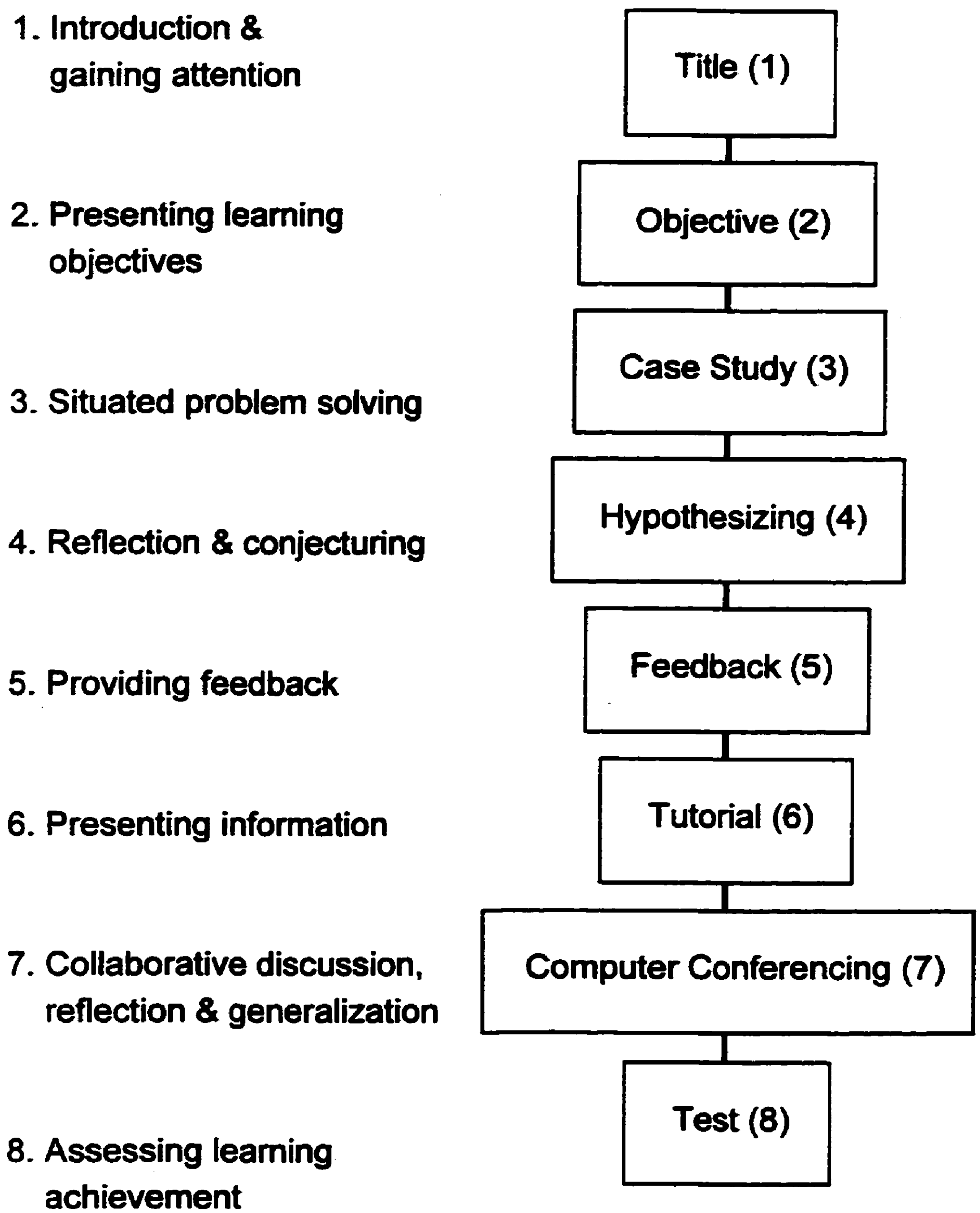
Development, Faculty of Medicine for two MAINPRO M-1 credits and was delivered over a two-week period during May of 1998.

Funding for the development of the courseware product was received from the pharmaceutical industry (Stieffel Incorporated and ICN Canada), the Canada/Newfoundland Cooperative Human Resources Development Agreement, and CANARIE Incorporated. The development and evaluation of the courseware product served as a research and development project to assess the utility of computer-based hybrid technologies for delivering continuing medical education at a distance to rural and remote physicians. The main goal of the courseware program was to introduce family physicians to common dermatological office procedures. Instructional objectives for the courseware are presented in the appendices.

A presentation flow template is a content-free framework and for the program developers designing the courseware evaluated in this study it was a convenient and reusable model for effectively and efficiently designing instructional courseware. A template is a skeleton without real content, which enables subject matter consultants to formulate and develop content to make deliverable courseware. Based on the theories and principles of micro and macro-design of hypermedia courseware (Yang & Moore, 1995) a prototype presentation flow strategy was developed for delivering multimedia enhanced, case-based CME instruction in a hybrid learning environment. The presentation template used in this courseware is presented in Figure 3.2.

According to Yang and Moore (1995) information format is an important category of effective micro-design in hypermedia learning environments. Micro-design is

Figure 3.2 Computer-Mediated CME Courseware Presentation Flow with Instructional Events



concerned with the creation of discrete information screens and is contrasted with macro-design which is concerned with the connections between them. Three general guidelines are suggested for designing information formats:

> "1) Graphics are more appropriate for representing concrete information; 2) multiple information formats are more effective and adaptive than single formats - multichannel, dual coding, and multiple intelligence theories all hold that information is communicated more successfully when it is presented in multiple formats; and 3) information must be consistent in different presentation formats." (Yang and Moore, 1995, p. 9)

Another important category of micro-design deals with the operation of instructional courseware. According to Yang and Moore (1995) operating directions should be clear and specific, and there should be "consistency of operation" (p.11). Courseware operating directions should be placed in the same screen area and the same operation should initiate the same function all the time. Screen layouts are an important feature as well. Screens should be designed to facilitate better learning, rather than impose further learning in how to navigate and use instructional courseware. Functional screen areas should be defined. The screen should be divided into several functional areas with each having a fixed layout and location on the screen. This standardized format provides extra structural information and facilitates learning the content rather than the unfamiliar.

Macro-design, on the other hand, is concerned with linkages between micro-design screen information nodes. According to Yang and Moore (1995) macro-design can be addressed from two perspectives: content organization and presentation flow. Content organization deals with the interrelationships of the connections between discrete information nodes. The presentation flow addresses the interrelationship on the instructional process. Content should be broken down into manageable information nodes which can be arranged according to an organizational structure.

There are three basic organizational structures: unstructured, hierarchical and network (Yang and Moore, 1995). Unstructured knowledge bases store all information fragments randomly. Hierarchical structures store information nodes in a fixed tree-like sequence. Strict hierarchical systems only allow learners to go one-step at a time. In a network knowledge base, an information node is linked to others based on the relationships between them. A mix between hierarchical and network structures was chosen for the design of this courseware.

As discussed in Chapter One the courseware was developed in a hypermedia learning environment. Hypermedia is characterized by its non-linear association of information. Learners are able to decide where to go, what to see, when to stop, and how to organize information. However, many learners can be overwhelmed and even find themselves lost in navigation in a hypermedia learning environment. Therefore, to achieve better results, presentation flows limit the learner control of navigation in hyperspace.

As demonstrated in Figure 3.2 the presentation flow began with a title screen which served to introduce the subject matter consultants, present an overview of the topic and the learning objectives for the CME courseware, and gain learners' attention. The objectives were clearly written to state the expected learning outcomes for the courseware. This page was accessible at any time from any other screen in the courseware.

Following this introductory screen participants were able to navigate to a case study screen and select a case from a case study content menu graphic. The individual case study screens presented information on the virtual patient, including graphical representation of the lesions to be investigated or managed. Upon review of the case, learners were queried (hypothesizing) on the correct management or investigative procedure for the particular case and were required to select the correct procedure from a multiple-choice list. Feedback to the physician was provided for correct and incorrect selections. Physicians were linked to a tutorial screen upon selection of the correct procedure. The tutorial presented information on the particular procedure in multiple communication channels (video, audio, text, graphics). These multimedia tutorials presented instruction on the correct processes for performing and applying the procedure in the management of appropriate lesions.

When learners were satisfied with their review of the tutorials they were free to return to study other case studies, link to the computer conference, or they could take the post-learning achievement test. The electronic posttest evaluated learning achievement and provided a summary of test scores and feedback to incorrect answers. A dermatological links page enabled the learners to link to other Web sites for further study and information. Navigation buttons common to all screens of the Web site included Instructions, Site Index, Office Procedures Index, and Related Dermatological Sites. These screens provided information related to the navigation features, content map and path histories, and links to the main title page.

Fifty-two subjects participated in this evaluation research study. Thirty volunteer physicians were recruited to participate in one of two computer-mediated CME experimental study groups. Recruitment for these computer-mediated CME treatment groups was criteria-based. Participants were expected to have access to a multimedia computer with CD-ROM capabilities and Internet access. Twenty-one (70 percent) of the treatment group subjects were recruited from the population of family physicians practicing in the province of Newfoundland and Labrador between January to February, 1998. These physicians were recruited through a mail-out letter to the general family physician population of Newfoundland and Labrador. Nine (30 percent) of the subjects in the treatment study groups were recruited from other Canadian provinces through an electronic mail posting on the Canadian Society of Rural Physicians list-server "RuralMed Electronic Mailing List" in February, 1998. A list-serve is a system for the electronic dissemination of all messages sent to it. This means that all messages sent to the list are automatically distributed to all the subscribers to the list, via electronic mail. The purpose of the "RuralMed Electronic Mailing List" is to support Canadian and non-Canadian rural practitioners, whether they be rural family physicians or rural specialists, by providing a forum for discussion, debate, and the exchange of ideas and information.

A print-based and a Web-based recruitment letter and recruitment submission form were developed. The recruitment submission form collected information on the computer type, browser specifications, and modem type of the subjects. These data provided an important summary of the hardware and software specifications of the participants, and were used by developers during the courseware design and production stages. All participants in the experimental study groups were required to review and sign an informed-consent agreement, acknowledging their informed and voluntary consent to participate in this study, and giving the investigator permission to use Computer Attitude Scale scores, learning achievement test scores, Course Evaluation Survey information, and computer conferencing transcript materials for evaluation research purposes.

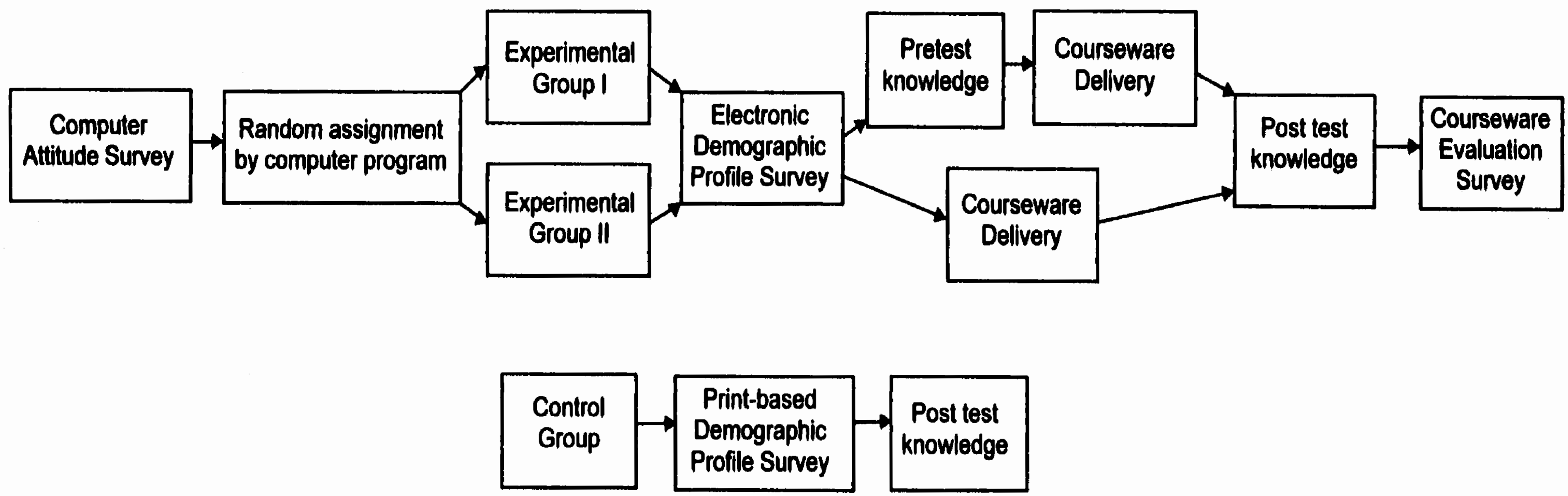
Twenty-two volunteer subjects were recruited to participate as a no-CME control group. These subjects were all Newfoundland and Labrador physicians and were recruited from a face-to-face continuing medical education workshop held in a rural area of the province of Newfoundland and Labrador in May, 1998.

Modified Pretest-Posttest Control Group Experimental Research Design

A modified pretest-posttest control group experimental research design was conceptualized and used in measuring the level of learning achievement resulting from participation or non-participation in this computer-mediated CME courseware. This modified experimental design combined the one-group pretest-posttest and the static-group comparison research designs. Figure 3.3 presents the modified research design. According to the design, subjects in experimental study group I underwent the one-group pretest-posttest design. Subjects in this group participated in the computer-mediated courseware and completed pre- and post-learning achievement tests.

The one-group pretest-posttest design presents a number of difficulties in interpretation, including possible internal validity threats due to maturation, history,

Figure 3.3 **Modified Pretest Posttest Control Group Experimental Research Design**



regression, instrumentation, selection, mortality, and interaction. In employing such a design, alternative explanations of the posttest results need to be minimized through logical reasoning at best. The interval between pre and posttesting was short, due mainly to the amount of subject matter covered in the courseware and the established schedule for the computer conferencing discussion. This short duration and the constant group size minimized several internal validity threats, namely maturation and mortality. The influence of history was also minimized by combining the static-group comparison method with the one-group pretest posttest design. The static-group comparison design controls for history threats since it is assumed that events occurring outside of the experimental setting will equally affect all groups similarly (Gay, 1996). Additionally, instrumentation as an internal validity threat was minimized through the use of identical items for the pre and post-learning achievement tests.

Experimental study group II and the control group comprised the static-group comparison design. In this static-group comparison design, the experimental study group participated in the computer-mediated courseware, while the control group received no instruction in the subject matter. Both groups were posttested. A significant problem with the static-group comparison design is equivalency between groups. The degree to which the groups are equivalent is the extent to which their comparison is reasonable. Tables 3.3 to 3.5 provide a summary of the characteristics of the members of each study group along the variables of practice experience, practice type, practice location, physician type, computer usage, computer ownership, computer access, computer experience, and Internet use. Experimental groups I and II may be compared along all of these aforementioned variables, whereas experimental study groups I and II and control group subjects may only be compared along the variables of practice experience, physician type, practice location, and gender. Computer characteristic information was not collected for control group subjects.

Table 3.3 presents the demographic characteristic data for subjects in each of the study groups. A majority of the physicians participating in the experimental and control study groups were practitioners with 11 years or greater practice experience (74.5 percent), males (80.2 percent), rural physicians (69.2 percent), and practitioners in a group setting (73.1 percent). Chi square analysis of the demographic variables of the study groups indicated no significant demographic differences between study groups at the <.10 level of probability.

Table 3.4 presents the computer usage, computer access, and computer experience characteristic data for subjects in the experimental groups. This information was collected using an electronic Demographic Profile Survey (DPS). A majority of participants in the experimental groups used computers for less than 6 hours per week (69.0 percent), and most reported having access to a computer at "home and work" (60.0 percent). A Chi square test of the variable "weekly computer usage" revealed no significant differences at the < .10 probability level.

The Chi Square test of "computer access" revealed a significant difference between the experimental groups at the < .10 level of probability. Experimental group II was comprised of a greater number of physicians with access to a computer at home. However, results of Kruskal Wallis tests of the relationships between individual physician

Table 3.3 Demographic Characteristics of Subjects by Study Group

Variable	Experimental Group I	Experimental Group II	Control Group	Total
Practice Experience				
N (Total)	16	13	22	51
10 years or less	6 (37.5%)	3 (23.1%)	4 (18.2%)	13 (25.5%)
11 years or greater	10 (62.5%)	10 (76.9%)	18 (81.8%)	38 (74.5%)
Chi Square $x^2 = 1.874$ df=2 Sig. = .392				
Gender				
N (Total)	16	14	22	52
female	2 (12.5%)	4 (28.6%)	4 (18.2%)	10 (19.2%)
male	14 (87.5%)	10 (71.4%)	18 (81.8%)	42 (80.8%)
Chi Square $x^2 = 1.269$ df=2 Sig. = .530				
Physician Type				
N (Total)	16	14	22	52
rural	11 (68.8%)	10 (71.4%)	15 (68.2%)	36 (69.2%)
urban	5 (31.3%)	4 (28.6%)	7 (31.8%)	16 (30.8%)
Chi Square $x^2 = .045$ df=2 sig. = .978				
Practice Type				
N (Total)	16	14	22	52
solo	2 (12.5%)	4 (28.6%)	1 (4.5%)	7 (13.5%)
group	12 (75.0%)	7 (50.0%)	19 (86.4%)	38 (73.1%)
neither	2 (12.5%)	3 (21.4%)	2 (9.1%)	7 (13.5%)
Chi Square $x^2 = 6.228$ df=4 sig. = .183				

Table 3.4 Computer Usage, Computer Access and Computer Experience Characteristics of Subjects by Study Group

Variable	Experimental Group I	Experimental Group II	Total
Weekly Computer Usage			
N (Total)	15	14	29
0 - 6 hours	10 (66.7%)	10 (71.4%)	20 (69.0%)
7 or more hours	5 (33.3%)	4 (28.6%)	9 (31.0%)
Chi Square $x^2 = .077$ df = 1 sig. = .782			
Computer Access			
N (Total)	16	14	30
home	5 (31.3%)	3 (21.4%)	8 (26.7%)
work	0.0	4 (28.6%)	4 (13.3%)
home and work	11 (68.8%)	7 (50.0%)	18 (60.0%)
Chi Square $x^2 = 5.279$ df = 2 sig. = .071			
Computer Experience			
N (Total)	15	14	29
a little - none	2 (13.3%)	4 (28.6%)	6 (20.7%)
moderate	8 (53.3%)	7 (50.0%)	15 (51.7%)
considerable - extensive	5 (33.3%)	3 (21.4%)	8 (27.6%)
Chi Square $x^2 = 1.200$ df = 2 sig. = .549			

computer characteristics (specifically computer access) and posttest learning achievement scores (see Chapter Four) revealed an insignificant relationship at the < .10 probability level (df = 2, p = .126). However, the results do suggest that a marginally positive relationship exists and this variable warrants further exploration in the future.

Table 3.4 also presents results for the variable "prior computer experience." Information on this individual characteristic was collected by asking physicians to rate their level of commuter experience along a Likert scale ranging from "a little experience" to "extensive experience." A small majority of participants in the experimental groups indicated that they had moderate computer experience (51.7 percent). Chi square analysis indicated no significant differences between experimental study groups I and II on computer experience ($x^2 = 1.200$, p = .549).

Table 3.5 presents results for participants' self-reported usage of computer software and Internet applications. Most experimental group subjects used software "regularly" or "often to very often" (37.9 percent and 34.5 percent respectively). Many of the physicians were also "regular" users of electronic mail (43.3 percent), and most used the Internet "regularly" or "often to very often" (36.7 percent and 23.3 percent respectively). Chi square tests indicated no significant differences between experimental study groups along these computer experience variables.

It is also possible to compare group equivalency between experimental and control group subjects according to pre-learning knowledge of the subject matter. As mentioned, identical twenty-item multiple-choice pretest and posttests were used in this study. The modified pretest-posttest control group experimental research design

Table 3.5 Computer Software, E-mail, and Internet Usage Characteristics of Subjects by Study Group

Variable	Experimental Group I	Experimental Group II	Total
Frequency of Software Usage N (Total)	15	14	29
not at all - occasionally	4 (26.7%)	4 (28.6%)	8 (27.6%)
regularly	4 (26.7%)	7 (50.0%)	11 (37.9%)
often - very often	7 (46.7%)	3 (21.4%)	10 (34.5%)
Chi Square $x^2 = 2.387$ df = 2 sig. = .303			
Frequency of E-mail Usage N (Total)	16	14	30
not at all - occasionally	5 (31.3%)	3 (21.4%)	8 (26.7%)
regularly	5 (31.3%)	8 (57.1%)	13 (43.3%)
often - very often	6 (37.5%)	3 (21.4%)	9 (30.0%)
Chi Square $x^2 = 2.068$ df = 2 sig. = .356			
Frequency of Internet Usage N (Total)	16	14	30
not at all - occasionally	5 (31.3%)	7 (50.0%)	12 (40.0%)
regularly	5 (31.3%)	6 (42.9%)	11 (36.7%)
often - very often	6 (37.5%)	1 (7.1%)	7 (23.3%)
Chi Square $x^2 = 3.880$ df = 2 sig. = .144			

conceptualized for this study included the pretesting of only one experimental study group, experimental study group I. Subjects in the control group were administered a posttest only, as called for by the static group research design. According to the static group research design, an experimental and control group are posttested only after a treatment has been delivered. However, in the modified pretest posttest control group design it was possible to substitute the posttest scores of the control group as pretest scores and to then compare these scores with the pretest scores of experimental group I.

The posttest of the control group is a posttest in a "semantic" sense only. Realistically, the control group was only tested once, and whether it is termed a pretest or posttest is irrelevant as these subjects were not presented any subject matter prior to, or during their completion of this test. Therefore, this "test" could be considered, alternatively, as a test of knowledge of the subject matter at the time the test was administered. While the test is termed a posttest in the static group design, the test scores may serve as a baseline for comparing control group subjects' knowledge of the subject matter with the pretest scores of experimental group I.

A Mann-Whitney test was used to assess the difference between the mean pretest scores of experimental study group I and the substituted mean pretest scores of the control group. These results are presented in Table 3.6. The mean pretest score of subjects in experimental group I was 10.44, while the mean substituted pretest scores for control group subjects was 9.68. The mean ranked score for experimental group I was 21.50 and 18.05 for the control group. A Mann-Whitney test of pretest scores indicated no significant difference at the $< .10$ probability level ($p = .337$). The results indicate no significant differences in pre-learning knowledge of the subject matter between experimental and control groups. These findings suggest that subjects in the control and experimental groups had similar knowledge of the subject matter before instruction occurred.

Table 3.6 Mann-Whitney Test of Pretest Scores by Experimental Study Group I and Control Group

	N	Mean	SD	Mean Rank	Sum of Ranks	Z	Sig.
experimental I	16	10.44	1.31	21.50	344.00	-.959	.337
control	22	9.68	2.51	18.05	397.00		
Total	38	10.00	2.11				

Many of the experimental designs documented in the literature are applicable to program evaluation study. However, because evaluation research often occurs in real life contexts, and not in a controlled laboratory setting, one is less likely to have much control over certain extraneous variables. As well, in evaluation studies of distance education, it is often difficult to achieve true randomization which is characteristic of the rigorous classical experimental designs. In evaluation research and distance education settings, the random selection required by classical experimental design is often impossible due to the characteristics of the study context, the subjects, and the necessity of using intact groups (Babbie, 1992; Dillon and Gunawardena, 1992).

In order to minimize possible logistical difficulties, a modified pretest-posttest control group experimental research design was used. Physicians are very busy professionals and have many responsibilities to the care of their patients and communities. The fieldtest of the evaluation model was designed to assess the effectiveness of the computer-mediated courseware as it occurred in the natural setting of the participants. Physicians were expected to participate in the computer-mediated courseware from their offices or homes, during their normal practice hours and personal time. And, while this flexibility is viewed as a significant advantage of computer-mediated learning, it does present obstacles and challenges in implementing a successful study. Therefore, potential challenges and problems apparent in the implementation of traditional experimental research designs had to be minimized. This was necessary to ensure annoyance with the evaluation process would not influence mortality and so the external validity threat of reactive arrangements could be reduced.

Instruments

1. Computer Attitude Survey (CAS)

The computer attitudes of physicians in the experimental study groups were measured using the Computer Attitude Scale (CAS) developed by Nickell and Pinto (1986). Nickell and Pinto (1986) developed the Computer Attitude Scale as an instrument for measuring the positive and negative computer attitudes of people across a variety of settings. The Computer Attitude Scale consists of 20 five-point Likert items (eight expressing positive attitudes towards computers and 12 items expressing negative attitudes). Possible scores on the CAS range from a minimum score of 20 (indicating an extremely negative attitude toward computers) to a maximum score of 100 (indicating an extremely positive attitude toward computers).

Nickell and Pinto (1986) tested the CAS survey on a variety of sample populations including secondary students, post-secondary students, and professionals. Analysis of the CAS data collected from these sample groups indicated an internal consistency (coefficient alpha) reliability coefficient of .81, while the test-retest data yielded a statistically significant, positive correlation ($r (45) = .86, p<.001$). The authors also measured the short-term predictive validity of the instrument by correlating CAS scores collected on a sample of college students beginning an introductory computer class with their overall final course grade. The result was a statistically significant, positive correlation ($r (80) = .32, p < .01$). Construct validity was also compared by correlating CAS with COMPAS, a Computer Anxiety Scale. The result was a significant, negative correlation ($r (47) = -.61, p < .01$) for the total COMPAS scale. Overall, the authors suggested that the CAS appeared to have high reliability and validity.

A majority of the attitudinal instruments reviewed for use in this study were designed for either specific populations (secondary students) or specific intents (the measurement of anxiety). In evaluating the psychometric properties of the CAS instrument, Nickell and Pinto had tested the Computer Attitude Scale with several different samples: college students from medium sized state universities and small liberal arts colleges, and computer operators for a large organization. It also appeared to have greater content validity than most other instruments. In the current study, responses to individual items on the Computer Attitude Scale were analyzed and the CAS was found to have a Cronbach Reliability Coefficient of .7353. Thirty physician CAS scores were included in the sample.

2. Demographic Profile Survey (DPS)

An electronic Common Gateway Interface (CGI) submission form and a print-based survey were developed for collecting demographic information, computer experience, computer access, and computer usage profiles of subjects. Subjects in the experimental groups were asked to complete an electronic Demographic Profile Survey, developed as a CGI form submission. CGI scripts are small computer programs, usually written in specialized scripting language, which are recognized by WWW servers and permit a "bi-directional" transfer of information, rather than simply sending a HTML page to a browser at a receiving end. The 10 items on the electronic DPS gathered information on demographic variables: years of practice, gender, physician type, and practice location; and, computer usage characteristics such as computer experience, weekly computer usage, computer ownership, frequency of software use, electronic mail use, and Internet usage. Subjects in the control group completed a 4 item print-based DPS. Information was collected on demographic areas only: years of practice, gender, physician type, and practice location.

3. Pretest and Posttest (Learning Achievement)

Identical 20-item multiple choice pre and post-learning achievement tests were developed by the clinical subject matter experts, with assistance from an instructional design specialist. The 20 test items were developed to measure physicians' knowledge of the subject matter before and after participation in the courseware. Test questions were based upon learning objectives which were developed by a CME Planning

Committee including the Director of CME, two rural physician representatives, two clinical subject matter experts, and an instructional design specialist. Questions on the achievement tests were reviewed and edited by planning committee members for accuracy and specificity. Two formats of the posttest were developed, an electronic CGI-based posttest (for experimental study groups) and a print-based posttest (for control group subjects).

4. Courseware Evaluation Survey (CES)

The Courseware Evaluation Survey (CES) was a quantitative, electronic CGI-based survey instrument. It was designed to collect information on participants' perceptions of the quality and effectiveness of the computer-mediated CME courseware and the computer conferencing system. The survey included 21 items, distributed along five evaluative categories. These categories included: content, graphics and media, navigation and organization, learner manual, computer conferencing, and overall impressions. Each category included one to five evaluative statements which participants were asked to rate along a five point Likert scale from 1 = "Strongly Disagree" to 5 = "Strongly Agree." All statements were positively worded.

The categories and the associated statements were developed following an extensive review of related research on the area of interactive multimedia and instructional software evaluation and evaluative instruments. The following instruments and studies were referenced for developing the evaluative categories and statements of the CES.

Spille et al. (1985) "Courseware Evaluation Checklist" was designed for the American Council on Education. The checklist is used by course and program providers for assessing the effectiveness of instructional software and interactive multimedia programs. The checklist includes questions which pertain to the dimensions of instruction directly related to the needs of adult learners. The checklist applied to adult learning courseware contains evaluative questions in 6 broad categories:

- Instruction Sheets (printed and/or on-line instruction manuals for using the courseware - a learner's guide)
- Sequencing (are instructional topics introduced in logical order - built-in flexibility)
- Roles of Instructors and Tutor-Mentors
- Feedback (extent and scope of mechanical feedback)
- User-Control (embedded flexibility for self-directed learning)
- Flexibility (user-friendly, opportunity for collaborative learning)

Another instrument, Weber's (1992) Survey of Interactive Communication Technology (ICT) was referenced as well. The items of this survey were based on a thorough search of the background literature on interactive, computer-based technologies and systems. The survey was tested against experts' perceptions of factors affecting the development and use of interactive technologies and the authors suggested that it has proven useful in evaluating the effectiveness of interactive technologies.

Wilkinson et al. (1997) developed an extensive categorical list of criterion indicators for assessing the quality of World Wide Web educational and information sites based on library reference and science literature. Eleven categories were identified and categorical indicators were rated by a panel of 30 authorities on Internet resources for relevancy. The results were used to develop a set of evaluative instruments for use by teachers, students, and others to evaluate resources that are located on the Internet.

Tolhurst (1992) also presented a variety of principles for evaluating instructional software on various aspects of instructional quality and effectiveness. The evaluation criteria include: implementation considerations (hardware requirements); documentation and packaging (operating instructions and secondary instructional materials); curriculum considerations (subject matter); and user interface (user control and navigation).

Several other instruments and evaluative criteria taxonomies were also consulted in developing the course evaluation instrument for this study including: Smith (1998), Grassin (1997), and Duke (1996). The courseware evaluation survey was reviewed for accuracy and relevancy by several post-secondary instructional design experts and a medical informatics specialist. As well, two drafts were pilot tested for readability and understanding by several physicians.

5. Interview Script

A semi-structured formal interview script was developed as a guide for eliciting information about participants' satisfaction, learning, and experiences in using the instructional courseware and computer conferencing system. An interview script was prepared by the evaluator and a random sample of rural and remote participants were interviewed by telephone. The interview questions were piloted on a family physician. The interview was semi-structured and the questions were developed to gather open-ended responses on the strengths and weaknesses of the various instructional components of the computer-mediated instructional system.

6. Post-Learning Performance Self-Assessment Survey

A retrospective pretest-posttest performance survey was developed to assess the affect of participation in the computer-mediated courseware on the self-reported knowledge and skills of participants. While the retrospective survey design has received little attention in the literature, it is useful and has been effective in previous distance education evaluation studies (Heinzen and Alberico, 1990). As well, Howard et al. (1979) have advocated the use of retrospective pretest-posttest designs in evaluation:

> "This design would simply improve a modification of Campbell and Stanley's Design 4 to include a retrospective pretest at the time of posttesting. This is accomplished by asking subjects to respond to each item on the self-report measure twice. First, they are to report how they perceive themselves at present (Post). Immediately after answering each item in this manner, they answer the same item again, this time in reference to how they now perceive themselves to have been just before the workshop was conducted (Retrospective Pre). Subjects are instructed to make the Retrospective Pre responses in relation to the corresponding Post response in order to insure that both responses are made from the same perspective. Each set of ratings is scored separately to yield a Post score and retrospective Pre score (pp. 130-131).

The impetus for using a survey instrument of this type arose, in principle, from the literature on the "Commitment for Change," particulary its emphasis and confidence in the validity of self-reports.

The commitment for change instrument has been used in evaluating the effectiveness of CME programming for many years. Its initial development grew out of frustration with the difficulties CME evaluators experienced in designing and executing evaluation studies to demonstrate the links between CME and changes in physicians' behaviors (Curry and Purkis, 1986). The detailed, systematic and rigorous evaluations of changes in physician behavior and patient outcomes proposed by some evaluators were often too expensive and time-consuming to use as routine evaluation for all CME courses.

In an attempt to overcome some of these obstacles, in the early 1980s the Division of CME at Dalhousie University developed an easily applied, inexpensive evaluation procedure called the "Commitment for Change" for CME short courses (Purkis, 1982). Purkis' preliminary trial of this instrument led to the conclusion that the procedure provided "indirect" but strong evidence that CME short courses changed physicians' behavior. The evidence was "indirect" because the procedure merely asked physicians attending a course to indicate which changes in patient management procedures they intended to make based on what they had learned in a CME program. Purkis performed a follow-up self-report survey with the physicians using the commitment for change instrument and the results indicated that 63 percent of the commitments were implemented and that an additional 27 percent were not implemented because of a lack of suitable cases.

The commitment for change instrument has, since that time, been piloted by several investigators as a means for promulgating and evaluating behavioral change (Pereles et al. 1997; Jones, 1990; Purkis, 1982; Mazmanian et al. 1997). These studies indicate that learners experience and report higher rates of successful change when commitments to change are secured, than when they are not. In one study, Crandall (1990) followed five physicians for six months after their participation in a 1-day continuing medical education conference on cardiac arrhythmia. Three physicians made commitments to nine changes. The results of Crandall's case studies suggested that physicians effect changes as a result of participating in traditional CME. Six of the nine changes physicians committed to make six months earlier were implemented.

Pereles et al. (1997) also studied whether physicians who committed to changes in clinical practice following a CME course were more likely to make changes in clinical practice than those not asked to make such a commitment. Twenty-six physicians registered for the short course under investigation. Seventeen agreed to participate in the study. They were randomly assigned to "commitment" and "no-commitment" groups. Sixteen were interviewed at 1 and 3 months after the course. Both groups made changes, with the commitment group making more.

The results of these studies were promising. However, long before the commitment for change procedure was considered a reasonably effective mechanism for CME evaluation and widely accepted, the validity of physicians' self-reports of change in their behavior needed to be established. Several authors had suggested that self-reports were a valid and effective means for evaluating performance change as a result of participation in a continuing medical education program (Abrahamson, 1984; Curry and Purkis, 1986; Mendenhall et al. 1978a, 1978b). Abrahamson (1984) has also suggested that the use of patient self-reports is a useful and proven technique for examining whether patients are being treated differently by physicians as a result of their having attended a continuing education program. However, it was not until Curry and Purkis (1986) conducted a rigorous study of the validity of self-reports as an evaluation method that reliable and irrefutable evidence for the instrument's utility was found.

The purpose of Curry and Purkis' (1986) study was to determine if the validity of self-reporting as a procedure was sufficiently valid to be recommended as a routine evaluation mechanism in CME courses. They sought to substantiate the validity of self-reports by comparing self-reported intentions of behavior change and self-reported behavior changes with physicians' actual behavior before and after a CME program. Sixty-one participants in a short CME course were followed for 6 weeks before and 16 weeks after participation. Curry and Purkis found no significant changes in prescribing behavior until the learners stated formal intentions to change. Carbon prescription pads were used to validate the findings. When the investigators' analysis included only those prescribing areas in which intentions to change were stated, there was strong evidence that the self-reports of behavior change accurately represented the actual behavior adjustment.

The utility of the retrospective pretest-posttest self-reporting performance survey developed for this study is based, for the most part, on this previous work. Self-reports of performance change are "indirect" measures, but the evidence suggested by previous investigations of the validity of the commitment for change and self-reports of practice-based behavior change do support self-reporting as a useful measurement. However, further research will be required to validate the methodology proposed in this evaluation as a reliable means for measuring performance change in the physician population.

The survey consisted of 13 performance statements which were based on performance objectives formulated and ascribed by the subject matter consultants and CME Planning Committee members during the initial phases of instructional planning for this courseware. These ascribed performance objectives formed the basis for the development of learning objectives and later, the instructional media and strategies used in the computer-mediated courseware. Subjects were asked to rate the level of their knowledge of, or skill in these performance areas using a five point Likert scale, 1 = "to no extent" and 5 = "to a large extent," both before and after their participation in the computer-mediated course. This survey was forwarded to participants 6 weeks after completing the courseware.

Data Analysis

The data collected in this study represented discrete nominal and ordinal measures, and interval and ratio measures. The sum computer attitude scale (CAS) scores represent ordinal measures because the expressed attitudes were scored along a Likert type scale. The data which was collected on the individual demographic and computer experience characteristic variables represent ratio, ordinal and nominal data. The data which was collected on the items, gender, computer ownership, physician type, and practice location represent nominal measures. The items, computer experience and computer software usage were measured at an ordinal level. Data collected on years of practice experience and weekly computer usage represent ratio measures. Where ordinal and ratio measures were employed a significant number of categories were used to decrease any distortion which could have been created by collapsing these categories into a smaller number. The ordinal and ratio categories for the variables, years of practice experience, number of hours of weekly computer usage, computer experience, and computer software and application usage were collapsed for data analysis.

The variables, pre and posttests of learning achievement represented a ratio level of measurement because there was a true zero score. The data for the courseware evaluation survey (CES) represented an ordinal level of measurement because items were ranked along a Likert type scale of "Strongly Agree" to "Strongly Disagree." The post-learning performance survey (PPS) collected retrospective data on the physicians' self-reported changes in clinical performance. This data represented an ordinal level of measurement as well. Participants were asked to rate their level of performance according to a list of performance statements for "before" and "after" their participation in the CME courseware program. Likert type scales were used for measuring this self-reported data.

Different tests of significance are appropriate for different sets of data. It is important that a researcher select an appropriate test as an incorrect test can lead to erroneous conclusions. The first decision in selecting an appropriate test of significance is whether parametric or nonparametric tests must be selected. According to Gay (1996) parametric tests are usually more powerful and generally to be preferred. By more "powerful" he is suggesting that parametric tests are more likely to reject a null hypothesis that is false; in other words, the researcher is less likely to commit a Type II error, less likely to not reject a null hypothesis that should be rejected.

Parametric tests, however, require that certain assumptions be met in order for them to be valid (Gay, 1996). A major assumption is that the data represents an interval or ratio scale of measurement. Another assumption is that subjects have been independently selected for a study. However, because this study used a volunteer sample of participants and a majority of the data types represented ordinal and nominal measures, nonparametric tests of significance were used. The following sections describe the various nonparametric tests used in this study.

The Kruskal-Wallis One-Way Analysis of Variance by Ranks was used as the nonparametric test for measuring the relationship and variance of scores. The Kruskal-Wallis takes into account the magnitude of each observation relative to the magnitude of every other observation (Daniel, 1995). The Kruskal-Wallis test is performed by combining each individual score from each sample into a single series of scores arranged in order of magnitude from smallest to largest. The observations are then replaced by ranks from 1, which is assigned to the smallest observation, to whatever the assigned value of the largest observation. When two or more observations have the same value, each observation is given the mean of the ranks for which it is tied. The ranks assigned to observations in each of the sample groups are added separately to the test equation to give group rank sums. The Kruskal-Wallis test statistic is then computed. This test for the analysis of variance by ranks was used for determining the significance of the relationships between the individual demographic and computer experience characteristic variables and computer attitude scale scores. These variables were represented by data which was of a nominal or ordinal type.

The Mann-Whitney was also used for data analysis. The Mann-Whitney is a nonparametric test which is based on the ranks of the observations. One major assumption underlying the Mann-Whitney is that the two samples selected for analysis have been independently and randomly drawn from their respective populations. As well, the measurement scale must be at least ordinal. The Mann-Whitney was used for analyzing the significance of the difference in learning achievement scores between the three study groups in this study. The first assumption, that the two samples selected for analysis have been randomly selected, is partly met. Subjects were randomly assigned to the respective experimental study groups. The learning achievement data was also represented by interval measures. The Mann-Whitney was computed by combining the two samples and ranking all observations from smallest to largest while keeping track of the sample to which each observation belonged (Daniel, 1995). Tied observations were assigned a rank equal to the mean of the rank positions for which they were tied. The test statistic was then computed.

The Spearman Rank Correlation Coefficient was used for measuring the correlation of computer attitude scores with the posttest learning achievement scores of physicians in this study. The Spearman Rank Correlation Coefficient makes use of two sets of ranks that may be assigned to the sample values of X and Y groups, the independent and continuous variables of a bivariate distribution (Daniel, 1995). The procedure involves ranking the values of the two samples from 1 to n (numbers of pairs of values of X and Y in the sample) and computing the r_s (correlation coefficient) value.

This test enabled an examination of the correlation between computer attitudes and post-learning achievement scores of physicians in a computer-mediated learning environment.

The Wilcoxon Signed-Rank test makes use of the magnitudes of the differences between measurements and a hypothesized location parameter rather than just the signs of score differences (Daniel, 1995). The Wilcoxon test was used for testing the difference between pre and posttest learning achievement scores of the computer-mediated experimental group. As well, it was used for testing the significance of the difference between self-reported "before" and "after" scores on the post-learning performance survey. In this survey, physicians were asked to report level of skill and knowledge before and after participation in the computer-mediated instructional program. A Likert type measurement scale was used and the two scores, before and after, were compared using the Wilcoxon test.

Chi-square, symbolized as X^2, is a nonparametric test of significance appropriate when the data are in the form of frequency counts (or percentages or proportions which can be converted to frequencies) occurring in two or more mutually exclusive categories (Gay, 1997). A chi-square test compares proportions actually observed in a study with proportions expected, to see if they are significantly different. Expected proportions are usually the frequencies which would be expected if the groups were equal. The chi-square value increases as the difference between observed and expected frequencies increases. The chi-square was used for testing the demographic and computer experience characteristic comparability of the experimental and control groups. Two limitations of this study were the use of volunteer samples and pre and quasi-experimental research designs. In using the chi-square test it was possible to determine the differences or similarities of the characteristics among groups. The more similar the groups (no or low significance) the greater their comparability.

When a researcher makes a decision to reject or not reject a hypothesis, she does so with a given probability of being correct. According to Gay (1997) this probability of being correct is referred to as the significance level, or probability level, of the test of significance. The level of significance selected determines how large a difference must be in order for significance to be declared. Gay (1997) reports that the most commonly used probability level is the < .05 level. However, he also reports that it is not uncommon for exploratory studies to use a probability level of < .10.

The significance level selected will determine the probability for committing either a Type I error, that is rejecting a null hypothesis when in fact it should be supported, or a Type II error, that is supporting a null hypothesis when in fact it should be rejected. The choice of probability level is in large part determined by the seriousness or the impact of committing a Type I error versus a Type II error. The current study was exploratory in nature because a new computer-mediated instructional program was being evaluated for its educational effectiveness. As well, the study also served to measure the relationships between various individual demographic and computer experience characteristics and computer attitudes. The seriousness of committing a Type I error was not great, because by labeling the instructional technology as an effective delivery means it would serve to stimulate further interest and research.

However, if a Type II error was committed and the technology was labeled as not being very promising then the promise for future research and investigation of the instructional technology would be minimized. If a Type I error was in fact committed in this study, the only real consequence would be that further research might disconfirm the findings of this study.

Further rationale for selecting a < .10 significance level was based on the use of the Computer Attitude Scale developed by Nickell and Pinto (1986). Nickell and Pinto originally tested the CAS survey on several large populations, much larger than the sample recruited for this study. Many of the significant relationships found in these studies were based on responses from hundreds of survey respondents. Given the difference in population size between the sample in the current study and the populations which Nickell and Pinto used for validating the CAS, the evaluator felt that a higher significance level would account for any statistical discrepancies caused by the limited number of subjects.

The interview data was analyzed using a typological analysis procedure described by LeCompte and Preissle (1993). This procedure involved the division of data into groups or categories on the basis of "theoretical frameworks, sets of propositions, or from common-sense mundane perceptions of reality" (p. 257). The data analysis began during the data collection stage and extended beyond it. Data was analyzed, coded, compared, contrasted, categorized, organized, and presented according to the themes and issues which emerged. Codes that reflected strengths and weaknesses of the effectiveness of the courseware and the computer conferencing were developed and assigned as appropriate to the data set. Once codes had been determined and assigned, the data set itself was sorted into subsets based on the categories. These subsets then became the focus for further analysis, including cross-coding until following extensive examination, comparison and contrasting, a comprehensive understanding of learners' perceptions and experiences with the instructional dimensions of the courseware was arrived at. Ethnograph 5.0, a software program for the management and analysis of text-based data, was used to assist in the coding and analysis of the data.

According to Guba (1981) in naturalistic investigation, methodological rigor is usually the primary source for verification of a study's findings. As well, triangulation, the cross-checking and verification of data through the use of a variety of data sources, should be used to verify evaluative findings (Guba, 1981). In this study, the results of the interview data were cross-checked with the results of the quantitative CES questionnaire. In addition, several instructional design specialists were consulted and asked to review the findings. These peer reviews were used to ensure the credibility of the interview findings. As well, the use of the Ethnograph 5.0 software enabled the compilation and management of an extensive interview database, which could serve as an audit trail were an external evaluator to examine the process of data analysis.

An analysis of the discussion which occurred through the computer conferencing system was also performed. A unique advantage of computer conferencing systems is they automatically provide an electronic transcript of discourse interchanges which can be analyzed using qualitative and quantitative methods. A thorough analysis of these interaction patterns enables an assessment of the types and levels of the many-to-many interactions which occur. The electronic transcripts were prepared by copying the

HTML discussion threads from the computer conferencing system and preserving linear sequences of strings of messages as generic text files. The discussion interchanges were then coded and analyzed in Ethnograph 5.0, each of which included the date of the posting and the name of the sender as a header to the message. The analysis of electronic transcripts of all computer conferencing activities enabled an assessment of the pattern of peer interaction in the computer-mediated environment.

Metaevaluation of the Evaluation Model

A metaevaluation, an evaluation of an evaluation, was conducted in order to address the research questions which were originally stated in Chapter One.

1. What were the strengths and/or weaknesses of the evaluation model?

2. Was the evaluation model useful in serving the information needs of the intended users?

3. Did the evaluation model follow practical and feasible means for collecting evaluative information?

4. Was the fieldtest of the evaluation model conducted in an ethical manner, with due regard for the welfare of those involved in the evaluation, as well as those affected by its results?

5. Did the evaluation model convey technically adequate information about the features that determined the worth or merit of the program being evaluated?

According to Stufflebeam (1981) metaevaluation is a process which involves:

> "delineating, obtaining, and using descriptive and judgmental information about the practicality, ethics, and technical adequacy of an evaluation in order to guide the evaluation and publicly to report its strengths and weaknesses." (p.151)

As discussed in Chapter Two, it was this conceptualization of metaevaluation that guided the development of the procedures for "evaluating evaluation" which were implemented in this study. In a similar way, Stufflebeam's (1978) definition of summative metaevaluation was also used to direct the actual metaevaluation process which was conducted:

> "...sums up the overall merit of an evaluation, and is usually done following the conclusion of a primary evaluation. It holds evaluators accountable by publicly reporting on the extent that their evaluation reports meet standards of good evaluation practice. Finally, summative metaevaluations help the audiences of primary evaluations determine how seriously they should take the primary evaluation's reported conclusions and recommendations."

The metaevaluation process was used to appraise the effectiveness and quality of the evaluation which was fieldtested in this study. It was applied as a mechanism for validating the evaluation model which was fieldtested upon the CME instructional courseware system. The procedures of the metaevaluation conducted in this study were based on the literature review presented in Chapter Two. Three different metaevaluation procedures were formulated and conducted. The first evaluation procedure was designed to address the first research question and entailed an analysis and self-report by the evaluator of the strengths and weaknesses of the evaluation fieldtested in the study. The self-report was a reflective-procedure in which the evaluator examined and critiqued the processes, methods and instruments which were implemented, and provided insight on ways that these elements could be improved.

This notion of "reflection-on-practice" is supported to a large extent by Collin's (1991) conceptualization of research on education versus research in education.

According to Collins, research in education, the prevalent paradigm, is conducted by an expert who defines the problems to be researched as well as the methodologies and the nature of the discourse through which they will be addressed. Research on education, on the other hand, envisages a less narrowly defined, professionalized orientation to research and practice. Research on education acknowledges the interconnectedness between research and practice, and invites practitioners to reflect carefully on what it is they do and what it is they are, without recourse to the constructs, constraints, and basic assumptions of professionalized pedagogues. This, Collins stresses, calls into play the need for a continuing reflection on everyday educational practice and its relevant implications for others.

> "A need for continuing self-conscious reflection in what adult educators should do and should be cannot be sensibly dispensed by the adoption of strategies emanating from the ideology of technique. Therefore, attention to ethical and practical matters (a serious consideration of "shoulds" and "oughts") ....is necessary to make up for shortfalls in a modern practice of adult education wedded to pedagogical strategies derived from a pervasive technical rationality." (Collins, 1991, p.43)

Cook and Gruder (1978) also supported the concept of an evaluator reflecting upon and self-reporting problems and/or difficulties which may have been encountered in conducting evaluation. They believed that evaluators should be encouraged to provide a detailed and honest outline of all the problems they were meeting, particularly those which related to implementing experimental designs and receiving cooperation from project personnel. The authors referred to this procedure as essay review and it encompassed "after-the-fact commentary on a single set of evaluation data that are not re-analyzed" (p.18). Cook and Gruder did stress that this essay review process should focus on certain significant elements of the design and implementation of the evaluation.

> "In assessing an evaluation, reviewers usually concentrate on the utility of research questions and on technical issues of sampling, measurement, data analysis, and the like." (P.19, Cook and Gruder, 1978)

The self-reporting procedure applied in this study involved a critical reflection on the difficulties which were experienced during the course of developing and administering evaluation instruments, and the analysis and reporting of the data which resulted from the use of these instruments. It also entailed a reflection on the main strengths of the evaluation procedures which were developed and applied in the evaluation. This "reflection-on-practice" was a retrospective process, meaning it occurred after the evaluation was conducted and the data collected and analyzed. In this manner, it closely resembled Cook and Gruder's notion of an essay review. The self-report also included recommendations for addressing the weaknesses which were inherent in the instruments and the evaluation model itself.

A second procedure was also used to answer the first question and involved a focus-group discussion with the courseware multimedia developers. A focus-group is a social research methodology and involves the gathering of a small group of people for the purpose of reflecting upon, exploring, and discussing some topic or issue of interest and/or concern. The main purpose of the focus-group facilitated in the metaevaluation process of this study was to examine the developers' thoughts and opinions regarding the usefulness of the formative evaluation methods conducted during the development stages of the courseware. Although unintended, the focus group also enabled the evaluator to explore the development teams' perceptions of the usefulness of the summative evaluation information as well. Discussion surrounding summative evaluation focused on the efficacy of the information which resulted from these methods and its utility in guiding or making revisions to the courseware which was evaluated.

The focus-group procedure was facilitated by the evaluator and the discussion was directed by the following open-ended question: "Reflecting on the formative evaluation methods which were conducted during the development of the CME courseware, in your opinion what were the strengths and/or weaknesses of these methods in providing information which you could use to improve or revise the instructional courseware?" The discussion which ensued was unstructured and the facilitator used open-ended probing questions to explore the perceptions of the program developers in greater depth and detail. The discussion was recorded by the evaluator and the transcripts were edited, combined, and categorized using Ethnograph 5.0.

The third metaevaluation procedure involved several structured interviews with a variety of key decision-makers and stakeholders. The individuals who were interviewed included the Vice-Dean of Professional Development, Director of CME, Coordinator of the Office of Professional Development, and the two Subject Matter Experts (Dermatologists). The structured interview questions were developed by identifying evaluation standards related to each of the research questions. These evaluation standards were based on the 1994 Program Evaluation Standards (2nd ed.) published by the Join Committee on Educational Evaluation. The 1994 standards included thirty specific standards categorized into groups relating to four attributes necessary and sufficient for sound and fair program evaluation: utility, feasibility, propriety, and accuracy. Each structured question was based on an educational evaluation standard. Several other questions not specifically related to the research questions, yet concerned with the effectiveness of the evaluation model were also asked during the interviews. The purpose of these questions were to explore the perceptions of the stakeholders' in terms of the broader applicability and effectiveness of the evaluation model. These questions and the responses from the interviewees are presented in Table 4.29 of Chapter Four, as are the questions and responses to all the interview questions.

These interviews were also of a retrospective nature and occurred after the evaluation was conducted and data reported to stakeholders and decision-makers. The interviews were recorded, transcribed and coded, categorized, and edited using Ethnograph 5.0. Many of the questions were of a closed-ended, structured nature so there was little opportunity for the individuals being interviewed to elaborate on their responses unless further articulation was necessary to express an opinion or perception. Therefore, the responses to many of the questions were simple yes or no replies. In these instances, where a yes, no or the same closed-ended reply was made by several of the respondents, answers were combined to ease the representation of data. As an example, instead of presenting a number of responses of "yes" or "no" only one response is presented meaning that a majority of respondents answered the question using the same response. Where dissimilar responses were made they are presented.

The results of these metaevaluation procedures are provided in Chapter Four.

Summary

Based on an extensive review of the literature, an evaluation planning matrix and an evaluation model were developed for assessing the effectiveness of computer-mediated CME instructional courseware. A courseware product under development by the Telemedicine and Educational Technology Resources Agency (TETRA) of Memorial University of Newfoundland was used to fieldtest the evaluation model which was developed using the planning matrix as a guide. A metaevaluation was conducted on the evaluation model as a means to assess the effectiveness of the evaluation model itself. The results for both the evaluation and the metaevaluation are presented in Chapter Four.

CHAPTER IV

RESULTS

Introduction

The results of the fieldtest of the evaluation model and the metaevaluation are presented in this chapter. The results of the fieldtest of the courseware evaluation model are presented first by matching the planning matrix evaluation questions with the results which were gathered through the implementation of the evaluation model. These results are grouped together according to the five types of data gathered during the summative stages of the fieldtest of the evaluation model, antecedent, reaction, cognitive, behavioral, and instructional transaction data. The other information presented in this chapter are the results of the metaevaluation of the evaluation model. The results for each of the metaevaluation procedures are outlined in relation to the study's five research questions which were originally stated in Chapter One.

Results of the Fieldtest of the CME Courseware Evaluation Model

The results of the fieldtest of the evaluation model described in Chapter Three are presented in this section. The evaluation questions originally presented in the evaluation planning matrix in Chapter Three are restated in conjunction with the data used to address the questions. The questions and results are arranged according to the different types of data collected, antecedent, reaction, cognitive, behavioral, and instructional transaction data.

Antecedent Data - Questions and Results

1. **How did the participants': self-reported years of medical practice; gender; practice location; and practice type influence their attitudes toward computers.**

The Computer Attitude Scale (CAS) measured subjects' attitudes toward computers. The CAS included 20 positively and negatively worded items, and respondents were asked to rate their level of agreement or disagreement with each statement on a scale of 1 = "strongly disagree" to 5 = "strongly agree." Scores for negatively worded items were inverted to calculate a total sum score ranging from a possible 20 to 100. Table 4.1 presents the total mean and range of scores for subjects in experimental groups I and II. The mean CAS score was 79.20 and scores ranged from a low of 68 to a high of 99.

Table 4.1 Mean Computer Attitude Scale Scores

	Mean	SD	Minimum	Maximum
Computer Attitude Scale Scores	79.20	8.23	68.00	99.00
N = 30				

A Demographic Profile Survey (DPS) was used to collect information on the demographic and computer experience characteristics of physicians participating in the study. Two versions of the DPS were developed, an online DPS for participants in the experimental study groups and a print-based DPS for control group subjects. The online DPS consisted of ten items which were divided into two sections: a demographic characteristics section and a computer experience characteristics section. The demographic characteristics section included four items which collected information on practice experience, gender, physician type, and practice location. The computer experience characteristics section included six items which collected information on subjects' computer experience, access to computers, and computer usage. Subjects in the control group completed a print-based version of the DPS which included the four demographic characteristic items only.

The original practice experience item on the electronic and print-based DPS measured practice experience along 11 categories. Ten of these categories were assigned practice experience scales of 3 year intervals, from "1- 3 years" to "28- 30 years" and "31 years and greater." During the analysis of the DPS data, this demographic variable was collapsed into two categories, physicians with "10 years or less" experience and physicians with "11 years or greater" experience. The frequencies of responses to the original scale were widely dispersed along all 11 categories. Collapsing the categories made interpretation easier.

There were two main reasons, based on logical reasoning, for creating these two categories. First, historically, medical informatics training was introduced into the Canadian undergraduate medical training curriculum during the mid 1980s, and at Memorial during the 1987-88 academic year (Interview with Dean of Medicine, Memorial University of Newfoundland). This curriculum change influenced the amount of computer training and computer application use medical students were exposed to during their undergraduate medical training. As well, computerized billing services were introduced between 1986-1988, requiring licensed physicians to use microcomputers for the preparation and submission of patient billing to provincial medical insurance agencies (Newfoundland and Labrador Medical Care Commission).

A Kruskal-Wallis Analysis of Variance of Ranks test was performed to determine the existence of relationships between the DPS variables (practice experience, gender, physician type and practice type) and physicians' attitudes toward computers. Table 4.2 presents the mean and mean rank computer attitude scores by individual demographic characteristics for the experimental study groups. The results of the Kruskal-Wallis tests indicated the presence of one significant relationship between individual demographic characteristics and computer attitudes at the < .10 level of probability and this was for gender. The individual demographic characteristics of practice experience, physician type and practice type did not reveal significant relationships with CAS scores.

Female physicians scored lower on the CAS (74.00), than their male counterparts (80.50). Mean ranked scores for male and female physicians were 16.92 and 9.83 respectively. The Kruskal-Wallis test revealed a significant difference between the computer attitudes of male and female physicians in this study ($p = .077$) indicating male physicians expressed more positive attitudes towards computers than female physicians. This result supports the findings of previous research literature cited in Chapter Two documenting the influence of gender on computer attitudes. However, it should be noted that the sample sizes in this study were very small and further investigation is warranted, particularly randomized survey research of larger physician samples.

Urban and rural physicians and physicians practicing in solo or group settings did not differ in their attitudes towards computers. As well, years of practice experience did not influence the computer attitudes of physicians participating in this study.

Table 4.2 Kruskal Wallis ANOVA of Computer Attitude Scale Scores by Demographic Characteristic Variables

	Experimental Groups I and II				Kruskal Wallis ANOVA	
Variables	N	Mean	SD	Mean Rank	df	Sig.
Practice Experience (Total)	29	79.38	8.31		1	.539
10 years or less	9	79.00	11.39	13.56		
11 years or greater	20	79.55	6.86	15.56		
Gender (Total)	30	79.20	8.23		1	.077
male	24	80.50	8.29	16.92		
female	6	74.00	5.97	9.83		
Physician Type (Total)	30	79.20	8.23		1	.230
rural	21	80.24	8.20	16.76		
urban	9	76.78	8.23	12.56		
Practice Type (Total)	30	79.20	8.23		2	.384
solo	6	79.50	3.08	17.17		
group	19	80.26	9.43	16.26		
neither	5	74.80	8.23	10.60		

2. How did the participants': self-reported computer usage; computer access; and computer experience influence their attitudes toward computers?

The second section of the online Demographic Profile Survey (DPS) for experimental study group subjects included items which collected information on participants' self-reported computer experience, computer access, computer usage, and frequency of use of software and Internet applications. Table 4.3 presents the results of Kruskal-Wallis Analysis of Variance of Ranks tests of CAS scores for the experimental groups by the variables of computer usage, computer access and computer experience.

Information on computer usage was collected by asking physicians to report their average weekly usage of computers (in hours). The original scale for this item included 11 categories, 10 of which were divided along "2 hour usage" intervals, from "0-2 hours" to "13-14 hours" and "15 hours or more." For data analysis, these 11 categories were collapsed into two weekly computer usage categories of "0-6 hours" and "7 hours or greater." These collapsed categories were created based on the computerized billing system in place for patient billing to the provincial medical insurance agency. Several interviews were conducted with family physicians and physician representatives prior to the creation of these collapsed categories. The findings of the information from these interviews suggested that the average weekly use of computers by physicians for preparing electronic billing reports was between 0 to 5-6 hours per week, approximately one hour per day. This was an "estimate" of the average amount of time that a physician could be expected to use a computer for billing purposes.

The mean CAS score for physicians in the 0-6 hour category was 78.05, while

Table 4.3 Kruskal Wallis ANOVA of Computer Attitude Scale Scores by Weekly Computer Usage, Computer Access and Computer Experience Characteristic Variables

	Experimental Groups I and II				Kruskal Wallis ANOVA	
Variables	N	Mean	SD	Mean Rank	df	Sig.
Weekly Computer Use (Total)	29	79.48	8.22		1	.085
0 - 6 hours	20	78.05	8.73	13.18		
7 hours or greater	9	82.67	6.26	19.06		
Computer Access (Total)	30	79.20	8.23		2	.681
home	8	77.38	8.18	13.44		
work	4	77.75	5.74	14.63		
home and work	18	80.33	8.87	16.61		
Computer Experience (Total)	29	78.76	8.00		2	.071
Little to None	6	73.33	5.13	9.00		
Moderate	15	78.27	6.69	14.97		
Considerable to extensive	8	83.75	9.68	19.56		

the mean CAS score for physicians using the computer for 7 hours or more per week was 82.67 per week. These scores indicated that physicians reporting higher levels of computer usage expressed marginally higher attitudes towards computers.

An assessment of the frequency distribution of subjects along the two collapsed categories indicated that a majority of physicians (69 percent) reported using computers between 0-6 hours per week. These findings suggest that the majority of reported computer usage could in fact be related to patient billing tasks (however it is difficult to fully support this hypothesis in the current study). The mean ranked score for physicians reporting 0 - 6 hours of computer use per week was 13.18, while the mean ranked score for those reporting computer usage of 7 hours or greater a week was 19.06. The Kruskal Wallis revealed a significant difference at the $< .10$ probability level between the expressed computer attitudes of physicians and their self-reported weekly computer usage ($df = 1$, $p = .085$). This result indicates that physicians reporting significant computer usage expressed more positive attitudes toward computers than physicians reporting less significant computer usage.

The DPS also collected data on physicians' reported access to computers. Access to computers was defined as either access to a computer "at home", "at work", or computer access at "home and work". The Kruskal Wallis test revealed that there was no significant relationship between access to a computer and attitudes toward computers at the $< .10$ level of probability ($df = 2$, $p = .681$). Physicians reporting access to a computer at home and office did not express more positive attitudes towards computers than physicians reporting computer access at only home or office locations.

Table 4.3 also presents the results for mean and mean ranked CAS scores by physicians' self-reported computer experience. This final item of the online DPS asked physicians to rate their experience with computers along a five-category Likert Scale. The original five category scale for this item was collapsed to three categories, including "little to none", "moderate", and "considerable to extensive" computer experience.

Physicians in the "little to none" and "moderate" computer experience groups had mean CAS scores of 73.33, and 78.27 respectively. Subjects in the "considerable to extensive" group had a mean computer attitude scale score of 83.75. These scores suggested that physicians reporting "moderate" and "considerable to extensive" experience expressed marginally higher computer attitudes.

Mean ranked scores for physicians reporting "little to none" computer experience was 9.00 and 14.97 for those reporting "moderate" computer experience. Subjects reporting "considerable to extensive" computer experience had a mean ranked score of 19.56. A Kruskal Wallis test was performed to determine the relationship between physicians' self-reported computer experience and their attitudes toward computers. The result reveals a significant difference at the $< .10$ level of probability between computer experience and physicians' attitudes towards computers ($df = 2$, $p = .071$). Physicians reporting extensive computer experience expressed more positive attitudes toward computers than physicians reporting less experience. This result also supports the research literature presented in Chapter Two regarding the relationship between computer attitudes and self-reported computer experience.

3. How did the participants': self-reported computer software and Internet application usage influence attitudes towards computers?

Three items in the computer experience section of the online DPS were constructed to collect information on subjects' self-reported frequency of computer software and Internet applications use. Table 4.4 presents the results of Kruskal Wallis tests of the relationship between these variables and CAS scores. For data analysis, the original scale of five categories, the same for all items, was collapsed into three. These three categories included "occasionally to not at all", "regularly", and "often to very often".

Physicians' mean ranked computer attitude scores by frequency of software use were 9.19 for "occasional to not at all" users, 16.41 for 'regular users", and 18.10 for "often to very often" users of software. The result of the Kruskal Wallis test revealed a positive relationship between frequency of self-reported software use and CAS scores. This relationship was significant ($df = 2$, $p = .068$) at the $< .10$ probability. Physicians reporting frequent use of computer software expressed more positive attitudes toward computers than physicians reporting less frequent use. However, there were no significant relationships between attitudes toward computers and reported frequency of Internet application usage. These results suggest that usage of computer software, possibly dependant on the complexity of the application, influenced participants' computer attitudes. This result supports, in part, the literature of Chapter Two regarding the relationship between prior computer experience and attitudes towards computers.

Table 4.4 **Kruskal Wallis ANOVA of Computer Attitude Scale Scores by Frequency of Computer Application Usage Characteristic Variables**

	Experimental Groups I and II				Kruskal Wallis ANOVA	
Variables	N	Mean	SD	Mean Rank	df	Sig.
Frequency of Software use (Total)	29	78.76	8.00		2	.068
occasionally - not at all	8	73.75	5.18	9.19		
regularly	11	79.71	8.48	16.41		
often - very often	10	81.50	8.14	18.10		
Frequency of Electronic Mail use (Total)	30	79.20	8.23		2	.315
occasionally - not at all	8	78.50	10.36	13.75		
regularly	13	77.08	5.60	14.00		
often - very often	9	82.89	9.01	19.22		
Frequency of Internet use (Total)	30	79.20	8.23		2	.216
occasionally - not at all	12	75.83	6.00	12.13		
regularly	11	81.00	8.99	17.18		
often - very often	7	82.14	9.37	18.64		

4. How did the participants' attitudes toward computers influence learning achievement in a computer-mediated instructional environment?

A twenty-item, multiple-choice posttest was used to measure physicians' knowledge gain as a result of participation in computer-mediated CME courseware. A Spearman's Rho Correlation test was performed to assess the existence of a relationship between Computer Attitude Scale scores and posttest learning achievement scores of subjects in experimental study group II. This analysis was only performed for experimental group II subjects because of a possible "testing sensitization affect" on experimental group I subjects.

Subjects in experimental group I received a pre and posttest, and as the results presented in Table 4.7 indicate the posttest scores for experimental group I subjects were significantly different than the posttest scores of experimental group II subjects. These results suggest that subjects may have been sensitized to the items on the posttest by the pretest. Experimental study group II subjects received a posttest only.

The results of the Spearman's Rho Correlation, presented in Table 4.5 revealed a marginally negative correlation between posttest scores and subjects' attitudes toward computers ($r_s = -.336$). However, the correlation was not found to be significant at the $< .10$ level of probability ($p = .241$). Physicians' attitudes toward computers were not positively related to learning achievement in a computer-mediated learning environment.

Table 4.5 Spearman's Rho Correlation of Posttest Scores by Computer Attitude Scale Scores for Experimental Group II

		Computer Attitude Scale Score	Posttest Score
Spearman's Rho Correlation Coefficient	Computer Attitude Scale Score	1.000	-.336
	Posttest Score	-.336	1.000
Sig. (2-tailed)	Computer Attitude Scale Score	.	.241
	Posttest Score	.241	.
N	Computer Attitude Scale Score	14	14
	Posttest Score	14	14

5. **How did the participants': self-reported years of medical practice; gender; practice location; and practice type influence performance on a pretest of dermatological office procedures?**

As discussed in Chapter Three, the modified pretest posttest control group experimental research design involved the pretesting of only one experimental study group -- experimental study group I. Subjects in the control group were administered a posttest only, as called for by the static group research design. However, the posttest of the control group is a posttest in a "semantic" sense only. The control group was only tested once, and whether it is termed a pretest or posttest is inconsequential as these subjects were not presented any subject matter prior to, or during their completion of this test. Therefore, this "test" could be considered, alternatively, a test of the participants' knowledge of the subject matter at the time the test was administered. While the test is termed a posttest in the static group research design, the test scores may serve as a baseline for comparing control group subjects' knowledge of the subject matter with the pretest scores of experimental group I.

Table 4.6 presents the Kruskal Wallis Analysis of Variance of Ranks test for experimental study group I and the substituted pretest scores for control group physicians by the demographic characteristic variables of practice experience, gender, physician type and practice type. The Kruskal Wallis test was conducted for each individual study group. The results reveal no significant relationships between the individual demographic variables and pre-instructional knowledge for either group at the < .10 level of probability.

These results suggest that the demographic variables of practice experience, gender, physician type and practice type did not influence physicians' prior knowledge of the subject matter of the computer-mediated learning program. And, the results lend further support to the claim made in Chapter Three regarding the comparability of the study groups participating in this study.

Table 4.6 Kruskal Wallis ANOVA of Pretest Scores by Demographic Characteristic Variables (Experimental I and Control Study Groups)

	Experimental Group I						Control Group					
Variables	**N**	**Mean**	**SD**	**Mean Rank**	**df**	**Sig.**	**N**	**Mean**	**SD**	**Mean Rank**	**df**	**Sig.**
Practice Experience (Total)	16	10.44	1.31		1	.265	22	9.68	2.51		1	.667
10 years or less	6	10.00	.89	6.83			4	10.25	2.99	12.75		
11 years or greater	10	10.70	1.49	9.50			18	9.56	2.48	11.22		
Gender (Total)	16	10.44	1.31		1	.370	22	9.68	2.51		1	.863
male	14	10.29	1.20	8.11			18	9.61	2.45	11.39		
female	2	11.50	2.12	11.25			4	10.00	3.16	12.00		
Physician Type (Total)	16	10.44	1.31		1	.907	22	9.68	2.51		1	.144
rural	11	10.45	1.44	8.59			15	10.20	2.70	12.87		
urban	5	10.40	1.14	8.30			7	8.57	1.72	8.57		
Practice Type (Total)	16	10.44	1.31		2	.367	22	9.68	2.51		2	.783
solo	2	11.50	.71	12.75			1	11.00		15.00		
group	12	10.33	1.37	8.04			19	9.74	2.40	11.53		
neither	2	10.00	1.41	7.00			2	8.50	4.95	9.50		

6. How did the participants': self-reported years of medical practice; gender; practice location; practice type; computer usage; computer access; computer experience; computer software; and Internet application usage influence performance on a posttest of dermatological office procedures?

A Kruskal Wallis test was performed to determine the relationship between individual demographic and computer experience characteristics and learning achievement in a computer-mediated CME courseware program (Table 4.7). This test was only performed on scores for experimental group II (posttest only study group). The results of the Kruskal Wallis tests revealed that only the computer experience characteristic of "frequency of software use" had a significant positive relationship with learning achievement in a computer-mediated instructional program. However, the variables "computer access" and "computer experience" also revealed marginally positive relationships, though not significant. Physicians' reporting frequent computer software usage performed higher in a computer-mediated learning environment than participants reporting less frequent use. The results also suggest the need for further investigation of these characteristics in the future because of the sample size and the use of a volunteer sample of physicians. Nonetheless, these results have significant implications for the influence of individual learner characteristics on physicians' achievement in computer-mediated learning environments.

Table 4.7 **Kruskal Wallis ANOVA of Individual Demographic and Computer Characteristics by Posttest Learning Achievement Scores (Experimental Group II)**

Experimental Group II									
Variables	**N**	**Mean Rank**	**df**	**Sig.**	**Variables**	**N**	**Mean Rank**	**df**	**Sig.**
Practice Experience	13				Computer Access	14			
10 years or less	3	9.33	1	.228	home	3	8.33	2	.126
					work	4	4.00		
11 years or greater	10	6.30			home and work	7	9.14		
Gender	14				Computer Experience	14			
male	10	6.70	1	.249	a little - none	4	9.00	2	.203
					moderate	7	5.57		
female	4	9.50			considerable - extensive	3	10.00		
Physician Type	14				Frequency of Software use	14			
rural	10	7.90	1	.565	not at all - occasionally	4	4.00	2	.083
					regularly	7	9.71		
urban	4	6.50			often - very often	3	7.00		
Practice Type	14				Frequency of E-mail use	14			
solo	4	8.75	2	.616	not at all - occasionally	3	7.00	2	.684
group	7	6.43			regularly	8	7.00		
neither	3	8.33			often - very often	3	9.33		
Weekly Computer Usage	14				Frequency of Internet use	14			
0 - 6 hours	10	7.60	1	.886	not at all - occasionally	7	7.00	2	.871
7 or more hours	4	7.25			regularly	6	7.83		
					often - very often	1	9.00		

Learning Outcome Data - Questions and Results

7. **Was there a significant difference in knowledge gain of dermatological office procedures between participants in a computer-mediated instructional treatment group and a no-CME control group?**

A Kruskal Wallis Analysis of Variance of Ranks test and the Mann-Whitney test were performed to determine the existence of significant differences between the posttest scores of the three study groups. Table 4.8 presents the results of posttest scores for experimental groups I and II, and the control group. The mean posttest scores for experimental study group I, experimental study group II, and the control group were 15.56, 13.29, and 9.68 respectively. These scores suggest that the computer-mediated learning groups performed higher on the posttest of learning achievement. The mean ranked score for experimental study group I was 40.44, 30.14 for experimental study group II, and 14.05 for the control group. The Kruskal Wallis test of these posttest scores revealed a significant difference at the $< .10$ level of probability ($df = 2$, $p = .000$). Several individual Mann-Whitney tests were then performed to determine the existence of significant differences between study group pairs.

The results of the Mann-Whitney test between experimental study groups I and II indicated mean rank scores of 19.19 and 11.29 respectively. There was a significant difference between the mean ranked posttest scores of these two study groups at the $< .10$ probability level ($p = .012$). A significant difference between the mean ranked scores of experimental study group I and the control group was also found at the $< .10$ probability level ($p = .000$). As well, the Mann-Whitney test of the mean ranked scores of experimental study group II and the control group also revealed a significant difference at

Table 4.8 Kruskal Wallis ANOVA and Mann-Whitney Tests of Study Groups by Posttest Scores

		N	Mean	SD	Mean Rank	df	Sig.
Study Group	experimental I	16	15.56	2.28	40.44	2	.000
	experimental II	14	13.29	2.16	30.14		
	control	22	9.68	2.51	14.05		

Mann-Whitney					
	N	Mean Rank	Sum of Ranks	Z	Sig.
experimental I	16	19.19	307.00	-2.498	.012
experimental I	14	11.29	158.00		

Mann-Whitney					
	N	Mean Rank	Sum of Ranks	Z	Sig.
experimental I	16	29.75	476.00	-4.875	.000
control	22	12.05	265.00		

Mann-Whitney					
	N	Mean Rank	Sum of Ranks	Z	Sig.
experimental II	14	26.36	369.00	-3.605	.000
control	22	13.50	297.00		

the < .10 level of probability (p = .000). Physicians participating in a computer-mediated instructional courseware program performed significantly better on a knowledge test of subject matter than physicians not receiving any continuing medical education.

8. Was there a significant change in the level of cognitive knowledge of dermatological office procedures from pretest to posttest for participants in computer-mediated instruction?

Table 4.9 presents the results of the pre and posttest scores for experimental group I. A Wilcoxon Signed Ranks Test was used to assess the difference in the mean ranked pre and posttest achievement scores for experimental study group I. The mean pretest and posttest scores for experimental study group I were 10.44 and 15.56 respectively. The difference in the mean scores on pre and posttests suggests that participants scored higher on the learning achievement test as a result of participation in the computer-mediated CME courseware program. The difference between the mean ranked scores of pre and posttests was 8.50. The Wilcoxon test revealed a significant difference between these achievement test scores at the < .10 level of probability (p = .000).

9. What was the participants' level of cognitive knowledge of dermatological office procedures at the beginning of the program?

A Mann-Whitney test was also used to assess the difference between the mean ranked pretest scores of experimental study group I and the substituted mean ranked pretest scores of the control group. In order to accommodate this analysis, the control group posttest scores were substituted as pretest scores. According to the static group research design, an experimental and control group are posttested only after a treatment

Table 4.9 Wilcoxon Signed Ranks Test for Significance between Pre and Posttest Scores of Experimental Group I

		N	Mean	SD	Minimum	Maximum
Experimental Study Group I	Pretest Score	16	10.44	1.31	8.00	13.00
	Posttest Score	16	15.56	2.28	11.00	20.00

Wilcoxon Signed Ranks Test

		Mean Rank	Sum of Ranks	Z	Sig. (2-tailed)
Experimental Study Group I	Post - Pre	8.50	136.00	-3.539	.000

Table 4.10 Mann-Whitney Test of Pretest Scores by Experimental Study Group I and Control Group

	N	Pretest Mean	SD	Mean Rank	Sum of Ranks	Z	Sig.
experimental I	16	10.44	1.31	21.50	344.00	-.959	.337
control	22	9.68	2.51	18.05	397.00		
Total	38	10.00	2.11				

has been delivered. However, as previously discussed, it is plausible to substitute the posttest scores of the control group as pretest scores. The result of the Mann-Whitney test revealed no significant difference at the $< .10$ probability level ($p = .337$). These results are presented in Table 4.10 and suggest that control group and experimental group subjects had comparable knowledge of the subject matter prior to the delivery of the program. This provides further support to the comparability of the study groups.

Reaction Data - Questions and Results

10. How satisfied were the participants with the subject matter, the instructional courseware, computer conferencing, and overall use of computer-mediated CME instruction at a distance?

The Courseware Evaluation Survey (CES) was a quantitative instrument, designed to collect information on participants' perceptions of the quality and effectiveness of different aspects and components of the computer-mediated CME courseware and computer conferencing system. Five evaluative categories were developed to measure these perceptions. These categories included: content, graphics and media, navigation and organization, learner manual, computer conferencing, and overall impressions. Each category included one to five positively-worded evaluative statements and subjects were asked to rate their level of agreement or disagreement with the statement along a five point Likert scale, with 1 = "Strongly Disagree" and 5 = "Strongly Agree."

Table 4.11 presents the mean scores for the three evaluative items designed to measure participants' attitudes toward the subject matter of the courseware and the instructional strategies used to present this material. Participants rated "relevancy of the subject matter" highest with a mean score of 4.26, SD = .59. Overall, all items in the content category had a mean score of 4.11 or higher.

Table 4.11 - Content

	N	Mean	Standard Deviation
I believe the subject matter of this courseware is relevant to my professional practice.	27	4.26	.59
The case studies were representative of possible clinical scenarios and encouraged professional problem-solving.	27	4.19	.56
The subject matter presented through this courseware enhanced my knowledge of dermatological office procedures.	27	4.11	.58

The graphics and media category included three evaluative statements designed to collect information regarding physicians' perceptions of the effectiveness of the media used, and the general attractiveness of the HTML pages, see Table 4.12 . Subjects rated their satisfaction with media components the highest with a mean score of 4.33, SD = .79. Attractiveness of web pages received a mean score of 3.82, SD = .68, and organization and satisfaction with screen layouts received a mean score of 3.69, SD = .93.

Table 4.12 - Graphics and Media

	N	Mean	Standard Deviation
The use of different media components (audio, video, text, images, photos) serves a clear purpose and presents the subject matter effectively.	27	4.33	.79
The web pages are attractive.	27	3.82	.69
The screen design and layouts were clear, uncluttered, and well-organized.	26	3.69	.93

Table 4.13 presents evaluative results for the third evaluative category: navigation, organization and instructional design. Items in this category were developed to collect information on physicians' reactions to aspects of learner control, interactivity, and the navigation features of the courseware. The results indicate that physicians' rated the courseware's ability to "provide learners with control over the rate of material presentation" the highest (M = 4.37, SD = .84). The level of "interactivity" was also rated very highly by participants, receiving a mean score of 4.14, SD = .77. Items which asked physicians to rate their reaction to, and satisfaction with "navigational aspects" of the courseware did not rate as highly, with mean scores of 3.50 or below.

Table 4.13 - Navigation, Organization, and Instructional Design

	N	Mean	Standard Deviation
I was able to control the rate of presentation of subject matter.	27	4.37	.84
This CME courseware provides interactivity which increases its instructional value.	27	4.14	.77
The web site is well organized for ease of use.	26	3.50	.91
It was easy to navigate, so I could concentrate on learning the material rather than learning to use the courseware.	27	3.48	.98
All links and navigation buttons are clearly labeled and serve an easily identified purpose.	27	3.30	1.10

One item, Table 4.14, was developed to measure physicians' satisfaction with the completeness and organization of the learner manual. The learner manual provided instructions and an overview of the components of the courseware, and how to use them. The mean score for this item was 3.70.

Table 4.14 - Learner Manual			
	N	Mean	Standard Deviation
The learner manual was complete, clearly organized and easily understood.	27	3.70	.95

The courseware developers viewed the computer conferencing as an important and unique instructional component of the courseware, enabling physicians to communicate asynchronously with one another and the subject matter consultants. Table 4.15 presents the results of items measuring physicians' reaction to, and satisfaction with computer conferencing. Participants rated the item "usefulness" of computer conferencing the highest with a mean score of 3.85. The mean score to the evaluative statement, "I found it informative and beneficial to be able to communicate with my peers and the instructors" was also scored fairly high, M = 3.80, SD = .82. The results appear to suggest that some participants may have experienced some difficulty in using the computer conferencing system, as the mean score to the statement "I found it easy to post, respond and reply to messages in the computer conferencing sessions" M = 3.12, SD = 1.24 indicates some dissatisfaction with the computer conferencing system. This "difficulty" was later confirmed in interviews conducted with participants.

Table 4.15 - Computer Conferencing			
	N	Mean	Standard Deviation
Computer conferencing was a useful component of this courseware system.	27	3.85	.82
I found it informative and beneficial to be able to communicate with my peers and the instructors.	25	3.80	.82
I found it easy to post, respond and reply to messages in the computer conferencing sessions.	26	3.12	1.24

The fifth category of the Courseware Evaluation Survey, Table 4.16, was designed to measure physicians' overall perceptions and satisfaction with the use of computer-mediated courseware for participating in CME at a distance. Two evaluative items scored very highly "I would participate in another CME course offering of this type" and "This courseware learning system is an effective way to participate in CME at a distance", M= 4.59, SD = .50 and M = 4.58, SD = .58 respectively. Two other items received high mean scores, M = 4.22 was received for "Overall, the instruction I received through this courseware learning system was appealing, interesting and motivating" and 4.19 for "This method of distance education compares favorably with other available means of obtaining a similar education."

Table 4.16 - Overall Impressions			
	N	Mean	Standard Deviation
I would participate in another CME course offering of this type.	27	4.59	.50
This courseware learning system is an effective way to participate in CME at a distance.	26	4.58	.58
Overall, the instruction I received through this courseware learning system was appealing, interesting and motivating.	27	4.22	.58
This method of distance education compares favorably with other available means of obtaining a similar education.	27	4.19	.79
This courseware was easy to use.	27	3.93	.96
This courseware makes me confident in using computers and other technology.	26	3.65	.75

Interviews were also conducted with a random sample of 12 participants in the computer-mediated instructional group. These interviews enabled the evaluator to collect in-depth and detailed information on the perceptions and experiences of the participants. The interview data was coded using Ethnograph 5.0 and categories describing and summarizing physicians' perceptions and experiences with computer-mediated CME were compared, contrasted, and identified. Four main classification categories were identified from the data: advantages of computer conferencing, disadvantages of computer conferencing, perceptions of multimedia courseware, and overall impressions of computer-mediated CME. These categories are presented in Table 4.17. These four classification categories were further divided into subcategories, detailing specific perceptions and participant experiences as they relate to the particular category. The

Table 4.17 **A Classification of the Major Categories and Sub-categories of Physicians' Experiences and Perceptions Relating to Computer-Mediated CME Courseware**

Categories of classification	Sub-categories of Evaluative Thoughts relating to each category	Frequency	Examples of Evaluative Thoughts or Experiences from the data that give rise to the sub-categories
Learner Perceptions of Computer Conferencing Refers to physicians' perceptions of their experiences with asynchronous computer conferencing	***Advantages***		
	(a) anonymity	5	"Oh, I think it's far less threatening. You don't have to stand up and risk making a fool of yourself with everybody looking at you. Anonymity, if you like." (F: Lines 149-168)
	(b) interaction	16	"I liked it in that I could post a question and then come back tomorrow and get the answer." (F: Lines 102-104)
	(c) reflection	17	"And it gave me time to think about it, and it'd also gave other people time to read, you know, look at your questions and give responses." (E: Lines 80-84)
	(d) a learning resource	9	"And it's there to be referred back to like some of the answers were, you know, where do you order this certain supply. That it can be referred back to." (I: Lines 62-66)

Table 4.17 (continued) **A Classification of the Major Categories and Sub-categories of Physicians' Experiences and Perceptions Relating to Computer-Mediated CME Courseware**

Categories of classification	Sub-categories of Evaluative Thoughts relating to each category	Frequency	Examples of Evaluative Thoughts or Experiences from the data that give rise to the sub-categories
Learner Perceptions of Computer Conferencing Refers to physicians' perceptions of their experiences with asynchronous computer conferencing	***Disadvantages***		
	(a) conference software interface and format	20	"The trail of messages was a little bit difficult to follow initially." (K: Lines 37-39).
	(b) tedious	7	"Posting wasn't too bad, but I must say just accessing the messages was very time consuming." (L: Lines 184 - 186)
	(c) immediacy of interaction	13	"It's a bit different than in classrooms, anyway, where you get an answer right away. (E: Lines 75-77)
	(d) impersonal	3	"When you're face to face with somebody there's a body language component. There's the ability to identify that you're going to say something, that you want to say something. You can carry on a more complex conversation extemporaneously." (G: Lines 111-117).
	(e) discussion relevancy	3	"That people posted and the replies. I don't know. It was just a lot of the same stuff. You know, how can you get liquid nitrogen and stuff like this, you know, from, you know, a source of supply to your office. Well, you know, there is so much difference in geography I mean you can't really answer a question on this kind of stuff, you know. Yeah, a lot of it wasn't really relevant." (A: Lines 149-167)

Table 4.17 (continued) **A Classification of the Major Categories and Sub-categories of Physicians' Experiences and Perceptions Relating to Computer-Mediated CME Courseware**

Categories of classification	Sub-categories of Evaluative Thoughts relating to each category	Frequency	Examples of Evaluative Thoughts or Experiences from the data that give rise to the sub-categories
Learner Perceptions of Hypermedia Courseware Refers to physicians' perceptions of their experiences with the courseware learning material (Web site on a CD-ROM)	(a) case studies	9	"I think that's how people get presented to us. And, you know, you always start off with a person in front of you and trying to arrive at what their problem and that's a logical way to approach it, you know." (E: Lines 834 - 843)
	(b) navigating and learner control	5	"It was very easy to find what you'd been looking for." (C: Lines 360-361)
	(c) self-paced	18	"I like being able to do it on the computer on my own time." (B: Lines 96-97)
	(d) technical constraints (video)	27	"The video wasn't great. It was too choppy." (A: Lines 43)
	(e) testing	25	"I mean, if you don't feel tested, you're not really paying as much attention. I enjoy the opportunity to be tested, and, as a matter of fact, it was very good for me because it made me pay a lot more attention." (B: Lines 492-501)
	(f) media	2	"Mostly I liked the graphics which gives very good exposure of particular items and magnifies very well certain pictures, that was very handy." (C: Lines 314-319)
	(g) content	22	"There is no problem at all learning from it but it is simplistic, but it's probably good that it's simplistic because I learned more from this than most things, you know." (A: Lines 574-579)

Table 4.17 (continued) **A Classification of the Major Categories and Sub-categories of Physicians' Experiences and Perceptions Relating to Computer-Mediated CME Courseware**

Categories of classification	Sub-categories of Evaluative Thoughts relating to each category	Frequency	Examples of Evaluative Thoughts or Experiences from the data that give rise to the sub-categories
Overall Impressions Refers to physicians' overall comments and perceptions of computer-mediated CME.	(a) becoming familiar, comfortable and adjusting to new form of communication (text-based)	9	"You have to be somewhat familiar and comfortable with the idea that you're talking to somebody who may not be listening right at that moment." (B; Lines 238-242)
	(b) prior computer experience	15	"The fact that I'm comfortable with computers made me more likely to go ahead and do this type of thing." (G; Lines 46-48)
	(c) cost savings	6	"It's also, you know, much cheaper. You don't have to transport bodies across the country." (B; Lines 676-678)
	(d) duration	27	"Well I think it's a reasonable time. I mean ideally it'd be nice to have it almost ongoing or, you know, for a significantly longer time, but I don't know how that would -- I mean you wouldn't probably get as much ongoing interaction." (I; Lines 414-420)
	(f) advice	10	"I think the first thing is you got to make sure that the software and everything is working before the day it starts." (A: Lines 518-521)
	(e) overall	14	"And particularly for those of us in outlying areas it's just the way to go. I think it's a great way to go. I would very quickly latch on to something like this as an ongoing way of maintaining my competence." (G; Lines 754-759)

category "advantages of computer conferencing" is sub-categorized into four areas: anonymity, interaction, reflection, and a learning resource.

Anonymity refers to physicians' perceptions of the communicative environment of computer conferencing. Physicians felt that the discourse of a computer conferencing environment was non-threatening. This perception of the computer conferencing environment was very dissimilar to face-to-face CME, which for several participants was often a threatening and intimidating communicative environment. As one physician noted, the anonymity of the computer conference encourages participation, questioning, and commenting:

> "Well if you're anonymous basically you feel free to ask questions. Or not anonymous but semianonymous, right. Whereas in face to face I mean it's really in sort of a group of people which might be quite large. Many people are shy to sort of speak up and ask questions." (L: Lines 119-126)

> "Although you felt I think you would probably get more input from people because at times in a workshop, a lot of people won't ask a question, depending on the size of the workshop. If it's a very large workshop, a lot of people, you know, will be intimidated by the size and probably not ask questions that they'd like to ask; and this gave you the ability to do this because it was directed, even though it was shared among everyone, it was, kind of -- it wasn't taking up anybody's time, as such. It really, I think it would probably -- people would ask more questions and give more input in this format than they would in a live setting unless -- And the only thing that would be better than this would be a very small workshop, obviously, and where you had, you know, very close contact; but this was, I'd say this way would elicit more responses than most educational sessions." (E: Lines 146-174)

Interaction was a second advantage of computer conferencing and this sub-category was related to perceptions of the interactive aspects of computer conferencing. Several physicians felt the ability to post and read messages, respond to comments or questions, direct questions to the consultants and peers, and share one's individual experiences or review the experiences of one's colleagues were beneficial aspects of the computer conference.

> "And the thing I liked was the ability to ask the questions and get the feedback because, otherwise, you know, you get frustrated if there's something there you don't understand or something you don't agree with and you can't -- there's no one there to ask. So having the course preceptors, or whatever, standing by for that period of time was excellent." (E: Lines 641-650)

> "I really liked the opportunity to have a direct line through to somebody who is able to share their experiences. I very much liked the opportunity to see the questions and the responses of my colleagues. And also the opportunity to have a little bit of direct interaction. I'm in a rural situation where I'm all by myself and physician contact is actually fairly infrequent. So just the opportunity to engage in that kind of discussion was really a benefit to me." (G: Lines 77-93)

Another advantage of computer conferencing was the perception that it enabled "more time" for reflecting on issues and more time for thinking about what one would like to say and how to articulate it. Participants also felt that this was another feature of computer conferencing which distinguished it from face-to-face CME workshops. Several physicians felt that discussion in face-to-face CME often results in feelings of being "pressured" or "uncomfortable" by not having enough time to formulate thoughts or time to think about what to say and how to say it. Participants also felt that this "reflective" nature of computer conferencing promoted a discourse which was more superior to that of face-to-face CME discussion.

> "On the other hand, it gives you time to formulate your question and to maybe rephrase it a second time if you didn't get the answer you wanted or if there are other aspects to it that come out. So it's, in many respects it was -- in some ways it was superior to a face-to-face setting because it gave you time to think about, digest the subject matter over a number of days and formulate different questions and come back to the same area if you had to." (E: Lines 105-121)

> "However, with this type of work, with this type of a conferencing you really also have the opportunity to sit and plan out what your thoughts are before you put them down on paper. And you have less chance - at least I do - less chance of shooting from the hip. You have more of a tendency to decide what it is you want to say and then get that across in the best way that you can." (G: Lines 118-127)

The text-based and static nature of computer conferencing was also viewed as an advantage of computer conferencing. Because computer conferencing is a text-based medium, it results in an electronic transcript of discussion, which can be used as a learning resource, both during the offering of the online course and in the future. Participants appreciated this aspect of the computer conference as they were able to read, review and learn from the discussion threads while the computer conference was facilitated, as well they felt that they could also refer back to the discussion at a later date if they wished.

> "I liked having something, you know, that you can then refer back to. You know, I always like handouts that CME and so on. And actually having the course material there that you can go back to even if not in solid form it can -- you can generate it in solid form if you want to." (H: Lines 154-161)

The second category of experiences and perceptions of computer conferencing focused on the perceived disadvantages of the computer conferencing environment. This category was divided into five subcategories which included: software interface, tedious, immediacy of interaction, impersonal, and discussion relevancy.

Several physicians experienced problems in navigating and using the computer conferencing software. These difficulties were related to attempts to follow the threads of discussion, the conceptual nature of computer conferencing, and the frustration in learning how to post messages. Many physicians found the software interface difficult to follow, which often resulted in further problems creating, editing and publishing messages in the conferencing system.

> "The format of the thing was rather confusing. It was difficult to go back and find stuff that you thought you'd read. Also, when you read through the postings you do a lot of sort of flipping from screen to screen to get to the next one. And then I went looking for something that I wanted to reply to and had a hell of the time finding it again." (H: Lines 39-49)

> "The only problem I saw with it is that sometimes it wasn't clear. The threads were not clear. Sometimes the people would post a message, and the answers to it, you weren't sure whether they were answering that query or another one; and if another person answered, you weren't sure whether they had read the response of Person B before they came on as Person C." (B: Lines 67-76)

> "It was just the actual mechanism of accessing the question and answers that I found rather cumbersome." (F: Lines 116-119)

Several physicians also commented on the tedious process for accessing the conference system, navigating through several levels of screens, typing in a message, and then having to edit and post it to the main conference page.

> "Unfortunately for those of us who are not expert typers, the actual mechanics of inputting questions and so on becomes more labored. And I think you're less likely to pose a question, to give a comment simply because it's an effort with the typing for those of us who are not, you know, really good typists." (G: Lines 128 - 137)

> "Posting wasn't too bad, but I must say just accessing the messages was very time consuming." (L: Lines 184 - 186)

Another disadvantage of computer conferencing was the perception that it lacked "immediacy of interaction" qualities. Several physicians felt that this was a significant disadvantage when comparing computer conferencing to the synchronous communicative nature of face-to-face discussion. The asynchronous nature of computer conferencing meant that users had to wait several hours or days for a reply to their comment or message.

"I guess the only weakness is not being able to have instant conversation. You know, if you're talking to a consultant face to face, you know, you can finish the conversation much more quickly than you can by this method." (B: Lines 683 - 693).

"It's a bit different than in classrooms, anyway, you get an answer right away." (E: Lines 75-77)

Physicians also felt that the computer conferencing environment was impersonal. It lacked the human contact that face-to-face CME workshops have. As well, the time delay, the lack of visual response, and the text-based nature were perceived disadvantages of the computer conferencing environment and may have influenced participation.

"When you're face to face with somebody there's a body language component. There's the ability to identify that you're going to say something, that you want to say something. You can carry on a more complex conversation extemporaneously." (G: Lines 111-117).

"Well since it's less personal I mean it was harder to get involved, yeah." (J: Lines 177 - 179).

The relevancy of the discussion was also perceived as a disadvantage. Unlike, face-to-face discussion, where you can "tune out" irrelevant discussion, computer conferencing discourse is text-based and requires the continuous review and reading of discussion threads, which in itself can be a very time consuming process. One physician, in particular, felt the discussion was not really relevant to his needs, and therefore he did not see a need to post any comments or participate in the ongoing dialogue.

A third classification category (see Table 4.18) was related to physicians' experiences and perceptions of the multimedia courseware. Seven subcategories were constructed within this category including: case studies, navigating, self-paced, technical constraints, testing, media, and content. A case-based instructional strategy was used for representing the subject matter in the courseware and for linking the content to the physician's practice context. Several physicians felt that this instructional strategy was appropriate because it reflected the cognitive problem-solving processes they were required to engage in everyday in their offices.

> "I think that's how people get presented to us. And, you know, you always start off with a person in front of you and trying to arrive at what their problem and that's a logical way to approach it, you know." (E: Lines 834 - 843)

> "We tend to think visually, and we tend to think in things that we can identify with. We've all had patients come in with things on their skin and how do we deal with this. So if you present it that way, it's much more real. And much more easily remembered." (B: Lines 447-458)

The navigational features of the courseware were based on a "learner-controlled" model of instructional design. Navigation features were developed to offer flexibility and control to the learner, to enable learners to dictate the pace at which they learned, and to control the presentation of content and the instructional strategies they wished to follow. A main feature of the courseware was it enabled and facilitated self-paced learning. Learners could study and review the instructional materials at a time and place which was convenient for them.

> "The aspect I liked the best was that I could work on it in my own time." (G: Lines 218-220)

> "Well I like anything that I can do at eleven o'clock at night because it seems to be that's the only time that I ever have time to do it I think. So that was nice. You could do this at two in the morning if you wanted. And not having to sort of turn up at a specific time it was useful." (H: Lines 261-269)

> "Well, the main thing is that you can do it when you have time; it doesn't have

> to be prescheduled. And you can -- the other thing is you can go at your own rate." (B: Lines 650-660)

Many physicians experienced problems with the video files which were coded in the HTML documents on the courseware CD-ROM. A majority of these problems were experienced by participants using Netscape Navigator Web browsers. Most of the problems were related to the "time to download" the video from the CD-ROM. As well, some participants found the video "choppy" and the audio and visual aspects of the video "out of sync." These problems were related to the digitization of the video files and the requirement to have digital video play at a frame rate of 15 fps (frames per second), which reduces file size and download time.

> "The video wasn't great. It was too choppy." (A: Lines 43)

> "I found that the quality of the video on my computer could have been a bit better." (D: Lines 200-202)

Physicians' perceptions of online testing were positive. Testing enabled self-assessment and assisted physicians in assessing areas of weakness and learning need. Several physicians felt that by being able to assess and identify areas of learning need they could focus their learning paths and attention on those needs. As well, testing was perceived as challenging, interactive and interesting. Participants felt that pretests and posttests should remain part of the courseware, and the idea of providing limited answer feedback to pretest responses, and detailed feedback on the posttest were useful aspects.

> "Like it's nice to have some challenge in it, you know, where you have to test yourself, and that also relates to the pre-test and the post-test. I found the pre-test was -- my first attitude was it's a bit of a nuisance, but I was shocked because, you know, I didn't make as good a mark as I should have made, and it made me more attentive when I was doing the case studies." (B: Lines 358-

368)

Physicians were also receptive to the use of multiple media for presenting the subject matter. They felt the graphics of various lesions and the use of video for demonstrating the office procedures were helpful for learning:

> "I liked the video part of it, you know. I mean, it was a procedure, sort of, oriented course, so it would've been kind of difficult to do without the video component, I think. So that was -- I mean, I think that's really the only way of, sort of, presenting that type of material in a sensible sort of way that you can understand, right." (D: Lines 115-120)

> "Whereas what kind of kept your eye was the demonstration, particularly helpful demonstration of liquid nitrogen and etcetera, these type of things, and curetting. So the discipline lends itself to this medium." (E: Lines 311-317)

There were varied opinions on the content of the courseware and how it was presented. Some physicians found the content of the courseware somewhat simplistic, but its simplicity also made it practical:

> "There is no problem at all learning from it but it is simplistic, but it's probably good that it's simplistic because I learned more from this than most things, you know." (A: Lines 574-579)

> "They didn't go into a lot of the academic stuff that we'd just ignore anyway. In the basal cell carcinoma one, if they went into the 18 subvarieties of basal cell carcinomas — that would be considered academic; and if they started showing histology of all the different substages, that's of no practical value at all. And for a rural physician, you just haven't time to go into the minutia. You just have too many areas in this life to consider. So the selection of which points to put on were all relevant to a rural physician." (B: Lines 397-434)

One rural physician felt the courseware should contain more evidence-based material:

> "Number one, everything that we do these days we try as much as possible to make it evidence based. And that is to say that it's proven in double-blinded random selected trials, placebo controlled where possible, you can't do that with procedures but ... and sometimes I got things like in my experience it has been okay, and that's great as one piece of evidence, but that is not best

available evidence. And it should have been more along the lines of in seven studies comprising 10,000 people the studies were unanimous to show that this procedure done in this way gives the most effective and least painful way of treating. That's what I'm looking for these days when I'm looking for information. I'm not looking for in my experience. That just means you've done it once. But the experienced side of it should be a corollary to or as an additional piece of information to what is actually proven in the literature. And that's most what I'm looking for in CME information." (G: Lines 404-448)

Another participant felt that rural physicians should play a larger role in presenting the subject matter:

"I mean in terms of content and presentation it might be nice to hear some of the information from the generalist or the rural person's perspective. Rather than just, you know, the subspecialist telling the, you know, the rural generalist this is what you do." (I: Lines 363-379)

The fourth category of perceptions and experiences was related to physicians overall impressions of computer-mediated CME. Table 4.18 presents an overview of physicians' overall thoughts and perceptions of computer-mediated CME. Six sub-categories of overall impressions are indicated including: becoming familiar and adjusting to text-based communications; prior computer experience; cost savings; duration; advice; and overall perceptions. Several physicians felt that computer conferencing was an innovative communications mode for engaging in discourse with colleagues and CME subject matter consultants. However, being "new" and "unusual," several participants felt that adaptation to this communications method would be a gradual process. They suggested that an adjustment and transition period occurs before an individual is able to fully engage and communicate effectively in this asynchronous computer-mediated communications environment.

> "So that was part of it. Like, I mean, not having done that sort of thing before, it took some time to actually familiarize myself with the process of using the thing." (D: Lines 672-676)

> "I think that over time the same group working for a longer period of time together would start to become more voluble in what they were saying. But I think just the - oh I don't know - the reticence, the unfamiliarity, the shyness, if you want, of people will show through in a short little time frame such as this." (G: Lines 168-180)

> "And I know when it first started it was very awkward in the fact that you pushed a button and you didn't know who else was pushing a button, but I must say after you get used to that you get used to it. And you just get into it. So I think you'd get used to it but it's different." (J: Lines 312-324)

Several physicians felt that prior computer experience would be helpful to physicians participating in computer-mediated CME courseware. This prior computer experience would equip learners with a level of computer knowledge and skill which would assist them in understanding the graphical user interface of the WWW. These physicians felt that their prior experiences with computer technology was helpful in understanding how to use the courseware and the computer conferencing system.

> "My impression is those of us who use a computer every day and use it for research every day are very comfortable with all of this stuff because it's stuff we do all the time. But I have colleagues who I know never touch a computer and the whole concept of doing something like this just bothers them." (G: Lines 61-71)

An important advantage of computer-mediated CME was it enabled rural physicians to participate in continuing education at a distance, regardless of geography or time. This means significant cost savings because rural physicians do not have to travel long distances to CME conferences in urban areas, do not have to secure replacement locums, and do not have to travel away from family. Several physicians also felt that computer-mediated CME was a "productivity" tool which enabled rural physicians to maintain their practice while pursuing self-paced and self-directed continuing professional education from their homes or offices.

> "Strength would be -- well, one, that it's presumably would be fairly inexpensive compared to especially for someone like me who's you know, usually have to go to St. John's or whatever, to do any kind of a course. So, I mean, doing it remotely is a lot cheaper, like, from that point of view." (D: Lines 552-560)

> "Well I think one of the major strengths is that there's the potential for real quality medical education where you didn't have to leave home. That it would be -- CME now for a physician is a very significant expense, not only from the point of traveling to where your course might have to be, but the distance has driven you from your business. And also the fact that it's probably the same across Canada, that its difficult now to get locums. And it's difficult to leave your practice. So this is a huge potential resource for docs of all ages." (K: Lines 403-419)

There were two contrasting perspectives on the time allotted for participating and completing the online courseware. Several physicians felt that a two week time period was adequate and provided ample time and opportunity for completing the components of the courseware, and for participating in the computer conferencing discussion. Some comments even suggested that a longer time period could be "demotivating" and could result in higher incompletion or attrition rates. However, other physicians felt that a longer time period for courseware completion would have been more beneficial, facilitated more "voluble" discussion, and provided more opportunity for studying and reviewing the courseware subject matter.

> "I think there was enough time given. I think you'd rather have shorter courses. Like, take less material and give it over less time, then have longer courses over longer time because people tend to give them up and go on to other things, and they're busy." (E: Lines 495-505)

> "I think if you have a longer time frame it's -- you're going to get more people participating. I understand, you know, that you got to cut things off at some point in time, but I think a week or two weeks is definitely too short. Our lives are just so full of stuff. And so that was the biggest problem. It didn't take me any time to do it. But in order to have some chance for feedback, for the discussion and for all of this to take place, it had to be during a time when I could find the time to do it." (G: Lines 285-307)

Participants also provided some advice for future computer-mediated CME learners. This advice included suggestions for ensuring that computer hardware and software was operating correctly, and also taking time to familiarize oneself with the interface of the World Wide Web (navigating).

> "I would just say make sure you have enough time to, first of all, get used to computers if you're not used to them already. But just to take the time to learn the format and be able to navigate around it. I imagine somebody sitting down at a computer for the first time might have some trouble getting around it, but it's pretty intuitive. You know, all the menus are there; once you get used to learning how to click the mouse, it's pretty hard not to catch on." (B: Lines 772-791)

> "They'd have you know, it would be nice to have posted to a bulletin board before, for example, and, you know, to, kind of, know how those things work because if you're new to them, they're pretty -- it takes you a while to figure them out sometimes. I mean, I know you can use the manuals, and that kind of stuff, but with people that are not too familiar with computers, I think, you know, it would be a bit of a challenge. So I would advise them to familiarize themselves with -- a little bit with the medium first." (D: Lines 753-771)

Overall, impressions of computer-mediated CME were very positive. Many rural physicians felt that computer-mediated CME, in combination with selected face-to-face rural medicine workshops, offered great potential and opportunity for the maintenance of competencies and skills.

> "And particularly for those of us in outlying areas it's just the way to go. I think it's a great way to go. I would very quickly latch on to something like this as an ongoing way of maintaining my competence." (G: Lines 754-759)

"I think it's essential. I think it's the only way to go eventually. It really needs to be developed so it's really strong, you know, I think at the moment, you know, doctors needs are met by journals and conferences and some people talk about sort of, you know, the Internet as a tool, right. But in the end it should be mainly the Internet, which is the way of accessing CME really cheaply. And then occasional conferences and some journals, right. So I think in years to come it will be a major component." (L: Lines 465-479)

"But if I have to express myself, I think this is extremely valuable item which the rural medicine doctors can use very wisely." (C: Lines 105-108)

Performance Data - Questions and Results

11. How did the participants rate the extent of knowledge gain or skill improvement they experienced as a result of participating in a CME computer-mediated courseware program?

A retrospective, pretest-posttest self-reporting performance survey was developed to assess the affect of participation in the computer-mediated courseware on the clinical knowledge and skills of participants. The survey consisted of 13 performance statements and subjects were asked to rate the level of their knowledge of, or skill in these performance areas using a five point Likert scale, 1 = "to no extent" and 5 = "to a large extent," both before and after their participation in the computer-mediated course. This survey was forwarded to participants six weeks after completing the courseware.

Twenty-four surveys were returned from participants in the computer-mediated treatment groups (24/30), a response rate of 80 percent. The results of Wilcoxon tests for all performance items revealed significant differences in self-reported performance change between the period before participation in the computer-mediated courseware and six weeks following participation. The results indicate that this significance existed for all 13 performance objectives at the < .10 level of probability. Table 4.18 presents the

Table 4.18 Wilcoxon Test for Significance Between Self-Reported Pre and Post-Learning Performance Change

Performance Objective	Mean (Pre)	Mean (Post)	SD	Mean Rank	Sum of Ranks	Z	Sig.
Identify lesions that are most appropriate for punch biopsy office procedure.	2.92	4.29	.71	11.50	253.00	-4.221	.000
Identify lesions that are most appropriate for cryotherapy procedure management.	2.79	4.17	.92	11.00	231.00	-4.122	.000
Perform curettage technique successfully for the removal of molluscum lesions.	2.46	3.79	1.09	8.50	136.00	-3.622	.000
Identify fungal infections using fungal culture and potassium hydroxide procedures.	2.46	3.58	1.12	8.50	136.00	-3.596	.000
Describe the purpose and possible complications of cryotherapy procedure.	3.33	4.33	.93	8.00	120.00	-3.487	.000
Perform cryotherapy procedure appropriately to manage selected lesions.	3.13	4.04	.97	8.00	120.00	-3.535	.000
Describe the purpose and procedure of fungal culture.	3.13	4.04	.88	8.00	120.00	-3.508	.000
Describe the purpose of the curettage technique for the removal of molluscum.	3.04	4.00	.99	7.50	105.00	-3.360	.001
Perform punch biopsy procedure successfully.	3.21	4.17	.99	7.50	105.00	-3.360	.001
Read and interpret a positive KOH clearing test.	2.13	3.04	.85	7.50	105.00	-3.391	.001
Describe the purpose of punch biopsy and haemostatsis procedures.	3.44	3.96	.59	6.00	66.00	-3.207	.001
Identify molluscum lesions.	4.17	4.58	.58	5.00	45.00	-2.887	.004
Perform haemostatsis procedure successfully.	3.65	4.04	.58	4.50	36.00	-2.714	.007

mean pre and post-learning performance scores, as well as the mean ranked difference between these scores for each performance statement.

The results for performance objectives: "Identify lesions that are most appropriate for punch biopsy office procedure" (Mean Rank = 11.50, p = .000), "Identify lesions that are most appropriate for cryotherapy procedure management" (Mean Rank = 11.00, p = .000), "Perform curettage technique successfully for the removal of molluscum lesions" (Mean Rank = 8.50, p = .000), and "Identify fungal infections using fungal culture and potassium hydroxide procedures" (Mean Rank = 8.50, p = .000) indicated the largest mean ranked differences between pre and post-learning performance periods. The results for performance objectives: "Describe the purpose of punch biopsy and haemostatsis procedures" (Mean Rank = 6.00, p = .001), "Identify molluscum lesions" (Mean Rank = 5.00, p = .004), and "Perform haemostatsis successfully" (Mean Rank = 4.50, p = .007) indicated the least largest differences in mean ranked scores between pre and post-learning performance periods.

Instructional Transactions- Questions and Results

12. How well did asynchronous computer conferencing function in facilitating collaborative learning and establishing a learning network?

Table 4.19 presents several numeric descriptions of the electronic transcript data. These descriptions include participation frequency for each discussion forum, mean participation rates per discussion forum, frequency and percentage of instructor initiated and participant initiated comments.

Table 4.19 Computer Conferencing Participation

Discussion Forum	Participants (N)	Total Messages	Instructor Initiated	Participant Initiated	Mean messages per participant # total participant initiated/ # of participants
Cyrotherapy Forum	8	31	14 (45%)	17 (55%)	2.1
Punch Biopsy Forum	9	26	13 (50%)	13 (50%)	1.4
Curettage Forum	7	13	5 (38%)	8 (62%)	1.1
Fungal Investigation & KOH Procedure Forum	7	14	7 (50%)	7 (50%)	1.0

Participants in the Cyrotherapy forum posted, on average (M = 2.1), more messages than any other discussion forum. As well, the Cyrotherapy conference had the highest number of total messages of all four discussion forums (N = 31). Participants' messages and questions were usually posted in response to instructor comments or to messages posted from other participants. Questions from participants focused on the techniques for performing the procedure and the materials or instruments needed for conducting the treatment. An interesting thread of discussion focused on the unavailability of materials for performing the Cyrotherapy procedure in rural areas. This resulted in a many-to-many discussion concentrating on the use of alternative instruments in performing the procedure.

The curettage forum focused on the presentation and the characteristics of lesions that one would treat using the curettage procedure. The content of the discussion also appeared to exhibit the "collaborative sharing" of professional clinical experiences with the curettage technique, and its success in managing different lesions. Participants also discussed issues concerning the purchase and use of equipment for performing the procedure, where to buy them, and substances for achieving haemostatsis (clotting of blood).

The punch biopsy forum focused on the use of the procedure and issues related to the treatment of lesions associated with the application of the punch biopsy procedure. Participation in this discussion also dealt with the sharing of professional experiences in using the technique and the use of alternative procedures. Some discussion centered around the problems rural physicians' experienced in using the procedure and the time to send and receive a pathological specimen report. Participants also discussed the type of instruments and equipment to use, and where to purchase them.

Table 4.20 presents a schematic participatory table for participation across all the discussion forums. This table provides a schematic overview of the level of participation for each physician partaking in the discussion forums. This table indicates that the level of participation across the four discussions varied according to the actual participants in the forum and the extent of their participation. This suggests "selective participation," perhaps based on the individual learning or information needs Generally, physicians were more inclined to participate in a particular discussion forum, rather than engage in active participation among all discussion forums. As well, higher levels of participation were documented for the Cyrotherapy and Punch Biopsy discussion forums.

The results of the content analysis of the computer conferencing discussion revealed that only half of the participants in the program participated in the online communicative environment. As well, of the physicians that did participate, only four posted three or more comments over a two week period. The majority of participants posted only one or two messages which, for the most part, were usually questions and responses directed to the subject matter experts.

The participants who were active in the computer conference were more likely to engage in collaborative discussion with other participants regarding the pros or cons of using a certain procedure or treatment. However, overall the conference functioned as a Q & A area in which there was collaborative sharing of information and experiences, but the depth of the discussion and extent of participation was limited. This may be

Table 4.20 Computer Conference Participatory Table

Participant ID. # (N=17)	Fungal Investigation & KOH Procedure Forum	Cyrotherapy Forum	Curettage Forum	Punch Biopsy Forum
16	●●●	●	●	●
07	●		● ●	
38	●			
12	●●	●		●
34		● ●		
04		●●		
32		● ● ●		
17		● ● ● ● ●		●
19		●●●		
08		●		●
13			●	
31			●	● ●
27			●	● ●
05			●	● ●
33			●	
18				● ●
37				●

● = Message posted to computer conference.

attributed to the structure of the conference, the perceptions of the users, and the schedule for courseware delivery. In summary, the conference served a limited function and its role could be improved in the future by focusing more closely on its relationship to the content of the courseware program and how it is designed to engage and facilitate a collaborative learning network.

In summary, a vast amount of information was collected to answer the evaluation questions of the evaluation of the CME instructional courseware in this study. The important aspects of the results have already been presented within the categories of antecedent, learning, satisfaction, behavioral, and instructional transaction data. The results presented to this point have demonstrated that CME courseware was effective in producing knowledge gain and improving the self-reported performance levels of participants. The courseware was also successful in providing continuing medical education at a distance to rural and remote physicians and because of this was well received. There were some interesting findings surrounding the relationship between individual learner characteristics, learning achievement, and computer attitudes. Prior computer experience appears to influence both achievement in a computer-mediated instructional environment and attitudes towards computers. These and other major findings of the fieldtest will be discussed in greater detail in Chapter Five.

Results of the Metaevaluation of the CME Courseware Evaluation Model

A metaevaluation, an evaluation of an evaluation, was conducted to address the original research questions stated in Chapter One. The summarized results of these metaevaluation procedures are presented in this section.

Metaevaluation Procedure - Evaluator Self-Report

The evaluator who designed and conducted the evaluation of the CME courseware prepared a self-report that described the strengths and weaknesses of the various methods and instruments used during the evaluation. As well, the evaluator also facilitated a focus group discussion with program developers regarding the strengths and weaknesses of the formative and summative evaluation methods and the usefulness of the information produced in allowing them to make improvements or revise the instructional courseware program.

1. What were the strengths and/or weaknesses of the evaluation model?

The major weakness of the fieldtest of the evaluation model was related to the development and validity of the pre and post-learning achievement test. Several difficulties were encountered during the development of the test items. The time allotted for the development of the test items was insufficient to allow the test to be piloted on a suitable group of physicians prior to its use with the participants in the study. As well, the two subject matter experts (Dermatologists) submitted items of varying multiple choice format (one correct answer and K-type). The K-type multiple-choice item format was very popular in medical and allied health testing in the past. However, Mehrans and Lehmann (1991) note than many investigators have found that K-type items, in contrast to comparable multiple-choice items, tend to be more difficult, less efficient to construct, and more laborious to read. Today, K-type are seldom used in achievement tests, yet the subject matter experts were adamant that they were accurate testing mechanisms and would best measure the participants' knowledge of the content presented in the courseware. Upon review of the items submitted the evaluator found several which were unrelated to the learning objectives originally formulated by CME Planning Committee members. Under the circumstances the evaluator and the CME Planning Committee members then attempted to rewrite many of the items, but the test was still not as well constructed as it might have been.

These development obstacles raised significant concerns about the validity of the achievement test items. And, although the participants' scores improved substantially from pretest to posttest, the percentage scores were lower than expected by the evaluator and the CME Planning Committee. As mentioned, the test items were originally authored by the subject matter experts, but were edited for clarity and readability by the evaluator and CME Planning Committee. However, several of the originally submitted items did not reflect the content nor the instructional objectives that were presented in the courseware. A close examination of the test items and the content of the courseware verified this. The test items were then revised to reflect the content of the program and this resulted in items that required mere rote recitation of facts or other comprehension-level activities. Thus, the practical focus of the content, application and problem-solving in novel patient management situations, was not measured by the test items.

In addition to an examination of the test items and the content of the courseware, an item analysis was conducted by the evaluator for each item on both the pre and post administration of the test. Item analysis is the process of examining the students' responses to each test item -- to judge the quality of the item. Specifically, what one looks for in conducting an item analysis is the difficulty and discriminating ability of the item.

A difficulty index was calculated for each item. The difficulty of a test item is determined by the percentage of learners who answer it correctly (Mehrens & Lehmann, 1991). An item difficulty index of .50 would indicate that 50 percent of the students answered that test item correctly. Kryspin and Feldhusen (1974) provided guidelines for assessing item difficulty levels and these are summarized in Table 4.21. Mehrens and Lehmann (1991) suggest that the ideal average difficulty for a maximally discriminating four-response multiple-choice test is .74.

Table 4.21 Difficulty Levels of Test Items

Difficulty Index	Difficulty Level
0 - .25	Hard
.26 - .74	Average
.75 - 1.00	Easy

The discrimination index for a given test item is performed by ranking the learners' sum test scores and then equally dividing the students into two groups, learners who performed well and learners who did not perform well. The discrimination index is then determined by subtracting the number of low learners who correctly answered the item from the number of high learners who correctly answered the item and dividing that figure by the number of learners assigned to each group. In summary, the index indicates how well the students who performed well on the test performed on an item in relation to the performance of students who scored poorly on the test.

If a test item truly discriminates, then the index should be positive indicating that students who performed well on the test performed better on the item than students who performed poorly on the test. A negative discrimination index indicates that more students near the bottom of the ranking fared better on the item than did students ranked near the top of the group. Kryspin and Feldhusen proposed the guidelines in Table 4.22 for assessing discrimination indices.

Table 4.22 Discrimination Levels of Test Items

Discrimination Index	Level of Discrimination
.40 - 1.00	Item discriminates well
.20 - .39	Item discriminates moderately well
.00 - .19	Item discriminates poorly
-1.00 - .00	Item discriminates negatively and needs revision or rejection

For the item-discrimination index, the value is expressed as a decimal and ranges from -1.00 to +1.00. If it has a positive value, the item has positive discrimination. This means that a larger proportion of the more knowledgeable students than poor students (as determined by the total test score) chose the right item. If the value is zero, the item has zero discrimination. If more poorer than better students get the item right, one would obtain a negative discrimination.

The higher the discrimination index of a test item, the better. However, according to Mehrans and Lehmann (1991) educators should try to have achievement tests that are of appropriate difficulty as well, because test difficulty is related to discrimination power. If an item is so easy that everyone answers it correctly, or so hard that no one can answer it correctly, it cannot discriminate at all and adds nothing to test reliability or validity. Although what is more important than the level of difficulty is to have a test that possesses adequate content validity. But, it is also important to have a test in which, for each item, a larger proportion of the better able than less able students can answer the item correctly.

Difficulty and discrimination indices should be considered together when analyzing test items. When the difficulty and discrimination indices for the items on the pre and post-achievement tests of this study were calculated, several poor items were identified. These items were either hard items on one or both the pre and posttest and/or also revealed very low, zero, or negative discrimination indices. The validity of two test items in particular were very questionable because the difficulty index decreased while the discrimination index increased from pre to posttest. These results suggest that the items may have been poorly written, ambiguous, or had poor content validity. The difficulty and discrimination indices for the tests are presented in Table 4.23.

A positive finding relating to the item analysis was that many of the items had average or near average difficulty indices between .50 to .80 and their discrimination indices were also of a high positive value. Nevertheless, the item analysis indicates that

Table 4.23 Test Difficulty and Discrimination Indices

Test Item	Pretest		Posttest	
	Difficulty	Discrimination	Difficulty	Discrimination
1	.50	.00	.07	.14
2	.75	.25	.90	.14
3	.50	.25	.93	.07
4	.88	.00	.90	-.14
5	.19	-.38	.50	.62
6	.81	.38	.77	.28
7	.38	.00	.80	.35
8	.38	.00	.68	.62
9	.38	.25	.77	.14
10	.88	.25	1.00	-.07
11	.44	.16	.53	-.14
12	1.00	.00	.90	.14
13	.63	.00	.53	.21
14	.50	.25	.40	.21
15	.38	.25	.90	.14
16	.38	.00	.97	-.14
17	.75	.38	.60	.35
18	.50	.00	.63	-.07
19	.25	.25	.90	.14
20	1.00	.00	.83	-.07
Mean	.57	.15	.73	.15

several of the test items were of little or no value and these findings raise sufficient evidence to question the validity of the results.

On the positive side, a major strength of the evaluation model is related to the electronic data collection and analysis procedures which were developed and used. For the computer-mediated learners, the completion of the Demographic Profile Survey, the Pre and Post-Learning Achievement Tests, and the Courseware Evaluation Survey were all conducted online. This minimized the use of pencil-and-paper tests, reduced the resources required to conduct the study, and enhanced the data collection and analysis processes. As well, another unique aspect of implementing evaluation research in an online WWW environment is the efficiency with which subjects may be randomly assigned to experimental groups. These features of the data collection and analysis processes are discussed in this section.

As discussed in Chapter Three, physicians in the two experimental groups were required to complete and submit an online pretest and/or a Demographic Profile Survey before they could link through their Web browser to the hybrid courseware. In order to facilitate these data collection procedures in an electronic instructional environment participants in the experimental study groups were directed to a universal resource locator (URL) or Web site address on the start date of the computer conference and the delivery of the courseware. A CGI PERL script program was developed for managing password controlled access to pretests and the DPS, and for recognizing the location of the CD-ROM drive on each learner's computer. Passwords were assigned to all learners upon registration for the evaluation study.

When learners accessed the hybrid Web site for the first time they were randomly assigned to either the Demographics Profile Survey or both a pretest and the DPS, depending on experimental study group assignment. Randomization was performed by the TGRADE software, designed by the program developers, and students were assigned to sub-lists in the roster on that basis. The CGI script would deliver the appropriate set of test/survey pages based on the label in the password file which the student had to enter when they reached the Web site.

The online posttest was also developed to be completed online and was automatically posted to the CME Web server by the TGRADE software after the computer conference had been facilitated for five working days. This was designed to encourage participants to logon to the online discussion and interact with peers, rather than reviewing the courseware case studies and learning tutorials, and submitting their posttest without having reviewed or participated in the discussion threads. The computer conference was open for two weeks, beginning on a Monday and concluding on the Sunday of the following week.

In order to produce the interactive pre and posttests and online survey components (Demographic Survey Profile and Course Evaluation Survey) of the courseware, developers used the Common Gateway Interface (CGI) protocols of WWW servers. World Wide Web servers recognize CGI script and enable interactive functionality, rather than simply sending a HTML page to a browser at a receiving end. CGI permits a bi-directional transfer of information. CGI scripts are small computer programs, usually written in specialized scripting language, either the C programming language or in PERL.

When a CGI script is activated, an HTML form is sent to a Web browser requesting the designated universal resource locator (URL) or Web site address which has the CGI reference in it. In this courseware, these were the files with the pre or posttests, or the demographic profile survey. These electronic survey and test forms had fill-ins for name and student number, and radio buttons for selecting responses to test or survey items. These HTML pages have CGI form structures within them, associated with elements of the questions which the learner manipulates. When all the blanks are filled and the buttons pressed, the learner clicks a final button to submit the form back to the server. The information about these responses are then processed by a specialized CGI program designed to extract information from the form.

The Exammail software, developed in the PERL scripting language, was used in this courseware for enabling form submission functionality. This software takes a HTML document which has been configured in a specified format with multiple choice questions and student name input, marks the quiz by comparing it against a configuration file which is maintained on the server, and sends back a HTML page grade report to the learner. If the learner has entered their e-mail address, it also sends them an e-mail message with the results of the test. Any particular test can be configured to send just the score, a table with the students' wrong answers compared to the right answers, or a table with all the students' answers compared to the right answers. An identical e-mail is sent to the course administrator or evaluator. The software can also be configured to have a link to a feedback page which is pointed to from the grade report. This HTML page has the full question, with a right answer, and a justification for the right answer.

The software TGRADE, an electronic grade-book with comprehensive reporting capability and TQUEST, a question bank program developed for managing test questions and assembling and printing paper-based examinations, were used in this study. As a first step in preparing for this evaluation study, extensions were written to TQUEST to automate the generation of the three files necessary in creating a test which could be executed by the Exammail script. TQUEST has the capacity to manage 11 different question types in the range from essay, short answer, fill-in-the-blank questions and several types of multiple choice questions for paper-based exams.

Software for automating the uptake and analysis of individual learner results to the grade book program was also developed. This software extracted files from an electronic mail box and reassembled the information into a database structure. The software program extracted learners' names and student identification information, quiz identifiers and answer keys, performed the computations, and recorded the responses in a standard TGRADE software file structure. The program would only allow one quiz execution to be recorded for each student. Once the quiz and survey response data was extracted by TGRADE, analysis and reporting of the responses permitted statistical reports on class performance with mean, standard deviation, standard error, and the generation of histogram tables of class responses for each question. An extension for export of response data to the software Statistical Program for the Social Sciences (SPSS) was also developed. The SPSS 8.0 software was used for summarizing and analyzing the quantitative data in this study.

In summary, the data collection and analysis procedures were greatly enhanced by the use of electronic CGI submission forms and the testing software which was developed for extracting and managing information in database formats. These procedures minimized the amount of disruption that the learners may have encountered were excessive numbers of paper-and-pencil tests and surveys forwarded to them for completion and return. The tests and surveys were completed and submitted online by the learners using their Web browsers. These procedures resulted in an efficient means for collecting data on achievement and self-reports and are a major advantage of the CME Courseware Evaluation Model presented in this study.

Another strength of the evaluation model is related to the introduction and integration of formative evaluation methods during the analysis and design phases of courseware development. The evaluation model advocates, in principal and practice, that formative evaluation should occur concurrently during CME planning and with a CME Planning Committee who may serve as product testers as well as content reviewers.

During the initial analysis stage of formative evaluation the CME Planning Committee members are jointly responsible for interpreting instructional needs and assisting the evaluator-as-practitioner and the subject matter experts in identifying a proposed content scheme which addresses the audiences' needs. Because the CME Planning Committee includes membership from the target audience (family physicians) they are able to comment and provide critical input on the type of information that is most applicable and the means through which it could best be communicated to enhance understanding.

In the formative design evaluation phase paper-based instructional strategies are prepared by the practitioner-as-evaluator and these are reviewed and critiqued by the CME Planning Committee members for content accuracy, relevance, comprehensiveness and bias. Once again, the knowledge of the members is pivotal to keeping the design on track. This phase is followed by a series of storyboard evaluation meetings in which an overview of the courseware's screen layouts, interface elements, and content presentation strategies are tried out with the CME Planning Committee members. The CME Planning Committee is utilized as much as possible at this phase, for trying-out the storyboards and for providing critical feedback on the draft courseware.

To summarize, the major weakness of the evaluation was related to the validity of the achievement test data. However, this is a weakness with the particular test and the items which were developed and used. It is not a weakness of the methodology nor the model. This weakness did not seem to be a major problem in implementing the evaluation model and was not raised as a concern by a majority of the decision-makers.

On the other hand, the major strengths of the evaluation were related to the utilization of electronic data collection and management procedures. These processes increased the efficiency of gathering and analyzing information, and suggests an evolution of data collection methods from the traditional paper-and-pencil instruments. Another major strength of the evaluation was related to the integration of formative evaluation procedures during the analysis, design, and development work of the CME Planning Committee.

Metaevaluation - Focus Group with Program Developers

A second metaevaluation procedure used to collect information to address the first question was a focus group meeting with the program developers of the courseware product. The purpose of the focus group was to explore the opinions of the program developers regarding the formative and summative methods used in evaluating the courseware product. The developers were also asked to comment on how useful they felt the methods were, as well as how effective the information from the summative evaluation would be in assisting them to make changes or revisions to the courseware in order to improve its quality. A summary of the responses from these individuals are presented in tabular form (Table 4.24) to facilitate data presentation. The responses were edited and combined for ease of representation.

Metaevaluation Procedure - Interviews with Key Stakeholders and Decision-Makers

The third metaevaluation procedure collected information to answer the remaining four research questions. For each question, one or more evaluation standards from The Program Evaluation Standards (1994) were selected and specific interview questions developed. Using these specific questions the evaluator interviewed five key stakeholders and decision-makers. The responses to the these questions served as the basis for formulating responses to research questions two through five.

A summary of the responses of these individuals to the specific questions are also presented in tabular form to facilitate data presentation. Additional questions were also asked during the interviews that were not directly related to any of the study's research

Table 4.24 Focus Group Comments: On the Summative Courseware Evaluation Survey and Interview Data

I'm wondering if the responses were related to the fact that there were a different number of people doing it that have a different experience with PC's. It might have been part of their attitude or part of their experience. I think like 70% of the people who were actually involved with the project had moderate or extensive use with PC's. And there was another 20% or so that had none or little. Also the quality of their machine probably would have effected what they received as well. That was a concern for us. I think personally on my side of it, obviously the things that are going to benefit us are the things that are directly referenced to appearance and interactivity. And I think that receiving number data like this doesn't really give us a place to step from to know exactly what we need to change.

It's an indication of their general rating. What's more useful is actual comments directed on specific components. And I don't know if there's anything discussed about the different types of PC's involved and stuff. Because that definitely did play a factor in the success for some of the people. For instance, I got a call from someone from ________ and he was on a 386. So I suggested that he use a PC elsewhere instead of his office PC because I don't think he ever used the Internet before and that was just a little bit too slow.

I think that in getting feedback from the learner we need to know certain things. Is the feedback coming from them having a low end machine and does part of the problem have to do with that. Or is it part of them actually being critical of what they're seeing on a high end machine. If you knew the type of PC they were using then the response would be useful to you. It would give a better idea of whether or not it's independent of the hardware. What they have and what they use. That's important in developing and trying to minimize any type of technical problems. Cause when you're sending stuff out everywhere, it's a challenge to try and make sure that you have everybody held in a net. You know, just in case anything goes wrong.

I think that the break down of the questions, like the different components, was appropriate. I don't know if it would help in giving an opportunity to have a comment box that related just to design maybe? Quality instead of quantity. And that would tell us if it is a machine related thing or if it is layout or quality on our end. So that's definitely something to look at. These do tell us something, these numbers. It doesn't really totally address exactly what the issue was, but it gives us a heads up so that we could critically look through what was developed and increase the usability of that.

Now I think the two things that really stood out for me were based on overwhelming responses from the participants. It wouldn't matter if they were using low or high machines because the percentage from what I remember are using low machines. But a lot of people replied that the video, they didn't enjoy the video. They enjoyed it but the actual presentation was not what they would like it to be. And they found that the conferencing system was a bit weird. They're the two things that really stood out. And that's useful in doing it over again because you'd probably be more inclined to use another type of conferencing software system.

Table 4.24 (continued) Focus Group Comments: On Formative Evaluation

I guess that when we get a reading back and you see that half the people had a problem with the navigation then yeah that would be the thing that while we're story boarding we'd look for ways to improve that. Also in finding those ways to improve that. When we actually start to develop them that we run them by our little test group that we select. And actually sit there and I think it would be important for us to be there to see what the person, the learner might be having issues with. I know that the only two people I know that really read through it were Dr. ______ and Dr. _______. And at that point I mean if someone was there with them and we could see where their questions were coming from. Where they were missing some things, or they skipped sections, or they weren't sure where to go.

I think that the people that were selected as far as Dr._______ and Dr. _______ we shouldn't select anyone that's ever come to us and asked us questions about how it's going. Obviously because they would have been given a heads up to what some of the abilities are. So it would be best to select some people from outside of our department first off.

I'd bring them in cold and you'd probably want three. You wouldn't want an even number. You'd want an odd number.

I think the best thing would be to let them go through it first and have them raise whatever issues come up. But if there are things they didn't touch on, or really didn't fit and you wanted to know about it then you could ask them. But I wouldn't ask them questions first. I'd let them go through and see what naturally comes up first.

Well maybe it would be better before the actual story boarding process to actually give someone our strategy and how we're going to approach it. Cause we'll obviously have some sort of strategy. Like we'll know what media we're going to use and what format we're going to deliver in. And maybe bounce if off of someone before we actually sat down.

questions. However, these questions did provide information about how well the evaluation model functioned during the fieldtest and this information has implications for the study.

The research questions, specific evaluation-standard referenced questions, and the summarized responses of the respondents are presented in Tables 4.25 to 4.28. The responses were edited and combined for ease of representation. The responses to the questions not related directly to any of the research questions are provided in Table 4.29.

In summary, the metaevaluation procedures successfully identified the strengths and weaknesses of the evaluation model. These strengths and weaknesses are briefly discussed in this section. As well, a number of the results also have implications beyond the study and are discussed in Chapter Five and Six. Table 4.30 contains the original research questions and a brief answer to each question. The answers are based on the results of the metaevaluation.

The metaevaluation data helped identify those procedures and information that were most useful to the program developers, stakeholders, and decision-makers. The preferred methods and information appeared to be the interview (reaction and satisfaction) data, the learning outcome data (including the study design), the formative evaluation procedures, and the behavioral data. The validity of the cognitive tests was questionable, but the majority of respondents felt that this information was useful to them in decision-making. There are areas for improvement during the formative and summative evaluation stages of the model and these are discussed in greater detail in Chapter Five.

Table 4.25 Responses to Research Question 2

Research Question 2 Was the evaluation model useful in serving the information needs of the intended users?	
Specific Questions	Stakeholders' Responses
Were all persons involved in or affected by the evaluation identified, so that their needs could be met? (U1)	Yes. But, a post-study meeting involving the people who advised and developed the product would be helpful, so we could all get together and have a discussion about the report and what we saw, and how we felt. It would have been nice to have a group session to discuss these things.
Was the information collected by the evaluation model broadly selected to address any pertinent questions you or other stakeholders had about the program? (U3)	Yes, there was a fair bit of information collected and we should be able to make good decisions based on the report. It looked at all the aspects that could have been evaluated, proving in each and every case that it was an effective form of education. And each time you did that it supported the other forms of evaluation. It was very extensive and thorough. Perhaps more tables as opposed to so much discussion. Only because discussion gets less well read. When I'm reading documents I tend to look at tables and take information from that rather than really lengthy discussion.
Did the evaluation report clearly describe the program being evaluated, including its context, and the purposes, procedures and findings of the evaluation, so that essential information was provided and easily understood? (U5)	Yes it did. I found it described the different target groups very clearly.
Was the evaluation planned, conducted, and reported in ways that encouraged follow-through by stakeholders, so that the likelihood that the evaluation would be used was increased? (U7)	Yes I believe so. When I read through it I was enthusiastic and that made me want to continue on.

Table 4.26 Responses to Research Question 3

Research Question 3 Did the evaluation model follow practical and feasible means for collecting evaluative information?	
Specific Question	Stakeholders' Responses
Were the evaluation procedures practical in that they kept disruption to a minimum while needed information was obtained? (F1)	Yes, from looking at the evaluation comment that went fairly smoothly for them. The online pre and post tests are a very unobtrusive evaluation method and participants were able to do it on their own time. And the interviewing for the evaluation was also conducted in a convenient way.
Was the evaluation efficient in that it produced information of sufficient value, so that the resources expended on it could be justified? (F2)	Absolutely. Yes definitely. It was a very comprehensive report. And I think that the funding that was received was spent very well.

Table 4.27 Responses to Research Question 4

Research Question 4	Was the evaluation model conducted in an ethical manner, with due regard for the welfare of those involved in the evaluation, as well as those affected by its results?
Specific Question	Stakeholders' Responses
Was the evaluation designed to assist the organization in addressing and effectively serving the needs of targeted participants? (P1)	Yes. Certainly from my office's perspective I think it provided an opportunity to look at the attitudes of the participants toward the program, content and the technology. It also provided information which would allow us to improve and expand on this concept in the future, so we could customize it towards the target groups that we're trying to educate. So yeah I think it definitely served that purpose.
Was the evaluation designed and conducted to respect and protect the rights and welfare of human subjects? (P3)	Yes.
Was the evaluation complete and fair in its examination of strengths and weaknesses of the program being evaluated, so that strengths could be built upon and problem areas addressed? (P5)	Yes, I think it was quite objective.
Were the evaluation findings made accessible to the persons affected by the evaluation?(P6)	Yes.

Table 4.28 **Responses to Research Question 5**

Research Question 5	**Did the evaluation model convey technically adequate information about the features that determined the worth or merit of the program being evaluated?**
Specific Question	Stakeholders' Responses
Was the program being evaluated clearly and accurately described and documented? (A1)	I would have to say yes. Absolutely.
Were the purposes and procedures of the evaluation described in enough detail so they could be assessed? (A3)	Yes.
Were the sources of information in the evaluation described in enough detail so that the adequacy of the information could be assessed? (A4)	Yes. You described why you were doing certain things and how you were doing it. I mean I'm not an education expert but at least I know that this is based on sound principles and you know research that's been done by experts in the area. And you followed those pre-existing guidelines.
Did it appear that the quantitative and qualitative information in the evaluation was appropriately and systematically analyzed? (A8) & (A9)	Yes.
Were the conclusions and recommendations presented in the evaluation report supported by the data? (A10)	Yes.

Table 4.28 (continued) Responses to Research Question 5

Research Question 5	Did the evaluation model convey technically adequate information about the features that determined the worth or merit of the program being evaluated?
Specific Question	Stakeholders' Comments
Were the information gathering procedures described in enough detail for their validity and reliability to be assessed? (A5 & A6)	The numbers may have been on the low side. And, I would have to question the validity of the questions on the achievement test because I wasn't sure where they came from and who developed them and how they were developed. You could improve upon that by having a broader group develop them and evaluate them before the program. You may ask other Dermatologists or you could look at a question bank. The behavioral survey was very helpful and I think that's where CME is moving towards, not so much what did I learn at the end of the program at minute 0, but what will I learn 6 weeks and 6 months and then 2 years down the road because that's really where you assess whether you learned anything at all. The validity of self-reporting is always a concern, but I think from the point of view of six weeks afterwards it's probably fairly reliable. I think when we get into six months there are so many other variables involved that it's difficult to say that this is because of your particular CME programming. Another survey at six months may also have been helpful because if it gave you negative results and said there had been no change then clearly you'd have to look at what you do to say well you know I'm not meeting the expectations in the long term.

Table 4.29 Responses to General Questions

General Question	How effective was the evaluation model when applied to a computer-mediated CME program?
Specific Question	Stakeholders' Responses
Would you change any aspect of the evaluation model or methods in the future?	It's a costly endeavor to do chart audits and observations of physicians in the field, especially when you are doing a rural outreach educational program. You would have to go into the practice and have full access to the charts on specific Dermatology cases without them knowing ahead of time. That would be ideal if you could get someone to agree to do it. You could also do a pre and post look at their practice patterns before they started the course and look at the practice patterns after they finished the course. But that type of information would not change, to a large extent anyway, the opinion I have formed of the effectiveness of the program from the evaluation information which has been provided to me. What I think would be more useful is to go back 6 weeks or 6 months later and ask them you know remember the CD-ROM, what other things would you have liked to have learned or what changes have you implemented now that you have the knowledge and you've used it. Or talk to them 6 months later and ask them point blank, "Has this CD-ROM or course changed your style of practice?" And I think most people will be honest and say "Yeah I recognized something that I wouldn't have recognized before." I've done a skin biopsy and I've never done one before" or "I felt comfortable using this drug or I take cultures now or I treat warts." And you could ask them "What changes have you made in you practice?" "I now do skin biopsies, I now do liquid nitrogen, I now do fungal culture. If you had lots of money and you lived in a perfect world then a chart audit would be a more reliable way to go. But do we live in a perfect world? No. I'm just not fond of multiple choice because the are really just testing recognition. So you may want to use other tests rather than that. You could ask somebody "What would you do with condition X" you know "how would you manage it?" To get something that's application rather than recall.

Table 4.29 (continued) **Responses to General Questions**

General Question	How effective was the evaluation model when applied to a computer-mediated CME program?
Specific Question	Stakeholders' Responses
What methods were the most useful?	I don't think that anything was most useful or least useful. I believe like it's different bits of information. So everything is useful. I think the pre and post testing and the structuring of the pre and post testing with the control group who received no CME, and then the pre and post test versus the just the pre-test group. I thought that was quite informative and useful.
What methods were the least useful?	None really.
What data would you most rely on?	The pre and post test information, the statistical data, and the comments. I think the comments that the target group made were most useful.
What data would you least rely on?	I think all of its important. But self reporting is probably the least reliable. So anything other than that. From an external perspective, I mean if I'm doing an evaluation on a person whose say doing a clinical traineeship, I would rely more on what the preceptors told me about the person's strength and weaknesses after the fact and what the person did.

Table 4.30 **Summary of Responses to Research Questions**

Summary of Responses to Research Questions		
Research Question		Response
1.	What were the strengths and/or weaknesses of the CME Courseware Evaluation Model?	The major weaknesses were related to the validity of the achievement test, short-term vs long-term evaluation, and interpretation of the quantitative CES reaction data. The strengths were it was thorough, provided in-depth qualitative responses, utilized electronic data collection and analysis procedures efficiently and effectively, and integrated formative evaluation during ISD.
2.	Was the evaluation model useful in serving the information needs of the intended users?	Yes, the evaluation was comprehensive and systematic in the collection and dissemination of information.
3.	Did the evaluation model follow practical and feasible means for collecting evaluative information?	Yes, the stakeholders and decision-makers felt that the resources spent in conducting the evaluation were justified.
4.	Was the evaluation model conducted in an ethical manner, with due regard for the welfare of those involved in the evaluation, as well as those affected by its results?	Yes, the evaluator was ethical and responsible in conducting the evaluation and the use of electronic survey mechanisms were unobtrusive and increased the efficiency of data collection and analysis.
5.	Did the evaluation model convey technically adequate information about the features that determined the worth or merit of the program being evaluated?	Yes, however there is room for improving the validity of cognitive test items as well as measures of behavioral change.

Among the types of data collected, the decision-makers and stakeholders felt that the qualitative interview responses and the behavioral survey data were most helpful to them. However, several of the respondents suggested that a longer-term investigation of the impact of the program on self-reported knowledge or skills would be helpful. They also reported that a major strength of the evaluation model was related to its comprehensiveness and that it collected information on reaction, cognitive, and behavioral data using a variety of collection methods. There were mixed results regarding the qualitative data. One of the stakeholders found the qualitative information of little use and overwhelming. However, the program developers and other decision-makers found this information very useful.

In conclusion, the results of the metaevaluation demonstrated that the evaluation model, as implemented in the fieldtest, was successful in providing useful evaluative information to program developers, stakeholders, and decision-makers. The model produced valuable information which could be used to make decisions about the effectiveness of the program.

Summary

The data gathered during the study were presented in this chapter. The results of the fieldtest of the CME Courseware Evaluation Model were presented in conjunction with the appropriate evaluation questions. These results were also arranged according to the type of data gathered, antecedent, reaction, learning, behavioral, and instructional transactions. The results of the metaevaluation of the evaluation model were also provided in this chapter. A self-report prepared by the evaluator and a focus group meeting with the program developers highlighted the major strengths and weaknesses of the model and served as a response to the first research question. Responses to the other four questions were provided based on information which was collected during interviews with the program's stakeholders and decision-makers. In Chapter Five of the dissertation the results of the fieldtest of the evaluation model and the metaevaluation are discussed.

CHAPTER V

DISCUSSION

Introduction

In this chapter issues related to the results of the fieldtest of the CME Courseware Evaluation Model and the metaevaluation are discussed. The major issues which are addressed focus on the antecedent data, cognitive data, satisfaction data, instructional transaction data, behavioral data, evaluation procedures, and the evaluation model. These seven issues are discussed relative to the results of the fieldtest and metaevaluation and to issues examined in previous chapters of the dissertation.

Antecedent Data

Antecedent data was collected during the fieldtest of the courseware evaluation model to assess the relationship between physicians' individual demographic and computer characteristics and attitudes toward computers. The literature presented in Chapter Two suggested that several individual learner characteristics, gender, age, prior computer experience, and computer ownership influenced learners' attitudes toward computers (Nickell and Pinto, 1986; Busch, 1995; Shashaani, 1994; Woodrow, 1994).

Many of these studies supported the existence of persistent gender-differences in attitudes toward computers. The results of this study support these findings (relative to a volunteer sample of physicians). A Kruskal Wallis Analysis of Variance of Ranks tested the relationship between gender and physicians' attitudes towards computers. A significant relationship at the < .10 probability level was revealed. Male physicians were found to express more positive attitudes toward computers than females.

The literature also suggested the existence of a relationship between computer experience and an individual's attitude toward computers. Both Busch (1995) and Shashaani (1994) found computer experience to be strongly correlated with computer attitudes, suggesting that students with high levels of computer experience tended to have more positive computer attitudes. Fann et al. (1989) also found that students with more computer experience were more likely to have positive attitudes toward computers than those with less experience. In this study a significant relationship was revealed between physicians' self-reported computer experience and frequency of software use and attitudes towards computers. Physicians reporting greater experience and more frequent use of computer software reported more positive attitudes towards computers.

Several investigators have also suggested that computer ownership and age influence learners' computer attitudes as well. Shashanni (1994) found that students with access to home computers demonstrated greater interest in participating in computer-related activities, higher attitudes towards computers, and greater knowledge of computers. Similarly, Nickell and Pinto (1986) found that younger people tended to have more positive attitudes towards computer technology than older student cohorts. However, the findings of this study indicated that access to a home computer and/or an office computer had no significant influence on the computer attitudes of physicians. As well, the number of years of physicians' practice experience did not affect their attitudes toward computer technology.

According to Busch (1995) attitudes towards computers and perceived computer self-efficacy expectations represent a very important issue in the area of computer-mediated instruction. Computer attitudes and perceived self-efficacy are believed to affect individuals' interest in using computers, enrollment in computer courses, and the motivation and persistence of learners to succeed in a computer-mediated learning environment. Ertmer et al. (1994) have suggested that individuals with less confidence or poor self-efficacy beliefs of computers are more likely to underachieve in computer-based learning environments. Therefore, to examine the influence of physicians' attitudes toward computers on achievement in computer-mediated instruction, the scores from the participants' Computer Attitude Scales were analyzed relative to their posttest learning achievement scores.

The findings of this analysis suggested that physicians' computer attitudes had little influence on achievement or success in the computer-mediated CME courseware on Dermatological Office Procedures. The analysis of the relationship between computer attitudes and the posttest learning achievement scores was non-significant. However, it should be noted that external validity or representativeness of the fieldtest results was acknowledged to be extremely low from the beginning of the study. Due to the design used for the fieldtest, there was little or no expectation that results could be generalized to other populations. Indeed, generalization of the results of the fieldtest was not a purpose of the study.

Nevertheless, this is a positive finding and it suggests that the computer-mediated instructional courseware which was delivered in this study was an equitable means for participating and succeeding in continuing education at a distance. But, the findings of several other antecedent characteristics may have implications for future courseware development and delivery efforts. For example, the analysis of the antecedent data did reveal a relationship between computer experience and learning achievement. This raises several issues, particularly if this form of CME is to have broad application for physicians with varying computer literacy skills. The findings do suggest that physicians with high levels of computer experience were more successful in cognitive learning and this lends greater support to claims made by many practicing physicians regarding the critical need for training in information technology.

These findings have implications for undergraduate, postgraduate and continuing medical education programming. Are CME offices, medical schools, and the rural physician themselves focusing enough attention on the enhancement of computer literacy skills? The implications for improvements in medical informatics training for rural and remote physicians are significant as well. The current trend in telemedicine and medical informatics, as this study clearly demonstrates, is towards increased usage and application of information technology in medical practice. The rural physician stands to gain the most from enhanced computer literacy skills given the future outlook for telemedicine and information technology in rural medicine. Information technologies will play a greater role in sustaining the activities of rural communities in the future, it is essential that rural health care practitioners are able to take advantage of these same information technologies for maintaining their knowledge and skills in rural medicine.

The antecedent information collected in the fieldtest of the CME Courseware

Evaluation Model was useful in examining the relationship between individual learner characteristics, computer attitudes, and learning achievement. However, given the findings of the fieldtest it is questionable whether the CAS survey is a necessary component for an effective evaluation of computer-mediated CME courseware. The findings of this study do warrant a need to examine the influence of computer attitudes and individual learner characteristics to a greater extent with larger randomized samples of physicians or other health care practitioners. Nevertheless, the data collected from the CAS survey did not provide any meaningful information which was useful to the decision-makers, stakeholders, or program developers regarding the effectiveness of the instructional courseware. It did serve a purpose for exploring the relationships between the various demographic and computer characteristic variables, but as a measurement instrument for informing stakeholders and program developers it was not helpful. The data collected from the Demographic Profile Survey was helpful in ascertaining the type of learner participating in the instructional courseware program, but only certain items provided relevant information for decision-making (computer experience, computer software and Internet application usage). In summary, the CAS survey provided some interesting "exploratory" antecedent information but did not provide data which could be utilized for improving the courseware product or decision-making. Several of the items of the DPS survey were useful in interpreting the results of the Courseware Evaluation Survey and it is suggested that these items be included as part of the CES in the future.

Cognitive Data

Did participants learn? This is a fundamental question for any systematic and comprehensive evaluation of educational effectiveness. Many of the evaluation approaches discussed in the literature review of Chapter Two identified the importance of an evaluative category which served to measure learning achievement or knowledge gain as a result of participation in an educational program (Coldeway and DeLisa, 1986; Cervero, 1988; Abrahamson, 1968; Reiser and Kegelmann, 1994; Gill et al., 1992; Dillon and Gunawardena, 1992). In particular, Cervero's (1988) fourth evaluative category of his continuing professional education evaluation framework focused on the assessment of changes in learners' cognitive, affective, or psychomotor competence. Cervero suggested that the pretest-posttest evaluation system was devised largely to assist in the determination of the extent of learning achievement as a result of participation in a continuing professional education program. The pretest is administered at the beginning of the program followed by the same measurement at the conclusion of the training. The difference between the scores allows judgements on the effectiveness of the program.

Following a similar line of thinking, the fieldtest of the evaluation model implemented a modified pretest posttest control group study design to assess the change in learning achievement of participants in computer-mediated CME, and also to compare that change with physicians receiving no-CME instruction. Learning achievement was measured by a 20 item multiple choice cognitive test and identical test items were used for pre and posttests. The findings of the fieldtest of the evaluation model revealed significant differences between the posttest scores of physicians participating in computer-mediated CME and physicians receiving no-CME instruction. Physicians in the computer-mediated CME study groups scored significantly higher than participants in the no-CME group. As well, a significant difference between the pre and posttest scores of the computer-mediated experimental study group revealed that a computer-mediated CME instructional courseware program was effective in producing knowledge gain of Dermatological Office Procedures.

Spencer (1991) has described the effect size (ES) as a "measure of the educational importance of any performance changes produced" (p.14). The effect size is estimated by using the average score difference between treatment and control groups and dividing it by the standard deviation of the control group. An ES of 1.0 means that the innovation has increased the performance of the group by an amount equal to one standard deviation unit of the control treatment. This would take an average student from a position in the middle of the control group to the position occupied by the top 20 percent of that group (Spencer, 1991). This is a large effect and is educationally significant. The ES between posttest achievement scores of participants in the no-pretest experimental group and the control group was ES=1.44 indicating educational significance of participation in a computer-mediated CME activity as compared to participation in no-CME instruction.

Among the cognitive results, the issue of the poor quality of the test was the primary concern. Certain logistical constraints, such as the lack of time to pilot the test prior to its use in the evaluation, were considered previously. Suggestions for improving the test are discussed in this section.

A major drawback of the cognitive achievement test, apart from the validity of the test items, was it did not test the participants' ability to apply or generalize the procedures or treatment techniques to other patient-management problem situations. A majority of the test items focused on the recall of facts and information. The test could be improved substantially if more of the test items were rewritten to test the interpretation of novel patient-management problems, the application and generalization of knowledge and understanding to those situations, rather than the recall of information.

One procedure for improving the test items could include the use of the interpretive exercise described by Mehrens and Lehmann (1991). The interpretive exercise consists of "either an introductory statement, pictorial material, or a combination of the two, followed by a series of questions that measure the student's ability to interpret the material" (p. 144). This procedure could be exploited best by the hypermedia capabilities of HTML and the WWW (integration and presentation of multiple media, text, audio, video, animation and graphics). As well, the interpretive exercise lends itself to patient management problem situations in which physicians often find themselves. It requires the learner to interpret the materials which are provided to them and extrapolate from the information which investigations, treatments or management protocols are required. According to Mehrens and Lehmann, such exercises are finding their way into more teacher-made tests because of the many advantages they have over traditional items:

> 1. The structuring of the problem assists both the examiner and examinee. Both approach the problem with the same frame of reference, which should help reduce ambiguity.
>
> 2. They lend themselves to, and place more emphasis on, the measurement of understanding, interpretation, and evaluation.

3. Complex material can be measured with a series of different items based upon a single introductory passage, graph, chart, or diagram.

4. They minimize the amount of irrelevant factual information.

5. They lend themselves to a variety of item formats and modes of presentation.

6. In contrast to the essay, complex achievement is measured in a more structured situation, but objective scoring is employed (p. 144).

Mehrens and Lehmann suggest that in order to have a valid and reliable interpretive exercise, the item writer must follow the basic tenets of good test construction, such as clearly communicating to the student the intent of the question and making certain that no irrelevant clues are provided. However, two additional tasks are also required: (1) the selection and/or preparation of the introductory material and (2) writing test items that are dependent on the introductory material and call on the higher mental processes to answer the questions. They offer a number of guidelines for preparing interpretive test exercises:

- Carefully select the material to be interpreted so that interpretations to be made will be significant and representative of course content and objectives;
- Keep the introductory material brief;
- Be novel and creative in preparing introductory materials;
- Base the items on the introductory materials;
- Write multiple-choice items to measure interpretive skills;
- If pictorial materials are used, they should be relevant and of high quality;
- If the interpretive exercise items use K-type make sure that the categories

are homogeneous and mutually exclusive (p. 147).

Overall, multiple-choice items are the most popular and flexible of the objective-type selection items. They are widely adaptable to different content areas and objectives, they can be used to measure rote memory as well as complex skills, and are ideally suited for measuring the higher mental processes. Many of the deficiencies in teacher-made multiple-choice items, such as ambiguity and imprecise wording, can be overcome by carefully preparing the test item. Instructors and evaluators need to pay close attention to many factors in preparing multiple-choice items. Mehrens and Lehmann suggest that the multiple-choice test items should avoid highly technical distracters, negative statements, textbook jargon, and all responses should be plausible and homogeneous. Multiple-choice test items should also have only one correct or best answer and the instructor should avoid providing clues to the correct answer by making it longer, by having overlapping items, and/or giving grammatical clues. As well, Mehrens and Lehmann advise against the use of "all of the above" or "none of the above," if at all. An independent review of the items by a subject matter expert or another evaluator will also generally improve their quality.

In summary, the cognitive data was useful to the decision-makers and stakeholders relative to decisions about the learning effectiveness of the computer-mediated instructional courseware. However, beyond that, because of the validity concerns with the test items, the cognitive data was not very useful for more extensive examinations of the relationship between content and learning achievement. Cognitive data can often be examined to measure and make inferences regarding the content areas in which learners are experiencing the most trouble. These findings will often lead an instructor or educator to reassess and improve the instruction and teaching of those concepts in future program delivery. It is important to note that the limitations of the test items which were discussed are related to the particular test (procedure) which was developed and implemented in the fieldtest of the evaluation model. It is not a weakness or limitation of the pre or posttesting methodology itself nor the CME Courseware Evaluation Model.

Reaction Data

Were participants satisfied with computer-mediated CME instruction? A main purpose of evaluation is to determine the worth or value of something and one way to do this is to ask participants about their perceptions of the strengths and weaknesses of an educational program. Reiser and Kegelmann (1994), Gill et al. (1992), and Dillon and Gunawardena (1992) indicated it was very important to examine learners' perceptions and experiences in using instructional courseware. But, learners tend to enjoy participating in continuing education programs and we generally anticipate positive reactions. Therefore, it is erroneous to infer from learner reaction data alone that a program was either effective because learner reaction was positive, or that a program was not effective because learner reaction was negative. Nevertheless, information from this evaluative category is useful to program developers in improving, enhancing, and fine-tuning aspects of an educational product. It also enables decision-makers to measure the level of participant satisfaction with a program and the likelihood that it would be well received again.

The CME Courseware Evaluation Model used two methods for measuring physicians' satisfaction with the instructional courseware. A quantitative Courseware Evaluation Survey collected information on five areas of instructional courseware effectiveness: content, graphics and media, navigation and organization, learner manual, computer conferencing, and overall impressions. As well, semi-structured qualitative interviews were also conducted with a random sample of 12 rural physicians who had participated in the courseware program.

Overall, the responses to the CES and the interview questions revealed that participants were generally satisfied with: the use of multiple media to present the subject matter; the design of the hypermedia courseware; the ability to participate in CME at a distance at times and places most convenient for them; the ability to pace the amount of learning they engaged in; and, they found computer conferencing to be an interactive, interesting, and important feature of online CME learning. The following section discusses the participants' reported dislikes and likes, and the features of the computer-mediated courseware which may have contributed to these perceptions.

The findings related to participants' satisfaction with the subject matter revealed that the content was highly relevant to clinical practice, practical, and was presented in an effective and interesting way. The use of multimedia, including graphics, video, sound, and text to present the content was rated very highly by participants. These findings are supported by several theories in the field of multimedia instruction. According to dual coding theory (Paivo, 1971, 1986, 1991; Clark and Paivo, 1991) information is processed through one of two generally independent channels. One channel processes verbal information such as text or audio. The other channel processes nonverbal images such as illustrations and sounds in the environment. Information processed through both channels is called referential processing and has an additive effect on recall.

Learning is better when information is referentially processed through two channels than when the information is processed through only one channel. Computer-based multimedia instruction tends to be highly interactive which appears to have a strong positive effect on learning. Najjar's (1996) examination of 75 learning studies indicated that people learn material faster and have better attitudes toward learning when they learn in an interactive instructional environment.

A case-based instructional approach was designed to elicit meta-cognitive problem-solving skills and to situate learning and the content of the courseware within the clinical practices of the health care provider. According to Jonassen (1995) computer-mediated learning technologies should be used to contextualize learning, to situate learning in a meaningful real-world task, simulated through a case-based or problem-based model. The findings of this study suggest that physicians' were very satisfied with the case-study design for orienting learning and bringing the content "home."

Similarly, Lawless and Brown (1997) have identified multimedia as a useful system for facilitating problem-based learning activities. Multimedia enables the representation of salient visual, auditory, and nonverbal cues using a variety of multiple information modalities. The representational richness of multimedia can assist learners in comprehending the situation and observing the relevance of various contextual elements. This simulates the real world more realistically than text-based cases.

A strong theme which emerged from the CES and interview data was physicians' perceptions that the courseware enabled them to pace their learning and to decide when and how much learning they wished to engage in. The hypermedia courseware was developed so learners could explore and control the presentation of the multimedia-enhanced content at their own pace. Within hypermedia courseware, each information node is linked to many others. Individuals use links (hypertext or hypermedia) to make choices about which nodes in a knowledge base to browse and learn from. A unique and empowering aspect of hypermedia systems is that, if designed effectively, they provide learners with opportunities for increasing their control over depth of study, range of content covered, number and type of alternative media selected for presentation, and time spent on learning.

Learner control in a hypermedia instructional context also allows students to tailor their instructional experiences to suit personal and professional needs and interests. Adult learners appear to prefer learner-controlled instructional materials. Learner control has been linked to a variety of favorable affective outcomes, such as increased levels of engagement, more positive attitudes, and decreased anxiety (Kinzie and Berdel, 1990). Many cognitive theorists contend that learners who are afforded the opportunity to direct their own learning can process information more deeply, and as such, obtain a better command of the information (Lawless and Brown, 1997).

By providing learners with opportunities to succeed in moderately challenging instructional situations, CME courseware can promote feelings of computer self-efficacy and encourage continuing motivation to learn. As well, furnishing learners with significant amounts of control and the ability to exercise it can contribute to positive learning outcomes. Several key strategies were used in this computer-mediated system to enhance continuing motivation. These included:

- expanding understanding by viewing video demonstrations;
- listening to spoken descriptions;
- viewing related photographs and graphics;
- exploring links to follow-up on information;
- providing a map of the courseware's content and where a learner has been;
- challenging learner's knowledge and providing appropriate feedback.

Learner curiosity was stimulated by challenging physicians' understanding and grounding content in simulated, real-life contexts. This was facilitated by encouraging problem-solving through the use of case studies and questioning the user on the use of correct treatment or office procedures. As well, to ensure successful learner control and self-regulation, and to minimize frustration a help feature explaining the navigation features was accessible on any page.

Navigation and user-friendly software interface features are an important part of effective instructional courseware design. The navigation features and interface of instructional courseware should be intuitive, meaning learners spend little time learning how to navigate through the courseware, thereby focusing more attention to the content. Yang and Moore (1995) propose that an important aspect of effective instructional courseware interface is its operational features. According to Yang and Moore operating directions should be clear and specific, and there should be "consistency of operation" (p.11).

Computer conferencing was also perceived as a useful communicative component for online CME. It enables physicians to interact with their peers and engage in discussion concerning important aspects of their clinical practices. However, while the process of being able to communicate asynchronously and interact with other learners was well received, the computer conference software was problematic for many users. The interface of the computer conferencing software "NetForum" was confusing and annoying. Participants felt the messages were all "jumbled" and not posted in correct order, and the effort and time to create and post a message was very tedious. One item on the CES asked physicians to rate the computer conference software: "I found it easy to post, respond and reply to messages in the computer conferencing sessions." This item received a mean score of 3.12, SD=1.24. The interviews provided a further opportunity to examine participants' experiences with the computer conferencing system. Several advantages and disadvantages were identified.

Nevertheless, physicians described the computer conferencing as important for interacting with their peers and the instructors, confirming beliefs and asking questions, and very helpful for providing a non-evaluative forum in which everyone could express their ideas, compare their responses with others, and rethink and expand upon their initial thoughts. Harasim (1986) also found similar results in a study of online learning. Learners felt the ability to post a message directly to the consultant and to receive a timely reply or feedback, was something not always available through face-to-face instruction, often due to time constraints (Harasim, 1986).

Anonymity emerged as a major theme from the interview data and was perceived as an advantage of computer conferencing. This supports the findings of several authors. The anonymity of the computer conferencing environment has been cited as one of its main advantages (Harasim, 1986; Ruberg et al., 1996; Huang, 1997). Ruberg et al. (1996) suggest that the computer conferencing context promotes a sense of anonymity and privacy, and because of this sense of "deindividuation" (loss of identity) individuals are less influenced by social conventions and more likely to participate in the online discourse without ever being known personally.

Several authors have also suggested that computer conferencing creates a learning environment which is more personalized than a traditional classroom (Davie, 1989). Research findings suggest that conferencing participants often become quite connected with their peers, and establish feelings of solidarity that may not occur in the competitive atmosphere of a traditional classroom (Feenberg and Bellman, 1990; Newman, 1990). Several physicians commented that the computer conference enabled them to participate in a discussion which was more personal and comfortable than the discussions they usually experienced at face-to-face CME conferences and workshops.

Collaborative learning is a well documented and studied feature of the online computer-mediated learning environment (Seaton, 1993; Rowntree, 1995; Steeples, 1993). The main advantage of a collaborative learning environment is that it permits learners to evaluate their ideas against the shared experiences of their peers (Seaton, 1993). In the collaborative learning context learners are able to learn as much from one another as from course materials or instructors. According to Rowntree (1995) what is learned is not so much a product (information) but a process, in particular the creative cognitive process of offering up ideas, having them criticized or expanded upon, and getting the chance to reshape them in the light of peer discussion. Adult learning becomes an interactive process in which mature adults with professional experiences are able to share a richness of perspectives and examples that would otherwise be impossible for instructors or course designers to provide in advance.

Several aspects of computer conferencing enable the establishment and implementation of collaborative learning environments. Most importantly, computer conferencing facilitates many-to-many peer interactions which support and encourage a sharing of individual professional experiences. Collaboration through computer conferencing also supports the active construction of knowledge and an engagement in a learning process with fellow adult learners. It may also lead to a sense of involvement and identification with a peer group, and this supports the human/sociological need to operate and interact with others (Steeples, 1993).

The interviews also revealed that physicians felt the computer conference allowed them more time to reflect on the discussion and the comments they would like to post, which they believed enhanced the quality of the dialogue which occurred. They felt that this aspect of computer conferencing was a significant advantage over the discussions which usually occur at face-to-face CME workshops.

The asynchronous mode of computer conferencing supports self-paced learning, reflection, and idea formation. Several authors have suggested that this leads to purposeful response construction and formulation processes which may very well enhance the quality of the discussion (Ross et al, 1995; Rowntree, 1995; Harasim, 1989; Burge, 1994). The text-based collaborative nature of computer conferencing discourse deepens student processing and requires an enhanced articulation to others, which furthers the development and refinement of understanding, and heightens learner reflectiveness (Ross et al., 1995). The initial, internal articulation or rehearsal of a point of view refines the expression of one's ideas more so than in a face-to-face discourse.

According to Rowntree (1995) computer conferencing supports reflective thought in a way the face-to-face classroom does not. The interaction is much different and learners have more time to reflect and give attention to their ideas than they normally can in classroom-based learning. More time for reflective thought and the formulation of one's ideas leads to an improvement in the quality of responses, perhaps of a higher quality than what would be given in a face-to-face discussion (Burge, 1994).

Computer conferencing also offers unique advantages over traditional classroom-based teaching. Computer conferencing is text-based and because of this on-going transcripts of the discussion are stored for future reference (Seaton, 1993; Harasim, 1986). Students in Harasim's study (1986) felt that these electronic transcripts were a wonderful learning tool, functioning as a current and future learning resource if required. Similarly, physicians participating in this online CME computer conferencing discussion felt that the conference threads and forums provided a useful resource, which they could read and review, and one individual even mentioned that the conference threads could very well be used as a learning resource for participants in subsequent courses.

Several authors have reported that students often report feelings of confusion and annoyance, and a lack of conceptual understanding of how conferencing systems and threading works with computer conferencing software (Ruberg et al., 1996; Burge, 1994; Seaton, 1993; Tagg and Dickinson, 1995). The findings of this study support these reports. Ruberg et al. (1996) found that students were uncomfortable with the interface of computer conferencing software, often becoming confused with what seemed like a lot of jumbled thoughts. As well, for some students, because of the lack of social cues and multiple threads of discussion, computer conferencing was confusing and inhibiting (Ruberg et al., 1996). Ruberg et al. contribute this anxiety and uncomfortableness to the extensive text-based interchanges occurring in computer conferencing, which participants often find difficult to adapt to.

Physicians' also reported that the "asynchronous" nature -- the time delay in messaging, and the irrelevancy of the discussion were disadvantageous. These perceptions and experiences support the findings of several other studies reported in the literature. In a qualitative study of learners' perceptions of computer conferencing, Tagg and Dickinson (1995) found that the asynchronous nature of computer conferencing resulted in feelings of "communication anxiety," attributable to the varying delays learners experienced between messages being sent and received. As well, learners expressed certain difficulties in assessing how one's contribution had been received in a medium where immediate feedback by way of facial expression, verbal response, or physical reaction was entirely absent (Tagg and Dickinson, 1995). Rowntree (1995) and

Harasim (1986) suggest that because computer conferencing lacks the visual and auditory cues on which we usually rely for interpreting other people's meanings, it is initially an impersonal social world for many participants. Burge (1994) also reported that graduate level students participating in an online education course indicated perceptions of "irrelevancy" of discussion topics and feelings of being "out of sync" with the discussion.

Several issues related to the CES and interview procedures and the information collected are important to discuss. First, both data types enabled data triangulation because the interview results could be checked and compared against the results of the CES and vice versa. This allowed the evaluator to confirm the accuracy of the results which were being received. A majority of the stakeholders found the qualitative data useful because it provided in-depth information on participants' satisfaction with the instructional courseware and its various components. Similarly, the program developers found the interview results very valuable because they were able to extract relevant information from the findings to assist them in improving the courseware. The CES data, on the other hand, had more limited application.

The program developers felt the CES data was useful, but not to the same extent as the qualitative interview information. The CES data did not allow the developers to pinpoint the participants' exact concerns or problems because the quantitative data reported limited closed-ended responses and was too restrictive. However, the developers noted that the results of the CES did highlight certain issues or raised "red flags" on elements which they would otherwise not have paid attention to during revisions.

One suggestion made by the program developers was to include comment boxes on the CES which would allow participants to elaborate on any problems or negative experiences they had with the courseware. Therefore, under each category on the CES submission form participants would be able to type in textual comments and submit these messages with their quantitative responses.

Instructional Transaction Data

Another strategy for measuring learning effectiveness involves the examination of the quality, extent, and type of interactions and communicative transactions which occur through computer-mediated instruction. The observation of discussion and interaction types, questions being asked, and the quality of dialogue relative to the objectives of a program provide an indication of collaborative learning-in-action. A unique advantage of computer conferencing systems is they automatically provide a text-based electronic transcript of discourse interchanges which can be analyzed using qualitative and quantitative methods. A thorough analysis of these interaction patterns enables an assessment of the types and levels of the many-to-many interactions which occur.

The findings of this study suggest that there was a high level of interaction between the participants, who participated, and the instructors. The level of interaction ranged from 38 percent "instructor-initiated (II)" to 62 percent "learner-initiated (LI)" comments, to 50 percent (II) -50 percent (LI) comments across the discussion forums. Harasim (1986) has suggested that the communication interchanges between instructors and students in traditional didactic instruction are 80 percent (II) to 20 percent (LI). The interchanges in this computer conference varied from direct questions and comments to the instructors, the sharing of professional experiences on the use of various procedures and instruments, to direct learner-learner interaction. Considering Harasim's (1986) reported interchange findings, the findings of this study suggest that there was a higher level of participation in this computer conferencing discussion context than would normally have occurred in a face-to-face CME didactic workshop.

A content analysis of the messages and comments indicated a discussion of topics and subject matter that was far beyond what was covered in the courseware. However, the main limitation to the computer conference was its format. The discussion was facilitated, for the most part, as a question and answer instructional activity. And, while some collaboration was observed, other formats may have produced greater interaction between learners, rather than the majority of learner-to-instructor based interactions observed in this discussion. Case-based discussions, which could have extended from or been based on the courseware material, may have been more effective. In this format, several case-studies could be presented in the computer conference. Physicians would be invited to discuss the cases and hypothesize investigations and proposed treatment protocols. The instructors would facilitate the case-study discussion by guiding participants, offering information and support on specific teaching points raised in the forums. The discussion would conclude with the reporting and a summary of the correct investigations and treatments, and a review of the main learning issues for each case. As well, a question and answer and socializing forum areas could be available within the conference for engaging in other discussion types.

Behavioral Data

Did the learning transfer into changed behaviors in the real-world? The application of learning after a continuing education program is the fifth category of Cervero's (1988) CPE evaluation model. This level of evaluation is concerned with measuring the extent to which the knowledge, skills, and attitudes taught during a learning activity are transferred to the professional's workplace. Evaluative information in this category is always collected after the program has been delivered and learners have the opportunity to return to the workplace and to perform in the natural work setting. Often times however, variables other than the CME program may influence whether or not new skills and knowledge are transferred and applied. Sometimes, institutional policies and cultures, and practice ideologies may discourage the use of new skills and techniques. The physician may be located in a practice which isolates him or her from instruments and equipment for performing the new skills or using the new knowledge. The patient population may not present with illnesses or problems which necessitate the application of what was learned in the CME program. As well, the physician themselves, their attitudes, and clinical experiences may influence the extent to which knowledge and skill is transferred to practice. Where performance change does not occur it is important to consider the influence of these and other extraneous variables upon the physician's overt behavior.

A retrospective pretest-posttest self-reporting performance survey was developed to assess the affect of participation in the computer-mediated courseware on the clinical knowledge and skills of participants. The survey consisted of 13 performance statements which were based on performance objectives formulated and ascribed by the subject matter consultants and CME Planning Committee members during the initial phases of instructional planning. This survey was forwarded to participants 6 weeks after they had completed the program.

Many of the issues surrounding self-report data have been discussed previously at various junctures throughout the dissertation. Self-report data has been criticized because of its lack of objectivity and invalidity. According to Abrahamson (1968) the only true measure of the effectiveness of CME is to indicate an improvement in the level of medical care. Yet, Davis et al. (1995) found that less than half of the studies they examined in their meta-review of CME evaluation studies revealed meaningful changes in patient or health outcomes. Several suggestions have been raised as to why CME has been less successful in affecting patient and health outcomes. The most prevalent suggestion relates to patient behavior, which is often beyond the control of the physician. Improving health outcomes, not changing patient behavior is the ultimate goal of continuing medical education.

However, before health outcomes could be improved, Abrahamson noted that a change in physicians' performance would have to occur -- what Abrahamson called "intermediate change." Several techniques were advocated for assessing physicians' behavior change including chart review, office visits and audits, and exploration of the office records of the practitioner (Abrahamson, 1968). These techniques usually meant having access to the physician and the opportunity to assess what a physician did in everyday activities. But, it was often beyond the resources of an evaluator to conduct a chart review or office visit of every physician who had participated in a CME program. In these instances evaluators often used interviews or questionnaires to examine physicians' self-reports of changes in performance.

According to Caplan (1973) providing evaluation results that are meaningful and statistically rigorous is difficult at best. This difficulty increases three-fold when the assessment of the performance of physicians practicing at a distance are introduced to the equation. Caplan suggested that anecdotal information, by way of physician self-reports following participation in a CME program provided results which offered important information on the success of a given CME activity. This information must not be overlooked simply because it does lend itself to objectivity or strong validity.

The findings from physicians' self-reports of clinical performance change before and following participation in the computer-mediated CME courseware "Dermatological Office procedures" suggests that the courseware was effective in producing self-reported change in clinical treatment and management behaviors.

Evaluation Procedures

The CME Courseware Evaluation Model fieldtested in this study used a variety of formative and summative evaluation methods and procedures to collect information. Consequently, another important area to discuss relates to the effectiveness of these various evaluation procedures as they were employed in the evaluation model. The issues related to evaluation procedures include preferred procedures identified by the stakeholders and program developers as a result of the metaevaluation, the issues surrounding formative evaluation, and qualitative versus quantitative data collection.

The preferred summative evaluation procedures identified by the stakeholders, key decision makers, and program developers during the metaevaluation were the cognitive learning achievement tests, the interviews with participants, and the retrospective pretest-posttest performance survey data. The modified quasi-experimental study design was also cited by several decision makers as an important aspect of the evaluation model. Although the procedure for analyzing the instructional transactions within the computer conference was not touched on during the metaevaluation, the procedure was helpful to the evaluator. This procedure involved an analysis of the transcripts of the online discussion to assess the effectiveness of asynchronous computer conferencing for facilitating collaborative learning. The procedure allowed the evaluator to assess the role of computer conferencing as an instructional component of the courseware.

The formative evaluation procedures were important components of the evaluation model as well. The methods which were most effective during the analysis and design phases of ISD were the paper-based instructional outlines and the story-board meetings. These formative evaluation procedures enabled the evaluator to receive critical feedback from CME Planning Committee members on learning needs and objectives, and the appropriateness of instructional content, media, and strategies. The division of tasks between the CME Planning Committee and the Program Development Team also made the formative evaluation process efficient and effective.

Although the formative evaluation procedures implemented in the fieldtest were useful to the program developers, several suggestions were made for improving the processes in the future, particularly during the development of courseware prototypes. One suggestion was to collect information on the learner's hardware specifications at the time they were completing the online Courseware Evaluation Survey. In order to situate this suggestion and the other issues raised by the developers within the context of the fieldtest evaluation of the hybrid courseware, it is necessary to discuss the development challenges which confronted the program developers before production began.

In assessing the available technologies for delivering multimedia-enhanced CME in a Web-based environment there were two major challenges facing program developers -- time and cost. The development of an interactive multimedia CD-ROM product using typical multimedia development tools is time consuming and expensive. As well, given the resources required in producing an interactive multimedia product, it is unlikely that revisions to a program can be made without further substantial costs. Because of the inherent restraints of interactive multimedia products, the understanding that the courseware under development was to be a prototype, and that significant changes and revisions to the courseware could result from the evaluation, the program developers conceptualized a hybrid WWW/CD-ROM product. The hybrid concept made it possible to rapidly prototype materials and then revise them without incurring substantial costs.

The concept of publishing HTML documents for delivery on CD-ROM is one which has value and numerous efficiencies in the context of courseware development and production. Developing courseware using HTML means a significant reduction in the work necessary to produce and deliver courseware materials. Numerous HTML editors, particularly those with WYSIWYG features (what you see is what you get) have simplified the process of publishing courseware for WWW delivery. These HTML editors simplify the process of specifying anchors and tags, footers and the inclusion of graphics in HTML documents. As well, platform independence of text and graphic information is a non-issue with HTML. Courseware need only conform to HTML document structuring rules to be operational on any platform (IBM or Apple). The graphics standard file formats of most WWW browsers -- Graphics Interchange Format (gif) and Joint Photographic Group (jpeg) are standard.

Nevertheless, the use of the WWW, CD-ROM and HTTP protocols as courseware delivery mediums were not without their problems. HTML documents are displayed in a browser by first downloading a marked-up text file and displaying it in its formatted state. The next step includes the downloading and display of each graphical or media element. Most browsers typically only support styled text and static graphics (gif and jpeg) within their overall operation. Continuous multimedia such as sound and video files are usually handled by "helper" applications or plug-ins. Plug-in applications are specified within a browser's preferences and are selected based upon the file extension of the media file to be displayed.

The program developers for the courseware selected .avi as the video file format for presenting the videos. The .avi file format was chosen for presenting the video because it is considered the standard application for viewing video in Windows 95 (providing that users have a video card) without users having to install a plug-in application, such as Quicktime. At the time of courseware development, approval for distribution and use of the Quicktime plug-in with the CME courseware was not approved by its manufacturers. Therefore, the courseware relied on the built-in avi player of Windows 95 for playing the videos. Avi files were embedded in the HTML documents.

However, Netscape Navigator users experienced significant problems with the downloading and playing of the avi video files. The download step took a considerable amount of time. The developers were unsure why this occurred, but felt it was a design issue related to that particular Web browser software. Some physicians resorted to opening and playing the video files through their Windows Explorer application, which in essence used the built-in avi video player of Windows 95 to the same extent intended for Netscape Navigator. In the future, Quicktime for Windows has been suggested as the video plug-in standard which will be used for displaying the video files. For participants who do not have this application already installed on their computer, the plug-in software will be available for download installation from the courseware CD-ROM. This separate, stand-alone plug-in application should provide consistency and efficiency across browser types, however there are possible areas where other problems may be encountered.

All media files are "downloaded" before display by a plug-in application, even those which are held on the local file system. Web browsers do not typically load multimedia files into a plug-in application without the download step. It is impractical to access large sound, video and other media/multimedia files directly from the CD-ROM due to response time and often local storage space considerations. The time required before the display of a media file is significant when a video of more than 1MB is to be viewed. The use of plug-in applications for presenting media files also becomes problematic due to the typical hardware configurations of learner's computers. Many older, low power computers have limited physical hard drive space and read-only access memory (RAM) available. To make efficient usage of a browser for viewing HTML, one needs enough physical RAM to support the WWW browser as well as any plug-in applications. Complex HTML documents using abundant amounts of graphics require a browser to be allocated a large amount of RAM, which may then not be available for plug-in applications required to display other media.

Netscape Navigator and the Internet Explorer WWW browsers appear to execute HTML documents and interact with the underlying operating system in different ways. These rendering inconsistencies range from incorrect translation of the HTML, font size and selection problems and the support provided for the different media types. Media standards for sound and video are platform specific. Video posed the biggest problem as indicated by the problems encountered by several physicians who were using Netscape Navigator WWW browsers. The video files downloaded very slowly in the Netscape browser application, and in some instances would not even download.

The fieldtest of the evaluation model included a one-on-one procedure for gathering feedback from a member of the target audience. The one-on-one method was useful in uncovering several design, technical, and interface problems, as well as the reactions of a user to problems if they were encountered. However, the program developers felt that a three-on-one method would be more useful and informative. This procedure would bring together three members of the target audience and at the same time, each individual would study the course materials. The evaluator would be present to assess the comments, concerns, and issues raised by the users as they worked through the courseware. Upon completing the courseware the evaluator would have the option of questioning the users on the quality and effectiveness of the instructional product using a prepared interview protocol. The program developers also felt that an expert review by an instructional designer would be helpful. This individual would review the storyboards and comment on the effectiveness of the instructional design features from his/her perspective.

Quantitative and qualitative measures were also used during the fieldtest of the courseware evaluation model. The Computer Attitude Scale, Demographic Profile Survey, learning achievement test, Courseware Evaluation Survey, and the retrospective pretest-posttest performance survey produced quantitative data only. The interviews with participants produced qualitative data and the content analysis of the computer conference instructional transactions provided both quantitative and qualitative information. If the preferred procedures identified in previous sections were used to re-evaluate the program, both quantitative and qualitative data would be gathered again.

CME Courseware Evaluation Model

One final area of discussion focuses on the evaluation model which was designed, developed, and fieldtested in this study. The following section discusses the advantages and disadvantages of the model as they were experienced in assessing the effectiveness of the courseware "Dermatological Office Procedures." These advantages and disadvantages are summarized in Table 5.1. The advantages reflect, for the most part, the rationale which were presented in Chapter Three and the literature discussed in Chapter Two. They require no further discussion. However, the disadvantages identified during the study, both within the model and the evaluation procedures, deserve elaboration.

Table 5.1 Advantages and Disadvantages of the Evaluation Model

Advantages	Disadvantages
Practitioner-as-evaluator orientation	Questionable validity of cognitive test
Evaluation of multiple evaluative criteria	Questionable accuracy and validity of self-reports
Eclectic evaluation which integrates quantitative and qualitative methods	No long-term behavioral or performance change assessment
Evaluation which is improvement, utilization, and decision-making oriented	No open-ended feedback from participants during CES submission
Evaluation as a concurrent ISD process	No small-group formative evaluation during courseware development
Prescriptive, but options for selecting data-gathering methods and evaluation questions	No formal action-planning on part of stakeholders or program developers based on evaluation results
Evaluation which is formative and summative.	

The focus of the CME Courseware Evaluation Model is to determine the effectiveness of instructional courseware. Effectiveness, in this regard, is concerned with enhancing or revising the quality of an instructional product while it is being designed and developed, as well as assessing the extent of instructional outcomes in terms of learning, satisfaction, and behavior. Program developers are usually more interested in formative evaluation data which enable them to improve the product while it is still being developed. Decision-makers and stakeholders, on the other hand, are interested in summative data surrounding the main outcomes of the instruction which was delivered.

Summative data can include the collection of information on the extent to which behavior or performance has changed. However, behavior change is something which occurs over a period of time, and can be the result of several variables interacting together to produce a change in the learner's performance. Several stakeholders felt that a longer term study of the impact of the program on the performance behavior of the physicians would have been helpful. Comments were also made about the validity and accuracy of the self-report information which was collected. Therefore, of all the procedures which were used in the evaluation study, the retrospective pretest-posttest performance survey was the weakest link.

Nevertheless, there are many obstacles and disadvantages in undertaking a more extensive analysis of behavior change. First, the implementation of an objective study of the behavior changes of distance education learners is a challenging feat. This becomes even more difficult when chart audits or observation measures are required. Second, with the addition of a six month behavioral evaluation procedure the collection and reporting of results to developers and stakeholders would also be delayed by as much as six to eight months following the program. Therefore, final results of the assessment would not be available for the decision makers until after the completion of the evaluation. Furthermore, performance change is influenced by an array of confounding variables which are usually beyond the control of the evaluator. Cervero (1988) has demonstrated that behavior change is influenced by a variety of factors including individual learner characteristics, the social system in which an individual practices, attitudes, and the quality of the continuing education program which was delivered.

Another disadvantage of the evaluation model is related to the lack of follow-up with stakeholders, decision-makers, and program developers regarding the results of the evaluation. Follow-up discussions can be used to discuss weaknesses and strengths and prescribe measures for improving the instructional material for future delivery. This "post-mortem" can serve to establish action-plans for producing revisions and enhancing the product and materials. It can also allow a means for metaevaluating the evaluation and the results, and for suggesting changes to the methods and instruments for future evaluations.

The remaining disadvantages are related to the procedures used during the fieldtest and have been discussed previously, particularly the cognitive test. The validity of the learning achievement test is questionable, but the lack of explicit criteria for measuring the cognitive test and behavioral results also constitute another weakness in the model. Without unambiguous criteria and cut-off scores for interpreting learning and performance outcomes, scores have little or no meaning to other individuals examining the results of the fieldtest. However, this weakness was in the procedures used during the fieldtest, not in the evaluation model.

The final issue to be discussed is related to future applications of the courseware evaluation model. The model was designed specifically for evaluating the effectiveness of computer-mediated CME courseware. The evaluation model was fieldtested on a

Dermatological Office Procedures courseware program for family physicians. To generalize the value of the evaluation model to a broader extent, it should be fieldtested with instructional courseware programs of varying content and audiences. It is possible that the evaluation model might be as effective or more effective with programs that are not of a hybrid computer-mediated format, as outlined in Chapter One. Nevertheless, based on the results of the fieldtest and metaevaluation it is difficult to generalize the model's usefulness to programs which are not of a computer-mediated hybrid courseware type.

Summary

In summary, a number of issues pertaining to the results of the fieldtest and metaevaluation were discussed in this Chapter. These issues were grouped into seven major categories, antecedent data, cognitive data, reaction data, instructional transactions data, behavioral data, evaluation procedures, and the evaluation model. Conclusions based on the issues discussed in this chapter and the findings of the study are presented in Chapter Six.

CHAPTER VI

CONCLUSIONS

Introduction

In this chapter the conclusions of the study are presented, recommendations for further research are suggested, and the implications of the fieldtest and metaevaluation results for continuing medical education practice are discussed. As a conclusion to the study and a call for further educational research a modified version of the CME Courseware Evaluation Model is presented which incorporates the major findings of the metaevaluation.

Conclusions

The following conclusions are drawn from the results of the study:

1. Hybrid Courseware Delivery Systems are an effective and efficient means for delivering self-paced computer-mediated instruction to rural and remote health care providers located in low bandwidth geographic regions.

2. The Computer-Mediated CME Courseware "Dermatological Office Procedures" had an impact on participants and the evaluation conducted during the fieldtest collected information which documented these effects.

3. The metaevaluation identified four evaluation procedures which were considered most effective during the fieldtest -- instructional outline review, storyboard assessment, participant interviews, and instructional transaction evaluation.

4. The evaluation model is effective in providing immediate formative evaluation information to program developers and summative outcome data to decision-makers.

5. Electronic data gathering procedures are an efficient and effective means for

collecting, compiling, and analyzing courseware evaluative information.

Recommendations for Further Research

The study described in this dissertation was designed to fieldtest and metaevaluate a model for evaluating the formative and summative effectiveness of computer-mediated CME instructional courseware. Because of this purpose the generalizations which can be made from the study to other content areas, situations, or populations are limited. Nevertheless, a number of the findings and issues raised by previous discussion in the dissertation warrant further examination. The recommendations for further research explored in this section are based on the issues which were generated from the fieldtest and metaevaluation results.

A primary recommendation for further research focuses on the courseware evaluation model itself. The evaluation model for computer-mediated CME courseware should be applied to other programs, content areas, and populations. The purpose for implementing the evaluation model in other contexts is to assess its ultimate value and usefulness in different situations with varying populations and subject matter. The disadvantages outlined in Chapter Five should also be scrutinized under varying circumstances to determine if refinements can be made to the model. As well, the advantages of an eclectic approach should also be assessed to establish if it is useful when applied in the evaluation of other programs.

The CME Courseware Evaluation Model should also be evaluated against other evaluation approaches. For example, a computer-mediated courseware program could be evaluated concurrently using the courseware evaluation model and one of the other approaches discussed in Chapter Two. The evaluative results of the two approaches could be assessed and compared using metaevaluation procedures similar to those applied in this study.

Additional research should also be conducted to determine the optimal number and type of formative and summative methods necessary to effectively and efficiently evaluate instructional courseware programs. Issues surrounding the validity and accuracy of the retrospective pretest-posttest design proposed by Howard et al. (1979) and modified in this study should be explored. Are physicians' retrospective self-reports of behavioral change a valid reflection of actual performance improvement? Do longer-term behavioral measurement procedures improve the effectiveness of the evaluation model? What are the most effective formative evaluation procedures for assessing instructional courseware? How do the individual demographic and computer usage characteristics of physician populations influence attitudes toward computers?

Future research should also consider the examination of participants' computer experiences as a component of evaluating the effectiveness of instructional courseware. The issue of frequent computer experience versus less frequent computer experience is an example of a factor which could be explored. As well, the influence of other antecedent factors should also be considered when and where resources and situations permit.

Finally, the procedures which were developed for the fieldtest should be tested further. In particular, the cognitive learning achievement test should be revised using the interpretive exercises discussed in Chapter Four to ascertain its effectiveness for assessing participants' achievement. As well, the achievement test and the retrospective performance survey procedures should be modified so evaluators can critically examine the success of instructional strategies for producing learning achievement and performance improvement. Content areas which exhibit deficiencies through achievement testing should be closely appraised and the instructional strategies improved for future delivery.

Study's Implications and Potential Use of the Model by Evaluators and Decision-Makers

The main implications of this dissertation originate from the results of the fieldtest and the metaevaluation. Evaluators, health care administrators, and program developers could benefit from the evaluation model, planning matrix, and evaluation procedures presented and discussed in this study. The eclectic planning matrix could be adapted or modified for use in a number of different continuing medical education programs. The evaluation model could be useful to other practitioners examining the effectiveness of variations of computer-mediated or computer-based learning programs. As well, the retrospective pretest-posttest performance survey, electronic data gathering procedures, Courseware Evaluation Survey, and formative evaluation methodologies are all possible practical outcomes of this study.

Evaluation is the determination of the worth, value, or merit of something. In education and training, evaluation is most often used to assess the effectiveness and/or efficiency of training, education, instructional materials or products. Evaluation can be course or program-specific, curriculum-specific, or based on regional, national, or international systems of human resources development. The evaluation model developed, fieldtested, and metaevaluated in this study is courseware-specific, meaning it is intended for the improvement of courseware products and the reporting of the educational impact or outcomes of these products. When viewed in this manner the evaluation model developed and validated in this study has several potential uses for other evaluators, program developers, and decision-makers in the education and training fields.

The model can be used to evaluate the effectiveness of computer-mediated instructional courseware programming at formative and summative levels. At a formative evaluation level, program developers and instructional designers can use the formative assessment procedures to make changes to instructional courseware while it is still in development stages. This can ensure that significant problems related to instructional design are addressed before considerable time and resources are devoted to producing a final educational product. This level of evaluation can also allow developers and designers to conduct formative assessment as instructional design is occurring. The involvement of content experts and representatives of the target audience during instructional design and formative evaluation ensures that instructional strategies are appropriate for the content area being developed, and that subject matter information addresses the learning needs of the potential recipients of the finished product.

The summative evaluation procedures induce information which reveals the effectiveness of the program in producing learning achievement and satisfying the learning needs and interests of its recipients. Summative evaluation can also collect information concerning the degree to which participants report significant changes in their performance as a result of participating in a training program. This information is of interest to stakeholders and decision-makers who need to be able to assess the value and worth of a program in terms of how well it increased learners' knowledge and impacted performance in the clinical setting. Evaluation information at this level also has uses for program developers. Information which has been collected through interviews with participants in a program and an analysis of instructional transactions reveals the effectiveness of instructional strategies and the design of the courseware program. The designers and developers may use this information to identify the strengths and weaknesses of the courseware and make revisions or changes to the courseware product based on this data.

The information which is collected through the evaluation model can serve the needs of a variety of individuals who have a stake in the success and effectiveness of a program. The metaevaluation results demonstrated that various stakeholders, decision-makers, and program developers found the information collected through the evaluation procedures useful in making decisions surrounding educational effectiveness and identifying program design and development strengths. The evaluation model may be used in its current state, or modified and adapted by others in the training and education field to improve instructional products and collect and disseminate information concerning the outcomes of educational programs.

In another way, the fundamental implications and value of the fieldtest results are based on the extent to which they can inform future courseware development efforts. In this study, a hybrid courseware delivery system was developed as a means for providing multimedia-enhanced computer-mediated instruction to rural and remote physicians in low bandwidth geographic regions. The results of the study revealed that a hybrid delivery system was an effective way for overcoming the telecommunication obstacles which are encountered in transmitting streaming multimedia over the Internet.

The results of the fieldtest also revealed that computer-mediated courseware was effective for providing continuing medical education at a distance. As well, the results of the fieldtest also demonstrated that several program design concepts can increase the instructional value of computer-mediated courseware:

- The sound design and development of an instructional software product that enables learner control over learning paths and the pace and presentation of material.

- The selection and presentation of information in a means appropriate to the knowledge and skills to be conveyed, based on an appropriate process for identifying the needs of the individual learners and the formulation of objectives for the courseware. This includes presenting information using a variety of media appropriate to the information type and learning objectives.

- The information also needs to be structured and organized in a way that it represents adequately the internal logic of the subject matter and meets the needs of individual learners. This includes the use of case-based learning activities.

- Opportunities for participating in collaborative discussion with peers. This collaborative learning may be facilitated through the application of computer-mediated communications. Students must be provided with suitable activities to assist their learning. They need to use their knowledge to analyze real world contexts and suggest alternative and appropriate ways of using the information they have acquired.

- Opportunity for self-assessment.

Asynchronous computer conferencing was a new communicative modality for many of the participants. McCreary and Van Duren (1987) suggest that successful participation in computer conferencing requires learners to have a fair grasp of the conceptual model of computer conferencing. However, the level of sophistication of the computer conferencing system or software will often predict the difficulties learners may experience in navigating the discussion threads and posting and replying to messages. The less intuitive or transparent a system the greater the confusion and frustration participants are likely to experience. The best advice for future CME courseware developers is to use a computer conferencing system that exhibits an intuitive interface and avoids complex threaded discussion displays.

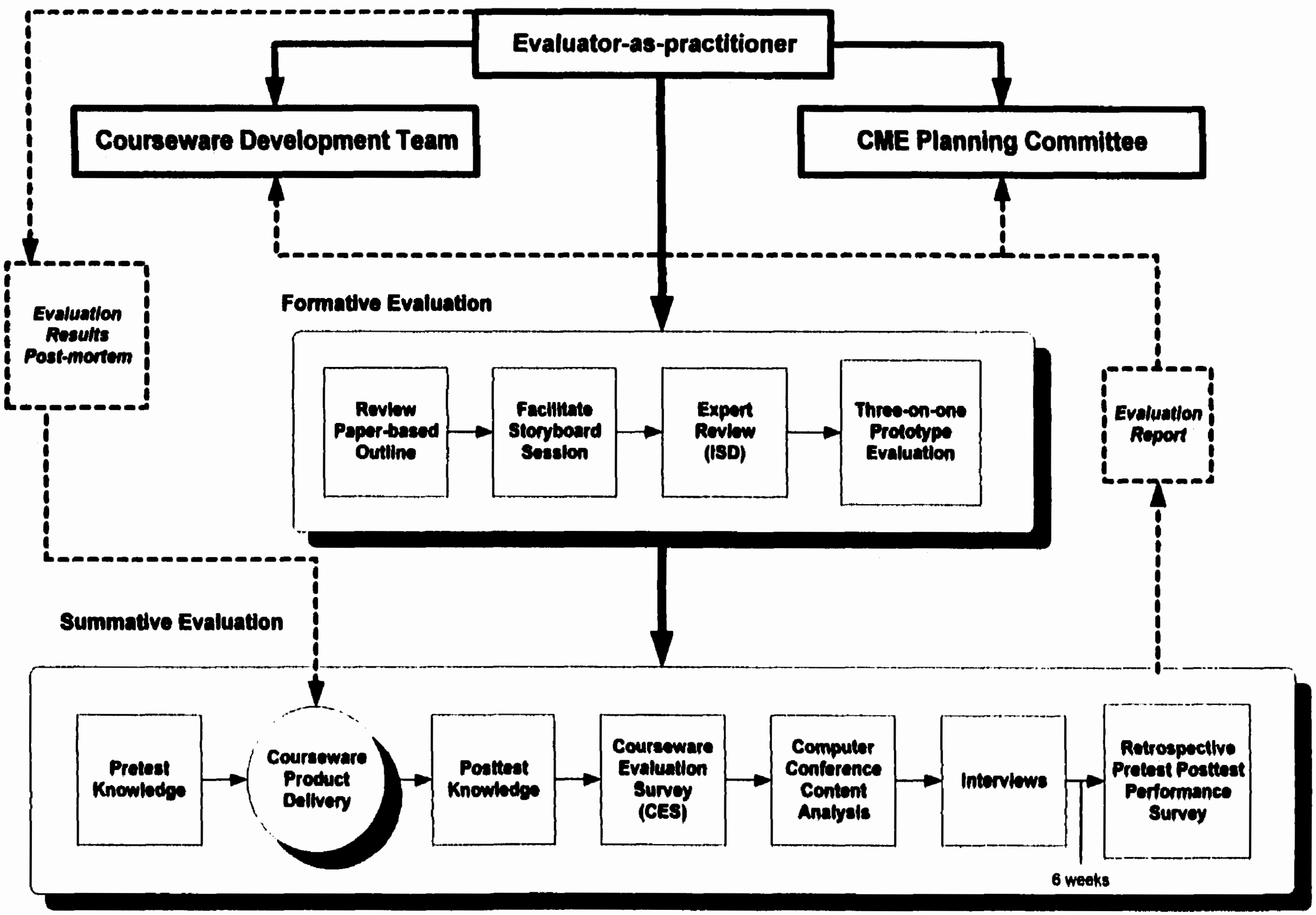
The use of asynchronous computer conferencing systems also has implications for CME instructors. Instructors will need to learn some skills that are not common to the instructional practices which occur in a face-to-face learning context. Successful facilitation of a computer conferencing discussion requires a change in the instructor's role, from information provider to a facilitator of learning networks. This involves the posing of questions instead of supplying answers, deflecting learner's questions to their colleagues, and learning when to remain silent so others may speak.

The literature review in Chapter Two indicated that a majority of studies of continuing medical education at a distance lacked systematic and comprehensive approaches to the evaluation of instructional effectiveness. Yet, evaluation is an important instructional design process which provides essential information for improving instructional materials and informing decision making. Distance education methods have proven to be cost-efficient and effective means for delivering instruction to rural and remote communities. And, governments and rural communities need guidance before they invest scarce resources in technologies that may not be appropriate or adequate for addressing rural and remote health care provider educational needs. The CME Courseware Evaluation Model fieldtested and metaevaluated in this dissertation is one approach for providing accountability and outcomes-based measures for key decision-making in the continuing education of our rural health care providers.

The courseware evaluation model was developed using the eclectic planning matrix designed for this study. The question of whether the evaluation model is superior to other approaches remains unanswered and will remain unanswered until additional courseware programs in other settings with different content and populations are evaluated using the approach. In the final analysis, the evaluation model was modified to reflect the results of the metaevaluation and issues discussed in previous sections of the dissertation. A revised CME Courseware Evaluation Model is presented in Figure 6.1. The revised model incorporates the major findings of the metaevaluation in its design and offers an alternative to the evaluation model presented in Chapter Three.

The modified evaluation model contains many of the same elements of the original courseware evaluation model fieldtested in this study. Formative and summative evaluation levels remain important components of the model and these elements distinguish it from other evaluation frameworks reviewed and discussed in the literature review.

Figure 6.1 Modified CME Courseware Evaluation Model



The participation of CME Planning Committee and Courseware Development Team members during the planning and formative evaluation stages of courseware development were also viewed as significant components of the model. Participation of both groups, including representatives of the target audience and subject matter experts, enabled the evaluator to conduct several instructional design steps concurrently with formative evaluation procedures. This process was seen to differ from traditional instructional design work in which formative evaluation occurred after a majority of the instructional design tasks had been completed. As a concurrent process, instructional design and formative evaluation occurred simultaneously and by that reduced the time and resources expended on post-design evaluation activities.

An expert review (Instructional Systems Design) and a three-on-on prototype review by representatives of the target audience were two new methods added to the formative evaluation stage of the courseware evaluation model. The three-on-one method replaced the original one-on-one formative evaluation procedure while the expert review procedure was a new addition to the model. The rationale for these revisions was based mainly on the opinions provided by the program development team members. These individuals felt that a small-group format, such as the three-on-one method, would provide more detailed and comprehensive information on the effectiveness of courseware prototypes. The three-on-one method involves three participants from the target audience completing the prototype of the courseware program with the evaluator present. Participants are invited to discuss problem areas and strengths with each other as they complete the prototype program and the evaluator can question, record, and explore the users' comments and perceptions as they proceed. Each participant should represent a different experience and/or skill level with both computers and the content of the program if possible.

The expert review method would involve the invitation of an instructional designer to critically review the initial paper-based storyboards of the courseware product. This expert would study the storyboards with the program developers and evaluator present so elaboration and discussion of certain features of the storyboard designs could occur. Reaction from this individual would focus on the distinct instructional design elements of the courseware program. The suggestions and recommendations from this individual could be used to revise or change elements of the courseware's instructional design before production of a prototype product occurred. The opinion from this expert would be considered and rated by the development team members before any revisions took place.

The summative evaluation level remains unchanged in the modified model. Pre and posttesting of learning achievement, evaluation of participant satisfaction and reaction to courseware instruction through interviews and survey methods, and a retrospective pretest-posttest performance survey were all highly valued components of the model as illustrated by the metaevaluation. In particular, the computer conference content analysis provided information which was very useful for assessing the effectiveness of computer conferencing for facilitating collaborative learning. The evaluator found this method to be of special significance in evaluating the extent and level of discussion, and for making judgements of the effectiveness of the design of the conference forums and the quality of the conferencing software.

One criticism of the original courseware evaluation model surrounded post-evaluation reflection. The original model provided no process for CME Planning Committee and Courseware Development Team members to meet, reflect upon, and discuss the results and findings of the evaluation. The original evaluation model did see the summative evaluation results reported to all courseware planning and development team members once the results had been compiled and analyzed. In the opinion of one stakeholder, a post-evaluation meeting was a key step in further planning and revision of the courseware product. In addressing the absence of this step an evaluation "postmortem" procedure was added to the modified model. This method would involve the gathering of all stakeholders and the discussion of summative evaluation results, identification of weaknesses and strengths, and the planning for changes and/or revisions to the courseware product. These changes and/or revisions would then be carried out to refine and improve the courseware product for future delivery.

Summary

This dissertation presented the concepts and methods underlying the planning, fieldtesting, and metaevaluation of a CME Courseware Evaluation Model. This model was designed and developed because there was a need for a comprehensive, systematic, and eclectic approach to evaluating the effectiveness of computer-mediated CME instructional courseware. An eclectic planning matrix was designed for formulating the model and a metaevaluation, an evaluation of an evaluation, was conducted to assess the effectiveness of the evaluation. The results of the fieldtest of the evaluation model revealed that computer-mediated CME was an effective distance education delivery system. Results of the metaevaluation indicated that the CME Courseware Evaluation Model was useful in providing information for decision making and product revision. In summary, the model has the potential to be a valuable tool for those responsible for planning, implementing, and evaluating computer-mediated instructional courseware.

REFERENCES

Abrahamson, S. (1968). Evaluation in Continuing Medical Education. *JAMA*, 206(3), 625-628.

Abrahamson, S. (1984). Evaluation of Continuing Education in the Health Professions. *Evaluation and the Health Professions*, 7(1), 3-23.

Babbie, E. (1992). *The Practice of Social Research*. Belmont, CA: Wadsworth Publishing Company.

Baker, J.R., Claus, R.N., & Manley, M. (1980). *Metaevaluation of the Saginaw Township Middle School Enrichment Center Project, 1979-80.* ED 206 720.

Bates, T. (1986). Computer Assisted Learning or Communications: Which Way for Information Technology in Distance Education? *Journal of Distance Education*, 1(1), 41-57.

Bandura, A., & Schunk, D.H. (1981). Cultivating Competence, Self-Efficacy, and Intrinsic Interest Through Proximal Self-motivation. *Journal of Personality and Social Psychology*, 41, 586-598.

Beaton, G.R. (1977). Continuing Education of Health Workers in Rural Area. *SA Medical Journal*, February, 201-202.

Benschoter, R.A., & Benson, S.J. (1992). Satellite System Addresses Rural Health Problems. *JBC*, 19(4), 26-30.

Berge, Z. (1995). Facilitating computer conferencing: Recommendations from the field. *Educational Technology*, 35(1), 22-30.

Berge, Z. (1996). *The Role of the Online Instructor*.
Available: http://cac.psu.edu/~mauri/moderate/teach_online.html.

Bhatara, V.S., Fuller, W.C., O'Connor-Davis, L., and Misra, L.K. (1996). Improving Job Satisfaction of Rural South Dakota Mental Health Providers Through Education: A Pilot Study. *South Dakota Journal of Medicine*, March, 93-96.

Bitt, H., Miles, D.A., Bridges-Webb, C.B., Neary, S., Charles, J., Traynor, V. (1993). A comparison of country and metropolitan general practice. *Medical Journal of Australia*, 159(suppl):S9-S64.

Black, D.P., & Dunikowski, L. (1985). Videotapes as Continuing Medical Education for Physicians in Isolated Communities. *Canadian Family Physician*, 31,1162-1163.

Blackwood, R., & McNab J. (1991). *A portrait of rural family practice: Problems and Priorities*: Toronto: College of Family Physicians of Canada,.

Bland, C.J., Ullian, J.A., and Froberg, D.G. (1984). User-Centred Evaluation. *Evaluation and the Health Professions*, 7(1), 53-63.

Boswell, E.J., Lorenz, R.A., Pichert, J.W., Schlundt, D.G., and Penha, M.L.I. (1994). Evaluation of a Home Study Continuing Education Program on Patient Teaching Skills. *The Journal of Continuing Education in the Health Professions*, 14(3), 155-165.

Bowes, W.A., Fryer, G.E., Raley, F.M. (1983). Rural Physicians and Continuing Education Programming. *Colorado Medicine*, March, 80-83.

Brigley, S., Littlejohns, Y., Young, Y., and McEwen, J. (1997). Continuing Medical Education: the question of evaluation. *Medical Education*, 31, 67-71.

Burge, E. (1994). Learning in Computer Conferenced Contexts: The Learners' Perspective. *Journal of Distance Education*, 9(1), 19-43.

Burleson, S., & Sugimoto, T. (1984). Cancer Education via Video-Teleconferencing: An Assessment of Participant Responses. *JBC*, 5-9.

Burnham, B.R. (1995). *Evaluating Human Resources, Programs and Organizations*. Malabar, FL: Krieger Publishing.

Busch, T. (1995). Gender Differences in Self-efficacy and Attitudes Toward Computers. *Journal of Educational Computing Research*, 12(2), 147-158.

Cairncross, R.G. (1985). Teaching and learning away from the medical centre. *Medical Education*, 19, 310-317.

Calhoun, P.S., & Fishman, E.K. (1994). Interactive Multimedia Program for Imaging the Spleen: Concept, Design, and Development. *Radiographics*, 14, 1407-1414.

Canadian Medical Association. (1992). Report of the advisory panel on the provision of medical services in underserviced regions. Ottawa: Canadian Medical Association.

Caplan, R.M. (1973). Measuring the Effectiveness of Continuing Medical Education. *Journal of Medical Education*, 48, 1150-1152.

Carter, R.G. (1987). The relationship between personal characteristics of physicians and practice location in Manitoba. *Canadian Medical Association Journal*, 136,366-368.

Cervero, R.M. (1988). *Effective Continuing Education for Professionals*. San Francisco: Jossey-Bass Publishers.

Chinien, C., & Hlynka, D. (1993). Formative Evaluation of Prototypical Products: From Expert to Connoisseur. *Educational Training & Technology International*, 30(1), 60-66.

Chouinard, J. (1983). Satellite contributions to telemedicine: Canadian CME experiences. *Canadian Medical Association Journal*, 128, 850-855.

Chow, Y.C., & Tan, K.M. (1994). Regional Electronic Bulletin Board System: Three Years of Experience with Continuing Medical Education. *The Journal of Continuing Education in the Health Professions*, 14(3), 166-172.

Clark, R.E. (1983). Reconsidering the research on learning from media. *Review of Educational Research*, 53, 445-459.

Clark, J.M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review*, 3, 149-210.

Coggins, C.C. (1988). Preferred learning styles and their impact on completion of external degree programs. *The American Journal of Distance Education*, 2(1), 25-37.

Coldeway, N.A. & DeLisa, J.A. (1986). Evaluation of Continuing Medical Education: Making a Difference. *Arch Phys Med Rehabil*, 67, 181-186.

Collins, M. (1991). Adult Education as Vocation: A Critical Role for the Adult Educator. Routledge: London & New York.

Collis, B.A. (1993). Evaluating Instructional Applications of Telecommunications in Distance Education. *Educational Technology & Training International*, 30(3), 266-274.

Cook, T.D., & Gruder, C.L. (1978). Metaevaluation Research. *Evaluation Quarterly*, 2(1), 5-51.

Craig, J.L., Gannon, P., and Brown, B.M. (1992). Providing Individualized CME to Geographically Isolated Specialists: A Feasibility Study. *The Journal of Continuing Education in the Health Professions*, 12, 15-23.

Crandall, S.J.S. (1990). The role of continuing medical education in changing and learning. *The Journal of Continuing Education in the Health Professions*, 10, 339-348.

Crandall, L.A., & Coggan, J.M. (1994). Impact of New Information Technologies on Training and Continuing Education of Rural Health Professionals. *The Journal of Rural Health*, 10(3), 208-215.

Cronbach, L.J. (1980). *Toward reform of Program Evaluation: Aims, Methods, and Institutional Arrangements.* San Francisco: Jossey-Bass.

Curry, L., & Purkis, I.E. (1986). Validity of Self-Reports of Behavior Changes By Participants After a CME Course. *Journal of Medical Education*, 61, 579-584.

Daniel, W.W. (1995). *Biostatistics: A Foundation for Analysis in the Health Sciences.* Toronto: John Wiley & Sons, Inc.

Davie, L. (1989). Facilitation techniques for the online tutor. In R. Mason & A. Kaye (Eds.), *Mindweave: Communication, computers, and distance education.* Elmsford, New York: Pergamon Press.

Davie, L., & Wells, R. (1991). Empowering the Learner through Computer-Mediated Communication. *Journal of Distance Education*, 5(1), 15-23.

Davis, D., Haynes, R.B., Chambers, L., Neufield, V.R., McKibbon, A., and Tugwell, P. (1984). The Impact of CME: A Methodological Review of the Continuing Medical Education Literature. *Evaluation and the Health Professions*, 7(3), 251-283.

Davis, D.A., Thomson, M.A., Oxman, A.D., and Haynes, R.B. (1992). Evidence for the Effectiveness of CME: A Review of 50 Randomized Controlled Trials. *JAMA*, 268(9), 1111-1117.

Davis, D., Thomson, M.A., Oxman, A.D., and Haynes, R.B. (1995). Changing Physician Performance: A Systematic Review of the Effect of Continuing Medical Education Strategies. *JAMA*, 274(9), 700-707.

Dawson, V.W. (1991). *Student Mediation Styles, Satisfaction and Achievement in an Electronic Distance Education Environment.* Unpublished Ph.D. thesis. Utah State University.

Dick, W., and Carey, L. (1985). *The Systematic Design of Instruction (2nd ed.).* Glenview, IL: Scott, Foresman and Co.

Dick, W., & Carey, L. (1990). *The Systematic Design of Instruction (3rd ed).* Glenview, IL: Scott, Foresman and Co.

Dillon, C., & Gunawardena, C. (1992). Evaluation research in distance education. *British Journal of Education Technology*, 23(3), 181-194.

Dorsch, J.L., & Landwirth, T.K. (1993). Rural GRATEFUL MED outreach: project results, impact, and future needs. *Journal of Medical Education*, 81(4), 377-382.

Doyle, J.D. (1996). Future Developments with the Internet: Some personal predictions. *International Journal of Clinical Monitoring and Computing*, 13, 95-98.

Draper, S.W., Brown, M.I., Henderson, F.P., and McAteer, E. (1996). Integrative Evaluation: An Emerging Role for Classroom Studies of CAL. *Computers in Education*, 26(1), 17-32.

Duffield, J. (1991). Designing Computer Software for Problem-Solving Instruction. *Educational Technology Research and Development*, 39(1), 50-62.

Duke, N. (1996). *Web Page Evaluation Worksheet.*
Available: http://www.duke.edu~de1/evaluate.html, St. John's University, Division of Library and Information Science.

Dunn, E.V., Acton, H., Conrath, D., Higgins, C., and Bain, H. (1980). The Use of Slow Scan Video for CME in a Remote Area. *Journal of Medical Education*, 55, 493-495.

Dunn, E.V., & Fisher, M. (1985). The use of freeze frame (slow scan) video for health professional education. *Medical Education*, 19, 148-154.

Ebbert, A. (1963). Two-Way Radio in Medical Education. *Journal of Medical Education*, 38, 319-328.

Ehrman, M. (1990). Psychological factors and distance education. *The American Journal of Distance Education*, 4(1), 10-24.

Eisner, E.W. (1975). *The perceptive eye: Toward the information of educational evaluation*. Invited address at the American Education Research Association, Washington, DC., In Burnham .(1995). *Evaluating Human Resources, Programs and Organizations*. Malabar, FL: Krieger Publishing.

Eisner, E.W. (1979). *The educational imagination.* New York: Macmillan.

Eisner, E.W. (1985). *The art of educational evaluation: a personal view*. London: Falmer Press.

Ellis, R.D., Jankowski, T.B., Jasper, J.E., and Abdul, A. (1996). Gero-Informatics and the Internet: Locating Gerontology Information on the World Wide Web. *The Gerontologist*, 36(1), 100-105.

Engel, C.E., Browne, E., Nyarango, P., Akors, S., Khwaja, A., Karim, A.A., and Towle, A. (1992). Problem-based learning in distance education: a first exploration in continuing medical education. *Medical Education*, 26, 389-401.

Ertmer, P.A., Evenbeck, E., Cennamo, K.S., and Lehman, J.D. (1994). Enhancing Self-Efficacy for Computer Technologies Through the Use of Positive Classroom Experiences. *Educational Technology Research and Development*, 42(3), 45-62.
Evans, C.E., Haynes, R.B., Birkett, N.J., Gilbert, J.R., Taylor, D.W., Sackett, D.L., Johnston, M.E., and Hewson, S.A. (1986). Does a Mailed Continuing Education Program Improve Physician Performance? *JAMA*, 255(4), 501-504.

Fann, G.L., Lynch, D.L., and Murranka, P. (1989). Integrating Technology: Attitudes as a Determinant of the Use of Computers. *Journal of Educational Technology Systems*, 17(4), 307-317.

Feenberg, A., & Bellman, B. (1990). Social factor research in computer mediated communications. In L. Harasim (Ed.), *Online Education. Perspectives on a new environment*. New York: Praeger Publishing.

Ference, P., & Vockell, E. (1994). Adult Learning Characteristics and Effective Software Instruction. *Educational Technology*, 34(6), 25-31.

Ferguson, K.J., Caplan, R.M., and Williamson, P.S. (1984). Factors Associated with Behavior Change in Family Physicians After CME Presentation. *Journal of Medical Education*, 59, 662-666.

Forti, E.M., Martin, K.E., Jones, R.L., Herman, J.M. (1995). Factors influencing retention of rural Pennsylvania family physicians. *Journal of the American Board of Family Practice*, 8, 469-74.

Francis, L.J., & Evans, T.E. (1995). The Reliability and Validity of the Bath County Computer Attitude Scale. *Journal of Educational Computing Research*, 12(2), 135-146.

Galbreath, J. (1992). The Educational Buzzword of the 1990's: Multimedia, or is It Hypermedia, or Interactive Multimedia or..." *Educational Technology*, 32(4), 15-19.

Galbreath, J. (1995). Multimedia on the Network: Has Its Time Come? *Educational Technology*, 35(4), 44-51.

Galbreath, J. (1997). The Internet: Past, Present, and Future. *Educational Technology*, 37(6), 39-45.

Galvin, J.R., D'Alessandro, M.P., Erkonen, W.E., Knutson, T.A., and Lacey, D.L. (1994). The Virtual Hospital: A New Paradigm for Lifelong Learning in Radiology. *Radiographics*, 14, 875-879.

Ganiats, T.G., & Groveman, H.D. (1986). Interactive Computer-Assisted Instruction on Cancer Screening Guidelines. *The Journal of Family Practice*, 23(3), 281-283.

Gay, L.R. (1996). *Educational Research: Competencies for analysis and application.* Englewood Cliffs: Prentice-Hall.

Gellman, E.F., & Franke, T.C. (1996). Experience with Four Years of CME Teleconferencing at St. Louis Children's Hospital, Washington University School of Medicine. *The Journal of Continuing Education in the Health Professions*, 16, 250-253.

Gill, B., Dick, W., Reiser, R.A., and Zahner, J.E. (1992). A new model for evaluating instructional software. *Educational Technology*, 32(3), 30-44.

Gill, D., & Game, D. (1994). Continuing medical education needs of rural GPs in South Australia. *Australian Family Physician*, 23(4), 663-667.

Glaser, B.G., & Strauss, A.L. (1967). *The discovery of grounded theory: Strategies for qualitative research.* New York: Aldine De Gruyter.

Golas, K.C. (1983). Formative Evaluation Effectiveness and Cost. *Performance & Instruction Journal*, 22(5), 17-19.

Gooler, D.D. (1979). Evaluating Distance Education Programs. *Canadian Journal of University Continuing Education*, 6(1), 43-55.

Gould, R.B, Basarab, D.J., McGuire, C., Robinson, P.C., Walser, F.L., & Wigdor, A.K. (1995). *The Development, Validation, and Applicability of "The Program Evaluation Standards: How to Assess Evaluations of Educational Programs."* ED 403 314

Gowin, D., & Millman, J. (1978). *Can Metaevaluation Give a Direction for Research on Evaluation*? Paper Presented at the Annual Meeting of the American Educational Research Association (62nd, Toronto, Ontario). ED 161 938.

Grassin, E. (1997). *Thinking Critically about World Wide Web Resources.* Available: http://www.library.ucla.edu/libraries/college/instruct/critical.htm, UCLA College Library, May 5, 1997.

Gray, J.D., Steeves, L.C., and Blackburn, J.W. (1994). The Dalhousie University experience of training residents in many small communities. *Academic Medicine*, 69(10):847-851.

Greenes, R.A. (1992). Medical Education and Decision Support Using Network-based Multimedia Information Resources. *Annals New York Academy of Sciences*, 670, 244-256.

Grunfield, A., Kendall, H., and Walls, R. (1996). The Internet: What's All The Fuss About? *The Journal of Emergency Medicine*, 14(6), 769-770.

Guba, E.G. (1981). Criteria for assessing the trustworthiness of naturalistic enquires. *Educational Communication and Technology*, 29(2), 75-91.

Guba, E.G., & Lincoln, Y.S. (1985). *Effective Evaluation*. San Francisco: Jossey-Bass.

Guba, E.G., & Lincoln, Y.S. (1989). *Fourth Generation Evaluation*. Sage Publications.

Hampton, C.L., Mazmanian, P.E., Smith, T.J. (1994). The Interactive Videoconference: An Effective CME Delivery System. *The Journal of Continuing Education in the Health Professions*, 14, 83-89.

Hannafin, M.J. (1983). Measuring the Importance of Learning from Instruction: The Outcome-Consequence Model of Evaluation. *Journal of Instructional Development*, 6(3), 14-18.

Harasim, L. (1986). Computer Learning Networks: Educational Applications of Computer Conferencing. *Journal of Distance Education*, 1(1), 59-70.

Harden, R.M. & Laidlaw, J.M. (1992). Effective continuing education: the CRISIS criteria. *Medical Education*, 26(5), 408-422.

Harper, G., & Kember, D. (1986). Approaches to the study of distance education students. *British Journal of Educational Technology*, 17(3), 2112-222.

Harrison, P.J., & Seeman, B.J. (1991). Development of a distance education assessment instrument. *Educational Technology Research and Development*, 39(4), 65-78.

Hartlaub P.P., Gordon R.L. (1993). Rural practice. Preference, perception, and reality. *Archives of Family Medicine*, 2(11):1198-201.

Hawkes, M.L. (1996). Evaluating School-Based Distance Education Programs: Some Thoughts About Methods. *NASSP Bulletin*, 80(582), 26-33.

Hayden, K.A. (1997). Internet Tools and Resources in Continuing Health Education. *The Journal of Continuing Education in the Health Professions*, 17, 121-127.

Hayes, K., & Lehman, C.U. (1996). The Interactive Patient: A Multimedia Interactive Educational Tool On The World Wide Web. *M.D. Computing*, 13(4), 330-334.

Hays ,R.B., Craig, M.L., Wise, A.L., Nichols, A., Mahoney, M.D., and Adkins, P.B. (1994). A sampling framework for rural and remote doctors. *Australian Journal of Public Health*, 18(3):273-6.

Health Canada. (1997). *Health Canada's Sustainable Development Strategy*. Sustainable Development Secretariat. Ottawa: Health Canada.

Heinzen, T.E., & Alberico, S.M. (1990) Using a Creativity Paradigm to Evaluate Teleconferencing. *The American Journal of Distance Education*, 4(3), 3- 12.

Hennen, B. (1997). Demonstrating social accountability in medical education. *Canadian Medical Association*, 156(3), 365-367.

Holmberg, B. (1995). *Theory and Practice of Distance Education*. London: Routledge.

Houle, C.O. (1980). *Continuing Learning in the Professions*. San Francisco: Jossey-Bass.

Houle, P.A. (1996). Toward Understanding Student Differences in a Computer Skills Course. *Journal of Educational Computing Research*, 14(1), 25-47.

House, E.R. (1978). Assumptions underlying evaluation models. *Educational Researcher*, 7, 4-12.

House, A.M. (1981). Telecommunications in health and education. *CMA Journal*, 124, 667-668.

House, E.R. (1991). Realism in research. *Educational Researcher*, 20(7), 2-22.

House, A.M., Roberts, J.M., and Canning, E.M. (1981). Telemedicine provides new dimensions in CME in Newfoundland and Labrador. *CMAJ*, 124, 801-804.

Howard, G.S., Schmeck, R.R., & Bray, J.H. (1979). Internal Invalidity in Studies Employing Self-Report Instruments: A Suggested Remedy. *Journal of Educational Measurement*, 16(2), 129-135.

Huang, A. (1997). Challenges and Opportunities of Online Education. *Journal of Educational Technology Systems*, 25(3), 229-247.

Irvine, H., Rowntree, C., and Thompson, J. (1994). *Needs of rural physicians.* Available: http:// www.agt.net/public/asrm/ruralama.htm, Regional Health Authority, David Thompson Health Care Region, Alberta. (June, 1997).

Jennett, P.A., Lakdal, O.E., Hayton, R.C., Klaassen, D.J., Swanson, R.W., Wilson, T.W., Spooner, H.J., Mainprize, G.W., and Wickett, R.E.Y. (1988). The effects of continuing medical education on family doctor performance in office practice: a randomized control study. *Medical Education*, 22, 139-145.

Jennett, P.A., Lockyer, J.M., Maes, W., Parboosingh, I.J., and Lawson, D. (1990). Providing Relevant Information to Rural Practitioners: A Study of a Medical Information System. *Teaching and Learning in Medicine*, 2(4), 200-204.

Jennett, P.A., & Swanson, R.W. (1994). Traditional and New Approaches to CME: Perceptions of a Variety of CME Activities. *The Journal of Continuing Education in the Health Professions*, 14, 75-82.

Joint Committee on Standards for Educational Evaluation (1st ed.). (1981). *Standards for evaluations of educational programs, projects and materials*. New York: McGraw-Hill.

Joint Committee on Standards for Educational Evaluation (2nd ed.). (1994). *The Program Evaluation Standards*. London: Sage Publications.

Jolicoeur, K., & Berger, D.E. (1988a). Implementing Education Software and Evaluating Its Academic Effectiveness: Part I. *Educational Technology*, 28(9), 7-13.

Jolicoeur, K., & Berger, D.E. (1988b). Implementing Education Software and Evaluating Its Academic Effectiveness: Part II. *Educational Technology*, 28(10), 13-19.

Jones, D.L. (1990). Viability of the Commitment-for-Change Evaluation Strategy in Continuing Medical Education. *Academic Medicine*, 65(9), S37-S38.

Jones, P.E., & Wall, R.E. (1989). Components of Computer Anxiety. *Journal of Educational Technology Systems*, 18(2), 161-168.

Jones, A., Scanlon, E., Tosunoglu, C., Ross, S., Butcher, P., Murphy, P., and Greenberg, J. (1996). Evaluating CAL at the Open University: 15 Years On. *Computers in Education*, 26(1), 5-15.

Jonassen, D.H. (1995). Supporting Communities of Learners with Technology: A Vision for Integrating Technology with Learning in Schools. *Educational Technology*, 35(4), 60-63.

Kamien, M., & Buttfield, I.H. (1990). Some Solutions to the shortage of general practitioners in rural Australia. *The Medical Journal of Australia*, 53, 168-171.

Kantrowitz, M.P., Vogel, I.W., and Wiese, W.H. (1979). *University of New Mexico's approach to continuing medical education for National Health Service Corps professional assigned to rural areas*. Public Health Reports: Hyattsville, 94:281-284.

Kember, D., Lai, T., Murphy, D., Siaw, I., Wong, J., and Yuen, K.S. (1990). Naturalistic Evaluation of Distance Learning Courses. *Journal of Distance Education*, 5(1), 38-52.

Kember, D., & Dekkers, J. (1987). The role of study centres for academic support in distance education. *Distance Education*, 8(1), 4-17.

Kerr, D., Toghill, P., and Klar, B. (1996). CME and the role of the Journal of the Royal College of Physicians of London. *Journal of the Royal College of Physicians of London*, 30(3), 246-251.

Kerr, B.D. (1997). *A responsive evaluation of a graduate distance education course offering*. Unpublished Thesis (M.Ed.), Memorial University of Newfoundland.

Kettle, B.W. (1994). *A study of the application of a modified transactionist program evaluation model to a non-formal distance education course for community-based adult literacy providers*. Thesis (M.Ed.), Memorial University of Newfoundland.

Kinn, S., & Jones, R. (1995). *Continuing Education in Health Informatics in the UK: The Need for Learning Materials*. MEDINFO 95 Proceedings, 1260-1264.

Kinzie, M. (1990). Requirements and Benefits of Effective Interactive Instruction: Learner Control, Self-Regulation, and Continuing Motivation. *Educational Technology Research and Development*, 38(1), 1-21.

Kinzie, M., & Berdel, R. (1990). Design and Use of Hypermedia Systems. *Educational Technology Research and Development*, 38(3), 61-68.

Kirkpatrick, D.L. (1976). Evaluation of Training., In Knowles, M.S. (1980). *The modern practice of adult education: from pedagogy to andragogy*. Chicago: Follett Pub. Co.

Knowles, M.S. (1980). *The modern practice of adult education*: from pedagogy to andragogy. Chicago: Follett Pub. Co.

Kripalani, S., Cooper, H.P., Weinberg, A.D., and Laufman, L. (1997). Computer-Assisted Self-Directed Learning: The Future of Continuing Medical Education. *The Journal of Continuing Education in the Health Professions*, 17, 114-120.

Kryspin, W.J., & Feldhusen, J.F. (1974). *Developing Classroom Tests: A Guide for Writing and Evaluating Test Items*. Minneapolis: Burgess Publishing Co.

Laporte, R.E., Akazawa, S., Drash, A., Gamboa, C., Lee, H.K., Libman., Marler, E., Orschiedt, S., Songer, T., and Tajima, N. (1995). Diabetes and the Internet. *Diabetes Care*, 18(6), 890-895.

Lawless, K. & Brown, S. (1997). Multimedia Learning Environments: Issues of learner control and navigation. *Instructional Science*, 25, 117-131.

Laxdall, O.E., Jennett, P.A., Wilson, T.W., and Salisbury, G.M. (1978). Improving physician performance by continuing medical education. *CMA J*, 118, 1051- 1058.

LeCompte, M.D. Preissle, J. (1993). *Ethnography and Qualitative Design in Educational Research.* Toronto: Academic Press.

Leduc, E. (1997). Defining rurality: A General Practice Rurality Index for Canada. *Canadian Journal of Rural Medicine*, 2(2):125.

Lee, V., Murphy,, D., Chan, C., and Chung, L. (1997). Computer-aided distance learning: a case study. *Open Learning*, 12(1), 58-62.

Lee, J.M., Koide, M.M., Salvador, E.M., Moura, J.M.A., Ramos, M.P., Uras, R., Matos, R.B., Ancao, M.S., and Sigulem, D. (1995). Educational Program on Opthamology. *MEDINFO 95 Proceedings*, 1241-1243.

Leist, J.C., & Kristofco, R.E. (1990). The changing paradigm for continuing medical education: impact of information on the teachable moment. *Bulletin of the Medical Library Association*, 78(2), 173-179.

Levin, J.A., Kim, H., and Riel, M.M. (1990). Analyzing Instructional Interaction on Electronic Message Networks. In L.M., Harasim (Eds.), *Online Education: Perspectives on a New Environment*, Praeger, New York.

Lindquist, J. (1978). Strategies for Change. In Dillon, C., and Gunawardena, C. (1992). Evaluation research in distance education. *British Journal of Education Technology*, 23(3), 181-194.

Lindsay, E.A., Davis, D.A., Fallis, F., Willison, D.B., and Biggar, J. (1987). Continuing Education through Telemedicine for Ontario. *CMAJ*, 137, 503-506.

Lloyd, J.S., & Abrahamson, S. (1979). Effectiveness of Continuing Medical Education: A Review of the Evidence. *Evaluation and the Health Professions*, 2, 251-280.

Locke, S.E., & Rezza, M.E.H. (1996). Computer-Based Education in Mental Health. *M.D. Computing*, 13(1), 10-17.

Lockyer, J.M., Parboosingh, I.J., and McDowell, A.C. (1987). Teleconferencing and The Continuing Education Preferences of Physicians and Nurses in Rural Areas. *MOBIUS*, 7(3), 53-58.

Lott, D.R. (1995). Obstacles to Self-Paced Learning. *The Journal of Continuing Education in the Health Professions*, 15, 203-208.

Madaus, G.F., Scriven, M..S., and Stufflebeam, D.L. (1983). *Evaluation Models, Viewpoints on Educational and Human Services Evaluation.* Boston: Kluwer-Nijhoff Publishing: .

Maddux, C., & Johnson, D.L. (1997). The World Wide Web: History, Cultural Context, and a Manual for Developers of Educational Information-Base Web Sites. *Educational Technology*, 37(4), 5-11.

Magnus, J.H., & Tollan, A. (1993). Rural doctor recruitment: does medical education in rural districts recruit doctors to rural areas? *Medical Education*, 25,250-253.

Mainous, A.G., Ramsbottom-Lucier, M., Rich, E.C. (1994). The role of clinical workload and satisfaction with workload in rural primary care physician retention. *Archives of Family Medicine*, 3:787-92.

Malaney, G.D., & Thurman, Q. (1989). Key Factors in the Use and Frequency of Use of Microcomputers by College Students. *Journal of Educational Technology*, 18(2), 151-160.

Mann, K.V. (1996). Part I: Connections across the Continuum of Professional Education. *The Journal of Continuing Education in the Health Professions*, 16, 133-135.

Manning, P.R., & Petit, D.W. (1987). The Past, Present, and Future of Continuing Medical Education. *JAMA*, 258(24), 3542-3546.

Manning, P.R. (1990). Continuing Education needs of health care professionals. *Bulletin of the Medical Library Association*, 78(2), 161-164.

Marquis, Y., Chaoulli, J., Bordage, G., Chabot, J.M., and Leclere, H. (1984). Patient-management problems as a learning tool for the continuing medical education of general practitioners. *Medical Education*, 18, 117-124.

Marshall, G. (1994). *The Concise Oxford Dictionary of Sociology*. Oxford University Press: Oxford, NY.

Mazmanian, P.E., Waugh, J.L., and Mazmanian, P.M. (1997). Commitment to Change: Ideational Roots, Empirical Evidence, and Ethical Implications. *Journal of Continuing Education in the Health Professions*, 17, 133-140.

McConell, D. (1994). *Implementing Computer Supported Cooperative Learning*. London: Kogan Page.

McCreary, E.Z., & Van Duren, J. (1987). Educational Applications of Computer Conferencing. *Canadian Journal of Educational Communications*, 16(2), 107- 115.

McDonald, C.J. (1983). Computer Technology and Continuing Medical Education. *MOBIUS*, 3(2), 7-12.

McDowell, C.A., Challis, E.B., Lockyer, J.M., White, L., Adams, K., and Parboosingh, I.J. (1987). Teleconferencing CME Programs to Rural Physicians: The University of Calgary Teleconference Program. *Canadian Family Physician*, 33, 1705-1708.

McLaughlin, P.J., & Donaldson, J.F. (1991). Evaluation of Continuing Medical Education Programs: A Selected Literature, 1984-1988. *The Journal of Continuing Education in the Health Professions*, 11, 65-84.

Mehrens, W.A., & Lehmann, I.J. (1991). *Measurement and Evaluation in Education and Psychology*. Toronto: Holt, Rinehart and Winston, Inc.

Meissner, F., & Vujicic, R. (1995). Critical Care Medicine in Cyperspace: Internet Resources and Techniques for the Intensivist. *Lijec Vjesn*, 117(2), 82-84.

Melton, R.F. (1995). Developing a formative evaluation system for distance teaching. *Open Learning*, June, 53-57.

Mendenhall, R.C., Girard, R.A., and Abrahamson, S. (1978a). A national study of medical and surgical specialities, I: background, purpose and methodology. *JAMA*, 240, 848-852.

Mendenhall, R.C., Girard, R.A., and Abrahamson, S. (1978b). A national study of medical and surgical specialities, II: description of the survey instrument. *JAMA*, 240, 1160-1168.

Meyer, T.C. (1983). Teleconferencing As a Medium for Continuing Education of Health Professionals. *MOBIUS*, 3(2), 73-79.

Money, P. (1986). Continuing Medical Education for Isolated Rural General Practitioners - A Journal Project. *Family Practice*, 3(2), 117-119.

Mooney, G.A., Bligh, J.G., Leinster, S.J., and Warenius, H.M. (1995). An electronic study guide for problem-based learning. *Medical Education*, 29, 397-402.

Moore, D.E., Green, J.S., Jay, S.J., Leist, J.C., and Maitland, F.M. (1994). Creating a New Paradigm for CME: Seizing Opportunities Within the Health Care Revolution. *The Journal of Continuing Education in the Health Professions*, 14, 4-31.

Moore, M., & Hartman, J.T. (1992). Information technology for rural outreach in West Texas. *Bulletin of the Medical Library Association*, 80(1), 44-45.

Muller, E. (1995). Application of Experimental and Quasi-experimental Designs to Educational Software Evaluation. *Educational Technology*, 25(10), 27-31.

Najjar, L. (1996). Multimedia Information and Learning. *Journal of Educational Multimedia and hypermedia*, 5(2), 129-150.

Nanberg, D.W. (1983). The Application of Satellite Television Broadcasting for Continuing Professional Education. *MOBIUS*, 3(2), 56-63.

Neuman, D. (1989). Naturalistic inquiry and Computer-Based Instruction: Rationale, Procedures, and Potential. *Educational Technology Research and Development*, 37(3), 39-51.

Newman, D. (1990). Cognitive and technical issues in the design of educational computer networking. In L. Harasim (Ed.), *Online Education: Perspectives on a new environment*. New York: Praeger Press.

Nickell, G.S., & Pinto, J.N. (1986). The Computer Attitude Scale. *Computers in Human Behavior*, 2, 301-306.

Northrup, P.T. (1995). Concurrent Formative Evaluation: Guidelines and Implications for Multimedia Designers. *Educational Technology*, 35(6), 24-31.

Nyarango, P.M. (1991). Distance Learning for Rural Medical Officers in Kenya: The First Pilot Project. *East African Medical Journal*, September, 741-743.

O'Dochartaigh, E. (1971). Continuing Education for Doctors in Rural Areas. *Journal of the Irish Medical Association*, 64, 484-486.

Oeffinger, J.C., Hiebeler, L., Sherman, T., Gaskill, M., Portante, T., Polasek, J., and Litterer, K. (1992). Innovative Desktop Learning Tools Implications for Rural Hospitals and Physicians. *Annals New York Academy of Sciences*, 670, 76-90.

Ohl, T., & Ward, M. (1997). Applying Metaphorical Interface Design Principles on the World Wide Web. *Educational Technology*, 37(6), 25-38.

Owston, R.D., & Wideman, H.H. (1987). The Value of Supplementing Panel Software Reviews with Field Observations. *CJEC*, 16(4), 295-308.

Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart and Winston.

Paivio, A. (1986). *Mental representations: A dual-coding approach*. New York: Oxford University Press.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology*, 45, 255-287.

Parker, L.A., & Baird, M.A. (1977). Continuing Education by Telephone: A Party Line for Professionals. *JAHA*, 51, 105-112.

Parlett, M., & Hamilton, D. 1972. Evaluation as illumination. In M.Partlett & G. Dearden (Eds.) (1981), *Introduction to illuminative evaluation: Studies in higher education* (pp. 9-29). UK: University of Surrey, Society for Research into Higher Education.

Parlett, M., & Dearden, G. (1977). *Introduction to Illuminative Evaluation: Studies in Higher Education*. Pacific Soundings Press.

Patton, M.Q. (1980). *Qualitative Evaluation Methods*. Beverly Hills: Sage Publications.

Pereles, L., Lockyer, J., Hogan, D., Gondocz, T., and Parboosingh, J. (1997). Effectiveness of Commitment Contracts in Facilitating Change in Continuing Medical Education Intervention. *The Journal of Continuing Education in the Health Professions*, 17, 27-31.

Piemme, T.E. (1988). Computer-Assisted Learning and Evaluation in Medicine. *JAMA*, 260(3), 367-372.

Pivalo, V. (1994). Outreach to Health Professionals in a Rural Area. *Medical Reference Services Quarterly*, 13(3), 19-26.

Pomeroy, E., & Detweiler, M. (1995). Compact Disc Interactive (CD-i) Multimedia Project. *Journal of Biocommunication*, 22(1), 7-13.

Purkis, I.E. (1982). Commitment for Change: An Instrument For Evaluating CME Courses. *Journal of Medical Education*, 57, 61-63.

Puskin, D.S. (1992). Telecommunications in Rural America. *Annals New York Academy of Sciences*, 670, 67-75.

Rankin, J.A. (1992). GaIN: On-Line Delivery of Medical Information to Physicians and Hospitals in Georgia. *Annals New York Academy of Sciences*, 670, 171-179.

Reiser, R.A., & Dick, W. (1990). Evaluating Instructional Software. *Educational Technology Research and Development*, 38(3), 43-50.

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.

Report of the Advisory Panel on the Provision of Medical Services in Underserviced Regions. (1992). Ottawa: Canadian Medical Association.

Rosenblatt R.A., & Moscovice I.S. (1982). *Rural health care*. Toronto: John Wiley & Sons.

Rosenblatt, R.A. (1984). The Microcomputer as a Vehicle for Continuing Medical Education. *The Journal of Family Practice*, 18(4), 629-632.

Rosenblatt, R.A., Whitcomb, M.E., Cullen, T.J., Lishner, D.M., and Hart, L.G. (1992). Which medical schools produce rural physicians? *JAMA*, 268:1559-1565.

Rosenthall, J.R., & Miller, T.C. (1982). Divergent Pediatric CME Priorities in a Rural Area. *Journal of Medical Education*, 57, 193-194.

Ross, J. (1996). The Influence of Computer Communication Skills on Participation in a Computer Conferencing Course. *Journal of Educational Computing Research*, 15(1), 37-52.

Ross, J., Crane, C., & Robertson, D. (1995). Equity of Access to Computer-Mediated Distance Education. *Journal of Distance Education*, 10(2), 17-32.

Rourke, J. (1988). Rural Family Practice Part II: Preferences in Continuing Medical Education. *Canadian Family Physician*, 34, 1035-1038.

Rourke J. (1991). Perspectives on rural medical care in Ontario. *Canadian Family Physician*, 37:1581-1584,1647.

Rourke, J. (1993). Politics of rural health care: recruitment and retention of physicians. *Canadian Medical Association Journal*, 148(8), 1281-1284.

Rourke, J. (1994). The Politics of Rural Medical Care. *Ontario Medical Review*. August 1994, 17-22.

Rourke J., & Strasser, R. (1996). Education for rural practice in Canada and Australia. *Academic Medicine*, 71(5):464-9.

Rourke, L., Graham, S., and Brown, J.B. (1996). Establishing personal and professional boundaries: the challenge for physicians in small communities. *Ontario Medical Review*, 63(5),74-5.

Rousseau N. (1995). What is rurality? *Occasional Papers of the Royal College of General Practice*, 71,1-4.

Rowntree, D. (1986). *Developing Courses for Students*. London: Harper and Row.

Rowntree, D. (1992). *Exploring Open and Distance Learning*. London: Kogan Page.

Rowntree, D. (1995). Teaching and Learning online: a correspondence education for the 21st century. *British Journal of Educational Technology*, 26(3), 205-215.

Rubenstein, L.Z., Rubenstein, L., Elkin, P., and Elkin, S.P. (1975). Determinants of the Choice of Rural Practice: A Study of Yugoslav General Practitioners. *Journal of Medical Education*, 50, 615-623.

Ruberg, L.F., Moore, D.M., and Taylor, C.D. (1996). Student Participation, Interaction, and regulation in a Computer-Mediated Communications Environment: A Qualitative Study. *Journal of Educational Computing Research*, 14(3), 243-268.

Russell, J.D., and Blake, B.L. (1988). Formative and Summative Evaluation of Instructional Products and Learners. *Educational Technology*, 28(9), 22-28.

Rutman, L. (1984). *Evaluation Research Methods: A Basic Guide*. Beverly Hills, CA: Sage Publications:

Salti, I.S. (1995). Continuing medical education. *Medical Education*, 29(supplement 1), 97-99.

Sanders, J., Brucker, P., and Miller, M.D. (1995). Using Telemedicine for Continuing Education for Rural Physicians. *Academic Medicine*, 70(5),457.

Sanmartin, C.A., & Snidal, L. (1993). Profile of Canadian physicians: Results of the 1990 Physician Resource Questionnaire. *Canadian Medical Association Journal*, 149(7), 977-84.

Scott, C.J. (1994). Applied Adult Learning Theory: Broadening Traditional CME Programs with Self-Guided, Computer-Assisted Learning. *The Journal of Continuing Education in the Health Professions*, 14, 91-99.

Screnci, D., Hirsch, E., Levy, K., Skawinski, E., and DerBoghosian, M. (1996). Medical Outreach to Armenia by Telemedicine Linkage. *Journal of Medical Systems*, 20(2), 67-76.

Scriven, M.S. (1967). The methodology of evaluation. *In perspectives of curriculum evaluation* (AERA Monograph series on Curriculum Evaluation, No. 1). Chicago: Rand-McNally.

Scriven, M.S. (1969). An Introduction to Metaevaluation. *Educational Product Report*, 2(5), 36-38.

Scriven, M.S. (1987). Pros and cons about Goal-Free evaluation., In B.R. Worthen & J.R. Sanders (Eds), *Educational evaluation: Alternative approaches and practical guidelines*. New York, NY: Longman.

Seaton, W. (1993). Computer-Mediated Communication and student self-directed learning. *Open Learning*, 8(3), 49-54.

Shashaani, L. (1994). Gender Differences in Computer Experience and Its Influence on Computer Altitudes. *Journal of Educational Computing Research*, 11(4), 347-367.

Shelstad, K.R., & Clevenger, F.W. (1996). Information retrieval patterns and needs among practicing general surgeons: a statewide experience. *Bulletin of the Medical Library Association*, 84(4), 490-497.

Shih, Y.F. & Alessi, S. (1996). Effects of Text Versus Voice on Learning in Multimedia Courseware. *Journal of Educational Multimedia and Hypermedia*, 5(2), 203-218.

Sheets, K.J. (1983). *The Design, Development, and Fieldtest of an Evaluation Framework for Short-Term Training Programs*. Unpublished Doctoral Dissertation, Michigan State University.

Sherwood, G.D. (1995). Creating Change through Partnership: Developing Education Resources in Rural Areas. *The Journal of Continuing Education in the Health Professions*, 15, 101-105.

Shortcliffe, E.H. (1983). Problems in Implementing the Computer for Continuing Education. *MOBIUS*, 3(2), 52-55.

Sibley, J.C., Sackett, D.L., Neufield, V., Gerrard, B., Rudnick, K.V., and Fraser, W. (1982). A randomized trial of continuing medical education. *The New England Journal of Medicine*, 306, 511-515.

Smith, A. (1998). *Criteria for evaluation of Internet Information Resources*. Available: http://www.vuw.ac.nz/~agsmith/evaln/, VUW Department of Library and Information Studies. (March 7, 1998).

Spencer, K. (1991). Modes, media and methods: the search for educational effectiveness. *British Journal of Educational Technology*, 22(1), 12-22.

Spille, H.A., Galloway, S.W., and Stewart, D.W. (1985). How to Evaluate Courseware Designed for Use in Education or Training Programs for Adult Learners. *Educational Technology*, 25(3), 40-42.

Stake, R.E. (1967). The Countenance of Educational Evaluation. *Teachers College Record*, 68, 523-540.

Stake, R.E. (1977). *Responsive Evaluation.*, In D. Hamilton, B. MacDonald, C. King, and M. Parlett (Eds), *Beyond the numbers game*. Berkeley, CA: McCuthcan Publishing Co.

Starr, R. (1997). Delivering Instruction on the World Wide Web: Overview and Basic Design Principles. *Educational Technology*, 37(3), 7-14.

Steele, S.M. (1973). *Contemporary Approaches to Program Evaluation and Their Implications for Evaluating Programs for Disadvantaged Adults*. Syracuse, NY: ERIC Clearinghouse on Adult Education.

Steeples, C. (1993). A Computer-Mediated Learning Environment for Adult Learners: Supporting Collaboration and Self-Direction. *Journal of Educational Multimedia and Hypermedia*, 2(4), 443-454.

Stein, L.S. (1981). The Effectiveness of Continuing Medical Education: Eight Research Reports. *Journal of Medical Education*, 56 (February), 103-110.

Steiner, V. (1995). *What is Distance Education.*
Available: http://www.fwl.org/edtech/ distance.html (October, 1995)

Sternberg, R.J. (1986). Electronic Conferencing for Continuing Medical Education. *MOBIUS*, 6(4), 39-48.

Stevenson, J.F., Longabaugh, R.H., & McNeill, D.N. (1979). Metaevaluation in the Human Services. In H.C. Schulberg & J.M. Jerrell (Eds.), *The Evaluator and Management, Sage Research Progress in Evaluation (Volume 4)*. Beverly Hills: Sage Publications.

Stocking, J.E., & Benjamin, P. (1995). Current Status of Interactive Multimedia Education in Medicine. *M.D. Computing*, 12(5), 373-380.

Storey, P.B. (1983). The Future of Computers in Continuing Medical Education. *MOBIUS*, 3(4), 56-63.

Strasser, R.P. (1992). Attitudes of Victorian rural general practitioners to country practice and training. *Australian Family Physician*, 21(7), 808-812.

Strasser, R. (1995). Rural general practice: is it a distinct discipline? *Australian Family Physician*, 24(5):870-6.

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Newbury Park, CA: Sage.

Stufflebeam, D.L. (1967). The Use and Abuse of Evaluation. In D.L. Stufflebeam and A.J. Shinkfield (Eds.), *Systematic Evaluation: A Self-instructional Guide to Theory and Practice*. Kluwer-Nijhoff Publishing, Boston.

Stufflebeam, D.L. (1971). *Educational Evaluation and Decision Making*. Itasca, ILL.: Peacock.

Stufflebeam, D.L. (1974). *Toward a Technology for Evaluating Evaluations*. Paper presented at the annual meeting of the American Educational Research Association, Chicago. ED 090 319.

Stufflebeam, D.L. (1978). Metaevaluation: An Overview. *Evaluation and the Health Professions*, 1(1), 17-43.

Stufflebeam, D.L. (1981). Metaevaluation: Concepts, Standards, and Uses. In R. Berk (Ed.), *Educational Evaluation Methodology: The State of the Art*. Baltimore: The John Hopkins University Press.

Stufflebeam, D.L., & Shinkfield, A.J. (1985). *Systematic Evaluation: A Self-instructional Guide to Theory and Practice*. Boston: Kluwer-Nijhoff Publishing.

Suchman, E. (1967). *Evaluative Research: Principles and Practice in Public Service Social Action Programs*. New York: Russell Sage Foundation.

Sweeters, W. (1994). Multimedia Electronic Tools for Learning. *Educational Technology*, 34(5), 47-52.

Tagg, A., & Dickinson, J. (1995). Tutor Messaging and Its Effectiveness in Encouraging Student Participation in Computer Conferences. *Journal of Distance Education*, 10(2), 33-55.

Taylor, H.G., & Mounfield, L.C. (1994). Exploration of the Relationship Between Prior Computing Experience and Gender on Success in College Computer Science. *Journal of Educational Computing Research*, 11(4), 291-306.

Tessmer, M. (1994). Formative Evaluation Alternatives. *Performance Improvement Quarterly*, 7(1), 3-18.

Thompson, R.D. (1987). Responsive, Formative Evaluation: A Flexible Means for Improving Distance Learning Materials. *Journal of Distance Education*, 2(1), 62-76.

Thompson, J.M., & McNair, N.L. (1995). Health care reform and emergency outpatient use of rural hospitals in Alberta, Canada. *Journal of Emergency Medicine*, 13(3):415-21.

Thompson, W., Dyke, J.P., and Buonocore, E. (1996). Using the World Wide Web to Train and Certify Physicians in the Safe Use of Fluoroscopy. *American Journal of Radiology*, 166, 1263-1264.

Thorpe, M. (1993). *Evaluating Open and Distance Learning.* London: Longman.

Tolhurst, D. (1992). A Checklist for Evaluating Content-Based Hypertext Computer Software. *Educational Technology*, 32(3), 17-21.

Tolhurst, D. (1995). Hypertext, Hypermedia, Multimedia Defined. *Educational Technology*, 35(2), 21-26.

Tovar, M., & Barker, N. (1986). Field Test Evaluation of Educational Software: A Description of One Approach. *CJEC*, 15(3), 177-190.

Tovar, M. (1989). Representing Multiple Perspectives: Collaborative-Democratic Evaluation in Distance Education. *The American Journal of Distance Education*, 3(2), 44-56.

Treloar, L.L. (1985). Facts About Teleconferencing for Staff Development Administrators. *The Journal of Continuing Education in Nursing*, 16(2), 47-52.

Tyler, R.W. (1950). *Basic Principles of Curriculum and Instruction.* Chicago: University of Chicago Press.

Umland, B., Waterman R., Wiese, W., Duban, S., Mennin, S., and Kaufman A. (1992). Learning from a rural physician program in China. *Academic Medicine*, 67,307-309.

Urban areas. Population and dwelling counts. (1992). Ottawa: Statistics Canada, (p. 62, Table 3).

Van Lint, S. (1994). QA and CE: a view from the country. *Australian Family Physician*, 23(9), 1648-9.

Vanzyl, A.J., & Cesnik, B. (1995). A Model for Connecting Doctors to University Based Medical Resources through the Internet. *Proceedings/The Annual Symposium on Computer Applications in Medical Care*, 517-521.

Verby, J.E., & Feldman, B. (1983). The Rural Physician Associate Program (RPAP). A Righton Program for all participants. *Minnesota Medicine*, 66, 181-185.

Vertecchi, B. (1993). A two-level strategy for mastery of learning in distance education. In K. Harry, M. John, & D. Keegan (Eds.), *Distance education: New perspectives* (pp. 126-136). London: Routledge.

Wagner, J.C. (1983). One-to-One and Small Group Formative Evaluation: An Examination of Two Basic Formative Evaluation Procedures. *Performance & Instruction Journal*, 22(5), 5-7.

Wagner, N. (1984). Instructional Product Evaluation Using the Staged Innovation Design. *Journal of Instructional Development*, 7(2), 24-27.

Walker, L.O., & Avant, K.C. (1995). *Strategies for Theory Construction*. Appleton & Lange: Norwalk, CT.

Weber, W.C. (1992). An Interactive Systems Model Derived from Experts' Perceptions of Human-Computer Issues. *Educational Technology Research and Development*, 40(1), 29-39.

Weston, C.B. (1987). The Importance of Involving Experts and Learners in Formative Evaluation. *CJEC*, 16(1), 45-58.

Whitener, L. (1996). Telecommunications and Rural Health Care. *The Journal of Rural Health*, Winter 1996, 67-71.

Whitten, P., Ford, D.J., Davis, N., Speicher, R., and Collins, B. (1988). Comparison of Face-to-Face versus Interactive Video Continuing Medical Education Delivery Modalities. *The Journal of Continuing Education in the Health Professions*, 18, 93-99.

Wilkinson, G.L., Oliver, K.M., and Bennett, L.T. (1997). *Evaluating the Quality of Internet Information Sources: Quality Indicators as Ranked by Experienced Internet Users*.
Available: http://itech1.coe.uga.edu/faculty/gwilkinson/rankings.html, Department of Instructional Technology University of Georgia, (May 19, 1997).

Willis, B. (1993). *Distance Education A Practical Guide*. Englewood Cliffs, NJ: Educational Technology Publications.

Willis, B. (1994). *Distance Education Strategies and Tools*. Englewood Cliffs, NJ: Educational Technology Publications.

Willis, B. (1995). *Distance Education at a Glance* (Online).
Available: http://www.uidaho.edu/evo/distglan.html, University of Idaho.

Wilson, A.L., Wellman, L.R., Fenton, L.J., and Witzke, D. (1982). Development of a Correspondence CME Course for Rural Physicians. *Journal of Medical Education*, 57, 635-637.

Wise, A.L., Hays, R.B., Adkins, P.B., Craig, M.L., Mahoney, M.D., Sheehan, M., Siskind, V., and Nichols, A. (1994). Training for rural general practice. *The Medical Journal of Australia, 161*, 314-318.

Woodrow, J.E.J. (1994). The Development of Computer-Related Attitudes of Secondary Students. *Journal of Educational Computing Research*, 11(4), 307-338.

Woolf, C.R. (1991). Comparison of the Perceived CME Needs of Semirural and Urban Physicians. *The Journal of Continuing Education in the Health Professions*, 11, 295- 299.

Worthen, B.R., & Sanders, J.R. (1987). *Educational Evaluation: Alternative Approaches and Practical Guidelines*. New York: Longman.

Yang, C.S., & Moore, D. (1995). Designing Hypermedia Systems for Instruction. *Journal of Educational Technology Systems*, 24(1), 3-30.

Young, C.Y., Chart, P., Franssen, E., Tippings, J., Morris, B., and Davis, D. (1988). Effective Continuing Education for Breast Disease: A Randomized Trial Comparing Home Study and Workshop Formats. *The Journal of Continuing Education in the Health Professions*, 18, 86-92.

Appendix A

Online Recruitment Letter and Recruitment Submission Form

Computer-Assisted CME at a Distance: An Evaluation Study

February 6, 1998

Dear Family Physician:

The Office of Professional Development and the Telemedicine Center, Faculty of Medicine are engaged in a development and evaluation project investigating the utility and learning effectiveness of continuing medical education delivered via a World Wide Web and CD-ROM learnware system. The lead evaluator for this project is Mr. Vernon Curran, Telemedicine Center and the Office of Professional Development, Faculty of Medicine.

We are currently compiling a list of volunteers who are interested in participating in this project. In order to participate as a subject in this project you will require:

- a multimedia computer workstation - CD-ROM, audio speakers, video and audio cards (either at home, office, a colleague's, community library);
- Internet connectivity; e-mail address; and a graphics capable Web Browser;
- be a rural or remote physician.

Participation will entail the completion of a computer-assisted program of continuing medical education delivered through a CD-ROM/World Wide Web hybrid learning system. This computer-assisted learning system will be accredited for MAINPRO M-1 study credit under the auspices of a CME Planning Committee of the Office of Professional Development, Memorial University of Newfoundland.

Participation in this pilot project will entail:

- completion of the self-paced computer-assisted instructional program over the period of seven days;
- completion and submission of on-line pre and post-achievement tests;
- participation in a 30 - 45 minute telephone or face-to-face interview with the evaluator or a designate;
- completion of a 3-month post-learning questionnaire.

The approximate time to participate in this self-paced computer-assisted learning project will be 1.5 - 2 hours. The time of completion is not important to the evaluation, but is mentioned here to give you some idea of the commitment involved. Your anonymity will be protected through the results of the evaluation being summarized with others, without reference to the name of the institution or individuals. All identifying information from achievement tests and interview notes will be removed and replaced with a code number. Participation in the study will have no effect on your work, progress, or other activities.

If you agree to participate, please complete and submit the attached form. If you wish to receive a copy of the evaluation results, please indicate on the attached form.

Thank you for considering this request to participate in this study.

Vernon R. Curran, M.Ed.
Program Manager
Telemedicine/TETRA
Faculty of Medicine
A1B 3V6

Phone # 709-737-6654

Fax # 709-737-7054

vcurran@morgan.ucs.mun.ca

Please complete the following form:

☑ **Yes I would be interested in participatng in this research study.**

Personal Information

Name

Address:

Telephone #:

Fax #:

E-Mail:

What are the specifications of your computer system (select those that apply):

Computer Type:
PC - 386

If you have a Mac, please enter the model number here:

Modem Speed:
9600 baud

CD-ROM Drive Speed:
2x

Web Browser:

Netscape 2.0 ▾

Other:

☐ **I would like to receive a copy of the research report.**

Submit the Form | Clear the Form

Appendix B

Consent Form

Memorial
University of Newfoundland

TELEMEDICINE/OFFICE OF PROFESSIONAL DEVELOPMENT
FACULTY OF MEDICINE

Consent To Participate In Continuing Medical Education Research

TITLE:	Evaluation of the Effectiveness of Distance Continuing Medical Education Delivered via a Computer-Mediated Learning System
INVESTIGATOR:	Mr. Vernon R. Curran, M.Ed.
SPONSOR:	Telemedicine, Office of Professional Development, University of Guelph (Rural Extension Studies)

You have been asked to participate in a research study. Participation in this study is entirely voluntary. You may decide not to participate or may withdraw from the study at any time. Information obtained from you or about you during this study, which could identify you, will be kept confidential by the investigator. The investigator will be available during the study at all times should you have any problems or questions about the study.

4. **Purpose of study:**

The purpose of this study is to examine the effectiveness of computer-mediated learning technologies (specifically the World Wide Web and interactive multimedia) for facilitating the delivery of continuing medical education at a distance.

5. **Description of procedures and tests:**

As a participant in this study you will be delivered continuing medical education at a distance through a computer-mediated learning system. This system will include the use of World Wide Web and interactive multimedia CD-ROM technologies. In order to participate in this computer-mediated learning group will need to have access to a multimedia computer with CD-ROM capabilities and Internet connectivity. You will be required to complete either: i) on-line pre and post-learning tests, or; ii) a post-learning test (depending on random assignment to a control or experimental group); and a post-learning performance change survey. As well, you may be invited to participate in a 30 - 45 minute telephone interview with the investigator. A survey will also be distributed to participants at a 3-month post-learning interval. It is anticipated that this survey will take approximately 10 minutes to complete.

3. **Duration of participant's involvement:**

Participants in the computer-mediated learning evaluation study will be required to complete the CME program over a two week (14 day) period. Requirements will include (estimated participation times are provided in brackets):

- the completion and submission of pre and post-tests or a post-test (depending on group assignment) (20 minutes);
- completion of the interactive learning modules (40 minutes);
- participation in an online asynchronous computer conference (15 minutes);
- *participation in a 30 - 45 minute post-learning interview with an investigator (30 - 45 minutes);
- and the completion of a post-learning survey at a 3-month post-course period (10 minutes).

The estimated time for participation in this study (including post-learning interviews): 2 hours and 10 minutes.

*A sample of participants will be contacted to participate in interviews. You may or may not be contacted by the investigator. Interviews will be conducted via telephone, at a time and place convenient to you.

6. **Possible risks, discomforts, or inconveniences:**

There are minimal risks, discomforts, and inconveniences for participants in the study. Participants in the study will have to fulfill several learning requirements in order to receive CME credit, ie, participation in computer conferencing, however all participants will receive CME credit at no cost to the participant.

7. **Liability statement.**

Your signature indicates your consent and that you have understood the information regarding the evaluation research study. In no way does this waive your legal rights nor release the investigators or involved agencies from their legal and professional responsibilities.

Participant's initials _______________

8. Any other relevant information:

If you would like to withdraw from this study at any time, please contact the investigator below:

Vernon R. Curran, B.A., Dip.Ed., M.Ed., Ph.D (candidate)
Telemedicine Centre
Faculty of Medicine
Memorial University of Newfoundland
St. John's, Newfoundland
A1B 3V6

Telephone #: (709) 737-6654
Fax #: (709) 737-7054 e-mail: vcurran@morgan.ucs.mun.ca

Participant's initials ____________

Signature Page

Title of Project: Evaluation of the Effectiveness of Distance Continuing Medical Education Delivered via a Computer-Mediated Learning System

Name of Principal Investigator: Vernon R. Curran, M.Ed.

To be signed by participant

I,____________________________, the undersigned, agree to my participation in the research

study described above.

Any questions have been answered and I understand what is involved in the study. I realise that participation is voluntary and that there is no guarantee that I will benefit from my involvement.

I acknowledge that a copy of this form has been given to me.

____________________________ ________________________________
(Signature of Participant) (Date)

____________________________ ________________________________
(Signature of Witness) (Date)

To be signed by investigator

To the best of my ability I have fully explained the nature of this research study. I have invited questions and provided answers. I believe that the participant fully understands the implications and voluntary nature of the study.

____________________________ ________________________________
(Signature of Investigator) (Date)

Appendix C

Computer Attitude Survey (CAS)

Computer Attitude Scale

NAME: ____

Directions to Participants: The following items describe some general ways that people may feel towards computers and computer technology. Please circle the number that best applies to that item: (5) if you strongly agree; (1) if you strongly disagree; or (4), (3), or (2) if it falls in between these extremes.

		Strongly Disagree 1	2	3	4	Strongly Agree 5
1.	Computers will never replace human life.	1	2	3	4	5
2.	Computers make me uncomfortable because I don't understand them.	1	2	3	4	5
3.	People are becoming slaves to computers.	1	2	3	4	5
4.	Computers are responsible for many of the good things we enjoy.	1	2	3	4	5
5.	Soon our lives will be controlled by computers.	1	2	3	4	5
6.	I feel intimidated by computers.	1	2	3	4	5
7.	There are unlimited possibilities of computer applications that haven't even been thought of yet.	1	2	3	4	5

8.	The overuse of computers may be harmful and damaging to humans.	1	2	3	4	5
9.	Computers are dehumanizing to society.	1	2	3	4	5
10.	Computers can eliminate a lot of tedious work for people.	1	2	3	4	5
11.	The use of computers is enhancing our standard of living.	1	2	3	4	5
12.	Computers turn people into just another number.	1	2	3	4	5
13.	Computers are lessening the importance of too many jobs now done by humans.	1	2	3	4	5
14.	Computers are a fast and efficient means of gaining information.	1	2	3	4	5
15.	Computers intimidate me because they seem so complex.	1	2	3	4	5
16.	Computers will replace the need for working human beings.	1	2	3	4	5

17.	Computers are bringing us into a bright new era.	1	2	3	4	5
18.	Soon our world will be completely run by computers.	1	2	3	4	5
19.	Life will be easier and faster with computers	1	2	3	4	5
20.	Computers are difficult to understand and frustrating to work with.	1	2	3	4	5

Appendix D

Online Demographic Profile Survey

Demographics Survey

Presented by ExamMail and TQUEST!

Please leave your name and e-mail address here:

To be recognized and credited, Name and Student # MUST be entered as below:

Jones, Henry B. (042)

E-mail (This is not essential. A written confirmation of your results will be sent to this address, if it is given):

If a survey question is Not Applicable, Leave it blank!

---

1. Theme: Practice Experience

Lead-in:

```
A. 1-3 years
B. 4-6 years
C. 7-9 years
D. 10-12 years
E. 13-15 years
F. 16-18 years
G. 19-21 years
H. 22-24 years
I. 25-27 years
J. 28-30 years
K. 31 years and greater
```

1. Years Practicing as a licenced physician

○ A ○ B ○ C ○ D ○ E ○ F ○ G ○ H ○ I ○ J ○ K

---

2. Theme:

Lead-in:

```
A. 0-2 hours
B. 3-4 hours
C. 5-6 hours
D. 7-8 hours
E. 9-10 hours
F. 11-12 hours
G. 13-14 hours
H. 15 hours or more
```

2. On average how many hours per week do you use computers? (At Work AND/OR at home)

○ A ○ B ○ C ○ D ○ E ○ F ○ G ○ H

---

3. Gender:

○ A. Female
○ B. Male

4. Are you a:

○ A. Rural Physician
○ B. Urban Physician
○ C. Neither

5. Practice Type:

○ A. Solo
○ B. Group
○ C. Not Applicable

6. Where do you have access to a computer?

○ A. Home
○ B. Work
○ C. Home and Work

7. How frequently do you use computer software for work-related or personal reasons (Word Processing, Personal Accounting, Spreadsheet)?

○ A. Not at all
○ B. Occasionally
○ C. Regularly
○ D. Often
○ E. Very Often

8. How frequently do you use e-mail?

○ A. Not at all
○ B. Occasionally
○ C. Regularly
○ D. Often
○ E. Very often

9. How frequently do you use the Internet/World Wide Web?

○ A. Not at all
○ B. Occasionally
○ C. Regularly
○ D. Often
○ E. Very often

10. Overall, how would you describe your experience with computers?

○ A. No previous experience
○ B. A little experience

- ○ C. Moderate experience
- ○ D. Considerable experience
- ○ E. Extensive Experience

Submit   Reset Quiz

Appendix E

Online Learning Achievement Pretest

Office Procedures Pretest

Presented by ExamMail and TQUEST!

Please leave your name and e-mail address here:

To be recognized and credited, Name and Student # MUST be entered as below:

Jones, Henry B. (042)

E-mail (This is not essential. A written confirmation of your results will be sent to this address, if it is given):

---

1. Which of the following procedures require local anaesthetic?

1. Fungal culture
2. Curettage
3. KOH preparation
4. Punch biopsy

○ A= 1,2, & 3 ○ B= 1 & 3 ○ C= 2 & 4 ○ D= 4 ○ E= All correct

---

2. When a patient presents with toe nail dystrophy, which of the following procedures are important as part of the routine investigation?

1. Examination of the web spaces of the feet.
2. Examination of the scalp, elbow, knees and finger nails.
3. Collection of nail debris for KOH and fungal culture.
4. Collection of nail debris for bacterial culture.

○ A= 1,2, & 3 ○ B= 1 & 3 ○ C= 2 & 4 ○ D= 4 ○ E= All correct

---

3. What features of a pigmented lesion are considered worrisome?

1. Asymmetry
2. Border irregularity
3. Colour variation
4. Diameter greater than 6 mm

○ A= 1,2, & 3 ○ B= 1 & 3 ○ C= 2 & 4 ○ D= 4 ○ E= All correct

---

4. Which of the following skin lesions are commonly confused with melanoma?

1. Basal cell carcinoma
2. Molluscum contagiosum
3. Seborrheic keratosis
4. Verruca vulgaris

○ A= 1,2, & 3 ○ B= 1 & 3 ○ C= 2 & 4 ○ D= 4 ○ E= All correct

---

5. Which of the following conditions can be associated with nail dystrophy?

1. Psoriasis.
2. Eczema.
3. Onychomycosis.
4. Lichen planus

○ A= 1,2, & 3 ○ B= 1 & 3 ○ C= 2 & 4 ○ D= 4 ○ E= All correct

---

6. With respect to basal cell carcinoma which of the following statements are true:

1. It is the most common cause of skin cancer.
2. Most lesions occur in the head and neck area.
3. High cure ratio with most treatment methods.
4. Higher recurrence rates for lesions in the nasiolabial folds.

○ A= 1,2, & 3 ○ B= 1 & 3 ○ C= 2 & 4 ○ D= 4 ○ E= All correct

---

7. Which of the following statements are true with respect to onychomycosis:

1. Dermatophytes are the most common cause of the the infection.
2. A negative culture means that a fungal infection is not the cause.
3. Candida accounts for 5% of the infection.
4. All patients with onychomycosis have a history of tinea pedis.

○ A= 1,2, & 3 ○ B= 1 & 3 ○ C= 2 & 4 ○ D= 4 ○ E= All correct

---

8. Which of the following treatments are useful for managing basal cell carcinoma?

1. Curettage and E.D.
2. Surgical Excision
3. Radiation Therapy
4. Office Cryotherapy

○ A= 1,2, & 3 ○ B= 1 & 3 ○ C= 2 & 4 ○ D= 4 ○ E= All correct

---

9. Onychomycosis may be associated with:

1. Tinea Pedis

2. Tinea Cruris
3. Tinea Corporis
4. Pain and discomfort of the affected nails

○A= 1,2, & 3 ○ B= 1 & 3 ○ C= 2 & 4 ○ D= 4 ○ E= All correct

---

10. Which of the following skin conditions would you treat with curettage alone?

1. Squamous cell carcinoma
2. Molluscum contagiosum
3. Basal cell carcinoma
4. Seborrheic keratosis

○A= 1,2, & 3 ○ B= 1 & 3 ○ C= 2 & 4 ○ D= 4 ○ E= All correct

---

11. In the treatment of onychomycosis, which of the following statements are true:

1. Terbinafine 250, PO daily x 3 months is effective therapy for toe nail infection.
2. Finger nail infection requires the same therapy as toe nails.
3. 200 mg Itraconazole PO BID x 7 days per month x 3 cycles, is effective therapy.
4. Topical anti-fungal cream is effective in onychomycosis and is an alternative when patients are unable to take oral therapy.

○A= 1,2, & 3 ○ B= 1 & 3 ○ C= 2 & 4 ○ D= 4 ○ E= All correct

---

12. Complications of the punch biopsy procedure could include:

1. Secondary infection
2. Bleeding
3. Lower limb biopsy can lead to chronic ulcer, especially in the elderly
4. Scarring

○A= 1,2, & 3 ○ B= 1 & 3 ○ C= 2 & 4 ○ D= 4 ○ E= All correct

---

13. Which of the following lesions would one likely biopsy using the punch biopsy technique?

1. Basal cell carcinoma
2. Lichen planus
3. Dermatitis Herpetiformis
4. Melanoma

○A= 1,2, & 3 ○ B= 1 & 3 ○ C= 2 & 4 ○ D= 4 ○ E= All correct

---

14. In an adult with molluscum contagiosum, which of the following tests would you consider performing?

1. CBC and differential
2. VDRL
3. HIV
4. IgE level

○ A= 1,2, & 3 ○ B= 1 & 3 ○ C= 2 & 4 ○ D= 4 ○ E= All correct

---

15. After a skin biopsy, what wound dressing is recommended?

○ A. Hydrocolloid dressing q 2 days
○ B. Plain gauze daily
○ C. Band-Aid and antibiotic ointment bid
○ D. Wet-to-dry saline dressings tid
○ E. None of the above

---

16. What type of virus causes molluscum contagiosum?

○ A. Togavirus
○ B. Papillomavirus
○ C. Poxvirus
○ D. Herpesvirus

---

17. What method is preferable in treating multiple molluscum contagiosum on the trunk of a four-year-old?

○ A. Cantharidin application
○ B. Electrocautery
○ C. Surgical excision
○ D. Curettage

---

18. When confronted with a patient with an unusual pigmented lesion one should:

○ A. Refer to a surgeon
○ B. Biopsy the lesion
○ C. Perform an excision of the lesion
○ D. Refer to a dermatologist

---

19. Which of the following statements regarding congenital nevi is true?

○ A. They have a 20% lifetime risk of developing melanoma
○ B. They should be excised as soon as possible, if possible.
○ C. They should be treated as if they were acquired nevi
○ D. They should be excised before adolescence, if possible

---

20. When treating warts with liquid nitrogen the recommended application time is:

○ A. 0-5 sec
○ B. 5-10 sec

- ○ C. 10-15 sec
- ○ D. 15-20 sec

Submit Reset Quiz

Appendix F

Paper-based Demographic Profile Survey and Learning Achievement Posttest (Control Group)

Demographic Survey

Instructions: ***Please circle the correct answer.***

1. Gender:

A. Female

B. Male

2. Are you a:

A. Rural Physician

B. Urban Physician

C. Neither

3. Practice Type:

A. Solo

B. Group

C. Not Applicable

4. Years Practising as a licenced physician:

A. 1- 3 years

B. 4 - 6 years

C. 7 - 9 years

D. 10 - 12 years

E. 13 - 15 years

F. 16 - 18 year

G. 19 - 21 years

H. 22 - 24 years

II. 25 - 27 years

J. 28 - 30 years

K. 31 and greater

Dermatological Office Procedures Achievement Test

1. Which of the following procedures require local anaesthetic?

   5. Fungal culture
   2. Curettage
   6. KOH preparation
   4. Punch biopsy

   A = 1,2, & 3   **B = 1 & 3**   **C = 2 & 4**   **D = 4 only**   **E = All correct**

2. When a patient presents with toe nail dystrophy, which of the following procedures are important as part of the routine investigation?

   1. Examination of the web spaces of the feet.
   2. Examination of the scalp, elbow, knees and finger nails.
   3. Collection of nail debris for KOH and fungal culture.
   4. Collection of nail debris for bacterial culture.

   A = 1,2, & 3   **B = 1 & 3**   **C = 2 & 4**   **D = 4 only**   **E = All correct**

3. What features of a pigmented lesion are considered worrisome?

   1. Asymmetry
   2. Border irregularity
   3. Colour variation
   4. Diameter greater than 6 mm

   A = 1,2, & 3   **B = 1 & 3**   **C = 2 & 4**   **D = 4 only**   **E = All correct**

4. Which of the following skin lesions are commonly confused with melanoma?

   1. Basal cell carcinoma
   2. Molluscum contagiosum
   3. Seborrheic keratosis
   4. Verruca vulgaris

   A = 1,2, & 3   **B = 1 & 3**   **C = 2 & 4**   **D = 4 only**   **E = All correct**

5. Which of the following conditions can be associated with nail dystrophy?

1. Psoriasis.
2. Eczema.
3. Onychomycosis.
4. Lichen planus

A = 1,2, & 3 **B = 1 & 3** **C = 2 & 4** **D = 4 only** **E = All correct**

6. With respect to basal cell carcinoma which of the following statements are true:

1. It is the most common cause of skin cancer.
2. Most lesions occur in the head and neck area.
3. High cure ratio with most treatment methods.
4. Higher recurrence rates for lesions in the nasiolabial folds.

A = 1,2, & 3 **B = 1 & 3** **C = 2 & 4** **D = 4 only** **E = All correct**

7. Which of the following statements are true with respect to onychomycosis:

1. Dermatophytes are the most common cause of the infection.
2. A negative culture means that a fungal infection is not the cause.
3. Candida accounts for 5% of the infection.
4. All patients with onychomycosis have a history of tinea pedis.

A = 1,2, & 3 **B = 1 & 3** **C = 2 & 4** **D = 4 only** **E = All correct**

8. Which of the following treatments are useful for managing basal cell carcinoma?

1. Curettage and E.D.
2. Surgical Excision
3. Radiation Therapy
4. Office Cryotherapy

A = 1,2, & 3 **B = 1 & 3** **C = 2 & 4** **D = 4 only** **E = All correct**

9. Onychomycosis may be associated with:

1. Tinea Pedis
2. Tinea Cruris
3. Tinea Corporis
4. Pain and discomfort of the affected nails

A = 1,2, & 3 **B = 1 & 3** **C = 2 & 4** **D = 4 only** **E = All correct**

10. Which of the following skin conditions would you treat with curettage alone?

1. Squamous cell carcinoma
2. Molluscum contagiosum
3. Basal cell carcinoma
4. Seborrheic keratosis

A = 1,2, & 3 **B = 1 & 3** **C = 2 & 4** **D = 4 only** **E = All correct**

11. In the treatment of onychomycosis, which of the following statements are true:

1. Terbinafine 250, PO daily x 3 months is effective therapy for toe nail infection.
2. Finger nail infection requires the same therapy as toe nails.
3. 200 mg Itraconazole PO BID x 7 days per month x 3 cycles, is effective therapy.
4. Topical anti-fungal cream is effective in onychomycosis and is an alternative when patients are unable to take oral therapy.

A = 1,2, & 3 **B = 1 & 3** **C = 2 & 4** **D = 4 only** **E = All correct**

12. Complications of the punch biopsy procedure could include:

1. Secondary infection
2. Bleeding
3. Lower limb biopsy can lead to chronic ulcer, especially in the elderly
4. Scarring

A = 1,2, & 3 **B = 1 & 3** **C = 2 & 4** **D = 4 only** **E = All correct**

13. Which of the following lesions would one likely biopsy using the punch biopsy technique?

1. Basal cell carcinoma
2. Lichen planus
3. Dermatitis Herpetiformis
4. Melanoma

A = 1,2, & 3 **B = 1 & 3** **C = 2 & 4** **D = 4 only** **E = All correct**

14. In an adult with molluscum contagiosum, which of the following tests would you consider performing?

1. CBC and differential
2. VDRL
3. HIV
4. IgE level

A = 1,2, & 3 **B = 1 & 3** **C = 2 & 4** **D = 4 only** **E = All correct**

15. After a skin biopsy, what wound dressing is recommended?

A. Hydrocolloid dressing q 2 days
B. Plain gauze daily
C. Band-Aid and antibiotic ointment bid
D. Wet-to-dry saline dressings tid
E. None of the above

A **B** **C** **D** **E**

16. What type of virus causes molluscum contagiosum?

A. Togavirus
B. Papillomavirus
C. Poxvirus
D. Herpesvirus

A **B** **C** **D**

17. What method is preferable in treating multiple molluscum contagiosum on the trunk of a four-year-old?

A. Cantharidin application
B. Electrocautery
C. Surgical excision
D. Curettage

A **B** **C** **D**

18. When confronted with a patient with an unusual pigmented lesion one should:

A. Refer to a surgeon
B. Biopsy the lesion
C. Perform an excision of the lesion
D. Refer to a dermatologist

A **B** **C** **D**

19. Which of the following statements regarding congenital nevi is true?

A. They have a 20% lifetime risk of developing melanoma
B. They should be excised as soon as possible, if possible.
C. They should be treated as if they were acquired nevi
D. They should be excised before adolescence, if possible

A **B** **C** **D**

20. When treating warts with liquid nitrogen the recommended application time is:

A. 0-5 sec
B. 5-10 sec
C. 10-15 sec
D. 15-20 sec

A **B** **C** **D**

Please complete the following if you would like to receive a research report of this study.

√ **Yes, I would like to receive a copy of a research report of this study.**

Name:
Address:

Comments:

Appendix G

Online Learning Achievement Posttest

Dermatological Office Procedures Post-Test

Presented by ExamMail and TQUEST!

Please leave your name and e-mail address here:

To be recognized and credited, Name and Student # MUST be entered as below:

Jones, Henry B. (042)

E-mail (This is not essential. A written confirmation of your results will be sent to this address, if it is given):

---

1. Which of the following procedures require local anaesthetic?

1. Fungal culture
2. Curettage
3. KOH preparation
4. Punch biopsy

○ A= 1,2, & 3 ○ B= 1 & 3 ○ C= 2 & 4 ○ D= 4 ○ E= All correct

---

2. When a patient presents with toe nail dystrophy, which of the following procedures are important as part of the routine investigation?

1. Examination of the web spaces of the feet.
2. Examination of the scalp, elbow, knees and finger nails.
3. Collection of nail debris for KOH and fungal culture.
4. Collection of nail debris for bacterial culture.

○ A= 1,2, & 3 ○ B= 1 & 3 ○ C= 2 & 4 ○ D= 4 ○ E= All correct

---

3. What features of a pigmented lesion are considered worrisome?

1. Asymmetry
2. Border irregularity
3. Colour variation
4. Diameter greater than 6 mm

○ A= 1,2, & 3 ○ B= 1 & 3 ○ C= 2 & 4 ○ D= 4 ○ E= All correct

---

4. Which of the following skin lesions are commonly confused with melanoma?

1. Basal cell carcinoma
2. Molluscum contagiosum
3. Seborrheic keratosis
4. Verruca vulgaris

○ A= 1,2, & 3 ○ B= 1 & 3 ○ C= 2 & 4 ○ D= 4 ○ E= All correct

---

5. Which of the following conditions can be associated with nail dystrophy?

1. Psoriasis.
2. Eczema.
3. Onychomycosis.
4. Lichen planus

○ A= 1,2, & 3 ○ B= 1 & 3 ○ C= 2 & 4 ○ D= 4 ○ E= All correct

---

6. With respect to basal cell carcinoma which of the following statements are true:

1. It is the most common cause of skin cancer.
2. Most lesions occur in the head and neck area.
3. High cure ratio with most treatment methods.
4. Higher recurrence rates for lesions in the nasiolabial folds.

○ A= 1,2, & 3 ○ B= 1 & 3 ○ C= 2 & 4 ○ D= 4 ○ E= All correct

---

7. Which of the following statements are true with respect to onychomycosis:

1. Dermatophytes are the most common cause of the the infection.
2. A negative culture means that a fungal infection is not the cause.
3. Candida accounts for 5% of the infection.
4. All patients with onychomycosis have a history of tinea pedis.

○ A= 1,2, & 3 ○ B= 1 & 3 ○ C= 2 & 4 ○ D= 4 ○ E= All correct

---

8. Which of the following treatments are useful for managing basal cell carcinoma?

1. Curettage and E.D.
2. Surgical Excision
3. Radiation Therapy
4. Office Cryotherapy

○ A= 1,2, & 3 ○ B= 1 & 3 ○ C= 2 & 4 ○ D= 4 ○ E= All correct

---

9. Onychomycosis may be associated with:

1. Tinea Pedis

2. Tinea Cruris
3. Tinea Corporis
4. Pain and discomfort of the affected nails

○ A= 1,2, & 3 ○ B= 1 & 3 ○ C= 2 & 4 ○ D= 4 ○ E= All correct

---

10. Which of the following skin conditions would you treat with curettage alone?

1. Squamous cell carcinoma
2. Molluscum contagiosum
3. Basal cell carcinoma
4. Seborrheic keratosis

○ A= 1,2, & 3 ○ B= 1 & 3 ○ C= 2 & 4 ○ D= 4 ○ E= All correct

---

11. In the treatment of onychomycosis, which of the following statements are true:

1. Terbinafine 250, PO daily x 3 months is effective therapy for toe nail infection.
2. Finger nail infection requires the same therapy as toe nails.
3. 200 mg Itraconazole PO BID x 7 days per month x 3 cycles, is effective therapy.
4. Topical anti-fungal cream is effective in onychomycosis and is an alternative when patients are unable to take oral therapy.

○ A= 1,2, & 3 ○ B= 1 & 3 ○ C= 2 & 4 ○ D= 4 ○ E= All correct

---

12. Complications of the punch biopsy procedure could include:

1. Secondary infection
2. Bleeding
3. Lower limb biopsy can lead to chronic ulcer, especially in the elderly
4. Scarring

○ A= 1,2, & 3 ○ B= 1 & 3 ○ C= 2 & 4 ○ D= 4 ○ E= All correct

---

13. Which of the following lesions would one likely biopsy using the punch biopsy technique?

1. Basal cell carcinoma
2. Lichen planus
3. Dermatitis Herpetiformis
4. Melanoma

○ A= 1,2, & 3 ○ B= 1 & 3 ○ C= 2 & 4 ○ D= 4 ○ E= All correct

---

14. In an adult with molluscum contagiosum, which of the following tests would you consider performing?

1. CBC and differential
2. VDRL
3. HIV
4. IgE level

○ A= 1,2, & 3 ○ B= 1 & 3 ○ C= 2 & 4 ○ D= 4 ○ E= All correct

---

15. After a skin biopsy, what wound dressing is recommended?

○ A. Hydrocolloid dressing q 2 days
○ B. Plain gauze daily
○ C. Band-Aid and antibiotic ointment bid
○ D. Wet-to-dry saline dressings tid
○ E. None of the above

---

16. What type of virus causes molluscum contagiosum?

○ A. Togavirus
○ B. Papillomavirus
○ C. Poxvirus
○ D. Herpesvirus

---

17. What method is preferable in treating multiple molluscum contagiosum on the trunk of a four-year-old?

○ A. Cantharidin application
○ B. Electrocautery
○ C. Surgical excision
○ D. Curettage

---

18. When confronted with a patient with an unusual pigmented lesion one should:

○ A. Refer to a surgeon
○ B. Biopsy the lesion
○ C. Perform an excision of the lesion
○ D. Refer to a dermatologist

---

19. Which of the following statements regarding congenital nevi is true?

○ A. They have a 20% lifetime risk of developing melanoma
○ B. They should be excised as soon as possible, if possible.
○ C. They should be treated as if they were acquired nevi
○ D. They should be excised before adolescence, if possible

---

20. When treating warts with liquid nitrogen the recommended application time is:

○ A. 0-5 sec
○ B. 5-10 sec

○ C. 10-15 sec
○ D. 15-20 sec

Submit Reset Quiz

Appendix H

Online Courseware Evaluation Survey

Dermatology Course Evaluation

Presented by ExamMail and TQUEST!

Please leave your name and e-mail address here:

To be recognized and credited, Name and Student # MUST be entered as below:

Jones, Henry B. (042)

E-mail (This is not essential. A written confirmation of your results will be sent to this address, if it is given):

If a survey question is Not Applicable, Leave it blank!

---

1. The case-studies were representative of possible clinical scenarios and encouraged professional problem-solving.

- ○ A. Strongly Disagree
- ○ B. Disagree
- ○ C. Neutral
- ○ D. Agree
- ○ E. Strongly Agree

---

2. The subject matter presented through this courseware enhanced my knowledge of dermatological office procedures.

- ○ A. Strongly Disagree
- ○ B. Disagree
- ○ C. Neutral
- ○ D. Agree
- ○ E. Strongly Agree

---

3. I believe the subject matter of this courseware is relevant to my professional practice.

- ○ A. Strongly Disagree
- ○ B. Disagree
- ○ C. Neutral
- ○ D. Agree
- ○ E. Strongly Agree

---

4. This CME courseware system provides interactivity which increases its instructional value.

- ○ A. Strongly Disagree

- ○ B. Disagree
- ○ C. Neutral
- ○ D. Agree
- ○ E. Strongly Agree

---

5. I was able to control the rate of presentation of the subject matter.

- ○ A. Strongly Disagree
- ○ B. Disagree
- ○ C. Neutral
- ○ D. Agree
- ○ E. Strongly Agree

---

6. The use of different media components(audio, video, text, images, photos) serves a clear purpose and presents the subject matter effectively.

- ○ A. Strongly Disagree
- ○ B. Disagree
- ○ C. Neutral
- ○ D. Agree
- ○ E. Strongly Agree

---

7. The web pages are attractive.

- ○ A. Strongly Disagree
- ○ B. Disagree
- ○ C. Neutral
- ○ D. Agree
- ○ E. Strongly Agree

---

8. The screen design and layouts were clear, uncluttered, and well-organized.

- ○ A. Strongly Disagree
- ○ B. Disagree
- ○ C. Neutral
- ○ D. Agree
- ○ E. Strongly Agree

---

9. All links and navigation buttons are clearly labeled and serve an easily identified purpose.

- ○ A. Strongly Disagree
- ○ B. Disagree
- ○ C. Neutral
- ○ D. Agree
- ○ E. Strongly Agree

---

10. The Web site is well organized for ease of use.

- A. Strongly Disagree
- B. Disagree
- C. Neutral
- D. Agree
- E. Strongly Agree

---

11. It was easy to navigate, so I could concentrate on learning the material rather than learning to use the courseware.

- A. Strongly Disagree
- B. Disagree
- C. Neutral
- D. Agree
- E. Strongly Agree

---

12. The learner manual was complete, clearly organized and easily understood.

- A. Strongly Disagree
- B. Disagree
- C. Neutral
- D. Agree
- E. Strongly Agree

---

13. Computer conferencing was a useful component of this courseware system.

- A. Strongly Disagree
- B. Disagree
- C. Neutral
- D. Agree
- E. Strongly Agree

---

14. I found it easy to post, respond and reply to messages in the computer conferencing sesssions.

- A. Strongly Disagree
- B. Disagree
- C. Neutral
- D. Agree
- E. Strongly Agree

---

15. I found it informative and beneficial to be able to communicate with my peers and the instructors.

- A. Strongly Disagree
- B. Disagree
- C. Neutral
- D. Agree
- E. Strongly Agree

---

16. Overall, the instruction I received through this courseware learning system was appealing.

interesting and motivating.

- A. Strongly Disagree
- B. Disagree
- C. Neutral
- D. Agree
- E. Strongly Agree

---

17. I would participate in another CME course offering of this type.

- A. Strongly Disagree
- B. Disagree
- C. Neutral
- D. Agree
- E. Strongly Agree

---

18. This courseware learning system is an effective way to participate in CME at a distance.

- A. Strongly Disagree
- B. Disagree
- C. Neutral
- D. Agree
- E. Strongly Agree

---

19. This method of distance education compares favorably with other available means of obtaining a similar education.

- A. Strongly Disagree
- B. Disagree
- C. Neutral
- D. Agree
- E. Strongly Agree

---

20. This courseware makes me confident in using computers and other technology.

- A. Strongly Disagree
- B. Disagree
- C. Neutral
- D. Agree
- E. Strongly Agree

---

21. The courseware was easy to use.

- A. Strongly Disagree
- B. Disagree
- C. Neutral
- D. Agree
- E. Strongly Agree

Submit Reset Quiz

Appendix I

Interview Script

Interview Script

Dr......Thank-you for participating in this interview with me today. I am interviewing a sample of physicians who participated in the Internet-based CME courseware "Dermatological Office Procedures." The purpose of these interviews is to evaluate from the physician's perspective the effectiveness of Internet-based CME. Your responses to my questions will allow us to assess and improve the effectiveness of CME delivered through computer-based technologies.

Our discussion will be anonymous and I will be the only person to have access to these tapes. Do you have any questions before we begin?

Computer Conferencing

The term computer conferencing refers to the Virtual Discussion component of the CME courseware. This was the area were you were able to post and respond to messages to and from your instructors and peers.

Did you post any messages or comments to the computer conference?

<table>
<tr><td rowspan="3">YES

Proceed to next.</td><td colspan="2">NO

Did you review the discussion?</td></tr>
<tr><td rowspan="2">YES

Why didn't you post any messages?</td><td>NO

Why didn't you review the discussion?</td></tr>
<tr><td>FINISH</td></tr>
</table>

Did you experience any problems posting or responding to messages?

YES	NO
How did you resolve these problems?	Proceed to next.

Do you think your knowledge of computers or your computer skills influenced your participation in the computer conference?

What aspects of the computer conferencing discussion did you like and/or dislike?

Do you feel a computer conferencing discussion differs from a discussion that would occur at a face-to-face CME workshop?

Did you find the discussion informative?

Is there anything else you would like to share about your experiences with the computer conferencing discussion?

Courseware Components

Courseware is a term I will use to refer to the CD-ROM learning materials.

What aspects of this continuing medical education courseware did you like and dislike?

Ease of Use Did you find the courseware convenient and easy to use?
Navigation Was it easy to navigate and locate information? Were the screens clear and uncluttered?
Content Was the subject matter of this courseware appropriate? Was the subject matter presented effectively? Did you find the courseware interesting to use?
Study Manual Was the study manual useful?

Testing

What is your opinion on being tested on what you have learned?

What other methods could be used to ensure individuals actually participate in self-directed learning?

Did you encounter any particular technical difficulties in using this courseware?

As a physician what do you think are the strengths and weaknesses of participating in continuing medical education through the use of computer technologies (such as the Internet and CD-ROM)?

Overall, how would you describe your experience with this form of computer-based learning?

If were to give advice to a physician preparing to participate in an Internet based CME course what would it be?

Do have any suggestions on what could be changed or done differently in the future?

Appendix J

Post-Learning Performance Self Assessment Survey

Post-Learning Performance Self Assessment Survey

The purpose of this survey is to assess the affect of your participation in the Internet-based CME courseware "Dermatological Office Procedures" on your clinical knowledge and skills. Your responses to this survey are anonymous and will only be used for evaluation research purposes.

For each of the following performance statements please identify the extent of your knowledge, skills and/or ability on a scale of 1 to 5 (To no extent - To a large extent) before and after your participation in this CME courseware.

1. Identify fungal infections using fungal culture and potassium hydroxide procedures.

BEFORE PARTICIPATING IN CME

1	2	3	4	5
To no extent				To a large extent

AFTER PARTICIPATING IN CME

1	2	3	4	5
To no extent				To a large extent

2. Read and interpret a positive KOH clearing test.

BEFORE PARTICIPATING IN CME

1	2	3	4	5
To no extent				To a large extent

AFTER PARTICIPATING IN CME

1	2	3	4	5
To no extent				To a large extent

3. Describe the purpose and procedure of fungal culture.

BEFORE PARTICIPATING IN CME

1	2	3	4	5
To no extent				To a large extent

AFTER PARTICIPATING IN CME

1	2	3	4	5
To no extent				To a large extent

4. **Identify lesions that are most appropriate for cryotherapy procedure management.**

BEFORE PARTICIPATING IN CME

1	2	3	4	5
To no extent				To a large extent

AFTER PARTICIPATING IN CME

1	2	3	4	5
To no extent				To a large extent

5. **Perform cryotherapy procedure appropriately to manage selected lesions.**

BEFORE PARTICIPATING IN CME

1	2	3	4	5
To no extent				To a large extent

AFTER PARTICIPATING IN CME

1	2	3	4	5
To no extent				To a large extent

6. **Describe the purpose and possible complications of cryotherapy procedure.**

BEFORE PARTICIPATING IN CME

1	2	3	4	5
To no extent				To a large extent

AFTER PARTICIPATING IN CME

1	2	3	4	5
To no extent				To a large extent

7. **Identify molluscum lesions.**

BEFORE PARTICIPATING IN CME

1	2	3	4	5
To no extent				To a large extent

AFTER PARTICIPATING IN CME

1	2	3	4	5
To no extent				To a large extent

8. Perform curettage technique successfully for the removal of molluscum lesions.

BEFORE PARTICIPATING IN CME

1	2	3	4	5
To no extent				To a large extent

AFTER PARTICIPATING IN CME

1	2	3	4	5
To no extent				To a large extent

9. Describe the purpose of the curettage technique for the removal of molluscum.

BEFORE PARTICIPATING IN CME

1	2	3	4	5
To no extent				To a large extent

AFTER PARTICIPATING IN CME

1	2	3	4	5
To no extent				To a large extent

10. Identify lesions that are most appropriate for punch biopsy office procedure.

BEFORE PARTICIPATING IN CME

1	2	3	4	5
To no extent				To a large extent

AFTER PARTICIPATING IN CME

1	2	3	4	5
To no extent				To a large extent

11. Perform punch biopsy procedure successfully.

BEFORE PARTICIPATING IN CME

1	2	3	4	5
To no extent				To a large extent

AFTER PARTICIPATING IN CME

1	2	3	4	5
To no extent				To a large extent

12. Perform haemostatsis procedure successfully.

BEFORE PARTICIPATING IN CME

1	2	3	4	5
To no extent				To a large extent

AFTER PARTICIPATING IN CME

1	2	3	4	5
To no extent				To a large extent

13. Describe the purpose of punch biopsy and haemostatsis procedures.

BEFORE PARTICIPATING IN CME

1	2	3	4	5
To no extent				To a large extent

AFTER PARTICIPATING IN CME

1	2	3	4	5
To no extent				To a large extent

Thank you.

Appendix K

Learning Objectives

Dermatological Office Procedures

The purpose of this dermatological office procedures learning module is to provide an overview of basic dermatological diagnostic and therapeutic office procedures. Upon completion of this self-directed instructional module you should be able to describe and perform the identified office procedures appropriately in a clinical office setting.

Learning Objectives

The learner will be able to:

III. Specimen Collection and Potassium Hydroxide Preparation in the Diagnosis of Fungal Infections:

- Describe the purpose of a fungal culture procedure
- Distinguish between the procedures for collecting skin and scalp lesion samples and nail samples
- Describe the purpose of the KOH procedure
- List the materials the required describe the KOH slide preparation procedure
- Identify an epithelial cell wall and a fungal hyphae cell wall of a positive KOH clearing test
- Determine the most appropriate cases for the use of the procedure.

II. Cryotherapy

- Describe the purpose of Cryotherapy
- List the lesions which Cryotherapy would be most useful in treating.
- Describe the liquid nitrogen treatment and the post-treatment procedure
- List and identify possible complications of Cryotherapy

III. Molluscum Management

- Identify molluscum lesions
- Describe the purpose of the curettage technique for the removal of molluscum

- Describe the cantharidin application and the post-treatment procedure.

IV. Punch Biopsy

- List materials required for the performance of the punch biopsy
- Describe the punch biopsy procedure
- Explain important points to consider in performing a punch biopsy
- List and identify lesions that one would punch biopsy.

Appendix L

Courseware Screen Captures

Courseware Pretest Screen

Office Procedures Pretest

Presented by ExamMail and TQUEST!

Please leave your name and e-mail address here:

To be recognized and credited, Name and Student # MUST be entered as below:

Jones, Henry B. (042)

E-mail (This is not essential. A written confirmation of your results will be sent to this address, if it is given):

---

1. Which of the following procedures require local anaesthetic?

   1. Fungal culture
   2. Curettage
   3. KOH preparation

Online Demographic Profile Survey Screen

Demographics Survey

Presented by ExamMail and TQUEST!

Please leave your name and e-mail address here:

To be recognized and credited, Name and Student # MUST be entered as below:

Jones, Henry B. (042)

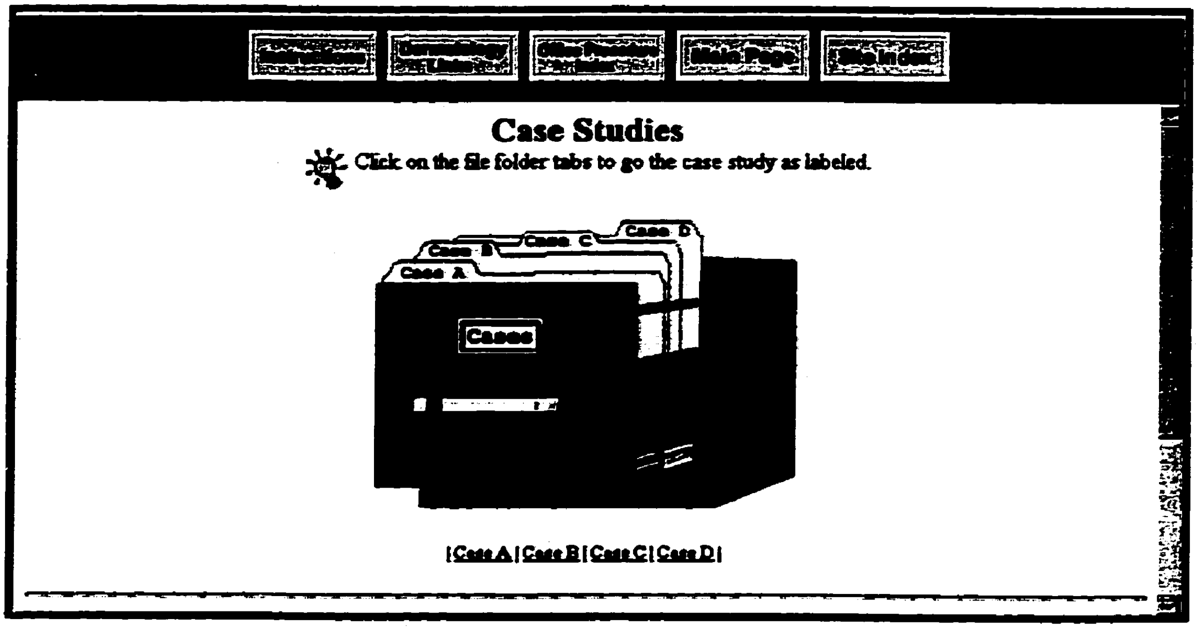
E-mail (This is not essential. A written confirmation of your results will be sent to this address, if it is given):

If a survey question is Not Applicable, Leave it blank!

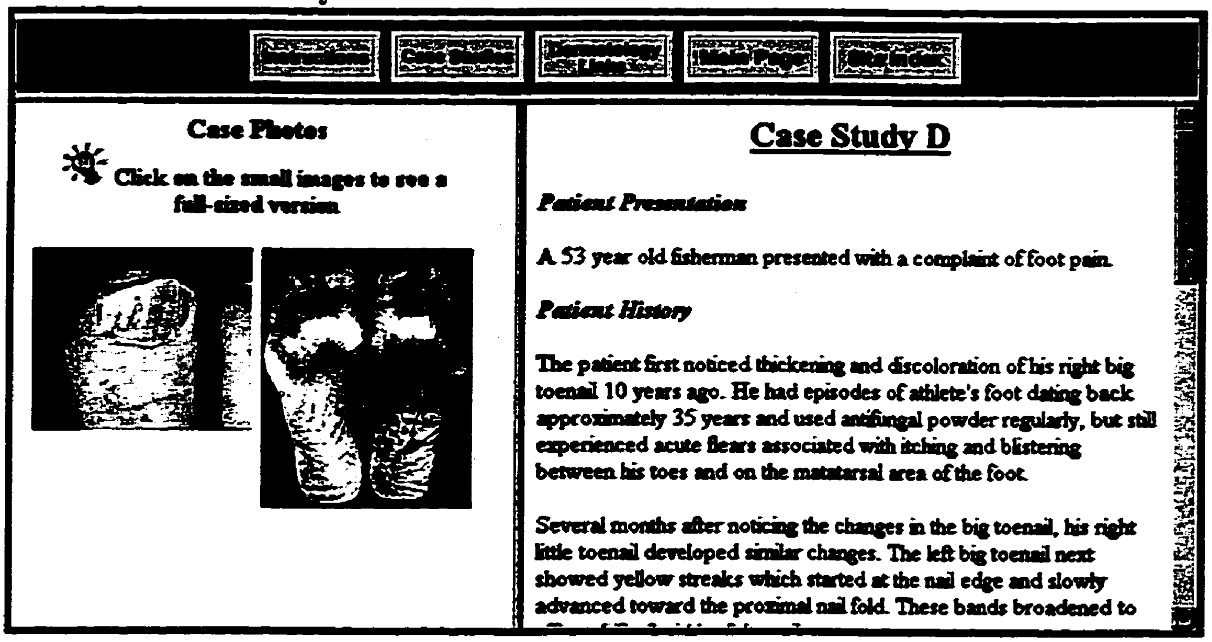
---

1. Theme: Practice Experience

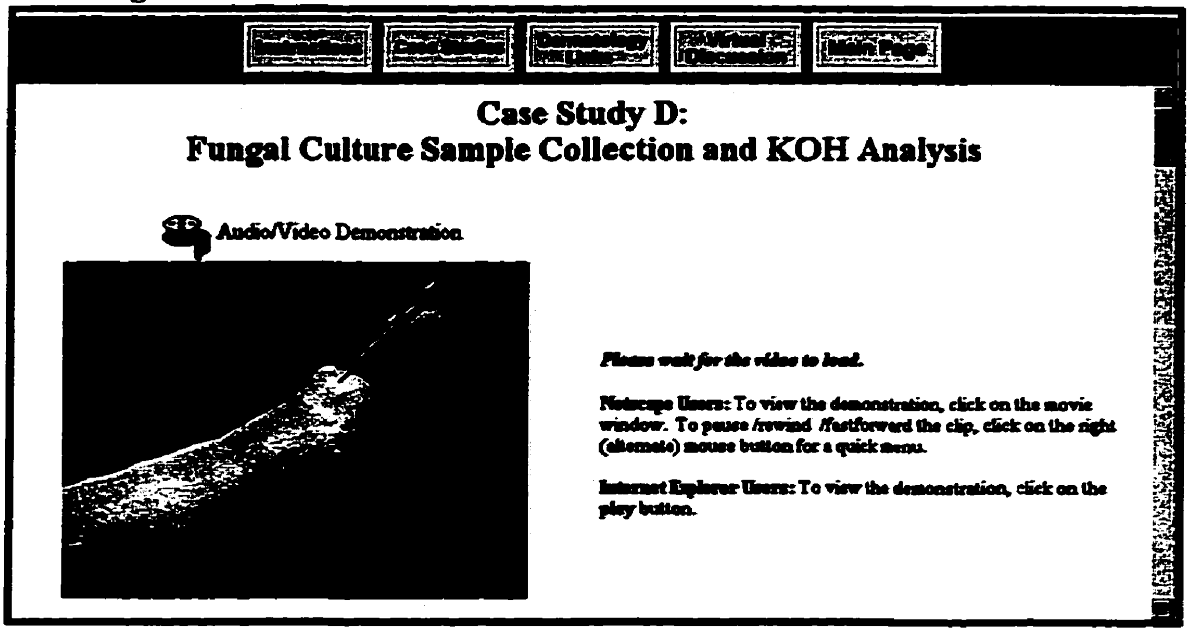
Case Studies Menu Screen



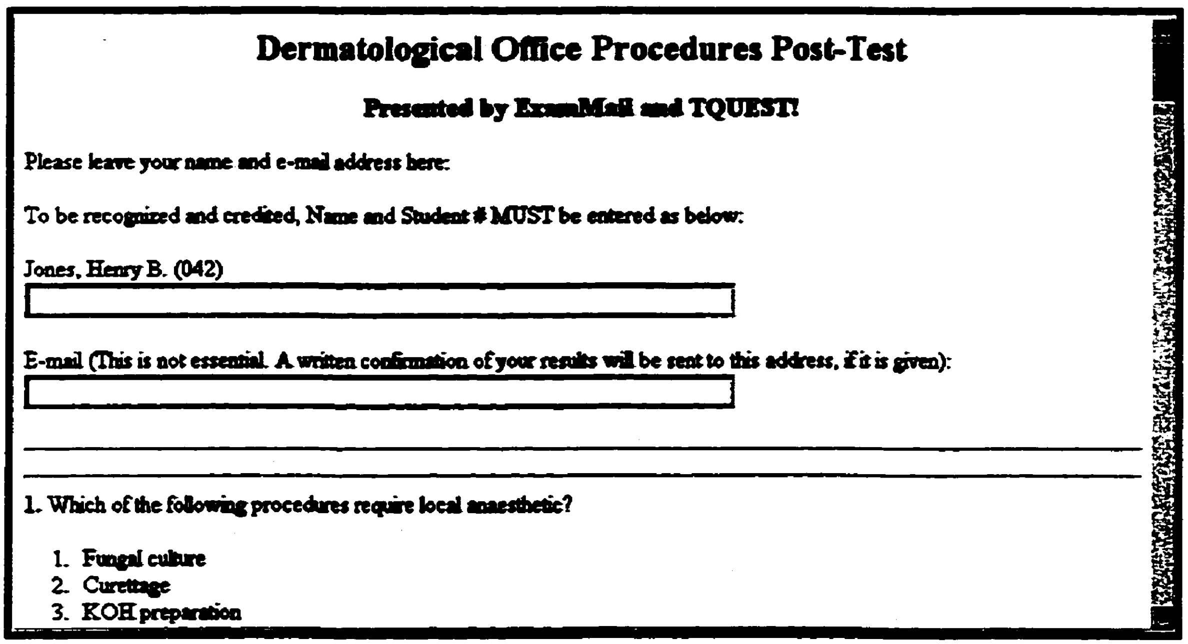
A Selected Case Study Screen



Learning Tutorial Screen



Online Posttest Screen



Online Courseware Evaluation Survey Screen

Dermatology Course Evaluation

Presented by ExamMail and TQUEST!

Please leave your name and e-mail address here:

To be recognized and credited, Name and Student # MUST be entered as below:

Jones, Henry B. (042)

E-mail (This is not essential. A written confirmation of your results will be sent to this address, if it is given):

If a survey question is Not Applicable, Leave it blank!

---

1. The case-studies were representative of possible clinical scenarios and encouraged professional problem-solving.